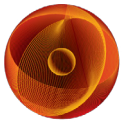


YST Centre *for* Music and Health

Supported By:

NATIONAL ARTS COUNCIL SINGAPORE

Better Everyday By Tote Board

Arts and Health Evaluation Toolkit (AHET)

A how-to guide to evaluate the health outcomes of arts programmes in the community

Acknowledgements

This toolkit was led by the Centre for Music and Health (CMH), at the Yong Siew Toh Conservatory of Music, National University of Singapore.

Authors:
Principal Investigator: Dr Kathleen Rose Agres, Centre for Music and Health, Yong Siew Toh Conservatory of Music, National University of Singapore
Co-Investigator: Dr Jean Liu, Centre for Evidence and Implementation
Collaborator: Dr Nisha Sajnani, Jameel Arts & Health Lab, New York University

Research Assistants (RAs), Student Assistants, and Volunteers:

Nik Ilya
Abielle Liu Shao Hua
Loh Jiayi
Peiyu (Evelyn) Zhang
Ye Xiaomeng
Lim Yu Juan
Sim Hsin Wei Lauren
Charlyn Boh
Seaw Jia Liew

Please cite this reference as:
Agres KR, Ilya N, Liu ASH, Loh J, Zhang P, Ye X, Lim YJ, Sim HWL, Boh C, Liew SJ, Sajnani N, Liu JCJ. Arts and Health Evaluation Toolkit: A preliminary guide to more rigorous evaluation of arts and health activities. Singapore; 2025. doi:10.17605/OSF.IO/4K9A8.

Acknowledgements:
The authors would like to thank and acknowledge the contributions of the team of students and CMH RAs, including Lee Liebin Daryl Renick, Raeanne Wong Chien Yin, Kishore Arushi Sanan, Foo Yi Yao, Seah Wei Hong, Shannon Chan, Laura Yee, Isabel Matthews, and Jovyn Lee for their help with the literature reviews for the Mental Health, Quality of Life, and Life Satisfaction outcomes.

Designer: Serene Soh

We would also like to thank our funders, the National Arts Council and Tote Board, for supporting this project.

Contents

SECTION 1

07 CHAPTER 1

Introduction

What is Arts and Health?

Why are Arts and Health programs not widely adopted?

Where does this Toolkit come in?

Who is this Toolkit for?

How to use this Toolkit?

Closing

12 CHAPTER 2

What is Evaluation

What is Evaluation? What is Research?

What are the different types of evaluation?

Why is it important to plan your evaluation?

How to map out your activities and outcomes?

Key concepts

21 CHAPTER 3

How to Design an Evaluation

Key things to consider before choosing a design

Types of Designs

What is an experimental design?

What are qualitatives study designs?

33 CHAPTER 4

Data Collection

Data collection

Conducting Quantitative Data Collection

Conducting Qualitative Data Collection

Combat Measures of Response Biases in Data Collection

SECTION 2

41 INTRODUCTION TO SECTION 2

44 CHAPTER 5

Measures of Mental Health Symptoms

Depression

Tools for measuring depressive symptoms

Anxiety

Tools for measuring anxiety symptoms

Summary of characteristics of all mental health questionnaires

How to choose an appropriate scale for your evaluation?

53 CHAPTER 6

Measures of Wellbeing and Related Outcomes

Quality of life, Life satisfaction, and Wellbeing

Tools for measuring quality of life

Tools for measuring life satisfaction

Tools for measuring wellbeing

Additional considerations when choosing an appropriate scale for your evaluation

65 CHAPTER 7

Measures of Social Connection

Social Connection, Loneliness, and Social Isolation

Tools for measuring social connection/connectedness

Tools for measuring loneliness

Tools for measuring social isolation

How to choose an appropriate scale for your evaluation?

76 CHAPTER 8

Physiological Measures

General considerations when choosing physiological measures

Physical fitness level

Motor function

Sleep

Stress response

88 **Appendix**

93 **References**

Executive Summary

The integration of arts in healthcare and public health has gained significant recognition for its potential to support the health and wellbeing of individuals and communities. There is a growing body of evidence suggesting that arts-based interventions can positively impact various health outcomes, including mental health, quality of life, and social well-being. However, despite the enthusiasm and promising results, a persistent challenge in the field of arts and health is the lack of robust evidence and rigorous evaluation practices.

While numerous studies claim benefits from arts-based interventions, the overall evidence base remains weak according to standard public health and medical evaluation criteria. This paradox of abundant yet methodologically limited evaluations has hindered the widespread adoption and integration of arts-based approaches in mainstream healthcare and public health initiatives.

This preliminary evaluation toolkit is designed to provide Arts and Health practitioners with guidance on conducting rigorous outcome evaluation of their programs and interventions. The toolkit aims to strengthen the evidence base for arts and health initiatives by promoting consistent and robust evaluation practices, enabling meaningful comparisons across studies and driving progress in the field.

The toolkit begins with an introduction that highlights the importance of evaluation in enhancing the effectiveness of arts and health programs. It then delves into the concept of evaluation, explaining its purposes and role in measuring the effects of a program. The toolkit also highlights study designs that can be more easily implemented by artist practitioners, such as quasi-experimental designs and mixed-methods approaches.

A key focus of the toolkit is on the use of validated tools/scales in quantitative study designs, emphasizing the importance of using established, reliable, and valid measurement instruments. Practitioners are guided through the process of selecting appropriate measures across various domains, including mental health, quality of life, social measures, and physiological measures. The toolkit also offers practical advice on implementing evaluation, covering data collection methods and best practices.

Overall, this toolkit aims to equip Arts and Health practitioners with the necessary knowledge and skills to conduct robust evaluations, ultimately enhancing the field's evidence base and supporting the integration of arts-based programs into mainstream health services.

CHAPTER 1

Introduction to the toolkit

The arts play an important role in supporting people's health and wellbeing. In recent years, arts-based health interventions have gained growing global recognition for their ability to enhance individual and community well-being. Despite the increasing number of Arts and Health studies being published, the field has been criticised for a lack of rigour in evaluation. In addition, many arts-based programs face challenges like limited funding and low participant numbers. These constraints make it even more important to share best practices for better and more consistent evaluation practices. This preliminary toolkit serves to provide these recommendations.

WHAT IS ARTS AND HEALTH?

The field of **Arts and Health** refers to an ecosystem of practices and professions that use arts, culture, and creative expression to address specific health concerns and promote wellbeing across the lifespan.[1]

The field of "Arts and Health" is also referred to as "Creative Health", "Arts in Health", or "Arts for Health". While each term may emphasise different approaches, they all refer to the use of arts to enhance health and wellbeing.[3]

These applications of arts across the health and care sectors are diverse. For example, the arts can support wellbeing in everyday life through regular creative activities; they can also be integrated into programmes for specific clinical populations or demographic groups, such as older adults, or applied in creative arts therapies, where trained therapists use art forms as part of clinical treatment.[2] In addition, the arts play a role in health promotion, helping to share health messages in engaging ways, and in healthcare environments, where art can make hospitals feel more welcoming and supportive.[2]

Arts and Health programs have a broad range of benefits for individual and community health, both as a means of prevention and treatment.

Impact of Arts on Health

Prevention and Promotion	Management and Treatment
Arts affect Social Determinants of Health	Arts help people experiencing mental illness
Arts encourage health-promotiong behaviours	Arts Support Care
Arts help to prevent ill health	Arts support end-of-life care
Arts support child development	Arts assist with management of non-communicable diseases
Arts support caregiving	Arts help support people with neurodevelopment and neurological disorders

Figure 1.1 Categories of impact of arts on health, based on Fancourt and Finn's 2019 scoping review.[4]

WHY ARE ARTS AND HEALTH PROGRAMS NOT WIDELY ADOPTED?

Despite growing interest, Arts and Health programs face barriers such as cultural perceptions of the arts as non-essential and a lack of rigorous evaluation data. Many funders and policymakers want clear evidence that Arts and Health programs "work".

This lack of rigorous data happens because:

Evaluation often gets overlooked in the arts. Many arts practitioners may not focus on formal evaluation for outcomes of their program.

Different goals, different measures. The arts and health sectors often have differing objectives and evaluation methods, making it hard to align the two.[5]

Limited formal research training. While arts practitioners may be experts in their field, many may not have formal training in research or evaluation methods. As a result, there's a need for more structured, evidence-based approaches to measuring impact.

Because of these challenges, structured outcome evaluation is especially important. It helps to strengthen the intersection of the arts and health-related sectors by generating credible, consistent data. For arts practitioners, structured outcome evaluation can:

- Facilitate access for Arts and Health interventions into the health system
- Build a stronger case for funding, long-term support, and buy-in from policy makers

WHERE DOES THIS TOOLKIT COME IN?

This toolkit guides you to implement effective outcome evaluations, addressing key challenges in measuring impact. In this toolkit, we provide:

An introduction to Outcome Evaluation. We answer key questions like “What is evaluation?” and “Where do I begin?” to help you assess the health-related outcomes of your programs.

Guidance on evaluation design. We provide practical advice for designing an evaluation that fits your project’s goals and context.

Recommendations for tools and measures. We offer recommendations for validated tools for accurate and consistent outcome measurement, guiding you on selecting the right ones for your project.

Directions to additional resources. We guide you to further materials that can support your evaluation efforts.

WHO IS THIS TOOLKIT FOR?

Anyone engaged in arts and health activities—whether you’re new to evaluation or looking to refine your approach—can make use of this toolkit! If you have no prior experience, we designed the toolkit to guide you step by step through designing and conducting evaluations in a **simple, structured way.** More experienced users can also use this toolkit to find additional resources to refine their evaluation design. We hope this toolkit encourages you to **incorporate evaluation into your practice** to demonstrate the value of your work.

HOW TO USE THIS TOOLKIT?

If you are just getting started…

Start with Section 1,

where we explain what outcome evaluation is and guide you through each step of the process. Each chapter in this section focuses on a specific stage of evaluation—from planning and design to choosing tools and analysing results—presented in a clear, structured way. This section is ideal if you’ve never done evaluation before.

If you’re familiar with evaluation…

Head to Section 2,

where we recommend specific, validated tools to help you measure common outcomes in arts and health programs, and present important considerations for their use.

CLOSING

By embedding evaluation into Arts and Health practice, we can ensure these programs continue to grow, evolve, and build a stronger case for funding. This is a step towards building a stronger, evidence-informed future for Arts and Health.

CHAPTER 2

What is Evaluation

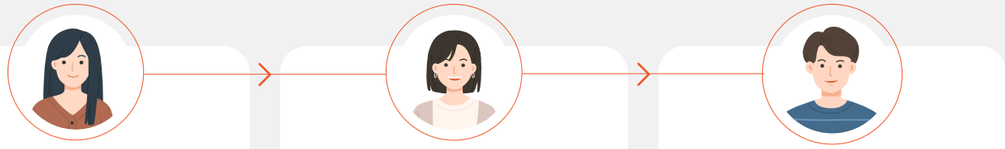
WHAT IS EVALUATION? WHAT IS RESEARCH?

Evaluation is the systematic process of collecting and analysing information to identify the components required for a programme or activity's success, uncover factors that may improve or hinder its implementation, and assess its effectiveness, value, or impact.[6] It helps to assess whether goals are being met, informs decision-making, and guides improvements. Evaluation is done within a specific context (e.g., within the aims and objectives of an organisation). In contrast, **research** is about making contributions to existing knowledge and finding new knowledge that is transferable to local contexts or generalisable to wider contexts. Evaluation findings are typically shared with stakeholders involved in a project, while research findings can be disseminated to the broader scientific community via publications.[7,8]

Evaluation and research are not mutually exclusive and can complement each other. For instance, research findings can inform areas to evaluate, and evaluation findings can inform areas that need further research. Regardless, you need to know your rationale for conducting evaluation and/or research before embarking on either.

WHAT ARE THE DIFFERENT TYPES OF EVALUATION?

Knowing why you are conducting an evaluation will help you decide the type of evaluation to undertake. Plan your project in a synchronous manner with your evaluation so that data can be collected at appropriate stages of your project.



If you have not started formalising your project and would like to know whether it is feasible, or if you need to understand **what elements are needed** for the intended audience, do **formative evalution.**

If your project is already underway and your wish to better understand whether it can be improved, you can do **process evalution.** Process evalution may also be used in the later or final stages of your project to **understand the reasons** why your project was successful or did not meet its intended goals.

If you would like to understand whether your project has achieved its desired goals, **outcome evaluation** can be done, which looks at immediate or **short-term effects** of your project and **whether the inteded objectives were achieved.**

Figure 2.1 Types of Evaluation

The different types of evaluation are presented in the table below:

Types of Evaluation	Definitions	Aspects
Formative Evaluation[7]	• Conducted before project implementation • Concerned with pre-implementation considerations such as the design and suitability of the project	• Situation Analysis (secondary data/existing findings and literature: Identify what/where/why is there a problem, who is affected, and what has already been done for the problem) • Stakeholder Analysis (identify stakeholders' interests in your project, their involvement/power in the project, and actions required with them)
Process Evaluation[9]	• Conducted while the project is being implemented (after project delivery has begun) • Seeks to explore considerations such as points of success and challenges	• Partnerships with stakeholders • Quality assurance (e.g., training of staff) • Coverage (e.g., outreach) • Utilisation (e.g., how many participants complete/engage/enrol in your project)
Outcome Evaluation[10]	• Conducted after the project has concluded • Typically examines short- and medium-term effects, and whether intended objectives were met. Long-term outcomes may be included if they are closely tied to the project goals and are measurable within the evaluation timeframe.	• Assess changes in knowledge/attitudes/behaviours/practices as a result of the activity or intervention that align with intended goals of your project • Note: Although outcome evaluation is usually conducted when your project has concluded, planning should start at the beginning of your project timeline, as baseline data (collected before the activity or during the Formative Evaluation stage) are beneficial for comparison with the data obtained at the end of your project

WHY IS IT IMPORTANT TO PLAN YOUR EVALUATION?

Before embarking on your project, it is important to establish *what* your project aims to accomplish and the rationale of *why* your project must be evaluated. Knowing *who* is involved in your project is also essential because projects involve various stakeholders, and may include external collaborators, internal team members, high-level decision-makers (e.g., policymakers, funders), and the recipients of a project (e.g., the participants or audience). Involving or consulting key stakeholders when planning and evaluating your project will help ensure that your project objectives are shared and achievable.[7] Setting objectives will help to establish clear and measurable project goals.[7,8]

To establish outcome evaluation objectives:
You may use the following template to guide the formulation of your outcome evaluation objectives: To evaluate the effectiveness of the [INSERT PROGRAMME] in improving the [INSERT OUTCOME] of [INSERT SAMPLE POPULATION]. Here are some examples:

To evaluate the effectiveness of a participatory music programme in improving the mental health of university students

To evaluate the effectiveness of the group dance programme on improving social connection among older adults

Examples of outcome evaluation questions according to the objectives above:
Consider the following template for writing your outcome evaluation questions: Do [SAMPLE POPULATION / PARTICIPANTS] report an improved [OUTCOME] after completing the [PROGRAMME]?

Do university students report improved mental health after completing the participatory music programme?

Do the older adults in the group dance programme exhibit improved social connection?

HOW TO MAP OUT YOUR ACTIVITIES AND OUTCOMES?

A **theory of change** is a visual pathway that maps out processes and steps required to achieve your project's desired outcome(s).[10,11] A theory of change consists of four main components - 1) the situation, 2) activities, 3) outcomes, and 4) the long-term desired change or impact. Outcomes can be segmented into short-term and medium-term outcomes if needed.

When designing and planning your project, the desired impact (i.e., long-term desired change) should be established first to define the project's main goal. Activities and short-term/medium-term outcomes can be subsequently planned out to identify the aspects required to achieve the project's main goal.[7] Hence, design and planning typically proceed in the opposite direction of the project's implementation process. The following diagram provides an overview of what a theory of change would look like (follow 'direction of implementation'), and the steps to take when designing one (follow 'direction of design and planning').[10,12,13]

Many practitioners and evaluators find it helpful to use a logic model[14], derived from components in the theory of change, to think through a planned project and map out indicators to measure the project's success. If you would like to explore this, we recommend consulting the Ministry of Social and Family Development of Singapore's resource on developing a logic model[15]. For a more elaborated account with interactive tools and examples, this resource[14] from the Community Tool Box may prove helpful.

Theory of Change (Overview of structure and planning)

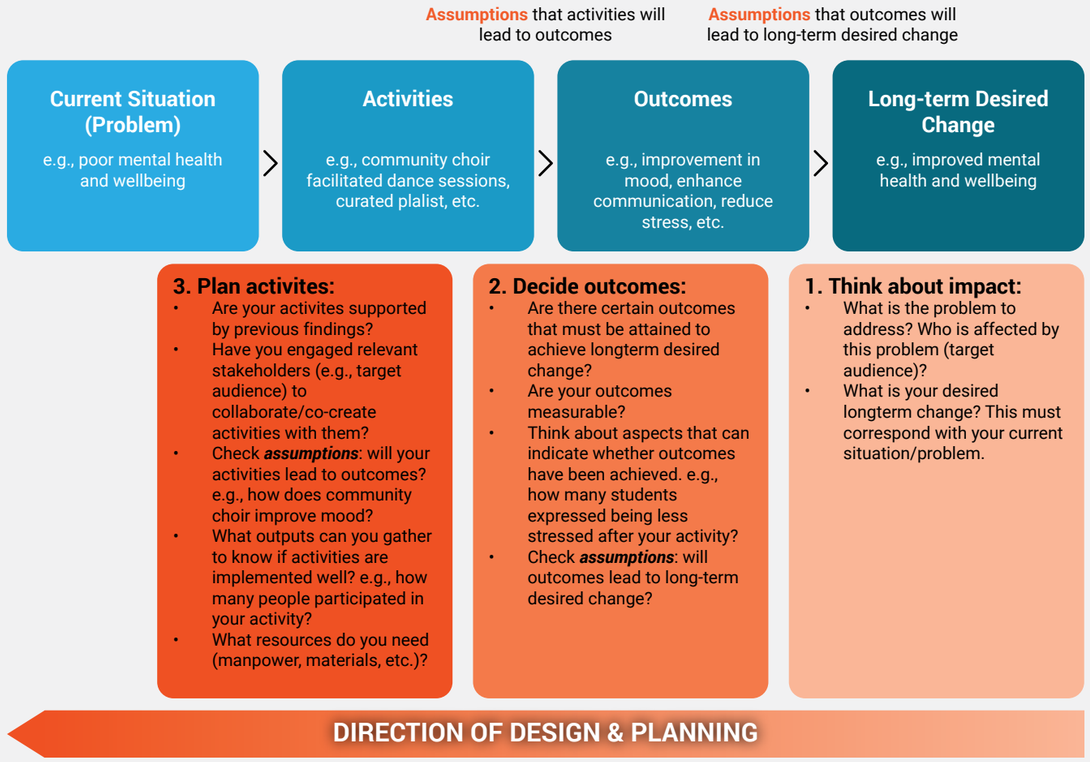


Figure 2.2 Theory of Change

KEY CONCEPTS

I. What are Quantitative and Qualitative Methods?

When conducting an evaluation, numerous methods of data collection can be employed.[7] However, this toolkit focuses primarily on *quantitative* and *qualitative* methodologies, with greater emphasis on the quantitative.

Quantitative methods deal with numerical data and focus on gathering measurable data from participants.[10,16] Surveys and questionnaires are often used for collecting data on psychological concepts like mental health and quality of life, although other methods, such as physiological data collection, may also be employed. Quantitative data generally measures the degree of an effect or compares numerical averages/ranges between different groups or time points. Hence, statistical analyses are required (e.g., calculating averages of the available data).

An example of an Arts and Health study employing quantitative methods is the interventional study conducted by Keisari and colleagues[17], which examined the effects of playback theatre on the mental health of community-dwelling older adults. Participants reported various aspects of their mental health through questionnaires, including a self-esteem scale and a loneliness scale.

On the other hand, qualitative methods involve non-numerical/textual data (e.g., interview transcripts) and focus on gathering subjective feedback.[10,15,18] They are useful for gathering lived experiences and perspectives of your project's participants, which are often not fully captured in surveys and standardised questionnaires, for a more holistic view of the program's impact.[10,16,18] Interviews, focus group discussions, open-ended surveys, and observations are some examples of qualitative data collection methods.[10,18,19] Statistical analyses are generally not required.

QUANTITATIVE METHODS

- Deductive
- Emphasis on measurable outcomes and facilitates measurement of items using the same scale
- Results independent of beliefs

QUALITATIVE METHODS

- Inductive
- Emphasis on meanings and interpretation
- Results dependent on perspectives

II. Mixed methods

The use of both quantitative and qualitative methods in the same study is called **mixed methods**[20]. It is important to note that mixed methods is more than just the usage of qualitative with quantitative methods; it is the integration of both methods to collect different types of data to capture additional nuances in your project's results.[21,22] This approach will allow you to tap into the respective strengths of quantitative and qualitative approaches, which can help provide a more holistic understanding of your project findings.[22,23,24]

III. Generalisability of study findings

Generalisability refers to whether the findings of a study, which uses a sample of the population, can be appropriately applied beyond the sample and circumstances of the study. While it is commonly emphasised in research, evaluations are generally less concerned about generalisability.

Generalisability is affected by many factors such as sampling (how participants were selected), the sample's representativeness, the research methodology, and the desired precision of the study (higher precision generally requires more controlled environments, which lowers generalisability).[25] These factors are affected by biases that may be inherent in the methodology of studies or sampling methods, or introduced by the way the data were collected. It is important to recognise these influences, mitigate them where possible, and acknowledge them in a section on your study's limitations. Some of these common biases will be discussed below in subsection VI.

An overview of methods and their generalisability is presented below:

	Quantitative	Qualitative	Mixed methods
Focus	Focuses on broad conclusions that can be applied to many situations with analysis of variables or factors for a wider range of cases	Focuses on background information or bigger picture through the in-depth analysis of holistic systems for a few selected cases	Integrates both quantitative and qualitative data collection and analysis for stronger inference and more holistic picture of a phenomenon
Generalisability	Usually has a larger sample sizes for the results obtained to be transferable to other settings/population/ cases **(External validity)**	Usually has a smaller sample size, results obtained cannot be easily be applied to other contexts and setting **(Transferability)**	A robust mixed- methods study with clearly defined aims and process would enable statistical generalisation from quantitative data and better contextualisation from qualitative data

IV. What is a validated scale?

Validated scales are questionnaires tested for psychometric properties of validity and reliability, and exhibit acceptable levels of both concepts.

Scale Validity and Reliability: Validity in a scale refers to the extent to which it measures what it is intended to measure (e.g., ensuring that a depression scale measures depression symptoms, rather than something else). **Reliability** refers to the consistency of results from a certain scale/tool.

V. Why use validated scales instead of bespoke scales?

Gathering data through validated scales enables comparisons to be made across different studies, whereas bespoke scales are generally created for a specific research study or evaluation. Additionally, bespoke scales are not often tested for their reliability, and thus may not be suitable for direct comparisons. Validated scales may also have benchmarks/cutoffs/ thresholds which allow users to interpret their scores (e.g., what a score of 7 out of 10 on a cognitive test means). Using a validated scale also helps reduce the possibility of measurement bias or error being introduced into your study through the choice of scale.

VI. Biases

Biases are errors in a study's design, implementation, analysis, or interpretation that are systematic (occurs consistently to favour one group or direction) and skew findings towards a certain direction. It is important to be aware of possible biases to better avoid them, or acknowledge them as limitations when reporting your findings. Two families of biases[26], are commonly recognised, with specific types of biases falling under them:

1. Selection bias arises from elements affecting the study population

a. Sampling bias: when individuals were selected in a way that results in a sample that does not represent the overall population.
b. Attrition bias: when dropouts from the study are systematically different from those who are retained or complete the study/programme. This usually occurs when a certain design feature of the study is not acceptable or accessible to certain demographics of the sample or population.
c. Volunteer bias: when people who volunteer for the study are systematically different (e.g., may already have an interest in the topic) from those who do not choose to participate.
d. Nonresponse bias: when potential participants who are unable or reluctant to participate in a study or respond to a survey are systematically different from those who do, resulting in a non-representative sample.

2. Informational bias when variables in the study are classified or measured inaccurately

a. Recall bias: when one group within the sample is more likely to remember, or remember in more specific terms, details relevant to the study than other groups in the sample. This occurs when groups of individuals have different abilities to recall past events.
b. Interviewer bias: when interviewers inadvertently influence participant judgements by asking questions or reacting differently to different participants.
c. Social desirability bias: when participants respond in a way that they believe is aligned with the beliefs of the researcher, other participants, or social norms.

CHAPTER 3

How to Design an Evaluation

KEY THINGS TO CONSIDER BEFORE CHOOSING A DESIGN

How to choose your sample?

Try your best to make sure that your *participants* are *representative* of your project's intended target audience to reduce the risk of bias and improve the accuracy of your evaluation's results/findings. For example, if you want to understand whether painting classes can support teenagers' mental health, your project should recruit teenage participants to explore whether painting helps them manage aspects of mental health, such as depression and anxiety.

Feasibility

Before deciding on the design and tools for your activity, you are encouraged to consider the following factors:

Number of participants: How many people typically attend your programme, and how many attendees do you expect to participate in your evaluation?

Duration and frequency: How many people typically attend your programme, and how many attendees do you expect to participate in your evaluation?

Intended outcomes: What are the main objectives of your activity, and what are the expected effects?

- Determine the objective(s) and outcome(s) that you want to measure
  - After doing so, you can select the most appropriate tool(s) to measure them
- Example: If your goal is to use singing to facilitate social connection — such as running a group singing programme for stay-at-home caregivers — you might expect the programme to reduce loneliness. In this case, a measurable indicator of that outcome could be participants' self-reported levels of loneliness before and after the programme.

Budget constraints: Budget may influence both *programme delivery* (e.g., the length of the recruitment period, number of study facilitators, and number of sessions) and *evaluation* (e.g., whether the scale used to assess the outcome measure requires licensing fees, or how many data collectors you can bring on board).

These factors may inform the type of design or tools that you would like to use, so it is important to keep them in mind before making any decisions. Along with the above list, the aims of your activity will help guide the design that is most appropriate for your project.

Considerations for feasibility:

- If you are conducting an activity to measure its *effects*, you might consider using a **quantitative** data collection method.
- If you wish to *gather opinions or feedback* regarding the activity, you could consider collecting **qualitative** data.
- If you wish to better understand how the effects of your activity have *affected* participants, you might want to employ a **mixed-methods** design, using quantitative data to measure/compare the effects between groups, while purposefully selecting some participants from your sample to collect qualitative data to gather in-depth perspectives and experiences.

Ethics

Ethics safeguards the rights and wellbeing of participants. Implementing ethical guidelines can ensure that the interests of each party are considered[27] and that ethical principles of fair treatment and practices are upheld. This facilitates *mutual respect*, fosters greater *trust* between stakeholders, and may increase *receptibility* of your project in the wider community.[28]

All data collection should be ethical, but not all evaluations require formal approval from an ethics board (e.g., Institutional Review Board or IRB). For example, evaluations carried out for internal review, or where there is no intention to publish or share the findings publicly, may not need to be submitted to an IRB — but this does not exempt them from ethical responsibility. Ethical practices should always be considered in evaluation and research.

If you are aiming to publicly disseminate or publish your findings, as in research studies, then ethics approval is needed before embarking on any study procedures (e.g., data collection, participant recruitment, etc.). In this case, an application will need to be submitted to the Institutional Review Board (IRB) for approval. Study personnel will also need to go through ethics courses to be certified to conduct research while complying with ethical guidelines. The Collaborative Institutional Training Initiative (CITI) program[29] is often used for ethics training in research.

I. Key ethical guidelines to consider:[30]

A. Informed Consent[31]: Inform participants about your project's objectives, the types of data collected, and any benefits or risks involved.[30] Ensure that participants have the freedom to choose to participate: Participants must not face repercussions if they decline or withdraw their participation at any point during the study.[32] It is strongly recommended to document consent obtained (i.e., with a copy of the consent form signed by the participant and researcher).

B. Data Confidentiality:

1. To safeguard participants' privacy, *de-identify* personal data from research data. Personal information such as names, contact details and health records can be removed/separated from research data collected.

Suggestion: An alpha-numeric code can be assigned to the data collected from each participant, so that anyone looking at the data will not be able to identify who it belongs to *(e.g., data obtained from the first participant of a study can be coded as P001; data from the second participant can be coded as P002, etc.)*

2. If personal data needs to be retained, provide justifications and ensure that personal data is handled responsibly and stored securely.

Suggestion: Sensitive data can be password-protected and given restricted access *(e.g., only the study's Principal Investigator is given access to personal data)*. It is a good practice to discard all personal data at the earliest possible instance, or once the project has been completed.

3. Data Credibility: Be transparent and comprehensive when reporting your results and limitations, and avoid overstating your findings. While best practice is to use blinding (a process where group assignment is concealed (i.e., hidden) from participants or study team members delivering the intervention, collecting data, or analysing the results), this is often not practical in Arts & Health projects. Instead, focus on clear, honest reporting to make sure your findings are trustworthy and meaningful.

II. Key Arts and Health Ethical Considerations:

Person-centred: Projects should be catered to the target audience.[33]

Equity: Ensure fairness, accessibility and inclusiveness for all parties involved, including participants, project staff and external collaborators.

Safety: Safeguard the wellbeing of participants and project staff.

Evidence-based practice: Projects should be supported with evidence (i.e., findings from other similar or previous projects) to maximise benefits, minimise risks, and prevent unnecessary use of resources or unwanted participant burden.

Culturally relevant: Consider the lived experiences, values, and expressive forms of the communities involved, ensuring that methods, analysis, and outcomes reflect and respect the cultural context.[34]

TYPES OF DESIGNS

How to choose?

When selecting a design, consider the purpose of your activity/programme and what you hope to achieve. The following decision tree may help guide your choice of study design to evaluate your programme.

Recommendations for Selecting an Evaluation Design

Figure 3.1 Decision tree for choosing a study design

WHAT IS AN EXPERIMENTAL DESIGN?

Experimental design refers to the various ways that experiments, or studies, are conducted to test hypotheses.

I. What is a Pure Experimental Design?

A pure experimental design examines cause and effect by changing one factor at a time while keeping everything else the same. The two key features of pure experimental designs are the use of a control group and random assignment of participants into study groups (experimental/intervention and control). One example of a pure experimental design is the **Randomised Controlled Trial** (see info box on the right).

A. Control Group

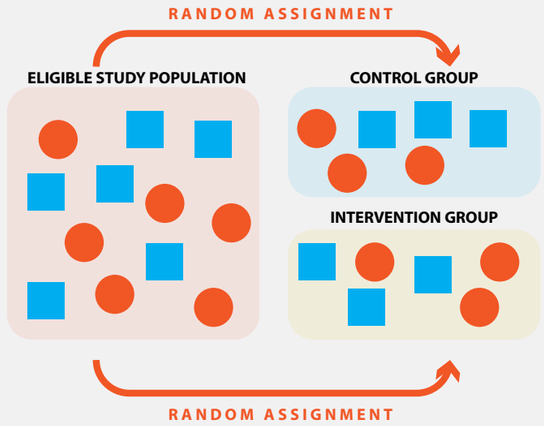
Control groups *do not receive the intervention and serve as a reference for comparison* with the intervention group. Broadly, there are two types of control groups (more details can be found in the Cochrane Handbook):

1. Passive control groups: May include wait-list controls (who do not receive the intervention immediately and are placed on a waiting list), groups that receive a placebo (i.e., an agent or procedure that has no therapeutic effect), or groups that do not receive the intervention at all.[35,36,37]
2. Active control groups: Involves a control group that engages participants in an activity or intervention that has been shown to be effective. Active controls are often used when it is unethical to withhold treatment or assign participants to a passive control group.[35,36]

Control participants should ideally share similar characteristics with the intervention group (e.g., age, gender, culture, geographical location, socioeconomic status, etc) to make both groups comparable, and ensure that any outcome difference obtained can be attributed to the intervention given.[38] Key characteristics between control and intervention groups can be matched to help make both groups comparable (e.g., ensure participants in both groups are 40 to 50 years old).[38]

B. Randomisation

Randomisation (also known as random assignment or allocation) can be done to randomly assign participants to control and intervention groups to balance out participant characteristics between groups and better establish causal effects of the intervention.[39]



Randomisation can be done at an **individual or cluster level**. At the individual level, randomisation can be done adaptively (i.e., participants are randomised as they enrol into the study) or separately (i.e., randomisation sequences/arrangements are decided beforehand) from participant enrolment.

Three common methods for individual-level randomisation are:

Simple random assignment: Use random number generators/tables that are computer-generated (e.g., from online sources) or use manual randomisation methods (e.g., tossing a coin or rolling a dice).

Stratified randomisation: Group individuals based on certain characteristics (e.g., gender or age category) before randomising individuals within each group.

Block randomisation: Randomly allocate individuals into blocks with an equal number of participants, before randomising within each block (note: it is recommended to have at least four participants in each block).[39,40]

Randomised controlled trials (RCTs) are often considered the *'gold-standard' in primary research for establishing causal relationships* due to its use of *randomisation* and *blinding*.[41,42] Clinical trials commonly use an RCT design. The CONSORT (Consolidated Standards Of Reporting Trials) statement is a checklist that consists of 30 criteria that help researchers formulate an RCT experimental design.[7,43,44] More information on the CONSORT 2025 statement can be found in the following resources:

SPIRIT (Standard Protocol Items: Recommendations for Interventional Trials)[45]
CONSORT 2025 statement: updated guideline for reporting randomised trials[46]

However, RCTs can require more time and financial resources to carry out. Therefore, we encourage you to keep these ethical and logistical considerations in mind if you want to implement RCTs:

- Justifications are needed for any risks or benefits that may be relevant for the intervention group compared to the control group. Reduce risks wherever possible. Ensure anonymity and confidentiality in data storage and reporting.
- Blinding is recommended for best practice. Ideally, 1) participants do not know whether they are or are not receiving the intervention, and 2) researchers collect data and analyse results without knowing which participants were in the experimental or control group.
- Group assignments should be impartial and random. Avoid being influenced by participants' preference for being in either experimental or control group.
- Experimental and control groups *should differ by the intervention participants receive.* For example, if the experimental group receives musical engagement as the intervention, then the control group should avoid engaging in music-related activities.
- Findings may be less generalisable or applicable to 'real-world' situations, which do not have the same artificial controls as the study.

Alternatively, **cluster randomisation** is a method for randomising at a broader level (i.e., by clusters), based on location, group or facility. This is used primarily in pragmatic trials (see infobox below) where it is difficult to conduct separate interventions within the same site, or when control group participants on site may be exposed to the intervention or experimental activity. A 2023 study by Clifford and colleagues is a recent example of a cluster-randomised pragmatic trial in arts and health.[47]

Pragmatic Trials

Pragmatic trials are often carried out in routine-practice settings to achieve real-world applicability.[49] Compared to RCTs, they can be more appropriate for arts and health projects that focus on evaluating impact in real-world conditions where randomisation may not always be feasible.[44,50]

The main pros and cons of pragmatic trials are outlined in the following diagram:[51,52,53]

PROS

1. Prioritises generalisability (more external validity) and applicability to real-world contexts.
2. Focus on effectiveness of interventions/ programs-more relevant to funders, policymakers, communities

CONS

1. Usually open-label (lack of blinding); subjected to biases-internal validity compromised
2. May require larger sample size to even-out heterogeneity in a study population (that is due to the pragmatic nature of the study design)

II. What is a Quasi-experimental design?

Just like experimental designs, **quasi-experimental designs** aim to test cause and effect, but with less control in deciding who goes into which group. As a result, participants are not randomly assigned to groups.[48] Groups tend to be pre-existing in quasi-experimental designs. In other words, participants are pre-assigned to groups based on non-random criteria (i.e., existing variables that cannot be manipulated). Control (or comparison) groups may or may not be utilised. Other methods are used to ensure acceptable levels of evidence are obtained, for example, through the use of pre-post tests.

To demonstrate the various quasi-experimental designs, let's consider a hypothetical scenario: You are interested in studying whether dancing can reduce symptoms of anxiety in community-dwelling older adults. Here is what a quasi-experimental design could look like given your study objective:

Quasi-experimental Study

Example: Can dancing reduce symptoms of anxiety in community-dwelling older adults?

The study observes two groups of people: a control group consisting of older adults with no symptoms of anxiety, and a 'treatment' group consisting of older adults with symptoms of anxiety. The results from both groups are taken and compared.

In this scenario, participants are pre-assigned to their groups as it is not possible for the experimenters to control whether a participant suffers from symptoms of anxiety (i.e., the experimenter cannot decide which participant goes to which group). Such a factor is known as a quasi-independent variable.

Randomisation is thus not possible. Accordingly, the study becomes a **quasi-experimental study**.

A. Single group pre-post

Single group pre-post tests only have a single group of participants exposed to the intervention with no second group for comparison. Data is collected from the same group of participants before (pre-test) and after (post-test) the intervention. This allows for comparisons between pre-test and post-test scores. A study by Kaimal and colleagues[54] is an example of this, where the study only implemented a single session of visual art making for participants to observe the effects of visual art making on cortisol levels, a common marker for stress. Participants' saliva samples (measuring cortisol present in saliva) before (pre-test) and after (post-test) the art-making session were collected.

However, with only a single group of participants and no other comparison group, it is hard to determine whether any changes observed are solely attributed to the intervention or due to other unaccounted and/or uncontrolled factors. If circumstances and resources allow, a multiple group pre-post design would be recommended instead.

Continuing with our hypothetical scenario, this is what a single group pre-post test design could look like to study the effects of dancing on anxiety symptoms:

Single Group Pre-post Test Design

Example: Can dancing reduce symptoms of anxiety in community-dwelling older adults?

In this scenario, a single group of older adults experiencing symptoms of anxiety would be observed.

To implement a pre-post design, a questionnaire to measure anxiety would be administered once before the intervention (for instance, weekly dancing sessions l asting 6 months) and once more after the intervention (at the end of the 6 months of weekly dancing sessions). The anxiety scores of participants would then be compared to find out the change from pre-intervention to post-intervention.

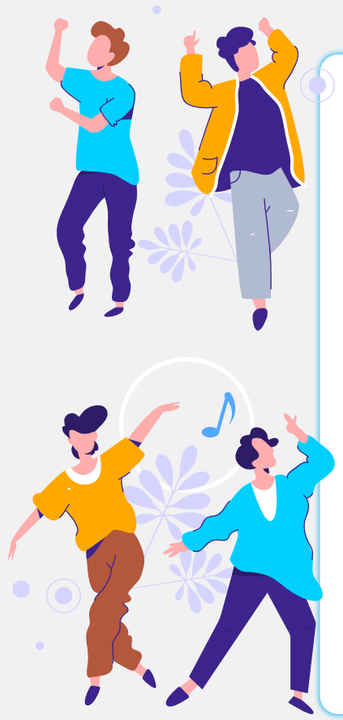
As randomisation is still not employed, the study is thus a **quasi-experimental study** that applies a **one-group pre-post test design**.

B. Multiple group pre-post

In a **multiple group pre-post design**, data is gathered from at least two groups of participants before (pre-test) and after (post-test) the intervention has been implemented. Compared to the single group pre-post study design, the multiple group pre-post design uses comparison groups, which can provide better causal attributions to the intervention if any changes are observed.

For instance, a two-group pre-post design allows for the comparison of pre-post scores between two different groups. A study by Mouriño-Ruiz and colleagues is an example of this.[55] The study's intervention group exposed participants to art-based interventions, while the comparison group was not exposed. Ten consecutive two-hour weekly sessions were held for the intervention group, and loneliness levels among participants in both intervention and comparison groups were measured before the intervention in April (pre-test) and after the intervention in June (post-test).[55]

Using our hypothetical study objective once more, here is how a multiple group pre-post test design could be employed to test the effects of dancing on reducing anxiety:



Multiple Group Pre-post Test Design
Example: Can dancing reduce symptoms of anxiety in community-dwelling older adults?

In this scenario, two groups are recruited: one group consisting of older adults with symptoms of anxiety and the other group consisting of older adults without symptoms of anxiety.

To implement a pre-post design, a questionnaire to measure anxiety would be administered to both groups, once before the intervention (weekly dancing sessions lasting 6 months) and once more after the intervention (at the end of the 6 months of weekly dancing sessions). The anxiety scores of participants would then be compared to find out the change from pre-intervention to post-intervention. This type of study design allows for comparison of scores between the two groups of participants, in addition to comparing within the individual groups.

The study is thus a **quasi-experimental study** that employs a **two-group pre-post test design**.

WHAT ARE QUALITATIVE STUDY DESIGNS?

Qualitative studies can be structured as a single study, relying only on qualitative data, or can be incorporated into a mixed-methods study that collects both qualitative and quantitative data. Often, we use qualitative methods to understand something (concepts, thoughts, experiences) while we use quantitative methods to confirm or test something (a theory or hypothesis).

Common qualitative study designs that may be used for arts and health interventions include case studies and descriptive qualitative studies (see infoboxes below). Note that it is also common to combine qualitative data collection methods, such as open-ended surveys, interviews, or focus group discussions (FGDs), with quantitative designs for a mixed-method approach, to obtain a richer understanding of the quantitative results and any nuances that may be present.

I. Examples of qualitative study designs

A. Case Studies

Case studies involve an in-depth examination of a single case or multiple cases (e.g., an individual, group, or programme) in real-life situations, providing detailed descriptions of how an Arts & Health intervention was delivered (e.g., for process evaluations) and experienced, within the broader context of its environment, the way activities unfolded, the lived experiences of participants, and the surrounding circumstances. When multiple cases are examined, comparisons can be made to identify similarities and differences. Case studies are commonly used to explore how an intervention is implemented, including the nuanced experiences of those who receive it, and to understand rare conditions or exceptional phenomena that are not easily generalised. You can find out more about case studies in this brief summary by Yin[56].

Case Study
Example: How does drawing reduce the anxiety of children affected by a traffic accident?

Two children who have experienced a traffic accident are recruited to participate in a 3-month drawing intervention. Data would be collected through interviews, observations and artworks at multiple time points (before, during, and after the drawing intervention) to explore the changes in emotional states throughout the intervention.

Given the **rarity** of such a study population, the case study approach is well suited to provide an **in-depth and contextualised understanding** of how art activities might help children reduce anxiety and express their emotions.

B. Qualitative Descriptive Studies

Qualitative descriptive studies aim to provide a comprehensive and straightforward summary of a phenomenon or experience, closely reflecting participants' own words and perspectives. They tend to involve more participants than case studies and are commonly used to identify shared patterns and elements across participants. Villamin and colleagues' paper[57] offers a helpful step-by-step guide for conducting qualitative descriptive studies.

Qualitative Descriptive Study

Example: How does drawing help children who have a sibling diagnosed with a chronic disease cope with anxiety?

16 children who have a sibling diagnosed with a chronic disease are recruited to participate in a 3-month drawing intervention. Interviews would be conducted at the end of the intervention to gain insights into their experiences with the drawing activities, including how drawing may have helped them manage and cope with anxiety.

Unlike a case study, which focuses on in-depth understanding of a small number of participants, this descriptive study explores the **broader experiences** shared by **a larger group** of children.

II. Implementing qualitative designs

Just like quantitative study designs, qualitative study designs have ethical and logistical considerations that we encourage you to keep in mind.

A. Ethical considerations

1. Anonymity and confidentiality: participants can still be identified after de-identification especially in small sample sizes. Generalising identifiable information can reduce the risk of dataset re-identification (e.g., changing specific profession to industry sector name). For more anonymisation techniques, refer to the Personal Data Protection Commission's Guide on Basic Anonymisation[58].

2. Misrepresentation, objectification and researcher's subjectivity: findings can be misrepresented when artistic value is prioritised over authenticity, especially when they are presented artistically. Researchers' personal beliefs, background and training may also influence how they design, conduct and present their study. It is important to engage in:

Reflexive practice: regularly reflect on how your personal perspective might shape the research in all stages of the research process.

Member checking: check your findings with participants to make sure their views are represented accurately and to strengthen the credibility of the research.

B. Logistical considerations

1. Time-intensive and deliberate processes: gaining ethical approval, converting speech into text, conducting member checking and analyses.

2. Rapport building: working with vulnerable populations (e.g., marginalised groups, children) or those working in sensitive contexts (e.g., healthcare facilities, crisis centres) requires experience/training and time.

3. Data privacy/protection: tools (e.g., image-blurring and voice modification software, institutionally supported cloud storage, encryption) may be required to protect participant confidentiality.

CHAPTER 4

Data Collection

DATA COLLECTION

A study's integrity depends on accurate data collection. Planning this process will help you collect accurate data, standardise practices across the study team, and mitigate biases, which all help to strengthen the findings of your evaluation.

What information can be collected for an evaluation?

Different types of data can be collected based on the study design and intended outcomes. Sources of information when doing an evaluation can include:

1. Existing information, such as local census data or public records
2. People, such as participants, funders, or collaborators
3. Pictorial records and observations, such as photos or video recordings

In this chapter, we cover commonly-used *quantitative* (focusing on numbers) and *qualitative* (focusing on text) data collection methods, with an emphasis on the former. The recommendations here are not mutually exclusive or exhaustive. For instance, methods such as observation and interview can be part of both quantitative and qualitative data collection.

Other methodologies, such as *Art-based methods* and *Participatory Action Research*, may be used to complement qualitative and quantitative methodologies.

- **Arts-based methods** involve creative self-expression such as story-telling, painting and drawing, photographic methods, film, dance, performance, and music in a research investigation.[59]
- **Participatory Action Research** involves a well-organised, structured partnership with participants who are not just involved in the program but also engaged in the study's data collection and analysis.[60]

What tools to use for an evaluation?

When planning a study, there are many important decisions you have to make, including which scales or tools to use for measuring the outcomes you aim to evaluate. Your choice of scale should be aligned with the nature of your project and its goals to ensure that the evaluation is logistically feasible. Further, choosing an inappropriate scale may lead to the potential benefits of your project being undetected (no significant findings). Some common considerations for choosing a scale are listed in the table on the next page:

Ask yourself the following questions when selecting a scale for research or evaluation. Keeping these in mind can help you plan your evaluation and increase the scientific rigour of your study!

Things to consider when choosing your scale	Questions you can use to check if the scale is right for your evaluation
Research/Evaluation	Does the scale measure what your programme intends to influence?
Recall Period	How far back does the scale require participants to remember? Does the scale cover the appropriate time frame you want to evaluate?
Training Requirement	Can the scale be self-administered? Do you need training to administer the scale?
Length/Time	How much time does it take to complete the scale?
Cost/Licensing	Do you need permission to use the scale? How much does it cost to use the scale?
Validity*	Is the scale validated for use in a population similar to your target demographic?
Longitudinal Validity*	Can the scale measure minimal important differences over time?
Sensitivity*	Can the scale measure subtle differences in the outcome?

*Scales recommended in Section 2 of this toolkit satisfy these criteria. However, it is important to consider these criteria if you are considering using scales not presented in the toolkit.

Data collection plan: The Who, When, Where, and How of collecting data

The following are some common questions that you may consider before implementing data collection:

Who will collect the data?

- Are you required to train personnel for collecting data?
- Is there training provided for people involved in data collection?
- Is there any relationship between the study/programme's team members and the participants?
- Any supervision and monitoring given to the team of data collectors to ensure quality?

When will the data be collected?

- Any time (Daytime, evening, weekdays or weekends)? Would the time chosen affect the response rate?
- Are there any seasonal variations which can affect the response?

Where will the data be collected?

- Which location is best suited for data location? You may want to consider aspects such as logistics, privacy, and efficiency.
- Will the presence of others influence how participants respond?

How will the data be collected?

- Select the most feasible data collection tool, e.g., validated questionnaire or open-ended survey
- How much time and funding are required to complete data collection, taking into account logistical considerations (manpower, venue, costs, etc.)?
- Are there additional measures required to align with the institutional or ethical guidelines?

Before implementing your plan, you should also consider what could go wrong. For example, if participants are in a hurry to leave a concert venue after the performance, would they be willing to complete your questionnaire? Are your participants able to understand the questions? Revisit your plans based on these practical considerations.

CONDUCTING QUANTITATIVE DATA COLLECTION

For quantitative studies, questionnaires are commonly used, and physiological measures, which are used to evaluate the physical or biological status of participants, have also been gaining popularity. More information on physiological measures will be covered in Chapter 8.

Validated scales for your main outcomes of interest

Demographic information is often collected, such as age/year of birth/age group, gender, ethnicity, housing, race, religion, occupation, education level, etc

Bespoke or self-created questionnaires are customised surveys designed specifically for your programme. They help you collect information tailored to your study aims — especially when you want to understand how particular parts of your programme contribute to certain outcomes. These surveys can complement validated scales by capturing data that may fall outside the scope of validated scales (see box to the right for more information).

More information on Bespoke Questionnaires

For example, if you are running a singing and dancing programme for older adults to improve social connection (primary outcome of interest), a bespoke questionnaire can help you explore secondary outcomes of interest (such as mood).

Programme: Singing and dancing for older adults
Secondary outcome: Mood, empowerment
How each tool helps you collect data:

Tool	What it measures
Validated scales	Overall change in social connection after taking part in the programme.
Bespoke questionnaire	How specific components of the the programme (e.g., singing or dancing) impacted social connection (primary outcome), or influenced other/secondary outcomes of interest, such as mood, empowerment, etc.

Tips for creating a good bespoke questionnaire:

- **Keep questions clear and neutral.** Avoid using biased language or leading questions that push participants toward a certain answer.
- **Pilot test your questions.** Try out the survey with a small group first and ask if the questions make sense or if anything feels unclear. Improve the survey based on the feedback before implementing it.

We recommend using a scale that has been validated where possible (see Chapter 2). However, data collection forms are usually created by researchers to collect the specific demographic data that may be relevant to the study. Bespoke scales may also be used to gather a rich, holistic understanding of participants' experiences, but we do not recommend using them as your primary quantitative outcome measurement tool.

Self-report questionnaires such as those mentioned above are generally easy to administer. However, it is important to be aware of the potential for response biases, mitigate their influence where possible, and acknowledge their presence in your reporting. See the table at the end of this chapter for some helpful ways to mitigate response biases.

CONDUCTING QUALITATIVE DATA COLLECTION

For qualitative studies, we can use participants' feedback, interviews, questionnaire responses, databases and reflective observation as data collection strategies.

I. Observation: direct observation of participants' behaviours, characteristics and other scenarios using different senses (visual, auditory, tactile, smell, etc). Note that while observations can also be quantitative, this approach is beyond the scope of this toolkit, which promotes the use of validated scales for quantitative data collection.

II. Interviews and focus group discussions (FGD)

A. Designing an interview guide can involve 4 phases

1. Align interview questions with study interest
2. Construct an enquiry-based conversation
3. Best practice: Receive feedback on interview guide
4. Best practice (optional but recommended): Pilot the interview guide

B. Considerations for focus group discussions/ interviews if you are an interviewer

1. Be aware of **groupthink**, which happens when a group simply goes with the flow and reaches a consensus with unchallenged discussion. This can lead to a lack of critical thinking, original inputs, and evaluation of alternatives.
    - As the moderator, you can play devil's advocate and encourage discussion in a new direction by providing an alternative view to explore different perspectives.
    - Further, to help prevent groupthink, you can consider having individual short engagement activity for participants to jot down their personal feedback, views and reflections prior to a group discussion to allow independent thinking.
2. Be aware of **power dynamics** among FGD participants, which can steer opinions to agree when not all hold the same view.
    - Make it a point to request input from every participant
    - Validate everyone's input with an encouragement
3. **Interviewer bias** occurs when the interviewer responds in a way that influences how participants respond. This can happen due to the interviewer's tone of voice, body language, leading questions, facial expressions, or reactions to participants' answers. Consequently, these subtle (or not-so-subtle) cues may prompt participants to say what they think the interviewer wants to hear, rather than sharing their true thoughts or experiences.
    - As the interviewer, try to keep calm and be present during the interview process to maintain a non-judgemental posture by being mindful of the verbal and non-verbal language you display.

To prevent or mitigate groupthink, power dynamics and interviewer bias, it is essential for one to be a neutral moderator to establish an atmosphere that enables open dialogue and dissent. To find out more about considerations when you are asking interview questions, refer to the twelve tips by McGrath and colleagues[61].

III. Qualitative surveys: Based on the Deakin University Library Guide on Qualitative Study Design, qualitative surveys commonly utilise open-ended questions to reveal opinions, experiences, narratives or accounts, usually as a precursor to interviews or focus group discussions for identifying initial themes or issues to explore further in research.

Best Practices for Designing Open-Ended Survey Questions (for more information, see Research Best Practices from Dig Insights)

- Use simple, clear language
- Avoid 'two-in-one' questions (e.g., How much do you enjoy your art session and working with the team members?) and instead ask one thing at a time
- Avoid double negatives (e.g., Did the art program fail to reduce your anxiety symptoms?)
- Use open-ended questions (e.g., How did you feel about the art program?)
- Avoid leading questions that may lead to biased responses (e.g., How helpful was the art program instructor?)
- Frame sensitive issues tactfully (e.g., Did the art program play a role in initiating your talk with your parents about their suspected onset of dementia? If the response is yes, ask "Could you share more about how these discussions begin?")
- Ensure the order of questions makes sense to participants logically

Find out more information about constructing online surveys from Thomas and colleagues[62].

COMBAT MEASURES OF RESPONSE BIASES IN DATA COLLECTION

All data collection methods, including those involving self-reported data (e.g., self-response questionnaires), documentation of behaviour observed, and interviews, are prone to bias. Taking measures to ensure respondents' anonymity is important to reduce response biases by assuring respondents feel safe and empowered to share openly. In addition, when reporting the outcomes of any research or evaluation, it is essential to acknowledge the limitations imposed by sources of bias inherent from the data collection method, and any measures and considerations in place to mitigate these response biases, for a critical evaluation of outcomes.

Here are some ways to reduce common biases (See Chapter 2 for more information):

Response Bias	Combat Measures
Social desirability bias (participants may want to look favourable to others)	For qualitative methods (e.g., focus-group discussion): • Create a comfortable atmosphere that facilitates frankness For quantitative methods (e.g., randomised controlled trial) • Use participant codes to ensure anonymity
Recall bias (e.g., when one group of participants is more likely to remember and provide relevant information than another)	For both quantitative and qualitative methods: • Rely on objective pre-existing records in addition to participant experiences, if possible • Ask participants to recall events over shorter periods (e.g., days or week) • Use standardised questionnaires and prompts for memory aid to assist recall
Extreme response bias (e.g., participants tend to select/ answer with the most extreme options even when they do not adopt an extreme opinion)	• Qualitative: use more sensitive words in the response set • Quantitative: use validated scales with clear questions and distinctive response options • Allow adequate time for participants to respond • Avoid pressuring them to complete the questionnaire within a stipulated time
Agreement bias (e.g., participants tend to select a positive response option without reflecting their true position)	• In Yes/No questions, especially in qualitative approaches when participants are asked whether they agree to a statement, make sure to be neutral in crafting questions and response options. - Avoid leading questions (e.g., "You enjoyed the programme today, right?") - If you provide a five-point scale, it is good practice to ensure that the response options are balanced from negative to positive, such as "strongly disagree," "disagree," "neutral," "agree," and "strongly agree". o E.g., if you want participants to rate how well they enjoyed an arts programme, '1' can represent "very little" and '5' can represent "very much". This is more likely to elicit more accurate, thoughtful responses. • As an interviewer, you may want to read the response options aloud to your participant and clarify what the values mean along the continuum of a scale. If you are using a scale that measures severity, it is good practice to explicitly highlight what each end of the spectrum represents. - E.g., "How would you rate your anxiety on a scale of 0 to 10, where '0' means no anxiety and '10' means the most anxiety you can imagine?" • Include open-ended questions, if you are using qualitative methods. • Keep questionnaires brief. Having a long list of questions can lead to decision fatigue as respondents become less focused and have the tendency for making quick decisions.

INTRODUCTION TO

Section 2

In Section 2, we will cover four common sets of outcome measures evaluated within the context of Arts and Health, including measurements for mental health, wellbeing (and related measures), social connection, and physical health. In each chapter, we provide an overview of the outcome(s) in focus, discuss quantitative measurement tools, and offer our recommendations based on scientific rigour and practicality.

HOW CAN WE MEASURE OUTCOMES QUANTITATIVELY?

We cover two common ways to measure the outcomes of a program using quantitative methods: using validated scales (i.e., questionnaires) and physiological measurements (e.g., heart rate or blood pressure measurements).

Both methods are valuable in their own ways. Although many people consider physiological measures to be more objective, aspects of health like mental wellbeing are subjective and not easily captured through physical indicators. In fact, physiological measures can complement validated scales by providing a more complete picture of the effects a programme has on the participants. For example, a validated scale might show that someone feels less stressed, and a physiological measure might show their heart rate has lowered — together, these give a fuller picture of their condition. Therefore, it is important to understand what these different tools aim to measure, in order to select or combine them to best evaluate your study outcome(s).

WHAT ARE THE TYPES OF VALIDATED SCALES INCLUDED IN THIS TOOLKIT?

To identify relevant scales for each set of health outcomes discussed in Section 2, we conducted several literature reviews (see the Appendix for further details on the methods used). These reviews helped us identify the most prevalent scales that met our predefined inclusion criteria (see infobox on the right). We then assessed these scales for their appropriateness for use in Arts and Health evaluations. Additionally, we provide practical guidance on how the scope and goals of your programme, along with the available budget and time allocated for data collection, may influence your scale selection, as scales vary in length and recall period (i.e., the time frame that respondents are asked to consider when answering the items in a scale).

To ensure scales are applicable to arts practitioners, we adopted the following "inclusion criteria" (required attributes for scales we recommend in Section 2):

- Do NOT require extensive training to administer (i.e., they do not require medically trained personnel), and in many cases can be *self-administered*
- Display adequate *validity* and *reliability* scores when used in English* and applicability in community and/or general populations.
- Are *responsive* and are able to detect minimal important differences, over their prescribed recall period (where applicable, e.g., not all scales have a recall period).

***Note:** Most of the recommended scales in this toolkit have been validated in Mandarin, as well as other languages such as Malay and Tamil. To check the scales' language availability, see the Appendix.

In addition to the considerations discussed under Data Collection in Chapter 4, we also recommend keeping the following factors in mind when selecting scales for the evaluation of your activity/programme: 1) Consider whether the interval at which you are collecting data/administering scales is sufficiently spaced apart according to the scale's recall period or other appropriate timing (e.g., during office hours versus after), and 2) check whether the scale is under copyright and requires payment for use (or is free but requires permission). To help guide your selection of the most appropriate scale for your evaluation, we include a summary table for each outcome that highlights key features and practical considerations for each recommended scale.

WHAT ARE THE TYPES OF PHYSIOLOGICAL MEASURES INCLUDED IN THIS TOOLKIT?

To identify tools that are typically used to assess different aspects of physical health in research, we conducted a broad literature review of physiological measures. Chapter 8 introduces a selection of more common physiological measures that can be used for evaluation of health-related outcomes. We include both relatively accessible tools and more advanced methods that can offer valuable insights into how arts activities affect health.

Because physiological measures are often more resource-intensive than scales, we primarily focus on non-invasive techniques (i.e., external methods that do not involve breaking the skin barrier), and those that do not require extensive training. Nevertheless, we also include relevant measures that are more appropriate to be undertaken in collaboration with domain experts. Where possible, we do recommend working with those experienced in collecting and analysing physiological data, as both processes often require specialised knowledge or training.

We hope this overview of common, relevant physiological measures helps Arts and Health practitioners become more aware of available options and better understand which approaches may be appropriate when evaluating the outcomes of their work.

CHAPTER 5

Measures of Mental Health Symptoms

Mental health is a dynamic state influenced by an interplay of individual, social, and structural factors that shift in response to life's changing circumstances. In recent years, mental health has been increasingly emphasised as a core component of overall health,[63] and it is now widely recognised that good mental health is not simply the absence of ailments and disorders, but should be characterised by a sense of flourishing and the capacity to adapt to and cope with life's challenges. In interventions that aim to improve mental health, including those in the field of Arts and Health, two of the most commonly assessed outcomes are symptoms of depression and anxiety.

DEPRESSION

What is Depression?

Depression is a persistent mental state that includes low mood and the inability to feel interest and pleasure (or the lack of these feelings; also known as anhedonia).[64] The symptoms of depression can be categorised into psychological and physical/behavioural. A list of symptoms can be found in Figure 5.1 below.

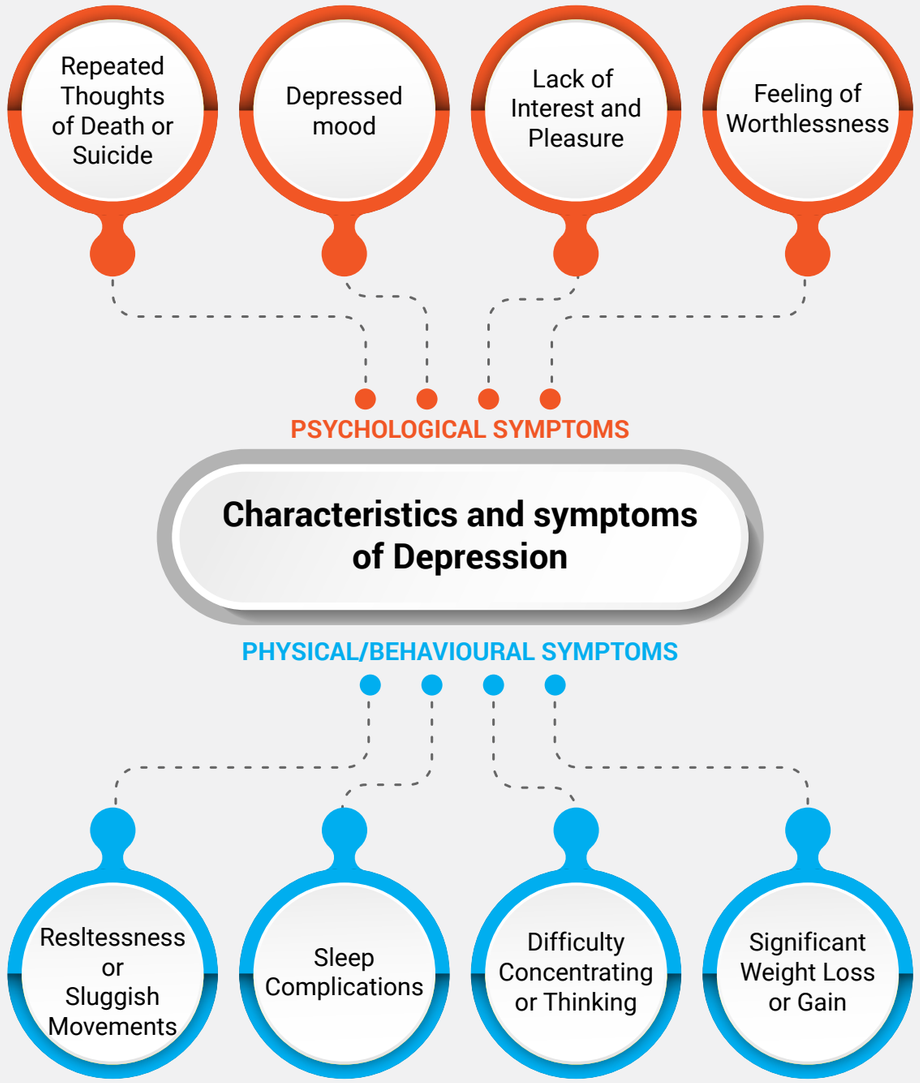


Figure 5.1 Characteristics and Symptoms of Depression

TOOLS FOR MEASURING DEPRESSIVE SYMPTOMS

Most standard questionnaires do not examine all the symptoms of depression listed above. Hence, it is important to first identify which specific symptoms your programme intends to influence. For instance, if you are planning to measure depressive symptoms as an outcome, first consider which symptoms your activity is likely to affect (e.g., low mood, sleep issues, loss of interest, etc.), and then select a questionnaire/scale that is capable of measuring these effects.

In our literature review, the 5 most popular scales measuring depressive symptoms were (in descending order):

1. Beck Depression Inventory (BDI)[65]
2. Center for Epidemiologic Studies Depression Scale (CES-D)[66]
3. Hospital Anxiety and Depression Scale (HADS)[67]
4. Patient Health Questionnaire-9 (PHQ-9)[68]
5. Depression, Anxiety, and Stress Scale-21 (DASS-21)[69]

If your programme…	You might prefer…
• Is large-scale with adequate time and funding • Seeks to measure severity of symptoms	**Beck Depression Inventory (BDI)** • Wide score range allows examination of small differences and to gain an understanding of severity
• Seeks to measure both depression and wellbeing	**Center for Epidemiologic Studies Depression Scale (CES-D)** • Developed to measure the emotional/affective/mood symptoms of depression present in the general population • Unlike other questionnaires measuring depressive symptoms, CES-D's scores represent a continuum from wellbeing to depression
• Targets the emotional aspects of depression • Has limited time to collect data	**Hospital Anxiety and Depression Scale (HADS) — Depression subscale** • Widely used in medical, occupational, and community settings • HADS' depression subscale focuses on the emotional or non-physical symptoms of depression • Quick to administer (7 items)
• Intends to influence many aspects of depression	**Patient Health Questionnaire-9 (PHQ-9)** • A short and concise screening tool based on the 9 diagnostic criteria (symptoms) of depression
• Intends to impact other aspects of mental health, including anxiety and/or stress • Targets symptoms related to the lack of interest and pleasure, feelings of worthlessness, sluggishness, and depressed mood	**Depression, Anxiety, and Stress Scale-21 (DASS-21) — Depression subscale** • Focuses on these symptoms of depression • Was adapted from a longer 42-item version, originally developed with research in mind to measure emotional states of depression, anxiety, and stress • There is also a 10-question version that is less popular

Figure 5.2 Guide to choosing a scale to measure depressive symptoms

ANXIETY

What is Anxiety?

Anxiety is characterised by the feeling of excessive and uncontrollable worry.[70] Additionally, many other symptoms can occur, ranging from physical (such as restlessness), to autonomic (unconscious functions that maintain the body's internal balance, e.g., heart rate, digestion), and psychological (e.g., irritability). The common (non-exhaustive) symptoms of anxiety are summarised in Figure 5.3 on the next page.

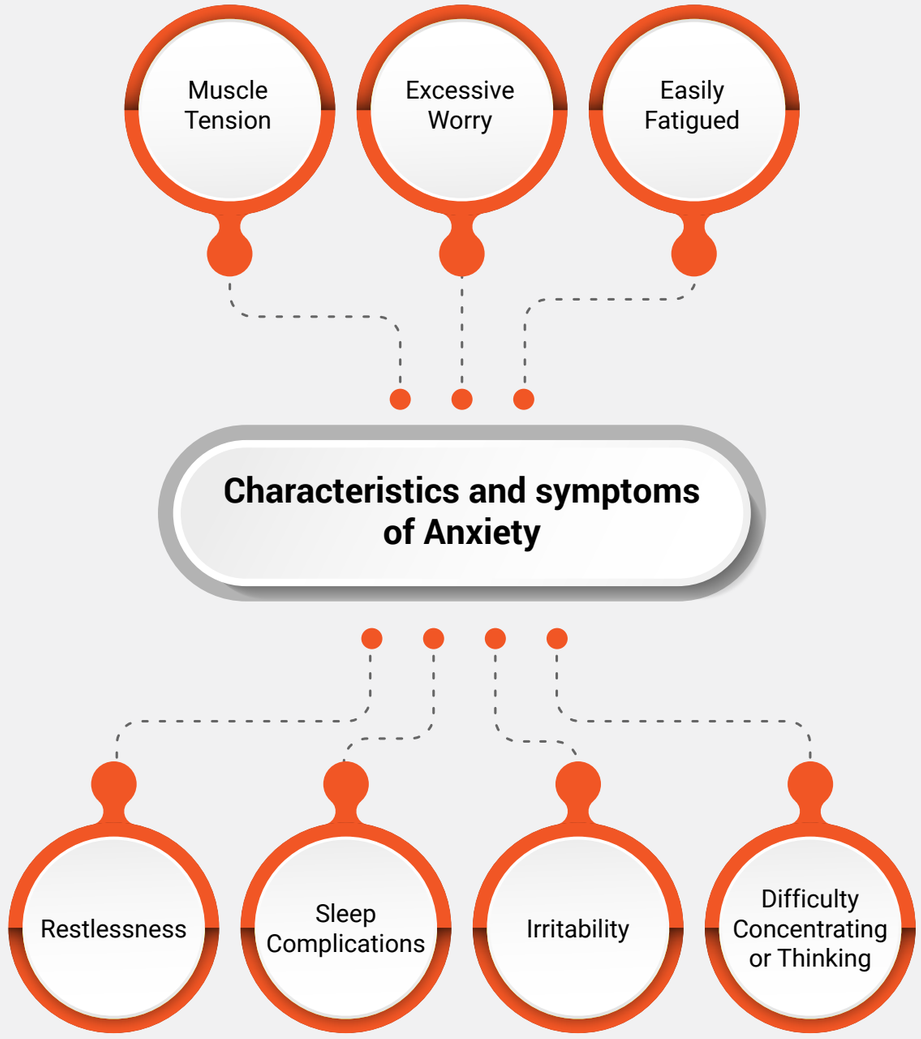
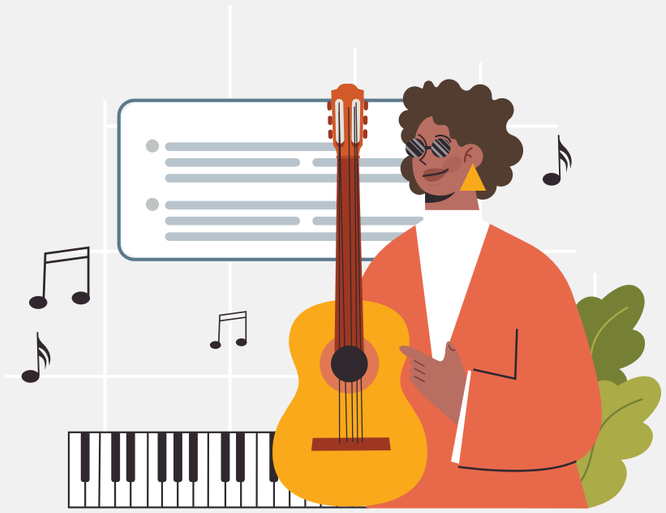


Figure 5.3 Common symptoms of anxiety



TOOLS FOR MEASURING ANXIETY SYMPTOMS

In our literature review, only 4 questionnaires measuring symptoms of anxiety met our inclusion criteria, which were (in descending order):

1. Beck Anxiety Inventory (BAI)[71]
2. Generalized Anxiety Disorder-7 (GAD-7)[72]
3. Hospital Anxiety and Depression Scale (HADS)[67]
4. State-Trait Anxiety Inventory (STAI)[73]

If your programme…	You might prefer…
• Is large-scale with adequate time and funding • Seeks to measure severity of symptoms	**Beck Anxiety Inventory (BAI)** • *Wide score range* allows examination of small differences and to gain an understanding of severity
• Is likely to have an effect on common anxiety symptoms	**Generalized Anxiety Disorder-7 (GAD-7)** • Developed to screen for anxiety and includes many common symptoms. • Focuses on worry and irritability
• Is collecting pre- and post-data between 1 to 2 weeks • Seeks to have an effect on depression and anxiety symptoms	**Hospital Anxiety and Depression Scale (HADS) — Anxiety subscale** • Focuses on the common psychological symptoms and includes symptoms like panic and fright
• Is a single-session or has a duration of under 1 week (but can also be used for longer programs)	**State-Trait Anxiety Inventory State Form (STAI-S)** • "Gold standard" for measuring anxiety symptoms • Measures state or transient anxiety symptoms
• Seeks to distinguish between temporary vs persistent changes	**State-Trait Anxiety Inventory Trait Form (STAI-T)** • "Gold standard" for measuring anxiety symptoms • Measures trait or characteristic anxiety symptoms

Figure 5.4 Guide to choosing a scale to measure anxiety symptoms

SUMMARY OF CHARACTERISTICS OF ALL MENTAL HEALTH QUESTIONNAIRES

The table below provides an overview of the characteristics and noteworthy features of the various mental health questionnaires mentioned in this chapter:

Scale	Noteworthy Features	Time to Administer / No. of Items	Symptoms Measured	Recall Period	Cost
Beck Depression Inventory (BDI)	Widest range of scores to compare subtle differences	~10 minutes / 21 items	All 9 symptoms	The past two weeks	Copyrighted, licence purchase required
Center for Epidemiologic Studies Depression Scale (CES-D)	Places wellbeing and depression on opposite ends of one dimension	~20 minutes / 20 items	Depressed mood, loss of appetite, sleep and psychomotor disturbances, worthlessness	The past week	Free
Hospital Anxiety and Depression Scale (HADS)	Fewest items for both subscales, focuses on specific symptoms of depression and anxiety	~2-5 minutes / 14 items	Depression: anhedonia, psychomotor disturbances Anxiety: worry, panic, restlessness, difficulty relaxing	The past week	Copyrighted, but free for research and clinical use upon request
Patient Health Questionnaire-9 (PHQ-9)	Examines 9 depressive symptoms	~2-5 minutes / 9 items	All 9 symptoms	The past two weeks	Free
Depression, Anxiety, Stress Scale-21 (DASS-21)	Fewest items for both depression and anxiety, more symptoms included than HADS; measures stress in addition to depression and anxiety	~5-10 minutes / 21 items	Anhedonia, depressed mood, hopelessness, worthlessness	The past week	Free for research and clinical
Beck Anxiety Inventory (BAI)	Widest range of scores to compare subtle differences	~5-10 minutes / 21 items	Autonomic symptoms (trembling, nausea, faintness, etc), nervousness, feeling scared, fear of losing control, fear of the worst occurring, fear of death	Past month	Copyrighted, licence purchase required unless permission is obtained
Generalized Anxiety Disorder Scale (GAD-7)	Comprehensive measure of general anxiety with 1 item per symptom	~10-15 minutes / 7 items	Worry, restlessness, irritability, trouble concentrating	Past 2 weeks	Free
State-Trait Anxiety Inventory (STAI)	STAI-S has the shortest recall period, making it suitable for single-session projects	~10-20 minutes / 40 items	Worry, nervousness, tension, stress, apprehension, autonomic activation, lack of calmness and security, lack of contentment and steadiness	State Form: current Trait Form: 2 months	Copyrighted, licence purchase required

Considerations

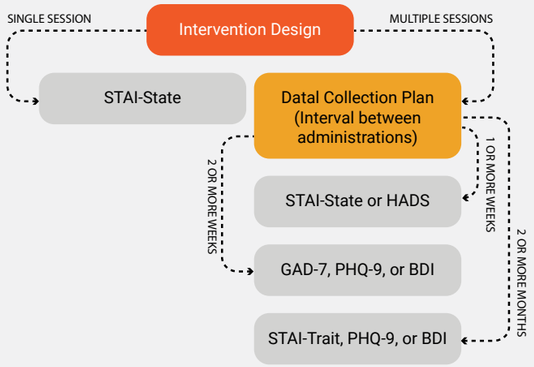
It is important to note that individuals who have been clinically diagnosed with depressive or anxiety disorders are considered to be a clinical population. These individuals should not be considered as similar to or homogenous with general population samples, as they may vary in a range of factors like severity, frequency, cause, or prognosis of their symptoms.[74] This toolkit, however, focuses on the general population, not clinical populations. Furthermore, these scales are used only for screening purposes, while diagnosing individuals must be done in a clinical setting by a trained mental health clinician.

Although these scales are used in the process of screening and diagnosing mental health disorders, they are also frequently used in public health research to monitor the presence of mental health symptoms within the population, and to use these trends to inform policy making. When evaluating mental health outcomes in your projects, you may wish to *exclude individuals who have been diagnosed with a mental health disorder* to ensure that your evaluation findings are not skewed by these clinical factors.

HOW TO CHOOSE AN APPROPRIATE SCALE FOR YOUR EVALUATION?

Many factors may influence the suitability of a scale for your evaluation. Here are some factors you may use to guide your decision.

A key feature that must be taken into consideration is the design or structure of your project or activity. Scales should be administered according to their stated recall period. The figure below may help you to narrow down the available scale options and select the most appropriate scale:



If you have logistical constraints, you may wish to consider the budget and time you have for data collection to help make your decision:

CHAPTER 6

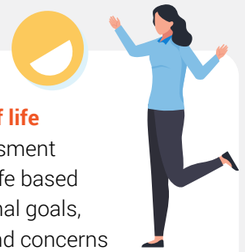
Measures of Wellbeing and Related Outcomes

Many art activities aim to improve the overall wellbeing of participants. Researchers often refer to this concept using three related terms – 'quality of life', 'life satisfaction, or 'wellbeing' – which are all frequently used as outcomes in studies. Broadly, they all refer to improving participants' overall state. It is helpful to consider which one of these three outcomes your activity aims to impact, in order to select the most appropriate questionnaire based on your art activity and study design. In the sections below, we'll explain how each term is used, and recommend widely-used and validated questionnaires for measuring wellbeing.

QUALITY OF LIFE, LIFE SATISFACTION, AND WELLBEING

How are these terms defined?

Quality of life, life satisfaction, and wellbeing are multidimensional constructs with overlapping definitions that broadly refer to how well individuals are faring in their daily lives.



Quality of life
An assessment of one's life based on personal goals, values, and concerns in relation to external conditions.[75]

Life satisfaction
A subjective and cognitive assessment of one's personal life situation compared to an ideal situation.[76]

Wellbeing
Wellbeing refers to how an individual feels and functions both on a personal and social level, and how they evaluate their life overall.[77]

Quality of life (QoL) refers to how one assesses their life, typically based on conventional ideas of success (such as education, housing, income), and is shaped by both health-related factors (e.g., chronic illness, mobility) and non-health-related ones (e.g., culture, norms, beliefs).[75, 78] QoL is examined in many fields, including public health, medicine, rehabilitation, and health economics.

In health economics, measures of QoL are used to calculate 'quality-adjusted life years' (QALYs) in order to account for quality of life as well as life expectancy. The value of a year of life is adjusted on a scale of 0-1 based on individuals' health state during that year, where 0 represents death and 1 represents perfect health. By comparing the cost of different health interventions to the QALYs they generate, researchers can use QALY to evaluate how cost-effectively these interventions promote healthy longevity. For instance, Coulton and colleagues[79] evaluated a community singing program for older adults in Britain by measuring improvements in participants' wellbeing, physical health, and the cost of the intervention, comparing these benefits with a control group performing usual activities.[79] The programme showed significant QoL and QALY benefits, suggesting that it may be a cost-effective option compared to usual activities at various levels of funding. This is an example of how QoL measures can support the case for evidence-based and economically efficient art-based intervention.

Life satisfaction involves subjectively and cognitively assessing one's life situation in comparison to an ideal state using their own criteria and values.[76] Compared to QoL, life satisfaction reflects an individual's overall contentment with their life based on personal values and internal judgment, rather than conventional standards of success or external factors. Studies in psychology, sociology, and economics frequently examine life satisfaction.

Wellbeing refers to how an individual feels, functions, and views their life overall.[77,78,80] It is a broader and more multidimensional concept than QoL or life satisfaction. Psychologists often define wellbeing using two dimensions:

- The affective dimension, typically referred to as *hedonic wellbeing*, focuses on needs, wants, and the experience of emotions and moods—essentially, the pursuit of pleasure and avoidance of pain.
- The functional dimension, typically referred to as *eudaimonic wellbeing*, emphasises a sense of control and self-realisation in the pursuit of meaning, purpose, self-potential, spirituality, and resilience.[81-84]

In addition to these dimensions, public health and policy research, such as work by the Organisation for Economic Co-operation and Development (OECD), often includes physical and social components when examining wellbeing.[84,85]

Arts and Health studies increasingly use quality of life, life satisfaction, and wellbeing as programme outcomes. While these concepts overlap in some ways, their use as outcome measures differs across fields. The table below summarises the key characteristics of each outcome and the fields in which the outcomes are commonly used:

Outcome	Key Characteristics	Primary Disciplines
Quality of life	Focuses on a person's health, physical abilities, and social circumstances	Medicine, Public Health, Nursing, Health Economics
Life satisfaction	Assesses overall life experience based on one's own values and goals	Psychology, Sociology, Economics, Public Policy, Arts and Cultural Studies
Wellbeing	Capture many dimensions, including affective, functional, physical, and social	Psychology, Sociology, Public Policy, Environmental Studies, Arts and Cultural Studies

TOOLS FOR MEASURING QUALITY OF LIFE

There are many different tools available to measure QoL, and they vary in what aspects of life they focus on measuring. QoL scales often include items on physical health, daily functioning, and sometimes social or environmental conditions. By considering the features of different scales, along with practical considerations (i.e., administration time and ease of use), you can select a QoL scale that best aligns with the intended outcomes of your programme.

In our literature review, the 5 most prevalent scales were (in descending order):

1. EuroQol five-dimensional instruments (EQ-5D-3L/5L)[86]

4. WHO Quality of Life-BREF (WHOQOL-BREF)[89]

2. 36-Item Short Form Survey (SF-36)[87]

5. Quality of Life Inventory (QOLI)[90]

3. 12-Item Short Form Survey (SF-12)[88]

If your programme…	You might prefer…
• Has a brief time window for data collection • Intends to identify different health concerns among participants based on the relative importance of different health dimensions	**EuroQol five-dimension instrument (EQ-5D-3L/5L)** • Simple to use with the shortest administration time (among these scales) to assess QoL • Assesses physical and mental functioning using two components: 1. Descriptive system: Assesses 5 key dimensions of health to reflect overall health status with 3 levels of severity 2. EQ-VAS: A 0–100 scale for individuals to rate their overall perceived health[92] • Most common QoL scale used in economic evaluation, with a focus on physical health status[91]
• Is expected to create subtle or incremental changes, especially among participants with similar characteristics • Has ample time for data collection • Is interested in assessing a wide range of health domains of QoL[93]	**36-Item Short Form Survey (SF-36)** • Wide score range for each of 36 items across 8 domains (see summary table on the next page) provides sensitivity in differentiating long-term health benefit [94] • Requires the longest administration time (among these scales) • Offers a comprehensive assessment of physical, mental, and social functioning which considers the interaction of physical health, emotional problems and social relationships
• Aims to consider a wide range of domains for health-related QoL (same as SF-36) with efficiency over depth • Has limited time for data collection	**12-Item Short Form Survey (SF-12)** • Uses fewer items (12) than SF-36 to estimate the same 8 domains of QoL for assessing physical, mental, and social functioning for brevity over a less comprehensive assessment of QoL • Commonly used in large-scale interventions for a snapshot of health-related QoL[95]
• Serves to improve participants' environmental conditions (e.g., physical surroundings, living conditions) in addition to physical and social functioning for improving QoL	**WHO Quality of Life-BREF (WHOQOL-BREF)** • Able to measure subtle changes in 4 domains (physical, psychological, social, environmental) • Favoured in global studies for cross-cultural comparison and community-based arts interventions, especially in resource-limited settings[89]
• Intends to gauge its effects from the participants' points of view • Collects QoL data at time points less than 2 weeks apart, but requires more extensive data than EQ5D	**Quality of Life Inventory (QOLI) (32-item inventory)** • Assessment of the current state of an individual's perceived quality of life (e.g., no minimum recall period) by rating the relative importance and life satisfaction for each of the 16 life domains (including love, work, life, and finances, etc)[96] • Commonly used in psychological or counselling-oriented arts interventions (e.g., art therapy with trauma survivors)

Figure 6.1 Guide to choosing a quality of life scale

The table below provides an overview of the characteristics and noteworthy features of the various quality of life questionnaires mentioned in this chapter:

Scale	Noteworthy Features	Time to Administer / No. of Items	Area(s) of Measurement	Recall Period	Cost
Euro-Qol (EQ-5D-3L/ EQ-5D-5L)	Simple and short for economic evaluation and cost-effectiveness analysis with a focus on physical health status	~2–5 minutes / 5 items	Mobility, self-care, usual activities, pain/ discomfort and anxiety/ depression	Recent (current state)	Copyrighted, but free for research and clinical use upon request
SF-36	More detailed and sensitive than SF-12; wide score range yields high sensitivity; suitable for long-term evaluation	~10–15 minutes / 36 items	Physical functioning, role physical, bodily pain, general health, vitality, social functioning, role emotional, and mental health	The past 4 weeks (except for general	Copyrighted, but free to use provided users agree to terms and conditions
SF-12	Shorter than SF-36; quick snapshot of health-related QoL that considers physical, mental and social functioning; good for large scale interventions	~2–5 minutes / 12 items	Same as SF-36	The past 4 weeks	Copyrighted, but fee waiver may be requested for scholarly projects

Scale	Noteworthy Features	Time to Administer / No. of Items	Area(s) of Measurement	Recall Period	Cost
WHOQOL-BREF	Captures effects of environmental conditions on QoL; culturally sensitive for cross-cultural comparison	~5–10 minutes / 26 items	Physical health, psychological health, social relationships, and environment (e.g., living conditions)	The past 2 weeks	Free
QOLI	Identifies strengths and areas of improvement for QoL by assessing which areas of life an individual considers most important and how satisfied they feel within each of those areas across 16 different domains	~5–10 minutes / 32 items	Assesses both life satisfaction and life priorities across 16 life domains including goals, self-esteem, health, relationships (friends, love, children, relatives), work/ retirement, play, helping, learning, creativity, money, and surroundings/ community.	Current state	Copyrighted, licence purchase required

TOOLS FOR MEASURING LIFE SATISFACTION

Based on our review, the Satisfaction with Life Scale (SWLS) stands out as the only widely validated tool for assessing life satisfaction.[97-100] It is especially well-suited for community and arts-based programmes seeking to understand how participants feel about their lives overall. Its brevity and ease of use make it quick to administer when time is limited or when working with diverse and non-clinical groups.

If your programme…

- Has a brief time window for data collection
- Aims to enhance participants' overall life satisfaction

We recommend…

Satisfaction with Life Scale (SWLS)

- Brief and simple to use with the short administration time
- Well-suited for community and arts-based programmes
- Widely used in both research and practice
- Valued for its flexibility and strong psychometric properties (strong reliability and longitudinal validity)[98,100]

Figure 6.2 Guide to choosing a life satisfaction scale

The table below provides an overview of the characteristics and noteworthy features of this scale:

Scale	Noteworthy Features	Time to Administer / No. of Items	Area(s) of Measurement	Recall Period	Cost
Satisfaction with Life Scale (SWLS)	Brief, validated, self-administered scale for general life satisfaction	~1 minute / 5 items	Overall life satisfaction	Recent (current state)	Free

While the Satisfaction with Life Scale (SWLS) is our primary recommended tool for assessing life satisfaction, we also wish to highlight the **Cantril Ladder** as an alternative option due to its simplicity, cross-cultural applicability, and widespread global use.[102] See infobox on next page for more details.

Cantril Ladder (Self-Anchoring Striving Scale)

The Cantril Ladder is a visual, self-anchored scale designed to assess an individual's overall life satisfaction and subjective wellbeing. Respondents are asked to imagine a ladder with steps numbered from 0 (worst possible life) to 10 (best possible life) and rate their current life position based on their own goals, values, and standards.

Noteworthy features:

- Single-item visual scale: respondents rate their life on a 0–10 ladder
- Simple and intuitive: minimal burden for both participants and administrators
- Widely used internationally: main measure in the World Happiness Report
- Applicable across cultures: adaptable for diverse populations
- Self-defined criteria: allows for individual interpretation of life satisfaction

Note: *While the Cantril Ladder does not offer detailed subscales or clinical diagnostic thresholds, it is valuable in contexts where quick, flexible, and broad assessments of subjective well-being are appropriate.*

We do not recommend using the Cantril Ladder if your main outcome of interest is life satisfaction, however, because the scale contains only one question, which means nuances in this outcome cannot be measured and sensitivity is limited. Furthermore, while generally reliable, the validity of Cantril Ladder's single-item measure may vary in certain populations, particularly those experiencing high stress or with specific social and cultural contexts.[103] Despite its long-standing use in large international studies like the World Happiness Report, we recommend using it mainly for a quick snapshot (a single time point to assess life satisfaction across different individuals) or make broad comparisons, rather than using it as the main tool for detailed evaluation to capture changes that arise from an intervention (e.g., tracking changes in pre- and post- studies).

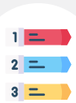
TOOLS FOR MEASURING WELLBEING

Questionnaires examining wellbeing vary in their focus—some concentrate on emotional experience, while others explore how individuals function in their daily lives, including their sense of meaning, purpose, and connection with others.

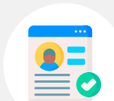
Across the literature, the most widely used and well-supported scales for assessing wellbeing are (in descending order):[84,104-107]

1. Warwick–Edinburgh Mental Well-being Scale (WEMWBS)[108]

4. 18-item Ryff's Psychological Wellbeing Scale (Ryff's PWBS 18-item)[111]



2. WHO-5 Well-being Index (WHO-5)[109]



5. PERMA-Profiler[112]

3. Personal Wellbeing Index-Adult (PWI-A)[110]

If your programme...	You might prefer...
• Focuses on enhancing the positive aspects of mental health • Is a larger-scale arts programme that is well-resourced	**Warwick-Edinburgh Mental Well-being Scale (WEMWBS)** • Primary focus on positive mental wellbeing, including optimism, functioning and relationships • Captures broad and sensitive changes in population's wellbeing
• Has limited time for data collection • Operates in multilingual environments with diverse populations	**WHO-5 Well-being Index (WHO-5)** • Only has five items (cheerfulness, calmness, energy, restfulness and interest) with primary focus on affective, functional and physical wellbeing (mood, vitality, interest) • Has been translated into over 30 languages, including Mandarin, Malay, and Tamil
• Serves to strengthen different aspects of wellbeing (see table below) based on participants' satisfaction in each aspect	**Personal Wellness Index - Adult (PWI-A)** • Assesses satisfaction about one's life with respect to different life domains, including standard of living, health, life achievements, relationships, safety, social connectedness, and future security
• Primarily aims to influence participants' personal development • Is well-resourced with funding and time available for data collection and analysis	**18-item Ryff's Psychological Wellbeing Scale (Ryff's PWBS 18-item)** • Assesses complex aspects of personal development over time, focusing on personal growth • May be less practical for quick evaluations, but is useful for in-depth and long-term understanding of overall mental health, life satisfaction, and psychological wellbeing
• Aims to enhance participants' emotional stability by focusing on their emotional wellbeing	**PERMA-Profiler** • Identifies risk with the inclusion of assessing negative emotions • Assesses both positive and negative emotion (i.e. sadness, anger, and anxiety) for a balanced evaluation of hedonic wellbeing along with engagement, relationships, meaning, and accomplishments

Figure 6.3 Guide to choosing a wellbeing scale

The table below provides an overview of the characteristics and noteworthy features of the 5 wellbeing questionnaires mentioned in this chapter:

Scale	Noteworthy Features	Time to Administer / No. of Items	Area(s) of Measurement	Recall Period	Cost
Warwick-Edinburgh Mental Well-being Scale (WEMWBS)	Captures a broad spectrum of positive mental functioning; suitable for population-level monitoring	~3–5 minutes / 14 items	Mental wellbeing, specifically subjective wellbeing (overall satisfaction, positive/ negative affect)'	The past 2 weeks	Copyrighted, but free with registration
WHO-5 Well-being Index (WHO-5)	Brief, sensitive to emotional state changes; ideal for rapid screening	<2 minutes / 5 items	Current mental wellbeing	The past 2 weeks	Free
Personal Wellbeing Index-Adult (PWI-A)	Focus on satisfaction within various life domains based on current circumstances and challenges	~2 minutes / 7 core + 2 optional items	Satisfaction across: standard of living, health, achieving in life, relationships, safety, community-connectedness, and future security	No specific recall period	Free
18-item Ryff's Psychological Well-Being Scale (Ryff's PWBS 18-item)	Focuses on deeper psychological growth (e.g. purpose, autonomy for eudaimonic wellbeing); flexible length options	~3-5 minutes / 18 items	Six aspects of wellbeing and happiness: autonomy, environmental mastery, personal growth, positive relations with others, purpose in life, and self-acceptance	No specific recall period	Free
Positive Emotion, Engagement, Relationships, Meaning, and Accomplishments Profiler (PERMA-Profiler)	Includes negative emotion items for a balanced assessment of the affective and functional dimensions of wellbeing	~5-7 minutes / 23 items (15 core + 8 optional items)	5 pillars of wellbeing: positive emotion, engagement, relationships, meaning, accomplishment, along with negative emotion and health	No specific recall period	Copyrighted, but free with registration for non-commercial research

In addition to the above scales, we would like to briefly mention the UCL Creative Wellbeing Measures Toolkit (see infobox on the right for more details). While this tool has been validated in the United Kingdom and widely adopted by artists and community practitioners,[113] it has not yet been validated in other languages and cultural contexts. Nevertheless, given its strong practical value, ease of use, and real-world relevance in arts and health programmes, we include it here as a supplementary recommendation. It may be useful for practitioners looking for a brief, user-friendly way to assess perceived wellbeing before and after arts-based activities, especially in non-clinical or community environments.

UCL Creative Wellbeing Measures Toolkit

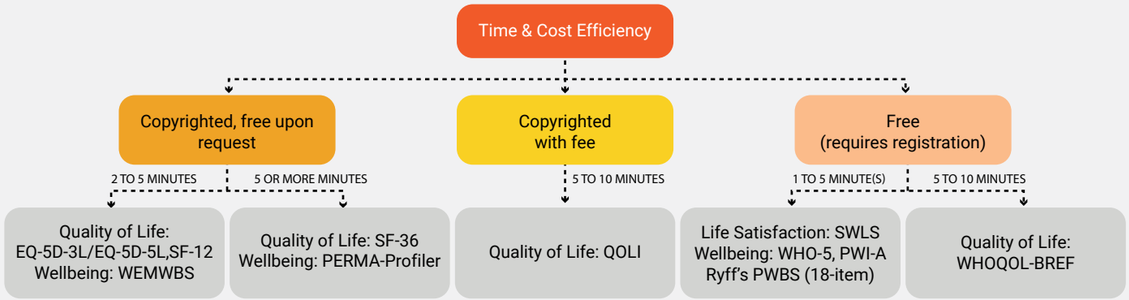
The UCL Creative Wellbeing Measures Toolkit is a practice-based evaluation tool developed to assess subjective wellbeing outcomes in arts and cultural participation contexts. It has been widely used by artists and cultural practitioners, particularly in community-based creative settings such as museums, workshops, and participatory arts programmes.

Noteworthy features:

- Designed for general populations in non-clinical, creative settings
- Low burden: brief, self-administered, and no specialist training required
- Mandarin version available (not yet psychometrically validated)
- Suitable for pre–post intervention use
- May be especially useful when presenting outcomes to funders or stakeholders outside of healthcare

ADDITIONAL CONSIDERATIONS WHEN CHOOSING AN APPROPRIATE SCALE FOR YOUR EVALUATION

There are different factors to consider in selecting a suitable scale, including but not limited to time and cost considerations. If you do have logistical constraints, you may wish to consider the budget and time you have for data collection to help make your decision:



CHAPTER 7

Measures of Social Connection

The World Health Organisation's (WHO) definition of health treats the social aspects of health as equally important as the physical and mental aspects, and the WHO has recently recognised social connection as a global health priority.[114,115] Studies consistently show that social isolation and loneliness can increase the risk of depression and anxiety[116], risk of early death[117], and cognitive decline in older adults[118]. Arts and Health interventions are excellent at bringing people together, promoting meaningful social connections and fostering social cohesion.[119] Given these benefits, many Arts and Health programmes now measure social connection as an outcome.

SOCIAL CONNECTION, LONELINESS, AND SOCIAL ISOLATION

How are these terms defined?

Social connection, sometimes called *social connectedness* or *belonging*, refers to the experience of having close relationships with others, such as family, friends, neighbours, and communities.[120] It comprises three main components: structure (who and how many people you know), function (what support they provide), and quality (how you feel about those connections).[121]

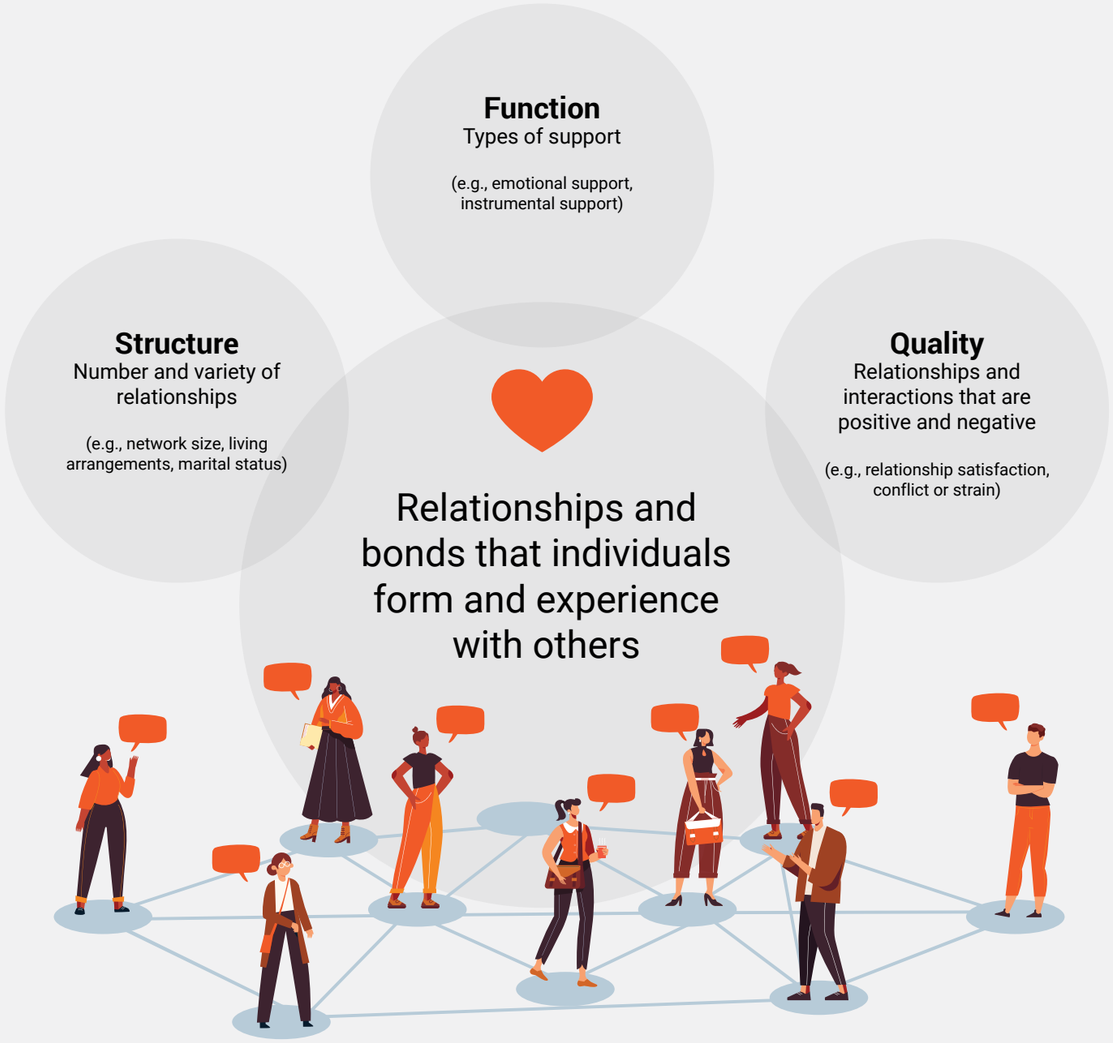


Figure 7.1 Components of social connection

While social connection and social connectedness are often used interchangeably, social connection is a continuum and social connectedness indicates where an individual falls on this continuum (i.e., how connected a person feels to others).[122] Poor social connection/ connectedness can lead to social isolation and loneliness.[123]

Loneliness (also known as perceived isolation) is the subjective feeling of having fewer meaningful relationships than desired.[124] In comparison, **social isolation** refers to a lack of contact with others.[124] It is measured objectively (e.g., size and diversity of social network, frequency of interactions) and is more frequently measured in older adults.

Social isolation can be a risk factor for loneliness but they are not the same – a person with few social contacts might not feel lonely, whereas someone with a large social network may still experience loneliness.[125] While increasing social contact may alleviate objective social isolation, addressing loneliness may require enhancing the quality and function of this contact.

As these definitions overlap across the literature, choosing the right tool starts with understanding the difference between the terms and how they relate to your intervention's goal:

- If your programme aims to create more social contacts or reduce barriers to interactions, you may want to choose a social isolation scale that captures these kinds of objective measures.
- If your programme targets social support or the function and quality of social connections, then we recommend using a social connection scale.
- If your programme aims to build meaningful relationships or reduce feelings of exclusion and lacking companionships, you may wish to use a loneliness scale.

The table below summarises the definition and key features of social connection, loneliness, and social isolation:

Term	Nature	Definition	Key Characteristics
Social connection/ connectedness	Objective and subjective	The experience of connecting to others, characterised by the structure, function and quality of the relationships	Umbrella term encompassing not only the quantity and structure of one's social network, but also the subjective experience of belonging, and feeling close to and supported by others
Loneliness	Subjective feelings	An unpleasant feeling arising from a disparity between existing and desired relationships	Unmet social needs and expectations; can occur even with frequent social contact
Social isolation	Objective	Limited social and communication, resulting from physical or social separation from others	Characterised by frequency of contact with others, or size and diversity of social network

TOOLS FOR MEASURING SOCIAL CONNECTION/ CONNECTEDNESS

Social connection scales typically measure a sense of belonging and social support, as well as feelings of loneliness and social network. In this section, we focus on the first two aspects, as scales specifically measuring feelings of loneliness and social networks will be covered in the subsequent subsections.

Most scales do not examine both belonging and social support. Hence, it is important to first identify which aspects of social connection your programme intends to influence. Based on our literature review, the 4 most prevalent scales for measuring social connection (in descending order) were:[126-131]

1. Social Connectedness Scale-Revised (SCS-R)[132]

2. Multidimensional Scale of Perceived Social Support (MSPSS)[133]

3. Social Provisions Scale (SPS-10)[134]

4. Duke Social Support Index (DSSI-11)[135]

If your programme...	You might prefer...
• Intends to foster a sense of belonging, perceived closeness to others, and connection to a wider social world	**Social Connectedness Scale-Revised (SCS-R)** • Efficient and general measurement of connectedness • Clear and concise questions to answer
• Involves family, friends, and/or significant other to improve social support • Is implemented across different ethnic groups	**Multidimensional Scale of Perceived Social Support (MSPSS)** • Able to discriminate social support from family, friends, and significant other • Available in a variety of languages, including English, Mandarin, Malay, and Tamil
• Aims to build different *types* of social support (i.e., attachment, guidance, reassurance of worth, social integration, and reliable alliance)	**Social Provisions Scale (SPS-10)** • Captures 5 key dimensions of perceived social support without distinguishing from whom (e.g., friends or family) • Slightly quicker to administer than other scales measuring social support
• Aims to create new social relationships or increase the frequency of social interaction (objective social connection), and/or strengthen the quality of existing relationships (subjective social connection) • Seeks to improve the quality of social support from family and friends, leading to greater satisfaction with social support	**Duke Social Support Index (DSSI-11)** • Measures social participation and satisfaction of perceived support from family and friends • Offers more detail than other scales on structural and quality components of social connection *Note: Another popular version, the DSSI-23, is validated in English and Mandarin (but not Malay) and includes an additional Instrumental Support Subscale to capture the types of practical support received from family and friends*

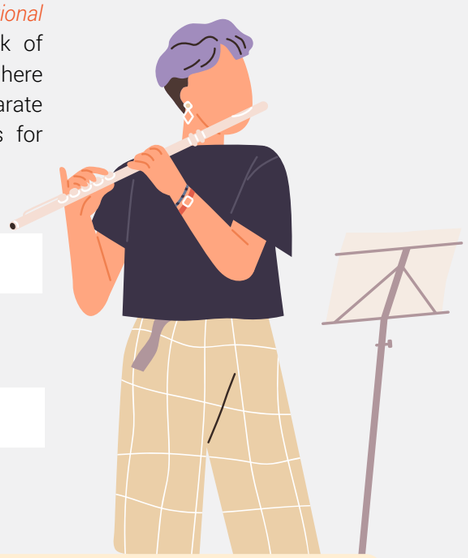
Figure 7.2 Guide to choosing a social connection scale

The table below provides an overview of the characteristics and noteworthy features of the social connection/ connectedness scales mentioned in this chapter:

Scale	Noteworthy Features	Time to Administer / No. of Items	Area(s) of Measurement	Cost
Social Connectedness Scale-Revised (SCS-R)	Comprehensive and general measurement of connectedness	~5-10 minutes / 20 items	Overall sense of belonging; perceived closeness to others; feelings of connectedness with the broader social world	Free, permission required
Multidimensional Scale for Perceived Social Support (MSPSS)	Discriminates support from family, friends, and significant other	~5 minutes / 12 items	Adequacy of social support from family, friends and significant other	Free
Social Provisions Scale (SPS-10)	Quick tool capturing 5 key dimensions of functional social support	~5 minutes / 10 items	Availability of functional social support across attachment, guidance, reassurance of worth, social integration, and reliable alliance	Free
Duke Social Support Index (DSSI-11)	Brief measurement of objective and subjective aspects of social connection	~5 minutes / 11 items	Social participation (network size, frequency); satisfaction with perceived social support	Free

TOOLS FOR MEASURING LONELINESS

Loneliness is commonly conceptualised using 2 dimensions: *emotional loneliness* (lack of intimate relationships) and *social loneliness* (lack of belonging to the broader social world).[124] The scales we recommend here examine these dimensions either as one overall score or as two separate scores. Based on our literature review, the 2 most popular scales for measuring loneliness (in descending order) were:[128, 136-140]



1. UCLA Loneliness Scale (Version 3; UCLA-LS3)[141]

2. De Jong Gierveld Loneliness Scale (DJGLS-6)[142]

If your programme...

- Aims to reduce loneliness in various social relationships
- Targets different age groups and/or populations

You might prefer...

UCLA Loneliness Scale (Version 3; UCLA-LS3)
- Widely recognised as a "gold standard" for measuring loneliness across diverse disciplines, age groups, and countries
- Provides a single, global score of loneliness, which is useful for tracking changes over time

If your programme...

- Is tailored to promote closeness (emotional connection) with others and/or increase social interactions, in order to reduce loneliness
- Mainly involves older adults

You might prefer...

De Jong Gierveld Loneliness Scale (DJGLS-6)
- Identifies whether emotional and social loneliness are present and provides different scores for these, allowing for separate analysis
- More commonly used for older adults, but also validated with large samples of other age groups

Figure 7.3 Guide to choosing a loneliness scale

The table below provides an overview of the characteristics and noteworthy features of the loneliness scales mentioned in this chapter:

Scale	Noteworthy Features	Time to Administer / No. of Items	Area(s) of Measurement	Cost
UCLA Loneliness Scale (Version 3; UCLA-LS3)	Gold standard; offers a broad and general snapshot of loneliness	~3-5 minutes / 20 items	Intensity of perceived isolation, lack of companionship, and feelings of being left out	Free
De Jong Gierveld Loneliness Scale (DJGLS-6)	Quick to administer; differentiates social and emotional loneliness in 2 different scores	~2 minutes / 6 items	Presence of emotional loneliness (feelings of emptiness, missing others, and rejection) and social loneliness (missing a wider network)	Free

TOOLS FOR MEASURING SOCIAL ISOLATION

Social isolation is typically measured through objective indicators such as size, frequency, and variety of an individual's social network. Improving these "structural aspects" may help facilitate social connection. In our literature review, Lubben Social Network Scale (LSNS-6) surfaced as the most widely used social isolation scale, followed by the Social Interaction Subscale of the Duke Social Support Index (DSSI-SIS), which may be used independently from the main scale.[130,131,136,137,140] These scales are typically used to assess social isolation in older adults, but can be used with other demographics.

1. Lubben Social Network Scale (LSNS-6)[143]

2. Duke Social Support Index-Social Interaction Subscale (DSSI-SIS)[135]

If your programme...	You might prefer...
• Seeks to increase the *frequency of contact* with friends and family, and enhance the perceived availability of support they provide	**Lubben Social Network Scale (LSNS-6)** • Focuses on social ties with family and friends • Clearly classified social isolation risk with established cutoff scores (LSNS-6 scores <12 indicate at risk of isolation)
• Seeks to increase the *number* of close and reliable relationships one has, and/or the *frequency* of participation in social activities	**Duke Social Support Index-Social Interaction Subscale (DSSI-SIS)** • Focuses on the frequency of participation in recent social activities • Can be used and analysed independently, although the entire scale is more commonly used for broader social support analysis

Figure 7.4 Guide to choosing a social isolation scale

The table below provides an overview of the characteristics and noteworthy features of the social isolation scales mentioned in this chapter:

Scale	Noteworthy Features	Time to Administer / No. of Items	Area(s) of Measurement	Cost
Lubben Social Network Scale (LSNS-6)	Focuses on size of social ties with family and friends	~5 minutes / 6 items	Size and frequency of social contact with family and friends; availability of social support received from them	Free, permission required
Duke Social Support Index-Social Interaction Subscale (DSSI-SIS)	Complements other subscales of DSSI to understand an individual's support system	~2 minutes / 4 items (for Social Interaction Subscale alone)	Frequency of social participation in various recent activities	Free

Emerging evidence suggests that one-session interventions can help participants feel less lonely or more connected, but meaningful social connection typically develops over time through repeated interactions. As such, these scales are more commonly used to evaluate multi-session programmes.

Structural measures like LSNS-6 or DSSI-SIS, which examine objective features such as the size, frequency, and variety of a social network, have specific recall periods (i.e., in the past week/month) for certain questions, and therefore may not be suitable for measuring change after just one session, as a regular/stable network takes time to build. Conversely, scales like SCS-R, MSPSS, SPS-10, DSSI-11, UCLA-LS3, and DJGLS-6 examine the respondent's current state/feelings, and do not have specific recall periods. While they can potentially capture subjective shifts in feelings following a single-session intervention, they are not specifically designed to measure immediate, in-the-moment feelings.

For *momentary* feelings of social connection, we highlight the UBC State Social Connection Scale (UBC-SSCS) as a promising option for future use.[144] See infobox below for more details.

UBC State Social Connection Scale (UBC-SSCS)

The UBC State Social Connection Scale is a brief (10 items) tool to assess momentary feelings of social connection. As a relatively new scale, it has not yet been widely adopted in the academic community, though early evidence supports its reliability and validity.[144,145] Designed for use in experimental and community settings, it captures state-level changes in social connectedness across short time frames, making UBC-SSCS a promising tool for evaluating the immediate impact of short-term interventions, such as community events or creative sessions.

Noteworthy features:

- Captures state-level social connection: ideal for capturing short-term or even moment-to-moment changes (e.g., in last 15 minutes)
- Strong psychometric properties: high internal reliability, construct validity, and sensitivity to changes in social connection

Note: Only English version is currently available. Adaptation would be needed for use in multilingual contexts.

HOW TO CHOOSE AN APPROPRIATE SCALE FOR YOUR EVALUATION?

As each scale is developed to measure different aspects of social connection, the choice of a scale for your evaluation depends on the goal of your activity or intervention. The scales recommended in this chapter are free to use but some may require permission from the author. Here is a flowchart to help you choose which scale is best suited to your needs:

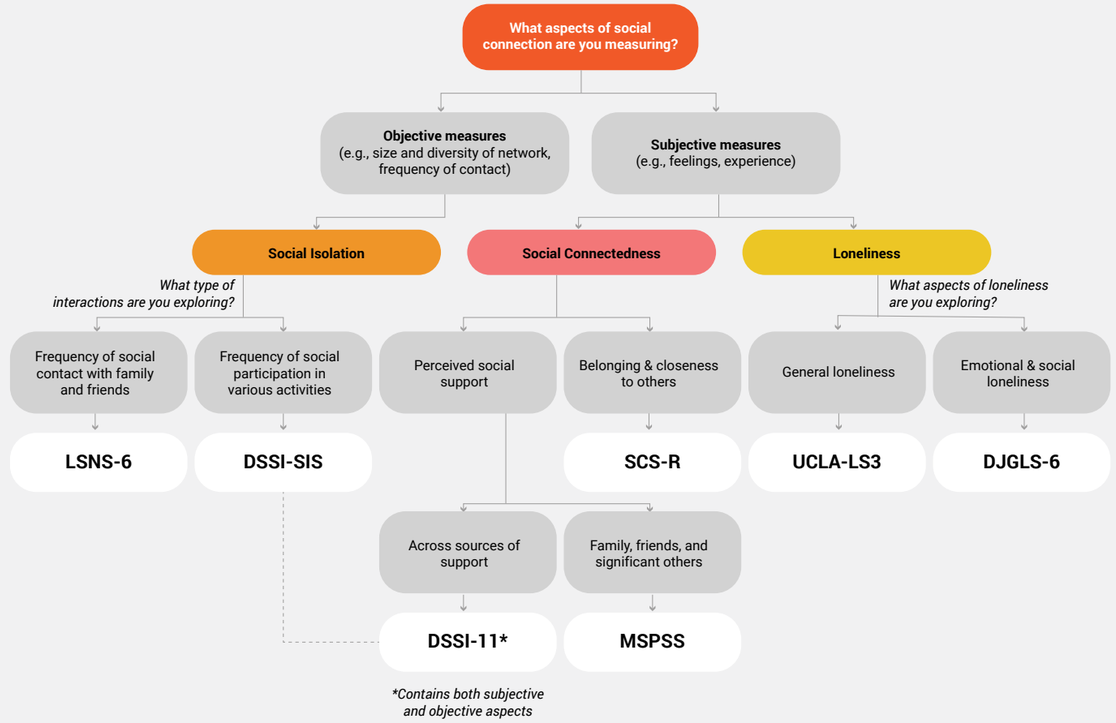


Figure 7.5 Flowchart to guide selection of scale

CHAPTER 8

Physiological measures

If you run an arts programme that aims to influence health outcomes, you might be interested in measuring physiological markers - that is, objective measures of bodily functions such as heart rate or blood pressure.[146] Physiological measures can reflect many different aspects of health and are useful for evaluating outcomes related to physical health and stress.

This chapter is a brief introduction to a selection of physiological measures you can consider for your evaluation. We focus primarily on outcomes related to overall physical health and tools that have been validated in research or validated against an established measurement tool (when, for example, a smart watch has been validated to be as accurate, or nearly as accurate, in heart rate measurements as an electrocardiogram (ECG) device). We have prioritised covering practical, non-invasive (procedures that do not directly interfere with the human body) methods of collecting physiological data that you can use without extensive training. Some measures, however, are more demanding in terms of collecting or analysing data. For these cases, we strongly encourage you to collaborate with research experts experienced in physiological measurement to ensure appropriate implementation and interpretation.

In this brief overview, we discuss some of the common outcomes used to assess physical health, including:

1. Physical fitness level

2. Motor function

3. Sleep, including sleep duration and sleep quality

4.Stress level and response

For a broader overview, we recommend comprehensive sources such as Fancourt and colleagues' paper[147] on how leisure activities, including the arts, affect health. This framework identifies many mechanisms of action, i.e., the ways in which engaging in leisure activities "leads to the prevention, management, or treatment of mental and physical illness."[147] Physiological responses are included among the many mechanisms provided (we recommend consulting the supplementary material for their comprehensive overview).

GENERAL CONSIDERATIONS WHEN CHOOSING PHYSIOLOGICAL MEASURES

Keep the following in mind to guide your choice of physiological measures:

- Do the data gathered by physiological measures help to evaluate my outcome(s) of interest?
- Do I need domain experts to help collect or interpret the data?
- Do I have sufficient funds in my budget to fund equipment, lab, or software (for data collection or analysis) purchases needed for physiological measures?

It is often a good idea to collect baseline data: Before your programme or activity starts, consider potential differences between your participants (if relevant), such as their age, health status, emotional state or personality, as these can influence physiological measures. Baseline data can help account for these individual differences.

You may also need to consider the environment or setting of your data collection. For example, the measurement accuracy of some tools is sensitive to noise and/or movement.

Another consideration to keep in mind is the grade of the tool you are using. Tools for physiological measures can be:

1. *Research-grade*, meaning that they are designed specifically for research purposes, or

2. *Commercial* or *consumer-grade*, which are produced for business or public consumption

Both grades are used in published research, but different tools tend to have different degrees of validity and accuracy. While research-grade tools usually offer more precision, accuracy, and validity, they are generally more expensive and require more specialised training.[148] However, it is important to keep in mind that many off-the-shelf technological devices may not be able to collect data reliably (i.e., some provide distorted data or inaccurate measurements).[149] As always, it is good practice to check the brand or model of your device to make sure that you are using a tool that has been validated.

PHYSICAL FITNESS LEVEL

Physical fitness refers to one's ability to perform physical activities and muscle endurance (i.e., the ability to perform physical activities for an extended period of time).[150] Research has shown that arts activities can have significant effects on people's physical health. For example, a review conducted by Fong Yan and colleagues[151] showed that structured dance activities improve participants' physical health.

There are many approaches to assessing physical fitness. Here are two common approaches:

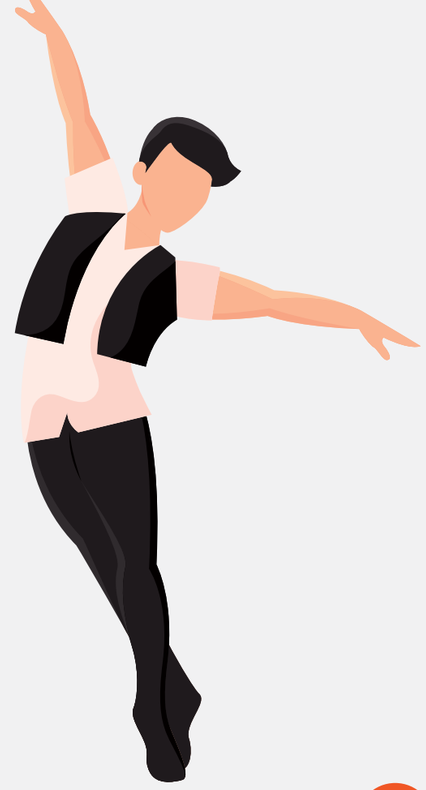
a. *Physical activity*, completed over a short (e.g., one dance session) or longer time (e.g., two months with weekly dance sessions)

b. *Cardiorespiratory fitness* (see definition below) over a sustained period of time or chronically.[150]

Physical activity

Physical activity includes any bodily movement that takes energy to complete (e.g., the number of steps completed after a dance session).[152] Daily steps provide an accessible measure of physical activity and its health-related benefits, as evidenced by the research of Kraus and colleagues[153]. In addition, a study by Bassett and colleagues[154] has linked step monitoring to physical activity and explored its health-related benefits. Step counters, or pedometers, can record more detailed and accurate data (note that accuracy depends on the specific device) compared to self-report questionnaires.[154]

While physical activity can be measured using *validated self-report questionnaire*s such as the Global Physical Activity Questionnaire (GPAQ)[155] and the International Physical Activity Questionnaire (IPAQ)[156], they can prone to recall and social desirability biases (e.g., participants might inflate their physical activity level because having more physical activity is considered more favourable). *Step counting* is a reliable and accessible alternative to self-report questionnaires.



Cardiorespiratory fitness

Cardiorespiratory fitness, also known as cardiorespiratory health, refers to the ability of the heart, lungs, or blood vessels (vascular system) to deliver oxygen to the muscles to sustain physical activity.[157] It is a well-established marker of physical fitness and a consistently strong marker of long-term health and longevity.[158]

Cardiorespiratory health can be assessed using indicators like resting heart rate (RHR) and blood pressure.[159,160] **Resting heart rate** is the number of times one's heart beats in one minute, usually while sitting or lying down, when feeling calm, and when not ill.[160] It is accessible to measure and is associated with cardiorespiratory fitness and life expectancy (e.g., reduction in resting heart rate is associated with an increase in life expectancy).[160, 161]

Blood pressure (BP) is the pressure applied to one's arteries by blood when the heart pumps.[162] It is indicative of the overall health of the vascular system, which tends to harden with age, causing high blood pressure, or hypertension, in older adults.[159] If your programme aims to improve the overall health of older adults, you can consider using blood pressure as an indicator for your main outcome. You can find more information, including how to understand blood pressure readings, in the American Heart Association's guide to understanding blood pressure.

VO_2max is the maximum amount of oxygen consumed per minute when someone engages in strenuous exercise[159,163], and can also be used as a measure of cardiorespiratory health. Because this measure requires intensive physical activity, blood pressure is usually used as an alternative, especially among older adults. You can find more information in Harvard's explainer on VO_2max.

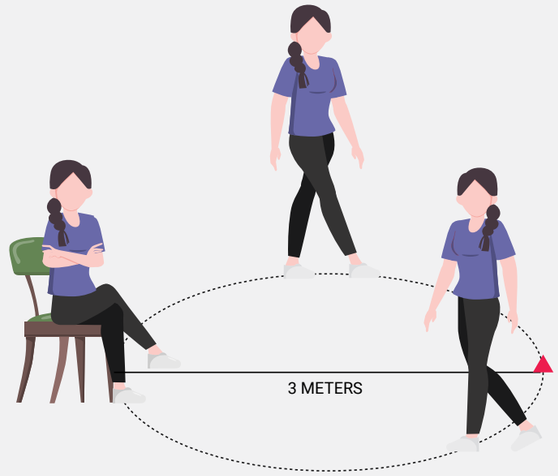
MOTOR FUNCTION

You may also be interested in evaluating how arts engagement affects participants' motor function when it comes to movement-based programmes. Physical fitness refers to one's overall physical condition, how well an individual performs physically, and their level of muscle endurance.[150] **Motor function**, on the other hand, focuses on assessing how effectively an individual is able to control their movements.[164] Motor function, especially combined with physical fitness, has been found to be a strong predictor of longevity.[165] Here is an example of how you may evaluate these two outcomes using a hypothetical scenario:

Suppose you are interested in evaluating the health effects of a 3-month dance program on a community of older adults.

1. If you are interested in the changes in participants' *physical fitness* levels, you can compare participants' blood pressure before and after the dance programme to evaluate whether the dance programme improved their capacity to handle sustained physical activity.
2. If you are interested in the changes in participants' *motor function,* you can compare participants' performance on the Timed Up and Go test before and after the dance programme to evaluate whether the dance programme has improved their coordination and control of their movements.

Motor function assessments tend to be more relevant for middle-aged and older adults. Therefore, if your programme is interested in studying how arts-based interventions can improve coordination, strength or help older adults move better, consider measuring this outcome. Here are a couple of accessible ways of assessing motor functions:



1. Timed Up and Go (TUG)[166] is a simple and widely used test of motor function, especially in assessing extremity function, mobility, and predicting fall risk for older adults. It has been validated for older adults with good reliability and the ability to detect subtle changes.[166,167] We also recommend that you run the test multiple times to prevent skewed results (inaccurate representation of how participants normally perform).[168] You can consult the Centers for Disease Control and Prevention's resource for a guide to using TUG.

2. The Berg Balance Scale (BBS)[169] is another commonly used test for balance and predicting fall risk. Like TUG, it is performance-based and has good validity and reliability.[169] BBS is less sensitive to changes in balance among healthy adults.[170] It is, however, good at identifying balance issues in adults and serves as a strong predictor of multiple fall risks.[170] You can find more information on using BBS on Brandeis University's data collection form.

If your intervention includes *full-body movements* (e.g., dance, theatre performance), TUG and BBS are useful tools to measure your programme's effects on improving coordination and balance.

3. Grip strength is another validated and widely used indicator of motor function, and has been linked to mortality and overall physical health.[171,172] It is commonly measured using a device called the handgrip dynamometer, and the Jamar *handgrip dynamometer* is a validated device for testing grip strength in middle-aged and older adults.[171,173] To conduct a grip strength test, participants are asked to hold the dynamometer (as shown above) at a comfortable position and squeeze the dynamometer as tightly as possible.

If your intervention is targeted at *movements* (e.g., drumming, painting), consider grip strength assessment to measure your programme's effects on improving overall strength and health.

SLEEP

Sleep has been established as an important indicator of overall physical health, with decades of research showing the harmful consequences of lacking or disrupting sleep.[174] Sleep is a well-recognised contributor to many aspects of health, including boosting immune functions[175] and ensuring good metabolism[176], as well as a strong predictor of mortality[177]. Although a relatively new research area, arts activities (e.g., listening to music or dance movement therapy) have been shown to improve different aspects of sleep, including *sleep duration* (the length of one sleep period) and *sleep quality* (how well an individual sleeps).[178]

While sleep can be measured subjectively using self-report questionnaires such as the Pittsburgh Sleep Quality Inventory (PSQI)[178], the reported results are often prone to biases.[179] Objective measures of sleep also face challenges of capturing and interpreting the complex data of sleep.[178]

Polysomnography (PSG) is a well-established method of measuring sleep. PSG is conducted in sleep labs using specialised equipment to assess sleep comprehensively and is considered to be the 'gold standard' for analysing various aspects of sleep, including sleep duration and quality.[180] If you wish to gain detailed insights into sleep and have adequate resources, we recommend collaborating with a domain expert in conducting sleep-related studies or PSG.

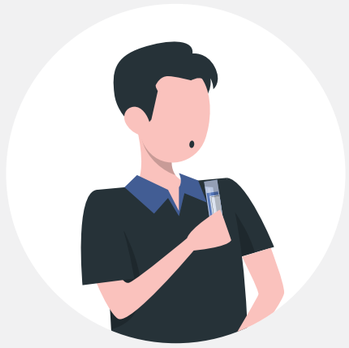
Practical alternatives, such as wearable sensors and actigraphy, that have been validated against PSG for both clinical and non-clinical populations are also available.[180,160] While they are more accessible, depending on the model, they may have lower reliability than PSG.

STRESS RESPONSE

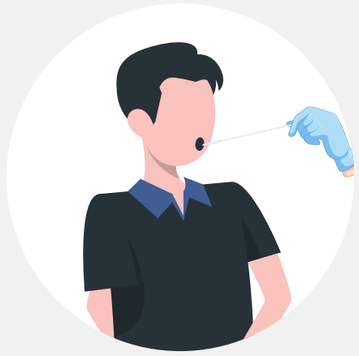
Stress response, or stress level, is part of one's emotional health, but is also a risk factor for mental health and many other aspects of health.[181] Stress can be measured using self-report scales such as DASS-21, but there are also physiological markers of stress, such as hormone levels and heart-related measures.

Cortisol

Cortisol is a steroid hormone that is involved in regulating the body's stress response. It is often used as an indicator of stress level,[182] although this measure is more sensitive to certain types of stressors than others (specifically, it is most sensitive to physical stressors such as an injury or illness, and psychosocial stressors such as public speaking or exams). Salivary cortisol is collected non-invasively and can be incorporated in a pre-post design to evaluate how arts engagement affects participants' level of stress.[183] A study by Kaimal and colleagues[184] is a good example of this. Common methods of collecting saliva samples include:[185]



a) Passive drooling: participants pool and drool saliva into a collection tube



b) Swab use: an absorbent swab is placed in a participant's mouth; a syringe is then used to squeeze saliva out of the swab

Saliva samples must be collected hygienically and stored properly. For more information, please consult the Salimetrics guide to saliva collection and handling. We also encourage you to consult a domain expert, as cortisol concentrations can vary throughout the day[186], to ensure that your saliva samples are collected at the appropriate times, given your study. You might also require help from the domain expert to extract and analyse results.

Heart-related measures

Heart-related measures, such as heart rate and blood pressure, are commonly used as indicators of stress.[187, 188] **Heart rate (HR)** measures the number of times the heart beats per minute and can reflect changes in stress response, e.g., an increase in stress is often associated with an increase in HR.[187] However, because HR can change for many reasons, we recommend not measuring HR on its own. It's more useful to combine HR with other measures to better understand a person's stress levels.

Blood pressure (BP) is another indicator of stress.[189] Elevated BP often signals stress, and the relationship between the two is especially strong in older adults.[190] Mastandrea and colleagues' study[191] of the effects of viewing artwork on stress reduction is a good example of using blood pressure to measure stress.

This table provides an overview of outcomes and potential tools that may be used for assessing overall physical health and stress response:

Measures	Validated Tools/ Methods	Noteworthy Characteristics of Tools/Methods	Considerations
Physical activity	Pedometer, or step counter	• Indicator of physical fitness • Generally an accurate and accessible measure (training often not required for data collection)	• Pedometers vary in accuracy based on where they are worn on the body (wrist or ankle-worn ones tend to have the most accurate and reliable measurements[154]) • May require domain expertise to analyse data
	Smart watch	• Accessible measure of step counts with generally high reliability • Convenient to collect and analyse data	• Reliability may vary depending on the device
Cardiorespiratory fitness	Smart watch	• Measures resting heart rate, which is associated with life expectancy • Convenient to collect and analyse data	• Participants should be resting (sitting up or lying down) during measurement period
	Blood pressure monitor, or sphyg-momanometer	• BP is strongly correlated with age • Easy to measure and analyse	• May consult domain experts for a more detailed analysis • More applicable to older adults
	VO_2max lab (e.g., exercise medicine lab)	• Considered the "gold standard" for measuring cardiorespiratory fitnes	• Requires going through intense physical activity to ensure measurement accuracy (not practical for older adults) • Requires specialised equipment and domain expertise to collect and analyse data

Measures	Validated Tools/ Methods	Noteworthy Characteristics of Tools/Methods	Considerations
Mobility, extremity function, balance	Timed Up and Go (TUG)	• Indicator of motor function • Predictor of fall risks and mortality • Validated for older adults with good validity and reliability • Easy to measure and analyse data without extensive training	• May require running the test multiple times and using an average score to prevent inaccurate results
	Berg Balance Scale (BBS)	• Particularly strong for predicting multiple fall risk in older adults • Easy to measure and analyse data without extensive training	• Less sensitive to subtle changes in healthy or younger adults
Grip strength	Handgrip dynamometer	• Strong measure of motor function • Predictor of longevity • Validated for middle-aged and older adults • Easy to measure and analyse data without extensive training	• May require domain expertise to analyse data
Sleep duration	Wrist-worn wearable	• Indicator of overall health • Linked to mortality risk	• Recommended to consult domain expert for detailed analysis

Measures	Validated Tools/ Methods	Noteworthy Characteristics of Tools/Methods	Considerations
Sleep quality	Polysomnog-raphy (PSG) or sleep lab	• Indicator of overall health • Provides accurate insights into sleep	• Requires domain expertise and lab to collect and analyse data
	Smart watch or ring for sleep tracking	• Easy to measure and analyse data • Does not require specialised lab equipment	• Measurements tend to be less reliable than PSG
Stress response	Saliva cortisol test	• Does not require extensive training or expensive equipment to collect data	• Requires proper data collection procedures, minding hygiene and temperature control • Requires lab and domain expertise to analyse data
	Smart watch	• Easy to collect and analyse • Can be monitored in real time	• Not recommended as a standalone measure for stress
	Blood pressure monitor, or sphyg-momanometer	• Particularly strong measure for indicating stress in older adults	• May require domain expertise for more detailed analysis

Appendix

APPENDIX A: IMPORTANT ETHICAL CONSIDERATIONS

It's important that you carry out Arts and Health activities in an ethical way, even if you don't need to apply for a formal ethics application. Here we flesh out important ethical considerations from Chapter 3:

1. Person-centered practice: For instance, co-creating programmes with the target audience can empower communities, enhance participant engagement, and contribute to long-term social change.[33]

> **Example:** Collaborating with seniors at Active Ageing Centres to co-create music programmes that resonate with their interests (e.g. performing music genres that they enjoy listening to, traditional songs, etc). could increase seniors' participation in community music activities that promote social interaction and bonding.

2. Equity: This is especially if your project involves 'vulnerable' populations (e.g., people with disabilities, illnesses, etc.). Perceptions of vulnerability may unintentionally reinforce stigmas or unequal social and power dynamics.[30] Thus, it is important to practice social responsibility, sensitivity and awareness for arts-based ventures in health.[30]

> **Example:** Music projects for people with dementia should respect individual differences and needs, while ensuring that dementia clients are not belittled for their health condition. This would allow music-making to be more accessible and enjoyable for different profiles of people.

For project staff and external collaborators, any conflict of interest should be declared, and fair policies and practices (e.g., clear contracts for various stakeholders) should be implemented.[33]

3. Safety: This can be done through setting ground-rules and having staff who are equipped with the knowledge and skills to handle any risks or unintended consequences encountered in a project.[33]

> **Example:** Ensure that sufficient and proper warm-ups are conducted before dance therapy sessions to prevent injuries.

APPENDIX B: VALIDATED LANGUAGES OF SCALES

In Chapters 5-7, we recommended scales for evaluating various outcomes relevant to Arts and Health. In the table below, we list which of the four official languages of Singapore each scale has been validated in:

Scale	English	Mandarin	Malay	Tamil
Mental Health				
Beck Depression Inventory (BDI)	✓	✓	✓	✓
Center for Epidemiologic Studies Depression Scale (CES-D)	✓	✓	✓	✓
Hospital Anxiety and Depression Scale (HADS)	✓	✓	✓	✓
Patient Health Questionnaire-9 (PHQ-9)	✓	✓	✓	✓
Depression, Anxiety, and Stress Scale-21 (DASS-21)	✓	✓	✓	✓
Beck Anxiety Inventory (BAI)	✓	✓	✓	
Generalized Anxiety Disorder Scale (GAD-7)	✓	✓	✓	✓
State-Trait Anxiety Inventory	✓	✓	✓	✓
Quality of life, Life satisfaction, and Wellbeing				
EuroQol five-dimensional instruments (EQ-5D-3L/5L)	✓	✓	✓	✓
36-Item Short Form Survey (SF-36)	✓	✓	✓	✓
12-Item Short Form Survey (SF-12)	✓	✓	✓	✓
WHO Quality of Life-BREF (WHOQOL-BREF)	✓	✓	✓	✓
Quality of Life Inventory(QOLI)	✓	✓	✓	✓
Satisfaction with Life Scale	✓	✓	✓	✓
Warwick-Edinburgh Mental Well-being Scale (WEMWBS)	✓	✓		
WHO-5 Well-being Index (WHO-5)	✓	✓		
Personal Wellbeing Index-Adult (PWI-A)	✓	✓	✓	
18-item Ryff's Psychological Well-Being Scales (PWBS)	✓	✓		✓
PERMA-Profiler	✓	✓		
Social Connection, Loneliness, and Social Isolation				
Social Connectedness Scale-Revised (SCS-R)	✓	✓		
Multidimensional Scale for Perceived Social Support (MSPSS)	✓	✓	✓	✓
Social Provision Scale (SPS-10)	✓	✓		
Duke Social Support Index (DSSI-11)	✓		✓	
UCLA Loneliness Scale (Version 3; UCLA-LS3)	✓	✓		
De Jong Gierveld Loneliness Scale (DJGLS-6)	✓	✓	✓	
Lubben Social Network Scale (LSNS-6)	✓	✓	✓	

APPENDIX C: METHODS FOR IDENTIFYING SCALES

In Section 2, we recommended scales for practitioners to use. In this Appendix, we describe how we chose these scales. For each outcome, a two-stage process was used to identify the best scales to recommend for use by Arts and Health practitioners. The first stage included a literature review, and the second stage involved investigating whether the identified scales met our inclusion criteria (see the Introduction to Section 2 for details). The literature review search process differed based on the outcome, as described below.

I. Mental Health

A. Depression and anxiety:
Review papers covering scales for measuring mental health outcomes were retrieved from the PubMed database, resulting in 46 review papers on depression and 38 for anxiety. A combined total of 34 papers were considered to be relevant and included in our review. From these 34 papers, 44 scales measuring depressive symptoms and 81 scales measuring anxiety symptoms were identified. The most commonly cited scales for each outcome were checked in relation to our inclusion criteria and the top 5 scales for each outcome were identified. For a more detailed description of the methods used, please see Ilya and colleagues[192].

II. Wellbeing and related outcomes

A. Quality of life and life satisfaction:
Review papers for scales measuring quality of life and life satisfaction were retrieved from PubMed and Scopus. A combined 151 review papers were retrieved from these searches and 67 papers were deemed relevant and included in our literature review. Based on the 67 retrieved review papers and the inclusion criteria stated in the Introduction to Section 2, the top 5 most commonly cited scales for assessing quality of life and life satisfaction were included in Chapter 6.

B. Wellbeing:
For wellbeing, a focused literature review was conducted on PubMed and Google Scholar, using the terms: “wellbeing”, “well-being”, “measure”, “measurement”, “systematic review”, “meta-analysis”, “scoping review”, “integrative review”, and “review”. From the review papers retrieved, we identified the most commonly-used scales, and recommended the 5 most prevalent scales for measuring wellbeing that satisfied our inclusion criteria.

III. Social connection, loneliness, and social isolation

To find review papers on measures of social connection, we searched the PubMed and Google Scholar databases using the terms: “social connection”, “social connectedness”, “social isolation”, “loneliness”, “social support”, “measure”, “measurement” “systematic review”, “meta-analysis”, “scoping review”, “integrative review”, and “review”. The most commonly cited scales were identified from the review papers, and the scales were then vetted based on our inclusion criteria. We recommended 4 widely-used scales for social connection, 2 scales for loneliness, and 1 for social isolation along with a subscale from a social connection scale.

References

REFERENCES

1. Jameel Arts & Health Lab. 2024 Annual Report [Internet]. Jameel Arts & Health Lab. 2024 Jun. Available from: https://crooked-antechinus.files.svdcdn.com/production/images/Updated-Final-Digital-Version_AW_JAHL_Annual-Report-2024_06.08.24_Inner.pdf?dm=1743438759.

2. Fancourt D. Arts in health: Designing and researching interventions [Internet]. Oxford: Oxford University Press; 2017. Available from: https://doi.org/10.1093/oso/9780198792079.001.0001.

3. White M. Arts development in community health: A social tonic. Oxford: Radcliffe Pub; 2009.

4. Fancourt D, Finn S. What is the evidence on the role of the arts in improving health and well-being? A scoping review [Internet]. Copenhagen: WHO Regional Office for Europe; 2019. Available from: https://www.who.int/europe/publications/i/item/9789289054553.

5. Huckaby S. Museums on prescription: Incorporating museum staff perspectives [master's thesis on the Internet] Seattle (WA): University of Washington; 2021. Available from: https://digital.lib.washington.edu/researchworks/items/0455a297-559e-4f35-9c99-7810b07fbeea.

6. United Nations Evaluation Group. Norms and Standards for Evaluation in the UN System [Internet]. New York: UNEG; 2018. Available from: https://thenoah.net/wp-content/uploads/2018/10/NOAH-Code-of-Ethics-and-Standards-for-Arts-in-Health-Professionals.pdf.

7. Warran K, Daykin N, Pilecka A, Fancourt D. Arts and Health Evaluation: Navigating the Landscape [Internet]. Social Biobehavioural Research Group; 2023. Available from https://sbbresearch.org/wp-content/uploads/2023/06/Arts-and-Health-Evaluation-report-UCLSBB-3.pdf.

8. Martino N, Fry R. Evaluation vs Research: What's the Difference and Why Does It Matter? [Internet]. ARTD Consultants; 2023. Available from: https://www.artd.com.au/news/evaluation-vs-research-whats-the-difference-and-why-does-it-matter/.

9. McQuade L, O'Sullivan R. Assessing the Impact of Arts and Creativity Interventions for Older People: A Public Health Evaluation Toolkit [Internet]. Institute of Public Health; 2024. Available from: https://www.artsandhealth.ie/guidelines/assessing-the-impact-of-arts-and-creativity-interventions-for-older-people-a-public-health-evaluation-toolkit/.

10. Public Health England. Arts for health and wellbeing-An evaluation framework [Internet]. 2016. Available from: https://assets.publishing.service.gov.uk/media/5c17bce1ed915d0c382e276b/PHE_Arts_and_Health_Evaluation_FINAL.pdf.

11. Pennsylvania Coalition Against Rape (PCAR). Theory of Change and Logic Models [Internet]. Enola (PA): Pennsylvania Coalition Against Rape; 2018. Available from: https://pcar.org/sites/default/files/resource-pdfs/tab_2018_logic_models_508.pdf.

12. Vogel I. Review of the use of 'Theory of Change' in international development [Internet]. London: Department for International Development; 2012. Available from: https://www.theoryofchange.org/pdf/DFID_ToC_Review_VogelV7.pdf.

13. Mackinnon A, Amott N. Mapping change: Using a theory of change to guide planning and evaluation [Internet]. New York: Grantcraft; 2006. Available from: https://learningforfunders.candid.org/wp-content/uploads/sites/2/2018/12/theory_change.pdf.

14. Milstein B, Chapel T. Community Tool Box [Internet]. Lawrence (KS): University of Kansas; n.d. Chapter 2, Section 1, Developing a Logic Model or Theory of Change. Available from: https://ctb.ku.edu/en/table-of-contents/overview/models-for-community-health-and-development/logic-model-development/main.

15. Ministry of Social and Family Development. Implementation Science and Evaluation # 10: Logic Model [Internet]. Singapore: Ministry of Social and Family Development; 2022. Available from: https://www.msf.gov.sg/docs/default-source/resources-document/poster10-logic-model-apr2022.pdf.

16. Zawadi NK. Research Methodologies for Arts-Based Health Intervention. RIJRE [Internet]. 2024 Nov;4(3):59–64. doi: 10.59298/RIJRE/2024/435964.

17. Keisari S, Palgi Y, Yaniv D, Gesser-Edelsburg A. Participation in life-review playback theater enhances mental health of community-dwelling older adults: A randomized controlled trial. Psychology of Aesthetics, Creativity, and the Arts [Internet]. 2022 May;16(2):302–17. doi:10.1037/aca0000354.

18. Staricoff RL. Arts in health: The value of evaluation. J R Soc Promot Health. 2006;126(3):116-120. doi:10.1177/1466424006064300.

19. Daykin N, Stickley T. The role of qualitative research in arts and health. In: Oxford Textbook of Arts, Health and Wellbeing. Oxford (GB): Oxford University Press; 2015. P.73-82. doi:10.1093/med/9780199688074.003.0010.

20. Wasti SP, Simkhada P, van Teijlingen ER, Sathian B, Banerjee I. The growing importance of mixed-methods research in health. Nepal J Epidemiol. 2022 Mar 31;12(1):1175. doi:10.3126/nje.v12i1.43633.

21. Fetters MD, Freshwater D. The 1+ 1= 3 Integration Challenge. J Mixed Methods Res. 2015 Apr;9(2):115-7. doi:10.1177/1558689815581222.

22. Archibald MM, Gerber N. Arts and Mixed Methods Research: An Innovative Methodological Merger. Am Behav Sci. 2018;62(7):956–77. doi:10.1177/0002764218772672.

23. Creswell JW, Plano Clark VL. Designing and conducting mixed methods research. 2nd ed. Thousand Oaks (CA): Sage; 2011.

24. Guetterman TC, Fetters MD, Creswell JW. Integrating Quantitative and Qualitative Results in Health Science Mixed Methods Research Through Joint Displays. Ann Fam Med. 2015;13(6):554–61. doi:10.1370/afm.1865.

25. Younas A, Durante A. The logics of and strategies to enhance generalization of mixed methods research findings. Methodology. 2023;19(2):170–91. doi:10.5964/meth.10863.

26. Scribbr. Research bias [Internet]. Amsterdam: Scribbr;Available from: https://www.scribbr.com/category/research-bias/.

27. NOAH National Organization for Arts in Health. Code of Ethics for Arts in Health Professionals and Standards for Arts in Health Professionals [Internet]. 2018. Available from: https://thenoah.net/wp-content/uploads/2018/10/NOAH-Code-of-Ethics-and-Standards-for-Arts-in-Health-Professionals.pdf.

28. Smith PT, Sonke JK. When Artists Go to Work: On the Ethics of Engaging the Arts in Public Health. Hastings Cent Rep. 2023;53(S2). doi:10.1002/hast.1530.

29. CITI Program. About CITI Program [Internet]. Miami (FL): CITI Program. Available from: https://about.citiprogram.org/.

30. National Healthcare Group. NHG Investigator Manual – Chapter 5: Informed Consent [Internet]. Edition 4. Singapore: National Healthcare Group Pte Ltd; 2023. Available from: https://ethics.gri.nhg.com.sg/files/Investigator%20Manual/IM_Chapter_5_Informed_Consent.pdf.

31. Laukkanen A, Jaakonaho L, Fast H, Koivisto T-A. Negotiating boundaries: Reflections on the ethics of arts-based and artistic research in care contexts. Arts Health. 2022;14(3):341–54. doi:10.1080/17533015.2021.1999279.

32. Gordon EJ, Prohaska TR. The Ethics of Withdrawal from Study Participation. Accountability in Research [Internet]. 2006 Dec;13(4):285–309. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9527709/.

33. Culture Health and Wellbeing Alliance. Creative Health Quality Framework [Internet]. Available from: https://www.culturehealthandwellbeing.org.uk/resources/creative-health-quality-framework.

34. Boardsworth K, Barlow R, Wilson B-J, Wilson Uluinayau T, Signal N. Toward culturally responsive qualitative research methods in the design of health technologies: Learnings in applying an Indigenous Māori-centred approach. Int J Qual Methods. 2024;23. doi:10.1177/16094069241226530.

35. Karlsson P, Bergmark A. Compared with what? An analysis of control group types in Cochrane and Campbell reviews of psychosocial treatment efficacy with substance use disorders. Addiction. 2015;110(3):420–8. doi:10.1111/add.12799.

36. Higgins JPT, Thomas J, Chandler J, Cumpston M, Li T, Page MJ, et al., editors. Cochrane Handbook for Systematic Reviews of Interventions version 6.5 (updated August 2024) [Internet]. Cochrane; 2024. Available from: https://www.training.cochrane.org/handbook.

37. Tock WL, Maheu C, Johnson NA. Considerations of Control Conditions Designs in Randomized Controlled Trials of Exercise Interventions for Cancer Survivors. Can J Nurs Res. 2022;54(4):377–91. doi:10.1177/08445621211062467.

38. Miller CJ, Smith SN, Pugatch M. Experimental and quasi-experimental designs in implementation research. Psychiatry Res. 2020;283:112452. doi:10.1016/j.psychres.2019.06.027.

39. Lim C-Y, In J. Randomization in clinical studies. Korean J Anesthesiol. 2019;72(3):221–32. doi:10.4097/kja.19049.

40. Berger VW, Bour LJ, Carter K, Chipman JJ, Everett CC, Heussen N, et al. A roadmap to using randomization in clinical trials. BMC Med Res Methodol. 2021;21(1):168. doi:10.1186/s12874-021-01303-z.

41. Hariton E, Locascio JJ. Randomised controlled trials – the gold standard for effectiveness research: Study design: randomised controlled trials. BJOG. 2018;125(13):1716. doi:10.1111/1471-0528.15199.

42. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: Updated guidelines for reporting parallel group randomised trials. Int J Surg. 2012;10(1):28–55. doi:10.1016/j.ijsu.2011.10.001.

43. Moher D, Boutron I, Chan AW, Hopewell S, Hróbjartsson A, Schulz KF. Improving the completeness and transparency of protocols and reports of randomized trials: SPIRIT 2025 and CONSORT 2025. Journal of Clinical Epidemiology [Internet]. 2025 Apr ;111792. Available from: https://www.sciencedirect.com/science/article/pii/S0895435625001258

44. Hotopf M. The pragmatic randomised controlled trial. Advances in Psychiatric Treatment [Internet]. 2002 Sep ;8(5):326–33. Available from: https://www.cambridge.org/core/journals/advances-in-psychiatric-treatment/article/pragmatic-randomised-controlled-trial/043E0319536DD5FC02D4AF5D7AA55052.

45. Chan AW, Boutron I, Hopewell S, Moher D, Schulz KF, Collins GS, et al. SPIRIT 2025 statement: Updated guideline for protocols of randomised trials. PLoS Med. 2025 Apr 28;22(4):e1004589. doi: 10.1371/journal.pmed.1004589.

46. CONSORT Group. CONSORT 2025 statement: updated guideline for reporting randomised trials [Internet]. BMJ. 2025 Apr 14;389:e081123. doi:10.1136/bmj-2024-081123. Available from: https://www.bmj.com/content/389/bmj-2024-081123.

47. Clifford AM, Ni Bhriain O, Byrne S, Cheung PS, Louw Q, Glynn L, et al. Music and Movement for Health: Protocol for a pragmatic cluster-randomised feasibility pilot trial of an arts-based programme for the health and wellbeing of older adults. HRB Open Res. 2023 Mar 29;5:42. doi: 10.12688/hrbopenres.13535.2.

48. Quasi-experimental and Single-case Experimental Designs. Chapter thirteen [Internet]. Available from: https://us.sagepub.com/sites/default/files/upm-binaries/89876_Chapter_13_Quasi_Experimental_and_Single_Case_Designs.pdf

49. Welsing PM, Oude Rengerink K, Collier S, Eckert L, Van Smeden M, Ciaglia A, et al. Series: Pragmatic trials and real world evidence: Paper 6. Outcome measures in the real world. J Clin Epidemiol. 2017;90:99–107. doi:10.1016/j.jclinepi.2016.12.022

50. Public Health England. Arts for health and wellbeing – An evaluation framework [Internet]. 2016. Available from: https://assets.publishing.service.gov.uk/media/5c17bce1ed915d0c382e276b/PHE_Arts_and_Health_Evaluation_FINAL.pdf.

51. Ford I, Norrie J. Pragmatic Trials. N Engl J Med. 2016;375(5):454–63. doi:10.1056/NEJMra1510059.

52. Lurie JD, Morgan TS. Pros and cons of pragmatic clinical trials. J Comp Eff Res. 2013;2(1):53–8. doi:10.2217/cer.12.74.

53. Patsopoulos NA. A pragmatic view on pragmatic trials. Dialogues Clin Neurosci. 2011;13(2):217–24. doi:10.31887/DCNS.2011.13.2/npatsopoulos.

54. Kaimal G, Ray K, Muniz J. Reduction of Cortisol Levels and Participants' Responses Following Art Making. Art Ther. 2016;33(2):74–80. doi:10.1080/07421656.2016.1166832.

55. Mouriño-Ruiz R, Serral G, Continente X, López MJ, Lapena C, Puigpinós-Riera R. Evaluation of effectiveness of an art-based museum intervention in reducing loneliness among older adults (ArtGran): A quasi-experimental study. Public Health. 2024;230:149–56. doi:10.1016/j.puhe.2024.02.024.

56. Yin RK. Applications of case study research. 3rd ed. Washington (DC): SAGE Publications, Inc.; 2012. Chapter 1, A brief refresher on the case study method; p. 3–19; Available from: https://study.sagepub.com/sites/default/files/a_very_brief_refresher_on_the_case_study_method.pdf.

57. Villamin P, Lopez V, Thapa DK, Cleary M. A Worked Example of Qualitative Descriptive Design: A Step by Step Guide for Novice and Early Career Researchers. J Adv Nurs. 2024;jan.16481. doi:10.1111/jan.16481.

58. Personal Data Protection Commission. Guide to basic anonymisation [Internet]. Available from:https://www.pdpc.gov.sg/-/media/files/pdpc/pdf-files/advisory-guidelines/guide-to-basic-anonymisation-(updated-24-july-2024).pdf.

59. Nathan S, Hodgins M, Wirth J, Ramirez J, Walker N, Cullen P. The use of arts-based methodologies and methods with young people with complex psychosocial needs: A systematic narrative review. Health Expect. 2023;26(2):381–394. doi: 10.1111/hex.13705.

60. Vaughn LM, Jacquez F. Participatory Research Methods – Choice Points in the Research Process. Journal of Participatory Research Methods [Internet]. 2020 Jul 21;1(1). Available from: https://jprm.scholasticahq.com/article/13244-participatory-research-methods-choice-points-in-the-research-process.

61. McGrath C, Palmgren PJ, Liljedahl M. Twelve tips for conducting qualitative research interviews. Med Teach. 2019;41(9):1002–6. doi:10.1080/0142159X.2018.1497149.

62. Thomas SL, Pitt H, McCarthy S, Arnot G, Hennessy M. Methodological and practical guidance for designing and conducting online qualitative surveys in public health. Health Promot Int. 2024;39(3):daae061. doi:10.1093/heapro/daae061.

63. World Health Organization. Mental health [Internet]. Geneva: World Health Organization; 2022. Available from: https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response.

64. Malhi GS, Mann JJ. Depression. Lancet. 2018 Nov 24;392(10161):2299-312. doi:10.1016/S0140-6736(18)31948-2.

65. Beck AT, Ward CH, Mendelson M, Mock J, Erbaugh J. An Inventory for Measuring Depression. Archives of General Psychiatry [Internet]. 1961 Jun 1;4(6):561–71. Available from: https://jamanetwork.com/journals/jamapsychiatry/article-abstract/487993.

66. Radloff LS. The CES-D scale: a self-report depression scale for research in the general population. Appl Psychol Meas. 1977 Jun;1(3):385-401. doi:10.1177/014662167700100306.

67. Zigmond AS, Snaith RP. The hospital anxiety and depression scale. Acta Psychiatr Scand [Internet]. 1983 Jun;67(6):361-70. Available from: https://pubmed.ncbi.nlm.nih.gov/6880820/.

68. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606-13. doi:10.1046/j.1525-1497.2001.016009606.x.

69. Lovibond SH, Lovibond PF. Depression Anxiety Stress Scales [Internet]. psycnet.apa.org. 1995. Available from: https://psycnet.apa.org/doiLanding?doi=10.1037%2Ft01004-000.

70. Kazdin AE, editor. Encyclopedia of psychology. 8 vols. Washington (DC): American Psychological Association; 2000. Available from: https://www.apa.org/pubs/books/4600100.

71. Beck AT, Epstein N, Brown G, Steer RA. An inventory for measuring clinical anxiety: psychometric properties. J Consult Clin Psychol. 1988;56(6):893-7. doi:10.1037/0022-006X.56.6.893.

72. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A Brief Measure for Assessing Generalized Anxiety disorder: the GAD-7. Archives of Internal Medicine [Internet]. 2006 May 22;166(10):1092–7. Available from: https://jamanetwork.com/journals/jamainternalmedicine/fullarticle/410326.

73. Spielberger CD, Gonzalez-Reigosa F, Martinez-Urrutia A, Natalicio LF, Natalicio DS. The State-Trait Anxiety Inventory. RIP/IJP [Internet]. 2017 Jul. 17;5(3 & 4). Available from: https://journal.sipsych.org/index.php/IJP/article/view/620.

74. World Health Organization (WHO). Mental disorders [Internet]. Geneva: World Health Organization; 2022. Available from: https://www.who.int/news-room/fact-sheets/detail/mental-disorders.

75. Felce D, Perry J. Quality of life: Its definition and measurement. Res Dev Disabil. 1995 Jan;16(1):51-74. doi:10.1016/0891-4222(94)00028-8.

76. Lee CW, Lin LC, Hung HC. Art and cultural participation and life satisfaction in adults: the role of physical health, mental health, and interpersonal relationships. Front Public Health. 2021 Jan 21;8. doi:10.3389/fpubh.2020.582342.

77. Jarden A, Roache A. What is wellbeing? Int J Environ Res Public Health [Internet]. 2023 Mar 12;20(6). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10049282/.

78. Teoli D, Bhardwaj A. Quality of Life [Internet]. PubMed. Treasure Island (FL): StatPearls Publishing; 2023. Available from: https://www.ncbi.nlm.nih.gov/books/NBK536962/.

79. Coulton S, Clift S, Skingley A, Rodriguez J. Effectiveness and cost-effectiveness of community singing on mental health-related quality of life of older people: Randomised controlled trial. British Journal of Psychiatry. 2015;207(3):250–5. doi:10.1192/bjp.bp.113.129908.

80. Michaelson J, Mahony S, Schifferes J. Measuring Well-being A guide for practitioners [Internet]. 2012. Available from: https://neweconomics.org/uploads/files/measuring-wellbeing.pdf.

81. Ruggeri K, Garcia-Garzon E, Maguire Á, Matz S, Huppert FA. Well-being is more than happiness and life satisfaction: A multidimensional analysis of 21 countries. Health and Quality of Life Outcomes [Internet]. 2020 Jun 19;18(192):1–16. Available from: https://hqlo.biomedcentral.com/articles/10.1186/s12955-020-01423-y.

82. Huta V. An overview of hedonic and eudaimonic well-being concepts. In: Reinecke L, Oliver MB, editors. The Routledge handbook of media use and well-being. London: Routledge; 2016. p. 14-33.

83. Stone AA, Mackie C, editors. Subjective well-being: measuring happiness, suffering, and other dimensions of experience. Washington, DC: National Academies Press; 2013.doi:10.17226/18548

84. Zhang W, Balloo K, Hosein A, Medland E. A scoping review of well-being measures: conceptualisation and scales for overall well-being. BMC Psychol. 2024 Oct 23;12(1). doi:10.1186/s40359-024-02074-0.

85. Mahoney J. Subjective well-being measurement: Current practice and new frontiers [Internet]. OECD. 2023. Available from: https://www.oecd.org/en/publications/subjective-well-being-measurement_4e180f51-en.html.

86. Devlin NJ, Brooks R. EQ-5D and the EuroQol Group: past, present and future. Appl Health Econ Health Policy. 2017 Feb 13;15(2):127-37. doi:10.1007/s40258-017-0310-5.

87. Ware JE, Snow KK, Kosinski M, Gandek B. SF-36 Health Survey: Manual and Interpretation Guide. 2nd ed. Boston (MA): The Health Institute, New England Medical Center; 1993.

88. Ware JE, Keller SD, Kosinski M. SF-12: How to score the SF-12 physical and mental health summary scales [Internet]. Boston (MA): The Health Institute, New England Medical Center; 1995. Available from:https://www.researchgate.net/profile/John-Ware-6/publication/291994160_How_to_score_SF-12_items/links/58dfc42f92851c369548e04e/How-to-score-SF-12-items.pdf.

89. Whoqol Group. Development of the World Health Organization WHOQOL-BREF quality of life assessment. Psychol Med. 1998 May;28(3):551-8. doi:10.1017/S0033291798006667.

90. Frisch MB, Cornell J, Villanueva M, Retzlaff PJ. Quality of Life Inventory (QOLI) [Database record]. APA PsycTests; 1988. doi:10.1037/t03748-000.

91. Xu RH, Sun R, Tian L, Cheung AW, Wong EL. Health-related quality of life in primary care patients: a comparison between EQ-5D-5L utility score and EQ-visual analogue scale. Health and quality of life outcomes [Internet]. 2024 Jan 3;22(1). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10765691/.

92. Devlin N, Parkin D, Janssen B. Analysis of EQ-5D Values [Internet]. www.ncbi.nlm.nih.gov. Springer; 2020. Available from: https://www.ncbi.nlm.nih.gov/books/NBK565677/.

93. Lins L, Carvalho FM. SF-36 total score as a single measure of health-related quality of life: Scoping review. SAGE Open Medicine [Internet]. 2016 Oct 4;4(4):1–12. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5052926/.

94. Contopoulos-Ioannidis DG, Karvouni A, Kouri I, Ioannidis JPA. Reporting and interpretation of SF-36 outcomes in randomised trials: systematic review. BMJ. 2009 Jan 12;338(jan12 1):a3006–6.doi: 10.1136/bmj.a3006.

95. Huo T, Guo Y, Shenkman E, Muller K. Assessing the reliability of the short form 12 (SF-12) health survey in adults with mental health conditions: a report from the wellness incentive and navigation (WIN) study. Health and Quality of Life Outcomes [Internet]. 2018 Feb 13;16(34). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5811954/.

96. Frisch MB. Quality-of-life-inventory. In: Michalos AC, editor. Encyclopedia of quality of life and well-being research. Cham: Springer International Publishing; 2024. p. 5794–7.

97. Tov W, Keh JS, Tan YQ, Tan QYJ, IAISB Aziz. Assessing subjective well-being: A review of common measures. Handbook of Positive Psychology Assessment[Internet]. Hogrefe. 2022. Available from: https://ink.library.smu.edu.sg/soss_research/3729.

98. Pavot W, Diener E. Review of the Satisfaction With Life Scale. Psychol Assess. 1993;5(2):164–172. doi:10.1037/1040-3590.5.2.164.

99. Hinz A, Mehnert-Theuerkauf A, Glaesmer H, Schroeter ML, Tibubos AN, Petrowski K, et al. Changes in life satisfaction over six years in the general population: a longitudinal study with the Satisfaction With Life Scale (SWLS). PLoS One [Internet]. 2025 Jan 13;20(1):e0316990-0. Available from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0316990

100. Diener E, Emmons RA, Larsen RJ, Griffin S. The satisfaction with life scale. J Pers Assess [Internet]. 1985;49(1):71-5. Available from: https://www.tandfonline.com/doi/abs/10.1207/s15327752jpa4901_13.

101. Chinni ML, Hubley AM. A research synthesis of validation practices used to evaluate the Satisfaction with Life Scale (SWLS). Validity and Validation in Social, Behavioral, and Health Sciences [Internet]. 2014 Jan 1;35-66. Available from: https://link.springer.com/chapter/10.1007/978-3-319-07794-9_4

102. Nilsson AH, Eichstaedt JC, Lomas T, Schwartz A, Kjell O. The Cantril Ladder elicits thoughts about power and wealth. Scientific Reports [Internet]. 2024 Feb 1;14(1):2642. Available from: https://www.nature.com/articles/s41598-024-52939-y#:~:text=The%20Cantril%20Ladder1%20is

103. Mohammed SEE, Warner G. The Cantril Ladder as a measure of well-being and life satisfaction among refugee youth experiencing symptoms of post-traumatic stress. J Immigr Minor Health. 2024 Jun;26(3):539-545. doi:10.1007/s10903-023-01563-5.

104. Ong ZX, Dowthwaite L, Perez Vallejos E, Rawsthorne M, Long Y. Measuring online wellbeing: a scoping review of subjective wellbeing measures. Front Psychol. 2021 Mar 11;12:616637. doi:10.3389/fpsyg.2021.616637.

105. Lindert J, Bain PA, Kubzansky LD, Stein C. Well-being measurement and the WHO health policy Health 2010: systematic review of measurement scales. The European Journal of Public Health [Internet]. 2015 Jan 22;25(4):731–40. Available from: https://academic.oup.com/eurpub/article/25/4/731/2398958.

106. Jarden R, Siegert RJ, Koziol-McLain J, Bujalka H, Sandham MH. Wellbeing measures for workers: a systematic review and methodological quality appraisal. Frontiers in Public Health [Internet]. 2023 May 24;11. Available from: https://www.frontiersin.org/journals/public-health/articles/10.3389/fpubh.2023.1053179/full.

107. Linton MJ, Dieppe P, Medina-Lara A. Review of 99 self-report measures for assessing well-being in adults: exploring dimensions of well-being and developments over time. BMJ Open [Internet]. 2016 Jul ;6(7):e010641. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4947747/.

108. Tennant R, Hiller L, Fishwick R, Platt S, Joseph S, Weich S, et al. The Warwick-Edinburgh Mental Well-Being Scale (WEMWBS): development and UK validation. Health Qual Life Outcomes [Internet]. 2007;5(1):63. Available from: http://www.hqlo.com/content/5/1/63.

109. Mortazavi F, Mousavi SA, Chaman R, Khosravi A. Validation of the World Health Organization-5 Well-Being Index; assessment of maternal well-being and its associated factors. Turk J Psychiatry [Internet]. 2013. Available from: https://turkpsikiyatri.com/Data/UnpublishedArticles/b8a5by.pdf.

110. International Wellbeing Group. Personal Wellbeing Index Manual: 6th Edition, Version 2, 190624. Cummins RA, editor. Geelong: Australian Centre on Quality of Life, School of Psychology, Deakin University – Melbourne Campus; 2024. p. 1-55.

111. Springer KW, Hauser RM. An assessment of the construct validity of Ryff's Scales of Psychological Well-Being: Method, mode, and measurement effects. Social Science Research [Internet]. 2006 Dec;35(4):1080–102. Available from: https://www.ssc.wisc.edu/~hauser/Springer_Hauser_PWB_MS_SSR_081805.pdf.

112. Butler J, Kern ML. The PERMA-Profiler: a brief multidimensional measure of flourishing. Int J Wellbeing. 2016 Oct 13;6(3):1-48. doi:10.5502/ijw.v6i3.526.

113. Thomson LJ, Chatterjee HJ. Measuring the impact of museum activities on well-being: developing the Museum Well-being Measures Toolkit. Museum Manag Curatorship. 2015 Jan;30(1):44-62. doi:10.1080/09647775.2015.1008390.

114. World Health Organization. Health and well-being [Internet]. Geneva: World Health Organization; 2022. Available from: https://www.who.int/data/gho/data/major-themes/health-and-well-being.

115. Adepoju P. WHO Member States Recognize Social Connection As A Global Health Priority - Health Policy Watch [Internet]. Health Policy Watch; 2025. Available from: https://healthpolicy-watch.news/who-member-states-recognize-social-connection-as-a-global-health-priority/.

116. Wickramaratne PJ, Yangchen T, Lepow L, et al. Social connectedness as a determinant of mental health: A scoping review. PLoS One. 2024 Nov 15;19(11):e0314220. doi:10.1371/journal.pone.0314220.

117. Holt-Lunstad J, Smith TB, Baker M, Harris T, Stephenson D. Loneliness and social isolation as risk factors for mortality: a meta-analytic review. Perspect Psychol Sci. 2015 Mar;10(2):227-37. doi:10.1177/1745691614568352.

118. Cardona M, Andrés P. Are social isolation and loneliness associated with cognitive decline in ageing? Front Aging Neurosci. 2023;15:1075563. doi:10.3389/fnagi.2023.1075563.

119. Sonke J, Pesata V, Colverson A, Morgan-Daniel J, Rodriguez AK, Carroll GD, et al. Relationships between arts participation, social cohesion, and well-being: an integrative review of evidence. Front Public Health. 2025 May 9;13:1589693. doi:10.3389/fpubh.2025.1589693.

120. Seppala E, Rossomando T, Doty JR. Social connection and compassion: Important predictors of health and well-being. Social Research: An International Quarterly. 2013;80(2):411-30. doi:10.1353/sor.2013.0027.

121. Holt-Lunstad J, Robles TF, Sbarra DA. Advancing social connection as a public health priority in the United States. Am Psychol. 2017 Sep;72(6):517. doi:10.1037/amp0000103.

122. Smith ML, Racoosin J, Wilkerson R, et al. Societal- and community-level strategies to improve social connectedness among older adults. Front Public Health. 2023;11:1176895. Published 2023 May 4. doi:10.3389/fpubh.2023.1176895.

123. Holt-Lunstad J. The Potential Public Health Relevance of Social Isolation and Loneliness: Prevalence, Epidemiology, and Risk Factors. Public Policy & Aging Report [Internet]. 2017;27(4):127–30. doi:10.1093/ppar/prx030.

124. De Jong Gierveld J, Van Tilburg T, Dykstra PA. Loneliness and Social Isolation. In: Vangelisti AL, Perlman D, eds. The Cambridge Handbook of Personal Relationships. Cambridge Handbooks in Psychology. Cambridge University Press; 2006:485-500. doi: 10.1017/CBO9780511606632.027.

125. Coyle CE, Dugan E. Social isolation, loneliness and health among older adults. J Aging Health. 2012 Dec;24(8):1346-63. doi:10.1177/0898264312460275.

126. Cordier R, Milbourn B, Martin R, Buchanan A, Chung D, Speyer R. A systematic review evaluating the psychometric properties of measures of social inclusion. PloS one. 2017 Jun 9;12(6):e0179109. doi: 10.1371/journal.pone.0179109.

127. Too EK, Chongwo E, Mabrouk A, Abubakar A. Adolescent connectedness: a scoping review of available measures and their psychometric properties. Frontiers in Psychology. 2022 May 18;13:856621. doi:10.3389/fpsyg.2022.856621.

128. Plackett R, Hulin J, Mukuria C, Clowes M, Ramsay SE, Spencer L, et al. Measures of social connectedness in adult populations: a systematic review. BMC Public Health. 2024 Dec 5;24(1). doi:10.1186/s12889-024-20779-0.

129. Boateng GO, Wachter K, Schuster RC, Burgess TL, Bunn M. A scoping review of instruments used in measuring social support among refugees in resettlement. International journal of environmental research and public health. 2024 Jun 20;21(6):805. doi:10.3390/ijerph21060805.

130. William S, McDonagh J, Allida SM, Kasa AS, Patterson C, Deek H, Moxham L, Wand T, Ferguson C. A systematic review to identify assessment instruments for social isolation or loneliness in adults with heart failure. npj Cardiovascular Health. 2025 Mar 7;2(1):10. doi:10.1038/s44325-025-00044-y.

131. Hinz A, Esser P, Richter D, Schönfelder A, Glaesmer H, Lehmann-Laue A, Heyne S, Leuteritz K, Sender A, Springer F, Götze H. Social support in a large general population sample over the course of six years. Scientific Reports. 2025 Feb 21;15(1):6405.doi: 10.1038/s41598-025-90703-y.

132. Lee RM, Draper M, Lee S. Social Connectedness, Dysfunctional Interpersonal Behaviors, and Psychological Distress. J Couns Psychol. 2001;48(3):310-8. doi:10.1037/0022-0167.48.3.310.

133. Zimet GD, Dahlem NW, Zimet SG, Farley GK. The multidimensional scale of perceived social support. Journal of personality assessment. 1988 Mar 1;52(1):30-41. doi:10.1207/s15327752jpa5201_2.

134. Caron J. A validation of the Social Provisions Scale: the SPS-10 items [article in French]. Sante Ment Que. 2013;38(1):297-318. doi:10.7202/1019198ar.

135. Koenig HG, Westlund RE, George LK, Hughes DC, Blazer DG, Hybels C. Abbreviating the Duke Social Support Index for use in chronically ill elderly individuals. Psychosomatics. 1993 Jan;34(1):61-9. doi:10.1016/S0033-3182(93)71928-3.

136. Majmudar IK, Mihalopoulos C, Brijnath B, Lim MH, Hall NY, Engel L. The impact of loneliness and social isolation on health state utility values: a systematic literature review. Quality of Life Research. 2022 Jan 24;31. doi:10.1007/s11136-021-03063-1.

137. Hoang P, King JA, Moore S, Moore K, Reich K, Sidhu H, et al. Interventions Associated With Reduced Loneliness and Social Isolation in Older Adults: A Systematic Review and Meta-analysis. JAMA Network Open. 2022 Oct 17;5(10). doi:10.1001/jamanetworkopen.2022.36676.

138. Maes M, Qualter P, Lodder GMA, Mund M. How (Not) to Measure Loneliness: A Review of the Eight Most Commonly Used Scales. International Journal of Environmental Research and Public Health. 2022 Aug 30;19(17):10816. doi:10.3390/ijerph191710816.

139. Bugallo-Carrera C, Dosil-Díaz C, Anido-Rifón L, Pacheco-Lorenzo M, Fernández-Iglesias MJ, Gandoy-Crego M. A systematic review evaluating loneliness assessment instruments in older adults. Frontiers in Psychology. 2023 Apr 25;14. doi:10.3389/fpsyg.2023.1101462.

140. Gardiner C, Geldenhuys G, Gott M. Interventions to reduce social isolation and loneliness among older people: An integrative review. Health & Social Care in the Community. 2018;26(2):147–57. doi: 10.1111/hsc.12367.

141. Russell DW. UCLA Loneliness Scale (Version 3): Reliability, validity, and factor structure. J Pers Assess. 1996 Feb;66(1):20-40. doi:10.1207/s15327752jpa6601_2.

142. De Jong-Gierveld J, Van Tilburg TG. A 6-item scale for overall, emotional, and social loneliness: Confirmatory tests on survey data. Res Aging. 2006 Sep;28(5):582-98. doi:10.1177/0164027506289723.

143. Lubben J, Blozik E, Gillmann G, Iliffe S, von Renteln Kruse W, Beck JC, Stuck AE. Performance of an abbreviated version of the Lubben Social Network Scale among three European Community–dwelling older adult populations. Gerontologist. 2006;46(4):503-13. Doi: 10.1093/geront/46.4.503.

144. Lok I, Dunn EW. The UBC State Social Connection Scale: Factor structure, reliability, and validity. Soc Psychol Personal Sci. 2023;14(7):835-44. doi:10.1177/19485506221132090.

145. Lok I. Harnessing social psychology to promote social interactions [dissertation]. Vancouver: University of British Columbia; 2022. Available from: https://open.library.ubc.ca/media/stream/pdf/24/1.0431103/4.

146. Webster JG, editor. The physiological measurement handbook. CRC Press; 2014 Dec 11.

147. Fancourt D, Aughterson H, Finn S, Walker E, Steptoe A. How leisure activities affect health: A narrative review and multi-level theoretical framework of mechanisms of action. The Lancet Psychiatry [Internet]. 2021 Feb 11;8(4):329–39. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7613155/

148. Sabio J, Williams NS, McArthur GM, Badcock NA. A scoping review on the use of consumer-grade EEG devices for research. PLoS One. 2024;19(3):e0291186. Published 2024 Mar 6. doi:10.1371/journal.pone.0291186

149. Bunde A. The different types of noise and how they effect data analysis. Chemie Ingenieur Technik. 2023 Nov;95(11):1758-67. doi: 10.1002/cite.202300031

150. Caspersen CJ, Powell KE, Christenson GM. Physical activity, exercise, and physical fitness: definitions and distinctions for health-related research. Public Health Reports [Internet]. 2024;100(2):126. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC1424733/?page=3.

151. Fong Yan A, Cobley S, Chan C, Pappas E, Nicholson LL, Ward RE, et al. The Effectiveness of Dance Interventions on Physical Health Outcomes Compared to Other Forms of Physical Activity: A Systematic Review and Meta-Analysis. Sports Medicine [Internet]. 2017 Dec 21;48(4):933–51. Available from: https://link.springer.com/article/10.1007%2Fs40279-017-0853-5.

152. World Health Organization. Physical activity [Internet]. Geneva: World Health Organization. 2024. Available from: https://www.who.int/news-room/fact-sheets/detail/physical-activity.

153. Kraus WE, Janz KF, Powell KE, Campbell WW, Jakicic JM, Troiano RP, et al. Daily Step Counts for Measuring Physical Activity Exposure and Its Relation to Health. Medicine & Science in Sports & Exercise [Internet]. 2019 Jun;51(6):1206–12. Available from: https://journals.lww.com/acsm-msse/Fulltext/2019/06000/Daily_Step_Counts_for_Measuring_Physical_Activity.15.aspx.

154. Bassett DR, Toth LP, LaMunion SR, Crouter SE. Step Counting: A Review of Measurement Considerations and Health-Related Applications. Sports Medicine [Internet]. 2016 Dec 22;47(7):1303–15. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5488109/.

155. Keating XD, Zhou K, Liu X, Hodges M, Liu J, Guan J, et al. Reliability and Concurrent Validity of Global Physical Activity Questionnaire (GPAQ): A Systematic Review. International Journal of Environmental Research and Public Health [Internet]. 2019 Oct 26;16(21):4128. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6862218/.

156. Sylvia LG, Bernstein EE, Hubbard JL, Keating L, Anderson EJ. Practical Guide to Measuring Physical Activity. Journal of the Academy of Nutrition and Dietetics [Internet]. 2014 Feb;114(2):199–208. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3915355/

157. Raghuveer G, Hartz J, Lubans DR, Takken T, Wiltz JL, Mietus-Snyder M, et al. Cardiorespiratory Fitness in Youth: An Important Marker of Health: A Scientific Statement From the American Heart Association. Circulation [Internet]. 2020 Aug 18;142(7). Available from: https://www.ahajournals.org/doi/full/10.1161/CIR.0000000000000866.

158. Lang JJ, Prince SA, Merucci K, Cadenas-Sanchez C, Chaput JP, Fraser BJ, et al. Cardiorespiratory fitness is a strong and consistent predictor of morbidity and mortality among adults: an overview of meta-analyses representing over 20.9 million observations from 199 unique cohort studies. British Journal of Sports Medicine [Internet]. 2024 Apr 9;58(10):bjsports-2023-107849. Available from: https://bjsm.bmj.com/content/bjsports/early/2024/04/09/bjsports-2023-107849.full.pdf.

159. Kokkinos P. Cardiorespiratory Fitness, Exercise, and Blood Pressure. Hypertension [Internet]. 2014 Dec;64(6):1160–4. Available from: https://www.ahajournals.org/doi/10.1161/hypertensionaha.114.03616.

160. Gonzales TI, Jeon JY, Lindsay T, Westgate K, Perez-Pozuelo I, Hollidge S, et al. Resting heart rate is a population-level biomarker of cardiorespiratory fitness: The Fenland Study. PLoS One. 2023 May 11;18(5):e0285272. doi: 10.1371/journal.pone.0285272.

161. Reimers A, Knapp G, Reimers CD. Effects of Exercise on the Resting Heart Rate: a Systematic Review and Meta-Analysis of Interventional Studies. Journal of Clinical Medicine [Internet]. 2018 Dec 1;7(12):503. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6306777/.

162. Singapore Heart Foundation. High Blood Pressure | Singapore Heart Foundation [Internet]. 2022. Available from: https://www.myheart.org.sg/health/risk-factors/high-blood-pressure/

163. Buttar KK, Saboo N, Kacker S. A review: Maximal oxygen uptake (VO2 max) and its estimation methods. International Journal of Physical Education, Sports and Health [Internet]. 2019;6(6):24–32. Available from: https://www.kheljournal.com/archives/2019/vol6issue6/PartA/6-5-57-922.pdf.

164. Hallemans A, Verbeque E, Van de Walle P. Motor functions. InHandbook of clinical neurology 2020 Jan 1 (Vol. 173, pp. 157-170). Elsevier. doi: 10.1016/B978-0-444-64150-2.00015-0

165. Buchman AS, Leurgans SE, Boyle PA, Schneider JA, Arnold SE, Bennett DA. Combinations of motor measures more strongly predict adverse health outcomes in old age: the rush memory and aging project, a community-based cohort study. BMC medicine. 2011 Dec;9:1-1. doi:10.1186/1741-7015-9-42

166. Herman T, Giladi N, Hausdorff JM. Properties of the 'timed up and go' test: more than meets the eye. Gerontology. 2011;57(3):203-210. doi:10.1159/000314963

167. Christopher A, Kraft E, Olenick H, Kiesling R, Doty A. The reliability and validity of the Timed Up and Go as a clinical tool in individuals with and without disabilities across a lifespan: A systematic review: Psychometric properties of the Timed Up and Go. Disability and rehabilitation. 2021 Jun 19;43(13):1799-813. doi: 10.1080/09638288.2019.1682066

168. Sanders DB, Guptill JT, Aleš KL, Hobson-Webb LD, Jacobus DP, Mahmood R, Massey JM, Pittman MM, Prather K, Raja SM, Yow E, Juel VC. Reliability of the triple-timed up-and-go test. Muscle Nerve. 2018 Jan;57(1):136–139. doi:10.1002/mus.25700.

169. Viveiro LA, Gomes GC, Bacha JM, Junior NC, Kallas ME, Reis M, Jacob Filho W, Pompeu JE. Reliability, validity, and ability to identity fall status of the Berg Balance Scale, Balance Evaluation Systems Test (BESTest), Mini-BESTest, and Brief-BESTest in older adults who live in nursing homes. Journal of geriatric physical therapy. 2019 Oct 1;42(4):E45-54. doi: 10.1519/JPT.0000000000000215

170. Muir SW, Berg K, Chesworth B, Speechley M. Use of the Berg Balance Scale for predicting multiple falls in community-dwelling elderly people: a prospective study. Physical therapy. 2008 Apr 1;88(4):449-59. doi: 10.2522/ptj.20070251

171. Bohannon RW. Hand-grip dynamometry predicts future outcomes in aging adults. J Geriatr Phys Ther. 2008;31(1):3–10. doi: 10.1519/00139143-200831010-00002.

172. Vaishya R, Misra A, Vaish A, Ursino N, D'Ambrosi R. Hand grip strength as a proposed new vital sign of health: a narrative review of evidences. Journal of Health, Population and Nutrition [Internet]. 2024 Jan 9;43(1). Available from: https://jhpn.biomedcentral.com/articles/10.1186/s41043-024-00500-

173. Biasini NR, Bannon B, Pellegrino MJ, Qaderi A, Trinh W, Switzer-McIntyre S, et al. Reliability and Validity of Shoulder and Handgrip Strength Testing. Physiotherapy Canada [Internet]. 2023 Feb 1;75(1):65–71. doi: 10.3138/ptc-2021-0041

174. Worley SL. The Extraordinary Importance of Sleep: The Detrimental Effects of Inadequate Sleep on Health and Public Safety Drive an Explosion of Sleep Research. P T. 2018;43(12):758-763.

175. Besedovsky L, Lange T, Born J. Sleep and immune function. Pflügers Archiv-European Journal of Physiology. 2012 Jan;463(1):121-37.

176. Sharma S, Kavuru M. Sleep and metabolism: an overview. International journal of endocrinology. 2010;2010(1):270832. doi: 10.1155/2010/270832

177. Cappuccio FP, D'Elia L, Strazzullo P, Miller MA. Sleep duration and all-cause mortality: a systematic review and meta-analysis of prospective studies. Sleep. 2010 May 1;33(5):585-92. doi: 10.1093/sleep/33.5.585

178. Buysse DJ, Reynolds CF, Monk TH, Berman SR, Kupfer DJ. The pittsburgh sleep quality index: A new instrument for psychiatric practice and research. Psychiatry Research [Internet]. 1989 May;28(2):193–213. Available from: https://www.sciencedirect.com/science/article/pii/0165178189900474?via%3Dihub.

179. Fabbri M, Beracci A, Martoni M, Meneo D, Tonetti L, Natale V. Measuring Subjective Sleep Quality: A Review. International Journal of Environmental Research and Public Health [Internet]. 2021 Jan 26;18(3):1082. Available from: https://www.mdpi.com/1660-4601/18/3/1082.

180. Fekedulegn D, Andrew ME, Shi M, Violanti JM, Knox S, Innes KE. Actigraphy-Based Assessment of Sleep Parameters. Annals of Work Exposures and Health [Internet]. 2020 Feb 13;64(4):350–67. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC7191872/.

181. Schneiderman N, Ironson G, Siegel SD. Stress and health: Psychological, behavioral, and Biological Determinants. Annual Review of Clinical Psychology [Internet]. 2005;1(1):607–28. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC2568977/

182. Burke HM, Davis MC, Otte C, Mohr DC. Depression and cortisol responses to psychological stress: A meta-analysis. Psychoneuroendocrinology [Internet]. 2005 Oct;30(9):846–56. Available from: https://www.sciencedirect.com/science/article/pii/S0306453005000831

183. Hellhammer DH, Wüst S, Kudielka BM. Salivary cortisol as a biomarker in stress research. Psychoneuroendocrinology [Internet]. 2009 Feb;34(2):163–71. Available from: https://www.sciencedirect.com/science/article/abs/pii/S0306453008002990.

184. Kaimal G, Ray K, Muniz J. Reduction of Cortisol Levels and Participants' Responses Following Art Making. Art Therapy [Internet]. 2016 May 23;33(2):74–80. Available from: https://www.tandfonline.com/doi/full/10.1080/07421656.2016.1166832

185. Fey JMH, Bikker FJ, Hesse D. Saliva collection methods among children and adolescents: a scoping review. Mol Diagn Ther. 2024 Jan;28(1):15–26. doi:10.1007/s40291-023-00684-9

186. Mohd Azmi NAS, Juliana N, Azmani S, Mohd Effendy N, Abu IF, Mohd Fahmi Teng NI, et al. Cortisol on Circadian Rhythm and Its Effect on Cardiovascular System. International Journal of Environmental Research and Public Health [Internet]. 2021 Jan 14;18(2):676. Available from: https://www.mdpi.com/1660-4601/18/2/676

187. Taelman J, Vandeput S, Spaepen A, Van Huffel S. Influence of Mental Stress on Heart Rate and Heart Rate Variability. IFMBE Proceedings [Internet]. 2009;22:1366–9. Available from: https://link.springer.com/chapter/10.1007/978-3-540-89208-3_324

188. Kim HG, Cheon EJ, Bai DS, Lee YH, Koo BH. Stress and Heart Rate Variability: A Meta-Analysis and Review of the Literature. Psychiatry Investigation [Internet]. 2018 Mar 25;15(3):235–45. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5900369/

189. Liu J, Sui X, Lavie CJ, Zhou H, Mark Park YM, Cai B, et al. Effects of Cardiorespiratory Fitness on Blood Pressure Trajectory With Aging in a Cohort of Healthy Men. Journal of the American College of Cardiology. 2014 Sep;64(12):1245–53. doi: 10.1016/j.jacc.2014.06.1184.

190. Gordon AM, Mendes WB. A large-scale study of stress, emotions, and blood pressure in daily life using a digital platform. Proceedings of the National Academy of Sciences [Internet]. 2021 Aug 3;118(31). Available from: https://www.pnas.org/content/118/31/e2105573118

191. Mastandrea S, Maricchiolo F, Carrus G, Giovannelli I, Giuliani V, Berardi D. Visits to figurative art museums may lower blood pressure and stress. Arts & Health. 2018 Mar 5;11(2):123–32. doi: 10.1080/17533015.2018.1443953.

192. Ilya N, Liu J, Agres K. Identifying and recommending validated measures to assess depression and anxiety outcomes in the field of Arts and Health. OSF [Internet]. 2025 Jun 4; Available from: https://osf.io/t97ky/