Social Service Research Centre
Faculty of Arts & Social Sciences

A Guide to Using Observation Rubrics for Evaluating Group Work Programmes

by Dr. Seah Lay Hoon, NUS Social Service Research Centre

Lui Hui Min, Fionn and Ang Jingyou, Ben; Thye Hua Kwan Moral Charities

Authors

1) Dr. Seah Lay Hoon

Social Service Research Centre, NUS

2) Lui Hui Min, Fionn

Thye Hua Kwan Moral Charities

3) Ang Jingyou, Ben

Thye Hua Kwan Moral Charities

Use the following to cite this publication:

Seah, L.H., Lui, H.M.F, Ang, J.B. (2024). A Guide to Using Observation Rubrics for Evaluating Group Work Programmes. National University of Singapore Social Service Research Centre

ABSTRACT

This research guide seeks to illustrate how observational data can be used to monitor and evaluate social service programmes, particularly group work. The guide will provide information on observation as a research method and how observational data, particularly in the form of rubrics, can be used to complement and triangulate other methods of data collection (e.g., pre- and post-self-reported survey) when monitoring the progress of group members and group processes. Using a group work support programme for men who engaged in family aggression as a case example, the guide will demonstrate how the observational indicators can be developed and used to support the observation of a programme that is aligned with its session objectives.

Introduction

Assessing behaviours is often an important part of monitoring and evaluating the outcomes of social service programmes, apart from assessing knowledge, skills, or attitudes of the participants. The go-to method to measuring behavioural change is often the pre- and post-self-reported survey as it is relatively easy to administer and is cost-effective. However, as with any evaluation method, it has its inherent limitations. This includes respondents who may not be conscious of their own behaviours. Their self- assessments are also prone to biases that have been well-documented by previous studies (Fix et al., 2022). Research has also documented disparities and contradictions between what respondents reported they believe or do and their observed behaviours (Mack et al., 2005). Direct observation can thus provide a powerful check against what participants report about themselves during surveys and interviews. Studies have also demonstrated the insights that can be gleaned from observation of intervention practices, especially insights that may be hard for participants to articulate (e.g., Evans & Lambert, 2008).

Naturalistic observation, with roots in traditional ethnographic research, is a practice that is part and parcel of every programme practitioner but often overlooked as a well- grounded evaluation method. As a way of gathering data in naturalistic settings, observation is done in a well-planned, systematic, and principled manner, with the observer knowing what and who to observe, why and when to conduct the observation, and the manner in which to document the observations (Centers for Disease Control and Prevention (CDC), 2008). A well-planned approach entails not just planning how behaviours are observed, but also how the observational data are organised and analysed.

In this guide, we describe the processes involved in planning and implementing observation for the purpose of evaluating group work. Evaluation here refers to ‘the process of obtaining formative, summative, or evaluative information about the group’ (Toseland & Rivas, 2017, p. 433). In particular, we zoom in on using observation to keep track of the progress of group members and group processes. Using the Brotherhood programme — a men’s groupwork programme aimed at family violence reduction — as a case study, we will demonstrate how observational rubrics provide a data source that enable not only consistent monitoring of the programme but also enhance its sustainability over the long run.

What is the Observation Method?

Observation is commonly used in qualitative research, which generates culturally specific and contextually rich data to provide insights into the perspective of participants. Effectiveness of social service interventions is contingent on how well they can account for socio-behavioural factors such as gender norms, ethnic identities, cultural norms, socio-economic status and stigma. Observational data can thus be useful to supplement quantitative outcome data (e.g., those recommended in NCSS's Sector Evaluation Framework) to gain a more complex and richer understanding of the workings of a programme in its localised context.

As a method, it allows you to directly document what people do, where and when an activity is occurring, rather than depend on what they say they did and does not rely on participants' ability to provide information (CDC, 2008). Suitable for a broad range of contexts, it is a less invasive data collection approach especially if the observer is able to blend into the observation setting and has minimal effect on the activities. It can be especially useful when sample size is small - which is typical of many group work programmes – as survey data may not yield significant results. Studies such as the ethnographic evaluation of an HIV (human immunodeficiency virus) intervention conducted by Evans and Lambert (2008, p. 467) have found that 'observation of actual intervention practices can reveal insights that may be hard for project staff to articulate or difficult to pinpoint, and can highlight important points of divergence and convergence from intervention theory or planning documents.'

Observation may be conducted by an evaluator who is either involved in the programme (participant observation) or not involved (non-participant observation). While the latter seeks to observe in the background without any attempt to influence the phenomenon being observed, the former is immersed in the context of the study and may even be actively influencing the phenomenon being observed.

Observation can be used at any stage of programme and for many purposes, and typically in conjunction with other data collection methods such as surveys and interviews. For the purpose of this guide, observation serves as a supplementary means to assess the outcomes of a group work programme in addition to the use of the pre-post- intervention surveys. Outcome indicators and rubrics are established, which enable the conversion of members' actions and behaviours into data that enable further analysis. The observational data provide a richer description of behaviours grounded in context and help in understanding how and why certain outcomes are realised, and in turn support decision-making about the programme.

Organisation of this guide

The following sections of this guide will provide a step-by-step guide on how observation can be planned, implemented and utilised. We will use exemplars from the Brotherhood Programme to further illustrate the steps described, including how observation conducted within this programme was further informed using participants' reflective journals.Insights gained from our experience in applying observation in the Brotherhood Programme will be shared. Ethical considerations associated with the method are then outlined. We will also identify strategies to ensure quality of the observational data collected and summarise the advantages and disadvantages of the method.

Planning Phase

Planning for systematic observation for monitoring and evaluating a programme involves determining the following:

- What, where, when, how, the frequency and duration to observe;
- How to document the observation (see Data Collection Tools in the next section);
- Who is responsible to observe and document;
- How to organize data;
- How/where to store the data;
- How can the data be used and analysed;
- Who can access the data;
- How can the findings be presented and communicated.

Data Collection Tools

There are several data collection tools, ranging from open-ended to highly structured, available for codifying what to observe and document (Fix et al., 2022). Documentation of observation makes data available for analysis, enable tracking of progress of group work members and enhance reflection of facilitators.

Examples of data collection tools include:

i) field notes: open-ended detailed records of observation that may include description of people, actions, interactions and observer's reflections;

ii) observation guides: semi-structured records of observation that include categories, themes and/or prompts to help the observers focus their observations on specific aspects of interest;

iii) checklists: highly structured records of observation that list specific behaviours or characteristics of phenomenon the observers need to observe (and count, if applicable);

iv) rubrics: similar to the checklists but include specific criteria and standards for assessing particular behaviours.

Determining which to adopt depends on the purpose of the observation and how much is known about the phenomenon to be observed. When information about what to observe is scant, more open-ended observation tools are preferred (Jorgensen, 1989). These tools would be particularly useful for new programmes in which behaviours and responses of the participants are less likely to be anticipated. When programmes are more mature, that is, have been conducted multiple runs, the participants' behaviours and responses can be specified with greater details and nuances and thereby facilitate the formulation of specific indicators, criteria and/or standards for the observation. Nonetheless, as most programmes have specific objectives, a preliminary checklist or rubrics can still be developed to guide the observation since the behaviours or actions to be observed from the members are bounded and dictated by the objectives and expectations of the programme. In particular, rubrics provide a consistent and principled tool for tracking the progress of group work members.

Observation Rubrics

Rubric use in peer-reviewed published program evaluation is still rather rare, more commonly found in education and health programme evaluation (Martens, 2018). Rubrics are useful tools that can assist evaluators in making evaluative judgements that are systematic and grounded in transparent reasoning. The reasoning takes the form of specified description of what quality of behaviour or “performance” is observed or looks like at the various defined levels. Rubrics has thus been commonly adopted to assess learner performance as a form of assessment in education. In social service programme evaluation, rubrics can similarly be used to assess the progress of members and monitor the effectiveness of the associated programme elements.

They are particularly useful for documenting the progress of members across group work sessions as they systematise the observations to be made and documented in the session notes without the need to write long-winded narratives of the members’ progress. It can be more challenging to distil trends in the progress of the members across multiple runs of the same programme from the open-ended observation notes than the data collected from the observation rubrics. Observation rubrics are also valuable when there is no available validated self-reported survey or observational checklist available for monitoring the progress of individual sessions. Where such tools are available, observation rubrics can either complement these tools or be replaced by less structured forms of observational data.

In this guide, we will illustrate how rubrics were developed for individual sessions to monitor a group work. For this purpose, the criteria of the members to be evaluated against are set and aligned with the session objectives. Specific, observable and measurable characteristics (or indicators) that provide evidence of whether the criteria have been met are then determined. Upon determining the indicator, the next step is to draft a description of the characteristic to be observed across different levels of performance. These descriptors help the observers to determine the level of performance that each member falls into for the criteria being evaluated. In other words, these descriptors set out examples of what quality would be like for each criterion at each performance level. Descriptors of performance are thus session-specific and typically phrased in absolute terms. The different performance levels are reflected in absolute terms as scores (e.g., 0-4 points), grades (e.g., A to D), evaluative terms (e.g., poor, good, excellent) or levels of met criteria (e.g., not met, partially met, fully met).

Determination of the descriptors and their respective performance levels can be based on the practitioners’ experience and actual observations and informed by theories and published studies. The descriptors and exemplars for each performance level can be provided in the observation template to guide the observation. This process establishes and ensures consistent standards for documenting outcomes, which in turn enhances trustworthiness in tracking and comparing outcomes over time within and between programme runs. The following section will illustrate the process of planning for the observation and developing the rubrics using the Brotherhood Programme as a case example.

Case Example: Brotherhood

Background: Brotherhood Programme

Initiated by the third author, the Brotherhood, a support group for men who have caused harm to their loved ones, was established within the Thye Hua Kwan Moral Charities in September 2019. Now in its fifth run, this groupwork was set up in response to a niche need arising from an increase in the number of Personal Protection Order (PPO) applications. Turning the dominant story of how these men are viewed by society as perpetrators on its head, the programme seeks to reauthor an alternate story of a better man, collaborator and partner of these men. Each run of this treatment group, informed by the Duluth framework (Pence et al., 2011) and family violence narrative framework (Scott, 2021), spans a total of 10 sessions with a primary focus on psychoeducation and support and concludes with a graduation cum family day. Members, referred to as Brothers, are encouraged to participate in subsequent runs either as experienced Brothers or trained mentors to reinforce their learning as well as to provide support for new Brothers. Such continued participation engages them as partners with the facilitators in advocating for respectful and violence-free relationships.

Planning for Observation

Using the prompts identified above, Table 1 below illustrates the plan for the observation to be conducted for the Brotherhood group work sessions.

Table 1. Observation plan for Brotherhood group work

Planning prompt	Observations to be conducted for Brotherhood
What, where, when, how, the frequency and duration to observe	*What*: Actions and speech of the Brothers during the sessions *Where*: At the family service centre where the sessions are conducted *When*: Group work sessions *How*: Observation to be made by two facilitators of the sessions and recorded at the end of each session *Frequency*: Every session conducted three weeks apart *Duration*: Throughout the session
How to document the observation	Semi-structured observation template that includes both rubrics and space for additional notes (see Table 4)
@Who is responsible to observe and document	Two designated facilitators to discuss and fill in the observation template at the end of each session. Whole team of facilitators to discuss the observation data collected during the debrief session
How can the data be used and analysed	Observation data to be used for - tracking the progress of individual Brothers across each run and subsequent runs; - identifying areas of strengths and for improvement of the group work sessions; - complementing the pre-and-post survey conducted on the behaviours and attitudes of the Brothers.
Who can access the data	- Facilitation team. - Social worker cum researcher. - Independent researcher.
How can the findings be presented and communicated	The findings would be reviewed amongst facilitators during the yearly meeting to finetune the details (e.g. removing or adding aspects) for the next run of the programme. Sharing of the learning from the findings would also be conducted internally with other group work programmes in the agency.

@Time gap between the observation and documentation should not be too far (within a day or two preferably) to ensure that the observers' memories are still fresh and that information recorded is as accurate as possible.

Developing the Rubrics

A collaborative inquiry approach (Shani, et al., 2023) was adopted to develop the rubrics. This approach tapped on the different expertise of those involved:

(i) the first author: an independent researcher provided guidance on the mechanics of data collection and rubrics development;

(ii) the second author: a social worker-cum-researcher conducted a literature review and came up with the initial criteria for the rubrics;

(iii) the third author and his facilitation team: the team of social worker-cum-facilitators of the programme provided input based on their observations of the Brothers in previous runs.

The authors and the rest of the facilitators would meet to review and refine the rubrics initially developed by the second author. Each group work session had between 2 and 3 numbers of observation criteria with the following performance levels and scores selected by the team:

- Most unfavourable outcome (-2);
- Less than expected outcome (-1);
- Expected outcome (0);
- More than expected outcome (1);
- Most favourable outcome (2).

This range of performance levels are anticipated to capture the wide range of behaviours observed among the Brothers with diverse backgrounds and varied experiences in the programme (ranging from 1 to 4 runs), and mirrored the Goal Attainment Scale illustrated in the Ministry of Social and Family Development (MSF)'s Group Work Practice Guide (2018).

Unlike the latter, the rubrics developed in Brotherhood focused on the objectives intended by the facilitators for each group work session rather than the collective goals determined in collaboration with group work participants to be achieved through the programme. For Brotherhood, the facilitators deliberately decided against determining the collective goals in the beginning, as the intention is to first build therapeutic alliance with the Brothers. Given the potentially complex circumstances participants are in, fixation on a goal for its own sake may also decrease affect, self-esteem and motivation, especially for those who experience failure to meet their expectations. Rather, the emphasis is on creating a supportive group environment and culture and helping participants develop their own intrinsic goals that are aligned with reducing recurrence of family violence as they assimilate more into the programme. The observation rubrics come in handy especially in this context in focusing the facilitators on what to look out for in the progress of the members. The rubrics also scaffold the process of documenting critical observations in their session notes, without the necessity to write out long-winded narratives, a data source that may demand more time for analysing and yielding insights about the effectiveness of the programme.

Example: Development of rubrics for Session 2

Objectives for Session 2:

(i) Develop group cohesion amongst Brothers and therapeutic alliance with facilitators.

(ii) Make connections and develop reasoning between personal history and present values.

(iii) Identify and deepen reflection of important values and preferred identity.

Table 2. Criteria and indicators developed for Session 2 of Brotherhood group work

Objective	Criteria	Indicators	Related theory
1	The extent to which brothers are able to develop group cohesion amongst Brothers and therapeutic alliance with facilitators.	- Nature of response to the group; - Level of contributions to tasks, processes and decision making during the session.	Tuckman's stages of group development (Tuckman, 1965)
2	The quality of the ways in which brothers drew connections between their values and valuable life lessons in their personal history.	- Extent of Brothers' attempts to reflect and trace their roots (history and influence) of where their values are coming from to relevant themes of their life and draw connections between the themes and examples of how their history influenced their values.	Narrative therapy (White, 2011)
3	The extent to which brothers can identify important personal attributes and preferred identity they wish to have as a better man.	- Number and nature of the helpful attributes written in Brother's reflective journal; - Extent of Brother's attempt to reflect on preferred identity.	Narrative therapy

As illustrated in Table 2 above, the criteria developed for each session were aligned with the objectives of the session. For example, one objective for Session 2 is to develop group cohesion amongst Brothers and therapeutic alliance with facilitators. The concept of group cohesion was first operationalised into specific actions based on the Brothers' nature of response to fellow group members during the discussion and the level of contributions to the activities and tasks during the session. The operationalisation of the objective into specific indicators entails reference to the literature and the facilitators' practice wisdom. In the case of objective 1, the literature on Tuckman's stages of group development (Tuckman, 1965) offers suggestions on what to look out for as positive signs of group cohesion. While the initial criteria were informed by the theory, the facilitators' input was invaluable in refining and establishing the descriptors and exemplars for the respective performance levels. The facilitators were prompted to think of the worst- and best-case scenarios observed from previous runs and reasonably expected of the Brothers in terms of their ability and willingness to form relationship with fellow members and facilitators.(i) These scenarios set out the boundaries of the behaviours anticipated, provided the exemplars and guided the formulation of the descriptors for the two extreme ends of the performance levels (i.e., -2 and +2). Discussion then focused on determining the descriptors and exemplars for the three performance levels in between. The descriptors and exemplars are outlined in the observation template to guide the observation and scoring during the session. The observation indicators for each objective are thus determined collectively by the facilitators as important and adequate for monitoring the extent to which the members have met the objective. Apart from the rubrics, the observation template also included open-ended space to allow facilitators to record observations that fall outside of the descriptors, whenever applicable.

The following is a list of prompts that can be helpful to derive the criteria for the observation rubrics:

1. What are the observable behaviours (e.g., verbal and non-verbal action, mannerism, facial expression) that could indicate the outcome of the objective?
2. Is quantity and/or the quality of the behaviour matter?
    a. If quantity, what would be the dimension (e.g., extent, frequency) that is valued in the behaviour to be observed?
    b. If quality, what would be the nature or characteristic/s (e.g., nature of response to others, level of contributions) that is valued in the behaviour to be observed?
3. What could be the best-case scenario (most favourable outcome) and worst-case scenario (most unfavourable outcome) for the behaviour that has been encountered or predictable by the facilitators?
4. What are the range of outcomes between the best- and worst-case scenarios? How could these outcomes be grouped into 2-3 levels, depending on the number of performance levels selected for the rubrics?
5. Are the descriptors for the different performance levels:
    a. able to capture the full range of behaviours that are relevant to the objective?
    b. easily distinguishable across the levels?

(i) For programmes which are relatively new, observation rubrics can still be developed. However, the indicators and criteria are likely to be refined over time as the facilitators become more familiar with how their members respond and behave during the group work.

Table 3 below illustrates the three criteria developed for the observation template for use in session 2 of the groupwork.

Table 3. Observational rubrics for Session 2 of Brotherhood

Objective	Most unfavourable outcome (-2)	Less than expected outcome (-1)	Expected outcome (0)	More than expected outcome (1)	Most favourable outcome (2)
1	Argued with other group members during discussion	Neutral during discussion	Demonstrated politeness during discussion	Listened to others during discussion	Actively listened to others during discussion
	Provided negative response to the group	Did not respond to the group	Provided neutral response to the group	Provided positive response to the group	Provided positive and constructive response to the group
	Obstructed defining tasks, processes and decision making to a great extent	Obstructed defining tasks, processes and decision making to some extent	Had no influence to defining tasks, processes and decision making	Contributed to defining tasks, processes and decision making to some extent	Contributed to defining tasks, processes and decision making to a great extent
2	No attempt to reflect on values by tracing roots (history and influence) of where their values are coming from	Some tracing of roots but weak connections to personal values & examples given are not relevant on how their personal history influenced their values	Tracing of roots to relevant sociocultural themes* but has difficulties giving relevant examples on how their history influenced their values *Themes include family upbringing, adverse childhood experiences, culture etc.	Tracing of roots to significant sociocultural themes with relevant examples on how their history influenced their values	Tracing of roots to the significant themes with relevant examples & strong connections made on the interrelatedness of their personal history influencing their values

3	Unable to list any helpful attributes they wish to have as a better man	Able to list 1-2 helpful attributes they wish to have as a better man But unable to share where the preferred identity comes from	Able to list 3 helpful attributes they wish to have as a better man But unable to share where the preferred identity comes from	Able to list 1-2 helpful attributes they wish to have as a better man Able to share where the preferred identity comes from	Able to list 3 helpful attributes they wish to have as a better man Able to share where the preferred identity comes from

Table 4 shows a template of how a table for each indicator looks like in the session plan. Spaces are provided to allow the observers to enter the name of the participants that belong to each score as well as additional qualitative feedback or remarks. It is important to highlight that the lowest scores are not indicators of "failure" or "deficit" on the part of the members but an indication of their journey of change for that particular indicator.

Table 4. A sample of the template for filling in the observational data

<table>
<tr><th rowspan="2">Objective</th><th rowspan="2">Model/ Theory</th><th colspan="5">Indicator</th></tr>
<tr><th>Most unfavourable outcome (-2)</th><th>Less than expected outcome (-1)</th><th>Expected outcome (0)</th><th>More than expected outcome (1)</th><th>Most favourable outcome (2)</th></tr>
<tr><td>Objective 1:</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Client(s) who achieve the outcome
Please write down the acronym of client(s) who achieve the score</td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td colspan="2">Facilitators' feedback
Please comment on the observations or additional information for each objective</td><td colspan="5"></td></tr>
</table>

Insights gained from developing the Rubrics

Refining the criteria for the rubrics frequently generated intense discussion that tapped into the professional expertise and experiences of the facilitators, raised questions previously not thought about and surfaced knowledge that is tacit and assumed to be shared among the facilitators. Interestingly, the discussion revealed that shared understanding among the facilitators cannot always be assumed. The facilitators remarked that they had gained insights that they were previously not aware of from the sharing by fellow facilitators. These insights pertain to the observations of individual brothers, expectations of the programme and its theoretical underpinnings. Such insights allow them to further elaborate and refine the programme's theory of change. Using the observation rubrics also facilitated their debriefing at the end of each session and make completion of the session notes more efficient than before.

However, the facilitators did encounter some challenges in implementing the rubrics. These are reflected in Table 5, together with the changes they made to mitigate them.

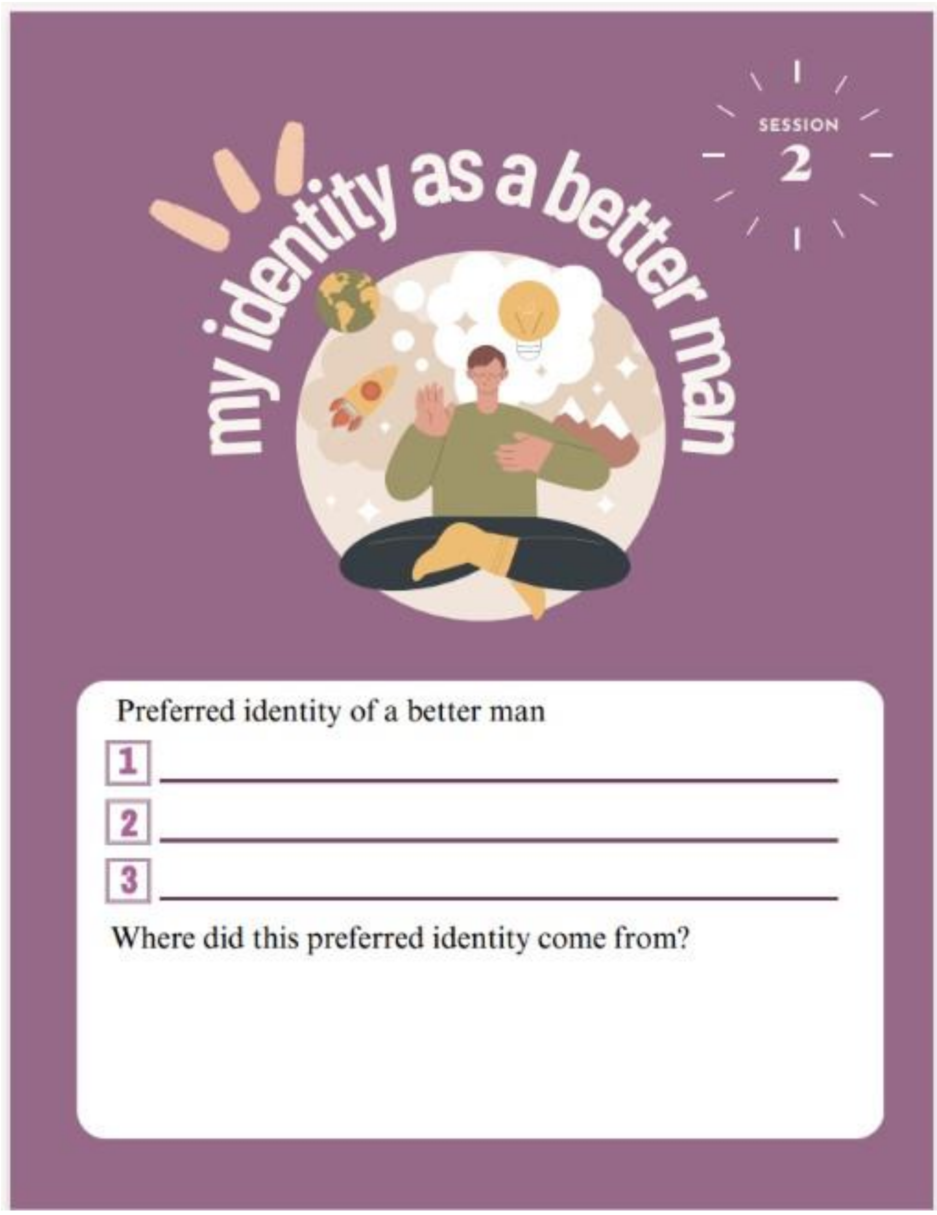
Table 5. Challenges and mitigating solutions in using the rubrics

Challenge	Mitigating solution
Brothers who felt less comfortable to talk may be more difficult to observe changes. It was challenging to distinguish whether they have processed the session objectives internally and do not want to share or have not processed them yet.	For members who are not ready yet to share, they are encouraged to write their inner thoughts out in a reflective journal (see the next section). For those who are not willing to share in any mode, spaces are provided in the rubrics table to capture the reluctance/resistance in sharing as this is an important observation in tracking the members' progress especially since the willingness to share is part and parcel of the therapeutic alliance between the members and the facilitators.
Differences between the indicators can still be quite grey when rating.	This invoked discussion between the facilitators about how to better distinguish the observations to be made across the various indicators. If the differences between indicators are still not apparent enough, the descriptors for the indicators were merged to constitute a single indicator.
Not enough time for all brothers to share during the debrief sessions.	Facilitators indicated their names in the feedback column that insufficient time was given rather than rating them. This would better distinguish members who are not willing to share versus those who are not given the opportunities to due to time constraint.

Use of the Reflective Journal

A reflective journal was developed alongside the generation of the observation rubrics. This journal was intended to engage the brothers in reflection during the group work. Prompts and spaces were provided in the journal to guide the reflection process. In the process, the journal also captured the thoughts and experiences of the brothers and supplement the facilitators' observations with additional data from the participants' perspective. Some items in the journal are designed to align with the observation rubrics and the output of the journal thus offers an alternative source of outcome data in its own right alongside the observation data. Figure 1 below shows a journal item of the Brotherhood programme that provides the data that can be used for assessing the indicator for objective 3 in Session 2. Reflective journaling is thus a useful tool for group work not only to deepen and reinforce the learning but also for monitoring and evaluation purposes.

Figure 1. An item in the reflective journal used in Brotherhood



Implementing the observation

Piloting

Piloting constitutes an important step in implementing the observation. Pilot testing of the data collection tool ensures that the observation plan works in practice. The process also serves as a form of training for the observers.

Piloting could reveal issues with the:

- Descriptors used in the rubrics (e.g., whether they are easily and consistently interpreted by the observers);
- Sampling plan (e.g., whether the behaviours occurred frequently enough);
- Staffing capacity (e.g., whether the number of members to be observed can be followed).

In our experience, piloting allowed the assigned observers to check the viability of the indicators against what they can observe and whether they share the same interpretation of the indicators and ratings. Piloting thus ensures that issues that may affect the evaluation integrity to be addressed and ensure that the data collection can be kept systematic and consistent over time. During piloting, additional nuances and unanticipated ways in which the members behave may be identified. Further refinement of the rubrics may ensue.

Implementing

There are additional considerations when implementing the observation protocol. These include:

- Whether and when to use pseudonyms when documenting the observations: Ideally, pseudonyms are used in the first instance of documenting the observations, example, in the session notes. This ensures the confidentiality of participants as individuals will not be linked to the data documented. However, if this is not practical, pseudonyms can be used when transferring the scores into Excel spreadsheets for analysis.
- Whether to use technology to facilitate the documentation and management of observational data: using digital forms to document the observation can enhance efficiency in data collection, management and analysis. This can potentially eliminate manhours incurred and errors due to transcription (i.e. when converting handwritten data into a digital form).
- What approach to use to ensure validity of the scoring:
  - One way is to estimate the degree of interrater reliability between two or more observers. This helps to ensure that the observers can reliably rate the observed behaviours.
  - Another way, which is the one adopted in Brotherhood, is the consensus approach, in which the observers come to a consensus of the final scoring through discussion. This allows for facilitators to share their insights about the members and ensures that the final scores consider multiple perspectives and "observational angles".

Analysis and Uses of the Data

Monitoring Progress of Participants

Using the rubrics to assess the members every session encourages the facilitators to look out for behaviours that matters to the members' developmental goals. Keeping track of the scores over time also allows the facilitators to monitor how the members have progressed over the course of a single run of the group work as well as multiple runs. Digression in the progress of an individual for example may signal a need to find out whether the participant might have faced new life challenges that might not be apparent to the facilitators.

Monitoring Effectiveness of Sessions

The observational scores obtained from the use of the rubrics are not intended for statistical analysis but to indicate patterns and trends. For example, the spread of scores of the participants for the various indicators in each session can give an indication of the effectiveness of the activities tied to the objectives as well as determine the emphasis needed for subsequent sessions. Uniformly high scores across participants for a particular criterion may suggest that the activity used to achieve the objective is effective. Facilitators may also consider whether less emphasis can be placed on the same objective in subsequent sessions. On the other hand, the reverse – uniformly low scores across participants - may signal the need for more reflections on the implementation of related activities, their appropriateness and whether more activities are needed to achieve the objective.

Monitoring and Improving Programme

The observational findings across the sessions can help facilitators track how the different sessions contribute to the overall programme goals and identify gaps and areas of improvement. It is thus important for the observational findings to be considered together with the programme's theory of change (if available). Apart from informing the programme design, the observational rubrics and the findings can also facilitate the training of new facilitators who join the group work on what to look out for during the sessions. Comparisons can also be made across programmes which share overlapping objectives and allow for learning across them. Enriching professional learning conversations can thus be triggered by the observational findings.

Besides providing an additional source of evidence for the programme outcomes, the findings from the observation rubrics can also inform the selection and/or design of the evaluation tools used to evaluate the programme, such as self-reported surveys and interviews. The observation rubrics allows for a more fine-grained assessment of the behaviours observed in the participants than what would be likely captured in typical pre- and post-self-reported surveys that tend to capture the generic progress of the participants. But this means that the data captured by the observations are also more likely to be affected by more immediate contextual factors, such as the moods of the participants during the sessions. Nonetheless, consistent monitoring of the participants across sessions and programme runs can help to mitigate such effects. Moreover, the observational data incorporate the observers' experiences and professional training that self- reported surveys do not. Long-term trends in the observational data can further augment the findings from the self-reported surveys when similar trends are observed by providing a richer and more nuanced account of the outcomes of the programme. Where discrepancies between findings are revealed, facilitators can consider more in-depth data collection method, such as interview, to unpack them. For example, when the observational findings are more positive than self-reported survey, interviews can be conducted to find out why participants are more reserved in their self-assessment. Similarly, interviews can also be used to uncover the reasons behind participants' more positive self-assessments than what was observed during the sessions. The resulting findings can in turn have implications on the validity and design of the survey and observation rubrics used to evaluate the programme.

Ethical considerations

The same ethical considerations applicable to other data collection methods for programme evaluation apply to the use of observation. These include ensuring that appropriate permissions and informed consent are obtained, as well as ensuring the confidentiality of the data. As observation puts the observer in a higher position of power relative to those being observed, it is important for the observers to be mindful of their responses and engagements with the members in order to preserve the integrity of the data collected.

Enhancing quality of the data collected

As making observations depend upon an observer's interpretation, this data collection method inevitably involves subjectivity to some extent. Such subjectivity can be reduced by the planning for and piloting the use of clear observation protocol and scoring rubrics and training observers to interpret the rubric and apply the rating scale consistently (Fix et al., 2022). Having two or more observers can reduce bias, preventing the results from being skewed, and reduce observations that might be missed by a lone observer. Regular debriefing meetings among the observers and programme team members during data collection can also ensure data quality by ensuring that everyone involved in the programme and evaluation share the same understanding of what is important to be observed and how to observe them. Collectively, the training and debriefing help ensure both intra-rater reliability (i.e., a consistent frame of reference is maintained for interpreting what the observer observes over time) and inter-rater reliability (i.e., same instance being interpreted and recorded in the same way by different observers) (NSW Department of Education, 2021). It is also important for observers to adopt reflexivity – by regularly reflecting on how their own assumptions, biases, experiences and life position might influence their observations - in ensuring the quality of the observation (Rubin & Babbie, 2017). Given their dual role, practicing reflexivity is especially important for observers who are also facilitators for the programme.

Advantages and Disadvantages of Observations

Advantages

The following summarises a non-exhaustive list of advantages that observations offer:

(i) Data can be collected wherever and whenever an event or activity is occurring, without considering sample size;

(ii) A less invasive data collection method especially if the observer/s are unintrusive and/or blend into the observation setting;

(iii) Not dependent on the participants' ability and willingness to provide data;

(iv) The explication and formalisation of the rubrics also ensured that the same expectations and standards of the members can be expected for subsequent programme runs regardless of any change in facilitators. This enhances the sustainability and scalability of the programme.

Disadvantages and Limitations

As with any data collection methods, observation comes with it disadvantages and limitations that need to be mitigated such as:

(i) Time-consuming, especially in developing the rubrics. However, once the rubrics are established, there could be time saving in documenting the outcomes of the sessions.

(ii) Inherently subjective. Observers need to be mindful of unintentional biases such as confirmation bias (looking out for things that confirm our thoughts), halo effect (initial impression of the people to be observed) and observation bias (looking in narrow places where we think we will find what we are looking for). Apart from having more than one observer, one important way to mitigate these biases is for the observers to distinguish between observations versus inferences made of the observations.

(iii) Does not necessarily increase our understanding of why members behave as they do (CDC, 2018).

(iv) In the context of group work, this method is only applicable to behaviours that can be observed during the group work but not outside of it.

Conclusion

The use of observation rubrics in monitoring and evaluating group work ensures that consistent standards can be applied across multiple runs and enhances the continuity of the programme. Observation indicators encapsulate the professional experience of the facilitators and establish the expectations that facilitators have of the members during the course of running the programme. The standards of the programme are thus more likely to be sustained even if there are changes in facilitators over time. These indicators are not intended to be cast in stone but can evolve as the needs and goals of the members change over time and hence can be particularly compatible with developmental evaluation (Patton, 2006).

REFERENCES

Centers for Disease Control and Prevention (CDC) (2018). Data Collection Methods for Program Evaluation: Observation. *Evaluation Briefs 16*. Retrieved 10th October 2023, from https://www.cdc.gov/healthyyouth/evaluation/pdf/brief16.pdf

Evans, C., & Lambert, H. (2008). Implementing community interventions for HIV prevention: insights from project ethnography. *Soc Sci Med*, *66*(2), 467-478. https://doi.org/10.1016/j.socscimed.2007.08.030

Fix, G. M., Kim, B., Ruben, M., & McCullough, M. B. (2022). Direct Observation Methods: a Practical Guide for Health Researchers. *PEC Innov*, *1*. https://doi.org/10.1016/j.pecinn.2022.100036

Jorgensen, D. L. (1989). *Participant Observation: A Methodology for Human Studies*. Sage.

Mack, N., Woodsong, C., MacQueen, K. M., Guest, G., & Namey, E. (2005). Qualitative Research Methods: A Data Collector's Field Guide. Retrieved 10th October 2023, from https://www.fhi360.org/sites/default/files/media/documents/Qualitative%20Research%20Methods%20-%20A%20Data%20Collector%27s%20Field%20Guide.pdf

Martens, K. S. R. (2018). Rubrics in program evaluation. *Evaluation Journal of Australasia*, *18*(1), 21-44. https://doi.org/10.1177/1035719x17753961

Ministry of Social and Family Development (MSF) (2018). Family Service Centre Code of Social Work Practice: Group Work Practice Guide. Retrieved 05 January 2024, from https://www.msf.gov.sg/docs/default-source/odgsw/2018-family-service-centre-code-of-social-work-practice-cswp-group-work-practice-guide9b2d044b-43ee-4068-8274-4d3508f37bb6.pdf

NSW Department of Education (2021). *Observation: a guide for use in evaluation*. Retrieved 10th October from https://education.nsw.gov.au/teaching-and-learning/professional-learning/pl-resources/evaluation-resource-hub/collecting-data/observation

REFERENCES

Patton, M. (2006). Evaluation for the way we work. *Nonprofit Quarterly*, *13*.

Pence, E., Payer, M., Wedge, L., Barnes, G., Jones-Schroyer, G., Miller, S. & Thompson, C. (2011). *Creating a Process of Change for Men Who Batter – Facilitator's Manual* (3rd ed.). Duluth, MN: Domestic Abuse Intervention Programs.

Rubin, A., & Babbie, E. R. (2017). Research methods for social work (9th ed.). Cengage Learning.

Scott, T. A. (2021). Safety and Repair: A Manual for Individual, Family & Group Conversations to Address Gender Based Violence. Nova Scotia.

Shani, A. B. (Rami), Coghlan, D., Paine, J. W., & Canterino, F. (2023). Collaborative Inquiry for Change and Changing: Advances in Science-Practice Transformations. *The Journal of Applied Behavioral Science*, 59(4), 541-555. https://doi.org/10.1177/00218863231204372

Toseland, R. W., & Rivas, R. F. (2016). *An Introduction to Group Work Practice*. Pearson Education. https://books.google.com.sg/books?id=JBWPCwAAQBAJ

Tuckman, B. W. (1965). Developmental sequence in small groups. *Psychological Bulletin*, 63(6), 384–399.

White, M. (2011*). Narrative practice: Continuing the conversations*. (D. Denborough, Ed.). W W Norton & Co.

ACKNOWLEDGMENTS

We would like to acknowledge the following who have contributed to our discussion one way or another: Iris Li, Mazlina Bte Salim, Rafi Bayhaqi, Edwin Tan, Chen Jie Ying, Meilinda Tan, Wilson Mack and the Brothers from Brotherhood.

If you have any research-related queries or if you are interested in learning more about developing observation rubrics for group work, please email lhseah@nus.edu.sg; For Brotherhood or programme-related queries, please email either fionnlui@thkmc.org.sg or benang@thkmc.org.sg

ACKNOWLEDGMENTS

The Publications Team

Review Editor: Dr Rosaleen Ow

Editorial Assistant: Nurul Fadiah Johari

NUS Social Service Research Centre
Faculty of Arts and Social Science
National University of Singapore

The Shaw Foundation Building, Block AS7,
Level 3, 5 Arts Link,

Singapore 117570

Email: ssr@nus.edu.sg
Phone: 6601-5019

Website: https://fass.nus.edu.sg/ssr/