PILOT STUDY PLAN

Becoming a Man (BAM)

Dartington Service Design Lab, University of Plymouth and University of Exeter

Principal investigators: Tim Hobbs and Nick Axford

YOUTH ENDOWMENT FUND

Pilot study plan

Dartington Service Design Lab, University of Plymouth, University of Exeter

Finlay Green, Nick Axford, Julia Mannes, Lynne Callaghan, Kate Allen, Vashti Berry, Tim Hobbs

Project title	*Becoming a Man (BAM)*
Developer (Institution)	*Youth Guidance*
Evaluator (Institution)	*Dartington Service Design Lab, University of Plymouth, University of Exeter*
Principal investigator(s)	*Tim Hobbs, Nick Axford*
Evaluation plan author(s)	*Finlay Green, Nick Axford, Tim Hobbs*
Evaluation setting	*School-based*
Target group	*12-14 year-old boys experiencing challenges with their social-emotional development*
Number of participants	*3 schools, ~80 students*

Study plan version history

Version	Date	Reason for revision
1.2 [*latest*]		
1.1		
1.0 [*original*]	22.11.21	*[leave blank for the original version]*

Any changes to the design need to be discussed with the YEF Evaluation Manager (EM) and the developer team prior to any change(s) being finalised. Describe in the table above any agreed changes made to the evaluation design, research questions and approach, and the rational for these.

Intervention

Introduction

Becoming a Man (BAM) is a selective school-based social-emotional learning intervention for 12-16 year-old boys.[1] It aims to improve school engagement and prevent or reduce interactions with the criminal justice system. It comprises four key activities: BAM Circles, special activities (group activities outside of school property/time), brief encounters (informal check-ins), and 1:1 support. Practitioners – known as 'BAM counsellors' – deliver the intervention in schools with groups of 8-12 participants over two school years.

Target group

BAM primarily targets adolescent boys living in deprived areas who are experiencing challenges with their social-emotional development. In this evaluation, participants are eligible if they are boys aged 12-14 years in one of three secondary schools in Lambeth and experiencing challenges in at least one area of their social-emotional development according to the Holistic Student Assessment (HSA): resiliencies (internal and external); relationships; and learning and school engagement (further information on the HSA can be found in the 'Methods' section).

Activities

BAM is made up of four activities:

- The BAM Circle (group sessions delivered in a school setting with 8-12 participants)
- Special activities (group activities outside of school property or school time)
- Brief encounters (quick, informal check-ins in public school spaces like the hallway or playground)
- 1:1 support (individualised support for those with greater levels of need)

BAM circles constitute the central element and are delivered over two years (50 1-hour sessions in total, ~25 per year). They are delivered by a prosocial male counsellor with QCF-6 level qualifications. Counsellors are required to have a clinical or therapeutic qualification and experience of working with young people. To ensure they are relatable, counsellors are recruited from the communities in which young people live and share some of their lived experiences. Sessions typically occur during school hours, substituting for a lesson. Activities

[1] In the project it will be implemented with young people in years 8 and 9 (ages 12-14 years), in line with the Youth Endowment Fund brief to work with young people aged 10-14 years.

delivered with BAM Circles include check-ins and check-outs to open and close sessions, role plays, group missions, video education, lectures, stories and homework.

Theory of change

The BAM theory of change was informed by: published research on BAM (Lansing et al. 2016, Heller et al. 2017, Heller et al. 2013); theoretical frameworks for group therapy (Yalom and Leszcz 2005), psychotherapy (Jung 1969) and youth development (Nagaoka et al. 2015); reviews of programme documents, including the BAM Circle curriculum; the observation of BAM circles in Chicago; workshops and interviews with staff from Youth Guidance (YG; US-based developers and purveyors of BAM) and the Mental Health Foundation (MHF; the lead organisation managing the implementation of BAM in Lambeth); and several rounds of iteration. A detailed version of the theory of change, developed and adapted in collaboration with the MHF and YG during the feasibility phase of this evaluation, can be found in Appendix 1. It can be summarised as follows.

- In the long-term, BAM aims to help young people engage in responsible decision-making, including improved educational attainment and the avoidance of/reduced involvement in youth violence.
- In the short-term, it does this by helping young people to internalise BAM's six core values (integrity, self-determination, positive anger expression, accountability, respect for womanhood and visionary goal-setting). These act as positive assets, which help to buffer young people from risks while empowering them to take advantage of opportunities and resources in their environment. Whether they do so depends on wider influences in young people's lives, and whether they reinforce or limit their efforts to apply the core values.
- BAM helps young people to internalise the core values through successful implementation (i.e., BAM is delivered with quality and fidelity to the intended target population, who attend activities regularly). This is because of the way in which young people actively experience each value in different ways before collectively reflecting on these experiences. Whether they engage in action and reflection depends on whether young people are sufficiently ready and able to establish and maintain healthy groups – those that are fun, safe spaces in which young people challenge themselves and each other to be open, honest and vulnerable.

Delivery model in Lambeth

Children are recruited to the intervention by each of the three secondary schools involved in the project. Counsellors work with senior leadership, heads of year and the pastoral team to

identify young people about whom they have concerns regarding their social-emotional development.[2] Counsellors also support school staff to recruit a range of young people with different types and levels of need, to prevent negative labelling of the intervention as a programme for 'bad' children. Specifically, all groups and school cohorts will aim for approximately a 15%/70%/15% split across the three respective tiers of the HSA:

- Tier 1: Students who are thriving and who exhibit primarily strengths and few challenges
- Tier 2: Students who have a balanced combination of strengths and challenges
- Tier 3: Students who are approaching crisis or are in crisis

Guidance for counsellors to support the conversations they have with school staff during the recruitment process can be found in Appendix 3.

Young people are ineligible if they are chronically absent or if they are a risk to themselves or others in a group setting. The needs of ineligible young people will continue to be supported by each school independently of BAM.

Training and support

Counsellors receive 300 hours of programme training. They are recruited, trained, coached and supervised by the MHF alongside YG.[3] An advisory council supports the implementation of BAM. In addition to supporting the financial sustainability of BAM, this forum acts as a two-way liaison with the local communities that BAM participants are based in, representing the interests, insight and expertise of these communities in the delivery of BAM while also representing BAM's interests in return. For example, the council can identify – and support the development of relationships with – local organisations working in the field who may be wary of BAM as an externally developed 'competitor'. They might also provide advice on how BAM should approach public relations in the local area.

[2] Counsellors explain that this could be owing to issues such as risk of exclusion, history of exclusion, poor educational attainment and mental health issues.

[3] Counsellors are trained and coached by the replication specialist. Training refers to the support counsellors receive as a group to understand and implement the curriculum as intended, and to a high standard. Coaching concerns the individual support provided by the replication specialist to help counsellors develop their competencies. YG's standard training model involves a monthly, full day, in-person session (adapted during COVID-19 to shorter (two hours) more frequent (once a week) online sessions (via Zoom). YG's standard coaching model involves meeting for half a day in person with the replication specialist every 1-2 months (adapted during COVID-19 to meeting every 2-3 weeks remotely for shorter periods).

Community partners include Black Thrive, a partnership of black communities and service providers for black wellbeing, and Colourful Minds, a partnership of South London and Maudsley (SLAM) employed psychiatrists and psychologists.

Evidence

BAM has a Level 4 evidence rating on the Early Intervention Foundation Guidebook, meaning that it has "evidence of a long-term positive impact on child outcomes through multiple rigorous evaluations".[4] Two randomised controlled trials have shown positive impacts on numbers of arrests (for violent/all crime) and school performance for students with a mean age of ~15 years living in racially segregated and deprived communities in Chicago (Heller et al., 2013, 2017).

Need for the pilot study

Evidence for the effectiveness of BAM comes exclusively from the US, and this is the first time the programme has been delivered in the UK. Given evidence of replication failure in Europe (including the UK) of multiple programmes developed and found to be effective in North America, it is necessary to understand the feasibility of delivering BAM in the UK and explore both (i) its potential to improve outcomes here and (ii) issues pertinent to further intervention delivery/development and a next-stage evaluation.

In response, we have designed a staggered, two-phase evaluation. The first phase (September 2020 to July 2022) is a feasibility study in which the evaluation team is supporting MHF, YG and partners to understand the conditions necessary to promote the successful implementation of BAM and identify areas for ongoing optimisation to increase likely impact and subsequent 'evaluability'. It was agreed that should BAM show strong feasibility with the first cohort by the end of the first year of delivery (July 2021), the evaluation of the second cohort (September 2021 to July 2023) would advance to a pilot outcomes evaluation. As outlined in the interim report of the feasibility study, this criterion was met.

Research questions and/or objectives

The pilot outcomes phase of this evaluation has two aims: (i) to establish whether BAM has promise; and (ii) to test aspects of evaluation design to inform a next-stage evaluation.

We will establish evidence of promise by gathering evidence on BAM's programme theory, including evidence of unintended consequences. A theory of change for BAM was

[4] https://guidebook.eif.org.uk/programme/becoming-a-man

developed in collaboration with stakeholders at the start of the feasibility study in 2020. This has been revised in the light of interim findings from year one of the feasibility study, which ended in July 2021. BAM's programme theory is made up of four sub-theories: implementation, intermediate outcomes, ultimate outcomes and unintended consequences. These sub-theories give rise to the research questions that will guide the design and delivery of the pilot.

1. To what extent is BAM being successfully implemented, with whom, under what circumstances and why?
2. To what extent is implementation contributing to improvements in young people's social-emotional development, for whom, under what circumstances and why?
3. To what extent are improvements in social-emotional development contributing to improved academic performance and behaviour and the avoidance of or reduced involvement in crime / anti-social behaviour[5], for whom, under what circumstances and why?
4. To what extent has implementation contributed to unintended consequences, for whom, under what circumstances and why?

Success criteria and/or targets

At the end of the pilot, evidence regarding intervention promise and the necessity, viability and nature of a next-stage evaluation will be interrogated by key stakeholders, including YEF, YG and MHF. The following criteria, which build on the pilot's research questions, are designed to support that decision-making process.

Evidence of promise

Observing some improvement in targeted outcomes during the pilot is a necessary condition for wider implementation and further evaluation of BAM. However, outcome change only becomes a sufficient condition for advancing BAM if the programme is deemed to have made a significant contribution to that improvement. To determine this, we will use contribution analysis (Mayne 2001).

Contribution analysis is a collaborative, systematic approach to establishing the extent to which a particular intervention contributed to change in a specific set of outcomes. It can support evaluators to collect and analyse data related to impact in projects that lack comparison groups and large samples. In contribution analysis, rigour is understood in terms of triangulation and the inclusion of multiple perspectives. Different data are collected from

[5] Subject to finding a suitably acceptable way to measure this.

different sources at different levels to support the accumulation of sound, plausible evidence at each stage of the theory of change. This evidence is interrogated iteratively by a broad collection of stakeholders to establish areas of weakness that can be addressed through further data collection and analysis.

The approach has six steps. The first two steps involve developing a theory of change – one that articulates not just activities and outcomes but also mechanisms, context, rival explanations and unintended consequences. Steps three and four involve gathering evidence on each of these elements within the theory of change, based on mixed methods data collection and analysis, to arrive at a first attempt at a 'plausible contribution story' – that is, one that enables *"a reasonable person'* to '*agree from the evidence and argument that the program has made an important contribution to the observed result"* (Mayne 2012, p.273). During step four, weaknesses in the contribution story are established in collaboration with wider stakeholders, which for this pilot, will include YEF, MHF, YG and the evaluation team. During step five, additional evidence is collected on these areas to inform revisions to the contribution story during step six.

Building from Mayne (2012), the contribution story assembled during this pilot will be deemed 'plausible' if it meets the following three criteria concerning implementation, outcomes and unintended consequences.

Implementation

The first criterion considers whether BAM's activities were implemented as outlined in the theory of change. There are targets for three areas of implementation: recruitment, attendance, and fidelity. These were informed by targets developed by YG, based on their experience and understanding of what constitutes successful implementation, and published research on BAM. They will be considered alongside evidence regarding whom BAM can be successfully implemented with, under what circumstances and why.

Implementation area	Target
Recruitment	Each counsellor is delivering five groups of 8 to 12 young people by early November 2021
	All young people who take part in BAM are experiencing challenges in at least one area of their social-emotional development, according to the Holistic Student Assessment (HSA; Allen at al., 2017)

	All three tiers of need are represented in each group and in each school according to HSA, with a split of ~15%/70%/15% for tiers 1, 2 and 3, respectively
Attendance	Young people attend a minimum of 13 BAM circle sessions on average every school year (26 in total)
	80% of young people receive a brief encounter every week during term time
Fidelity	Counsellors deliver all lessons from the 30-lesson BAM manual (roughly two sessions per lesson) with each of their groups
	Counsellors deliver 25 BAM circle sessions per year with each group

Outcomes

The second criterion considers whether the theory of change is verified by evidence. This means that the chain of expected results occurred, culminating in improvements in social-emotional development, academic performance and behaviour, and offending / anti-social behaviour.[6] The assumptions for each link in the theory of change have occurred and together provide a reasonable explanation for the results. The relative contribution of contextual factors and rival explanations is recognised.

Unintended consequences

The third criterion considers unintended consequences resulting from BAM, specifically evidence that either (i) these consequences did *not* materialise, or (ii) they *did* materialise but their relative contribution is recognised and their severity is deemed acceptable by stakeholders.

Evaluation design

We will report on the extent to which outcome measures can be applied successfully (completion rates at baseline and follow-up), including extent to which outcome measures are acceptable to participants and data collection procedures are efficient and minimally burdensome (from discussion with counsellors / school staff)

[6] Subject to finding a suitably acceptable way to measure this.

Methods

Background

There are two cohorts of young people in the evaluation, constituting a total of ~145 young people.[7] The first cohort started BAM in September 2020 (ending July 2022). The second cohort started in September 2021 (ending July 2023), although recruitment to the programme is not finalised until the second half-term. The evaluation of the first cohort is a feasibility study, while the evaluation of the second cohort is a pilot outcomes study. The second cohort involves students in three secondary schools in Lambeth (two of which are the same as those in the first cohort).[8] Each school has a dedicated BAM counsellor delivering the sessions.

Developing the programme theory

Combining theories of change with realist evaluation

To shed light on evidence of BAM's promise, our approach to contribution analysis must be informed by a '*reasoned*' theory of change. This is one in which "*the chain of results, and the assumptions behind why the intervention is expected to work are plausible, sound, informed by existing research and literature and supported by key stakeholders*" (Mayne 2012, p.272). Evaluations based on theories of change have been criticised for failing to develop and test the causal logic of interventions in sufficient depth (Breuer et al. 2016). Realist evaluation is well-placed to address this challenge (Mackenzie and Blamey 2007; Rolfe 2019). It generates hypotheses regarding how programme mechanisms (M) lead to certain outcomes (O) in particular contexts (C) (Pawson and Tilley 1997). These context-mechanism-outcome configurations, or 'CMOs', act as the fundamental building block of realist research.

Learning from year one of the feasibility study

Our experience of combining both approaches during the feasibility study has helped us to strengthen the reasoning underpinning the theory of change. The version that informed year one of the study combined formal scientific theory and local stakeholder expertise to articulate high-level predictions regarding how BAM is supposed to work, for whom, under what circumstances and why. We then used the CMO heuristic to generate hypotheses that captured the causal links between implementation, mechanisms, outcomes, context and BAM's implementation teams. Using Jackson and Kolla's (2012) concept of CMO dyads, we used each hypothesis to articulate the relationship between two items within these five

[7] This figure was agreed during the grant application process.
[8] The Pupil Referral Unit involved in the first cohort has been replaced for the pilot study by a third school.

categories (for example, between an aspect of implementation and a mechanism, or a contextual factor and a mechanism). We generated 86 hypotheses in total.

The findings from year one informed amendments to the theory of change, the latest version of which can be found in Appendix 1. As well as revising content, we also changed the structure. First, the revised theory of change is divided into sub-theories – implementation, intermediate outcomes, and ultimate outcomes – with mechanisms and contextual factors articulated at each stage based on the findings from year one. Each sub-theory sets the conditions for the one to follow, a common feature in realist evaluations of complex interventions (e.g., Shaw et al. 2018).

Second, since 86 hypotheses were too many and too specific, each of the sub-theories within the revised theory of change is now structured as a realist matrix, in a similar fashion to those developed by Ebenso et al. (2019). Each sub-theory constitutes one candidate theory, articulated as a CMO statement in the 'if-then' format typical of realist evaluations (Pawson and Manzano 2012).

Unintended consequences

Drawing on the findings from year one of the feasibility study, the concerns of delivery and community partners and wider research – particularly Evans et al.'s (2015) study into the unintended consequences of targeting SEL interventions – an additional sub-theory was developed for unintended consequences. The evaluation will also consider the potential harmful effects resulting from participants missing lessons to attend BAM.

Recruitment

Young people and their parents/carers will be recruited to the intervention by the BAM counsellors and colleagues in the MHF. Participants are eligible if they are boys aged 12-14 years in one of three secondary schools in Lambeth and experiencing challenges in at least one area of their social-emotional development according to the HSA. However, all groups must incorporate a range of young people with different types and levels of need. Specifically, all groups and school cohorts will aim for approximately a 15%/70%/15% split across the three respective tiers of the HSA. This is to prevent negative labelling of participants. Counsellors will also work with school leadership to ensure that, as far as possible, groups are not made up of factions of young people.

All participants in the intervention (~80 young people) are eligible to participate in the evaluation. At the point of enrolment to BAM, parents/carers are given a Privacy Notice by the MHF and asked if they agree to their contact details being shared with the evaluation team. We will send parents/carers who agree an information sheet and consent form for them and the young person in their care and arrange either for (i) the parent/carer and young person to complete and return it to the evaluation team or (ii) a call (online/phone)

during which the evaluation team can obtain verbal informed consent from the carer and young person.[9]

Data collection

As in our approach to the feasibility study and how other evaluators have applied contribution analysis (e.g., Dybdal et al. 2010; Delahais and Toulemonde 2012), the BAM theory of change will be converted into a data collection table. This table, which outlines what data will be collected from whom for each element in the theory of change, will act as an operational framework to guide the data collection process. The table in Appendix 2 provides a high-level summary. Further detail on the use of qualitative and quantitative data collection methods is provided in what follows.

Quantitative data: Outcomes

Based on the aims of BAM, it is planned to collect data on the following outcomes for young people:

1. Socio-emotional skills/competencies (including behaviour)
2. Engagement with school and learning
3. Offending / anti-social behaviour

The first outcome area will be tested via the youth self-report version of the Strengths and Difficulties Questionnaire (SDQ; Goodman, 1997; www.sdqinfo.com). The SDQ is a 25-item self-report tool with five 5-item subscales on hyperactivity, conduct, emotions, peer relations and prosocial behaviour. Each subscale is scored 0-10, with higher scores indicating greater difficulties (except for the prosocial behaviour subscale, where higher scores indicate greater prosocial behaviour). A total difficulties score[10] scored 0-40, with higher scores indicating greater difficulties. An impact supplement measures whether reported difficulties upset or distress the child and the extent to which they interfere with home life, friendships, classroom learning and leisure activities (scored 0-10, higher scores indicate greater impact).

The first and second outcome areas will be measured using data from the Holistic Student Assessment (HSA; Allen et al., 2017), a validated 61-item self-report assessment tool that measures social-emotional skill development and learning and school engagement. The HSA assesses respondents against 14 sub-scales grouped into three categories of life skills: (i) Resiliencies (internal and external); (ii) Relationships; and (iii) Learning and school

[9] The parent/carer and young person do not necessarily need to be contacted at the same time but the young person's participation in the evaluation requires that both parent/carer *and* young person provide consent.
[10] The sum of four subscales: conduct problems, hyperactivity, emotions and peer relations.

engagement. Each of the 14 sub-scales is deemed to be a 'strength' or a 'challenge' for a young person if their score is at least one standard deviation away from the mean of a normative sample developed by the assessment developers, the Pear Institute.[11] YG uses the HSA for BAM owing to (i) its alignment with BAM's core values and (ii) the useful and usable insights that the HSA can provide counsellors and schools.

For the second outcome area we will additionally use data on participants routinely collected by schools, including attainment, attendance, and behaviour (specifically records of internal and external exclusion, both temporary and permanent).[12] This data will also be used to inform understanding of the extent to which missing lessons as a result of BAM has a negative influence on educational performance.

For the third outcome area, we had planned to use the 19-item Self-Reported Delinquency Scale (SRDS), which focuses on criminal acts committed in the last 6/12 months (criminal damage, stealing, robbery etc.) (Smith and McVie 2003). However, the MHF raised strong reservations about the use of this measure in the evaluation, in particular the concern that it would stigmatise participation in BAM, undermine the counsellor-participant relationship and reduce young people's engagement in the intervention, all of which could potentially reduce the effectiveness of the programme. Prior to its use in the evaluation, therefore, it has been decided in collaboration with the MHF add YEF to explore options for measuring this outcome in a manner that is acceptable and rigorous (see below).

Quantitative data: Implementation

We will draw on existing YG and MHF systems and measures to track implementation. This will include: dose (number of BAM circle/brief encounter/one-to-one sessions delivered/attended, length of sessions); adherence to content (counsellor-completed implementation checklists after each session, formal assessments of counsellors by the replication specialist); quality (including counsellors' core competencies, assessed by the YG replication specialist);[13] and recruitment (proportion of participants, groups and schools meeting eligibility criteria, retention[14]).

[11] This sample is based on a stratified random sample of 9,000 male and female participants aged between 9 and 19 years old, drawn from a population of 27,808 HSA respondents in the US (Allen et al. 2017).
[12] This will be subject to the availability and quality of this data at the three participating schools.
[13] There are five core counsellor competencies: systemic leadership (how counsellors engage school staff and parents); clinical skills; modelling; group work; and youth engagement. Youth Guidance developed a six-step coaching model in partnership with the National Implementation Research Network at the University of North Carolina. Counsellors are given a score for each competency against a five-point metric by the replication specialist via observation of BAM circles at the beginning and end of the school year. The five points are: Basic Knowledge; Novice; Intermediate; Advanced; and Expert.
[14] For example, the proportion of participants attending 80% of BAM Circles (the target threshold).

Quantitative data: Other

MHF will share socio-demographic data on the backgrounds of students and counsellors, including age (and year group of students) and ethnicity. Depending on the quality and availability of the data, we will also draw on information collected by schools and shared via MHF regarding free school meals eligibility (FSM), attendance, attainment and behaviour.

MHF also carry out a feedback survey with all young people on BAM at the end of the school year, to inform ongoing improvements to the service. This is a light-touch online questionnaire that asks young people to share their feelings about their counsellor and their favourite and least favourite things about BAM.

Qualitative data

Qualitative data collection methods will include counsellor case notes, interviews (with young people, parents/carers, school staff and other stakeholders) and focus groups. In line with standard realist evaluation practice, we will adopt a purposive approach to sampling, based on which stakeholder is deemed best placed to speak to a particular element of our programme theory (Manzano 2016).

Drawing on learning from year one of the feasibility study, qualitative data collection with young people will involve interactive methods that cater to a variety of preferences, including role play, drawing and written responses to questions. A consultant facilitator will be hired who will share the ethnicity, gender and some of the lived experiences of young people participating in BAM.

We will apply realist interview principles to qualitative data collection, by explicitly orienting interviewees to the theory being evaluated and asking them to speak to that theory, based on their experience (Manzano 2016). To aid this process, topic guides for qualitative interviews and focus groups will be informed by those developed by Brand et al. (2019). Interview questions will feature alongside associated theory, which will be included as references in all topic guides. These topic guides will also include open-ended exploratory questions to ensure that the evaluation remains open to alternative and unanticipated contextual factors and unintended consequences. A group of young people from London of a similar age to those on BAM will be invited to scrutinise the topic guide for interviews with programme participants, to ensure the language and format are appropriate.

Process for piloting delinquency measure

Given the concerns raised by the MHF about use of the SRDS measure we will:

1. Convene a group (or groups) of young people (ideally some from BAM, but if not then young people similar to BAM participants), explain the issue and discuss the SRDS and alternative measures and options for their use. This includes

experimenting with different ways of improving the acceptability of measuring involvement in crime, violence and anti-social behaviour. This may include: removing the title of the assessment; combining the assessment with other survey questions (especially more strengths-based questions); and ensuring that researchers – rather than BAM counsellors – facilitate the assessment's completion. This will help us to understand young people's reactions to the assessment. Where possible, we will seek to recreate the conditions that would characterise the completion of measures in real life; however, some options – for instance, asking a whole year group (as opposed to BAM young people only) to complete the measure(s) – will not be possible in practice and will instead need discussion with students.

2. Report back to YEF with a recommendation.

3. Either:

    a. apply the measure in the evaluation (at two time points if possible)

    b. consider other options for use, as needed, in a next-stage evaluation

Who collects what data

The MHF will collect HSA and SDQ outcome data from young people participating in BAM at two timepoints with both cohorts, as part of their standard monitoring and evaluation procedures. Young people will use the IT facilities available in their school to complete the assessments online during a BAM circle session. BAM counsellors will facilitate the session with support from the Senior Evaluation Officer at MHF. The first timepoint is once young people have been formally recruited and before the intervention proper starts (i.e., following the first 3 orientation lessons of the curriculum). The second timepoint is during the final month of BAM Circle sessions.

BAM counsellors also collect routine data on programme implementation (recruitment, attendance, and fidelity). Data collection tools for this purpose have been developed and refined over many years by YG. These tools were adapted to the context of London as part of a collaborative effort involving YG and the Senior Evaluation Officer for BAM at MHF. They support counsellors to collect both qualitative and quantitative data:

- Quantitative data: demographic information on participants; details on the delivery of each of the four main activities (BAM circles, brief encounters, 1:1 support, special activities[15]), including when a particular session takes place, who attends and what the session covered (including implementation checklists after each session)

- Qualitative data: Any adaptations to the BAM circle curriculum; counsellor case notes and reflections on group behaviour; and information related to the counsellor's school, including records of meetings as well as assessments of the degree to which the school represents an enabling context for the implementation of BAM.

A BAM coach at YG will collect data on the development of BAM Counsellors. This covers how BAM counsellors progress through BAM's coaching model, including recordings of meetings between counsellors and coaches (date, time, topic and case notes), the coaching plans for each counsellor drafted by the replication specialist, and assessments by both the counsellors and the replication specialist of each counsellor's progress against the core counsellor competencies. During these assessments, counsellors are given one of five grades across each of the competencies: basic; novice; intermediate; advanced; and expert. Neither sums nor averages are taken of these grades; instead, areas for improvement are identified and prioritised for the next phase of each counsellor's development. This information will be stored on YG's Case Management System.

Schools will share data on the academic attainment, attendance and behaviour of participants with MHF.[16]

Methods for minimising bias

Counsellors will receive training from both the MHF Senior Evaluation Officer and the PEAR Institute (which developed the HSA) in how to collect data as part of routine service delivery (e.g., not asking leading questions, not influencing answers). The quantitative measures have proven validity and reliability. Interviews will be conducted by experienced and trained members of the evaluation team. Participants will be informed that the information they provide will be confidential and their comments reported anonymously.

Data analysis

In contribution analysis, data analysis occurs at step four (drafting the contribution story, based on evidence collected during step three) and step six (finalising the contribution story, based on supplementary evidence collected during step five). As previously discussed,

[15] Special activities were not delivered in the first year of the feasibility study owing to COVID-19.
[16] This will be subject to the availability and quality of this data at the three participating schools.

the findings from the feasibility study will constitute step four, while their interrogation by stakeholders and subsequent data collection during the pilot will be step five.

Each element of the theory of change will be systematically analysed to determine the extent of its contribution to the next link in the results chain. The robustness of a contribution claim will depend on whether the intended change occurred and whether the supporting evidence is strong, convergent and triangulated (Delahais and Toulemonde 2012).

As a mixed methods study, all findings will be considered together to inform a decision about the need for and value and nature of further intervention development and evaluation.

Quantitative data: Outcomes

Data from pre-post measures will be scored and subjected to descriptive and inferential statistical analysis (using paired sample t-test). This will show whether outcomes improve during the intervention. We will also calculate: standard deviation and correlation between pre-post scores (with 95% Confidence Intervals); reliability data for the measures; and the extent to which outcome scores are similar for young people within a school or BAM group (Intraclass Correlation Coefficient; ICC).

Quantitative data: Implementation

Summary statistics will be used to describe implementation in terms of recruitment, attendance, fidelity and quality:

- Recruitment – group size, proportion of young people in each HSA tier (1, 2, 3) in each BAM group
- Attendance – number of BAM circles held, mean number of BAM circles attended, percentage of BAM circle sessions attended, proportion of young people receiving ≥1 brief encounter each week, proportion of young people receiving a 1:1 each month
- Fidelity – progression through BAM manual (number of lessons reached per group) and adherence to the BAM manual. Quantitative data on adherence will be captured and analysed using two sources. First, counsellors will use implementation checklists to record how many activities they completed out of all those recommended within each lesson. Greater completeness will indicate greater adherence. Second, replication specialists will formally assess counsellors' ability to deliver the curriculum with fidelity. Higher ratings will also indicate greater adherence.

- Quality – extent to which formal assessment of counsellors by the replication specialist indicates that counsellors are on track with their development (i.e., performing as well as would be expected of a counsellor with three years experience).[17]

Quantitative data: Other

Summary statistics will be used to describe the number and socio-demographic characteristics of young people recruited to BAM. Inasmuch as data permit, these will be compared with school and neighbourhood statistics.

Qualitative data

The qualitative analysis will build on the approach we used during year one of the feasibility study. This includes: an ontological position associated with retroduction (Jagosh 2020) – i.e., using both top-down, theory-driven reasoning and bottom-up, data-driven reasoning to understand and unearth causal chains and connections that are not immediately observable (Greenhalgh et al. 2017); the application of framework analysis (Ritchie and Spencer 1994; Brand et al. 2019); and the use of the NVivo 12 software package[18] to build and apply our working analytical framework.

Interpretation

As in year one of the feasibility study, we will facilitate one or more feedback sessions with programme staff from YG and the MHF to give delivery partners the opportunity to interrogate and validate preliminary findings and to address inconsistencies in data.

Outputs

Results from the feasibility and pilot outcomes phases of the evaluation will be written up in a final report for the YEF.[19] This will include recommendations regarding future intervention development and evaluation.

These are in addition to two earlier reports produced as part of the BAM evaluation: a report on the process and outcome of adapting BAM for the UK; and an interim report on the first year of the feasibility phase of the evaluation.

Ethics and registration

[17] This will be discussed and agreed with the replication specialist.
[18] NVivo qualitative data analysis software; QSR International Ply Ltd. Version 12, 2018.
[19] If the YEF prefers, two separate reports could be produced.

The Warren House Group Ethics Committee approved the ethics submission for the feasibility phase of the evaluation (Ref: 19/20-1301, dated 17th December 2020). An ethics amendment for the pilot outcomes evaluation phase has been prepared and will be submitted to the same committee for Chair's Action once this protocol has been finalised.

Data protection

Legal basis

The legal basis to collect and process personal and sensitive information for this project is 'public task' i.e., research carried out in the public interest. In the UK, section 8 of the *Data Protection Act* 2018 says that the public task basis can cover processing that is necessary for, among other things, 'the exercise of a function of the Crown, a Minister of the Crown or a government department.' This project is funded by the Home Office (via the YEF) in the exercise of their statutory powers to assist victims, witnesses or other persons affected by offences. On this basis, the YEF has recommended the use of public task as the lawful basis for all evaluations of their grantees.

Confidentiality

Each individual participant will be assigned a Pseudo ID known only to the evaluation team. The Master Index linking PseudoIDs to personal identifiers (first name; family name) will be stored in a file separate to the evaluation data on the secure shared drive.

It is a requirement of the funding that limited identifiable data are shared with the Department for Education (DfE) for the purposes of data archiving. Specifically, the research team must share a dataset containing the following data: child's name, DOB, gender, UPN, level of participation in the intervention (e.g., sessions completed), and outcome data (pre/post). Once this reaches the DfE it will be pseudonymised. Once the dataset with identifiable data has been securely transferred to the DfE we will delete it.

Anonymity

For young people and their parents/carers, no real names or other identifiers/distinguishing features of participants will be used on any reports, presentations or papers. Discussions with MHF, YG and other project partners will not contain specific details of cases but will focus on issues raised to preserve anonymity.

For interviews / focus groups with the project team, school staff and community partners, it will be possible to anonymise findings that are shared with the YEF, wider stakeholders, the general public and young people and their families. However, it will be difficult to

anonymise findings shared with the project team responsible for implementation (the counsellors and other programme staff from MHF and YG). With those participants for whom complete anonymity in all instances will be difficult to guarantee (i.e., the project team and some school staff/community partners), as part of the consent procedures the researcher will clarify what will or won't be possible regarding the anonymity of their views with different audiences. It will be made clear that while we will preserve as much anonymity as we can, using our judgement about what is helpful, their views may be attributed to them in reports to the project team. However, anonymisation will be assured for any documents or reports shared with the YEF, wider stakeholders, the general public or young people and their families participating in the evaluation. Their names and other identifying details will not be attributed to any quotes or feedback about BAM for these audiences unless they explicitly give us permission to do so.

Regarding data in the YEF archive, no-one who looks at information in the archive will know the identity of participants.

Data quality

Procedures for collecting routine data (outcomes, implementation) as part of service delivery are overseen by the Senior Evaluation Officer at the MHF. In practice, this includes providing counsellors with training in (i) how to administer measures to minimise bias and (ii) how to enter data into the MHF's Case Management System (developed by Social Solutions through their Apricot software but managed and maintained by programme staff at MHF). Data are analysed on a monthly basis and any issues identified so that they can be rectified as quickly as possible.

Regarding primary data collected by the evaluation team, staff have previous experience of using such methods and receive additional training on the specific tools. Where necessary, initial data collections will be shadowed by a more experienced team member.

Data sharing and storage

Data will be shared by the MHF with the Dartington Service Design Lab according to a data sharing agreement. Primary data will be stored in password protected files on secure servers at the Lab and the Universities of Plymouth and Exeter. Data will only be accessible to evaluation team members and from password protected computers. Any hard copy data (e.g., consent forms) will be stored in a locked filing cabinet in a locked office on the property of the Lab, University of Plymouth or University of Exeter. After the research is complete, a dataset (containing outcome data (pre/post), level of participation in the programme and identifying information (child name, gender, date of birth, Unique Pupil Number) will be transferred securely to the Department for Education for deposit (in pseudonymised form) in the ONS Secure Research Service.

Personnel

Delivery team

Jane Caro (Associate Director of Programmes, MHF)

Ntale Eastmond (Project Manager, MHF)

Mariyam Farooq (Program Advisor, MHF)

Victoria Zamperoni (Senior Research Officer, MHF)

Dean Idoniboye-Obu (BAM counsellor, London Nautical School)

Hugh Mayers (BAM counsellor, City Heights Academy)

Kohliah Roberts (BAM counsellor, Saint Gabriel's College)

Antony Di Vittorio (Program Founder, YG)

A. J. Watson (BAM National Director, YG)

Wendy Fine (Director of BAM Research and Evaluation, YG)

Jason Story (Replication Specialist, YG)

Christopher Jaffe (Senior Manager, Partnerships and Operations, YG)

Michelle Morrison (CEO, YG)

Michael Bergstrom (Director of New Site Development, YG)

Rebecca Clarkin (COO, YG)

Roles

Recruitment to BAM: BAM Counsellors

Delivery in schools (BAM circles, 1:1 etc.): BAM Counsellors

Routine data collection: BAM Counsellors,

Data management and quality: VZ

Counsellor training and support: JC, NE, MF, VZ, JS, AV

Evaluation team

Dr Tim Hobbs, Dartington Service Design Lab

Dr Nick Axford, University of Plymouth

Finlay Green, Dartington Service Design Lab

Julia Mannes, Dartington Service Design Lab

Dr Lynne Callaghan, University of Plymouth

Kate Allen, University of Exeter

Dr Vashti Berry, University of Exeter

Roles

Co-Principal Investigators (responsible for oversight of study): TH, NA

Project Manager (responsible for day-to-day running of the project): FG

Study design and planning: TH, NA, FG, VB

Recruitment and data collection: JM, KA, FG

Quantitative analysis: FG, KA

Qualitative analysis: FG, LC, JM

Report-writing and dissemination: TH, NA, FG, LC, VB

Risks

The table below indicates risks to the project, their likelihood/impact, and steps we are taking to mitigate them.

Risk	Likelihood	Impact	Mitigation
Covid-19 and/or other issues lead to delay or other complications with BAM delivery	Medium	High	BAM counsellors follow Covid-19 precautions in school. Means of delivering BAM online learnt in feasibility phase could be used again.
Low sign-up to evaluation by parents/carers	Medium	High	Parents evenings hosted by evaluation team (opportunity for questions). Information sheets and consent forms made as simple as possible. Persistence by evaluation team. Support from BAM counsellors where necessary.
Low participation in data collection by students	Medium	High	Good use of routinely collected data. Students given options for qualitative data collection (in-person, online, telephone).
Measurement of crime/violence/anti-social behaviour has adverse effects on young people and/or their participation in BAM	High	High	Work with funder, MHF and community partners to find a mutually acceptable way to measure these outcomes.

Timeline

The table below outlines when each activity in the pilot outcomes evaluation will take place and who completes each task.

Dates	Activity	Staff responsible/ leading
By end of Dec '21	Set-up, including ethics approval and refinement of design for pilot outcome evaluation phase (inc. this protocol)	Evaluation team
Dec '21 / Jan '22	Baseline data collection of routine outcome data (HSA and SDQ) with young people formally recruited to BAM	MHF / BAM counsellors
Dec '21 / Jan '22	Recruitment to the evaluation of parents/carers and students formally in the BAM intervention	Evaluation team
Feb-Apr '22	Exploration of measures of crime, violence and ASB Qualitative data collection with BAM counsellors re. cohort 2 (especially recruitment phase)	Evaluation team
May-Jul '22	Contribution analysis 'step three': Midpoint HSA / SDQ Possible 'baseline' of SRDS (or similar)	MHF / BAM counsellors for HSA/SDQ Evaluation team for SRDS (or similar)
May-July '22	Contribution analysis 'step three': Qualitative data collection (interviews / focus groups) with cohort 1 (BAM counsellors, parents/carers, students, other stakeholders)	Evaluation team
Aug-Dec '22	Contribution analysis 'step four': Analysis and write-up of feasibility data	Evaluation team

Jan-May '23	Contribution analysis 'step five': Qualitative data collection (interviews / focus groups) with cohort 2 (BAM counsellors, parents/carers, students, other stakeholders) + analysis of the data	Evaluation team
Jun/Jul '23	Contribution analysis 'step five': End point HSA / SDQ Possible follow-up of SRDS (or similar)	MHF / BAM counsellors for HSA/SDQ Evaluation team for SRDS (or similar)
Aug-Oct '23	Contribution analysis 'step six': Analysis of cohort 2 HSA and SDQ Analysis of cohort 2 SRDS Write-up of pilot outcome phase, leading to final report Dissemination	Evaluation team

References

Allen, P. J., Thomas, K., Triggs, B., Noam, G. G. (2017). *The Holistic Student Assessment (HSA) Technical Report.* Belmont: MA, The PEAR Institute: Partnerships in Education and Resilience.

Blamey, A., & Mackenzie, M. (2007). Theories of change and realistic evaluation: Peas in a pod or apples and oranges? *Evaluation*, 13(4), 439-455.

Brand, S. L., Quinn, C., Pearson, M., Lennox, C., Owens, C., Kirkpatrick, T., et al. (2019). Building programme theory to develop more adaptable and scalable complex interventions: Realist formative process evaluation prior to full trial. *Evaluation*, 25(2), 149-170.

Breuer E, Lee L, de Silva M, et al. (2016) Using theory of change to design and evaluate public health interventions: A systematic review. *Implementation Science,* 11: 63.

Dalkin, S.M., Greenhalgh, J., Jones, D. et al. (2015). What's in a mechanism? Development of a key concept in realist evaluation. *Implementation Science,* 10: 49.

Delahais, T. and Toulemonde, J. (2012) Applying contribution analysis: Lessons from five years of practice. *Evaluation*, 18(3), 281-293.

Dybdal, L, Nielsen, SB, Lemire, S (2010) Contribution analysis applied: reflections on scope and methodology. *Canadian Journal of Program Evaluation* 25(2), 29-57.

Ebenso B, Manzano A, Uzochukwu B, et al. (2019). Dealing with context in logic model development: Reflections from a realist evaluation of a community health worker programme in Nigeria. *Evaluation and Program Planning,* 73, 97-110.

Evans, R., Scourfield, J. and Murphy, S. (2015). The unintended consequences of targeting: young people's lived experiences of social and emotional learning interventions. *British Educational Research Journal* 41(3): 381-397.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry,* 38(5), 581-586.

Greenhalgh, T., Pawson, R., Wong, G., Westhorp, G., Greenhalgh, J., Manzano, A., Jagosh, J. (2017) Retroduction in Realist Evaluation. *The RAMESES II Project.* https://www.ramesesproject.org/media/RAMESES_II_Retroduction.pdf

Jackson, S. F. and Kolla, G. (2012) A new realistic evaluation analysis method: linked coding of context, mechanism, and outcome relationships. *American Journal of Evaluation*, 33(3), 339-349.

Jagosh, J. (2020) Retroductive theorizing in Pawson and Tilley's applied scientific realism. *Journal of Critical Realism*, 19(2), 121-130.

Heller, S., Pollack, H. A., Ander, R., & Ludwig, J. (2013) *Preventing Youth Violence and Dropout: A Randomized Field Experiment*. Cambridge MA: National Bureau of Economic Research.

Heller, S. B., Shah, A. K., Guryan, J., Ludwig, J., Mullainathan, S., & Pollack, H. A. (2017). Thinking, fast and slow? Some field experiments to reduce crime and dropout in Chicago. *Quarterly Journal of Economics*, 132(1), 1-54.

Jagosh, J., Bush, P.L., Salsberg, J. *et al.* (2015). A realist evaluation of community-based participatory research: partnership synergy, trust building and related ripple effects. *BMC Public Health* 15: 725.

Joyce, A. S., MacNair-Semands, R., Tasca, G. A., & Ogrodniczuk, J. S. (2011). Factor structure and validity of the therapeutic factors inventory–short form. *Group Dynamics: Theory, Research, and Practice*, 15(3), 201-219.

Jung, C. G. (1969). *Archetypes and the Collective Unconscious* [*sic*], Collected Works of C.G. Jung, Volume 9 (Part 1). Princeton, NJ: Princeton University Press.

Lansing, J., & Rapoport, E. (2016). *Bolstering Belonging in BAM and Beyond: Youth Guidance's Becoming a Man (BAM) Program Components, Experiential Processes, and Mechanisms*. A Report to Youth Guidance. Chicago, IL: Chapin Hall at the University of Chicago.

MacNair-Semands, R. R., Ogrodniczuk, J., & Joyce, A. (2010). Structure and initial validation of a short form of the therapeutic factors inventory. *International Journal of Group Psychotherapy*, 60(2), 245-281.

Manzano, A. (2016) The craft of interviewing in realist evaluation. *Evaluation*, 22(3), 342-360.

Mayne, J. (2001) Addressing attribution through contribution analysis: using performance measures sensibly. *Canadian Journal of Program Evaluation*, 16(1), 1-24.

Mayne, J. (2012) Contribution analysis: coming of age? *Evaluation*, 18(3), 270-280.

Nagaoka, J.; Farrington, C.; Ehrlich, S. and Heath, R. (2015). *Foundations for Young Adult Success: A Developmental Framework*. The University of Chicago Consortium on Chicago School Research. Available at: https://consortium.uchicago.edu/publications/foundations-young-adult-success-developmental-framework

Orfanos, S., Burn, E., Priebe, S. and Spector, A. (2020) A systematic review and quality assessment of therapeutic group process questionnaires. *International Journal of Group Psychotherapy*, 70(3), 425-454.

Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. London: Sage.

Pawson, R. and Manzano-Santaella, A. (2012) A realist diagnostic workshop. *Evaluation* 18(2), 176-91.

Rolfe, S. (2019) Combining theories of change and realist evaluation in practice: lessons from a research on evaluation study. *Evaluation*, 25(3), 294-316.

Shaw, J., Gray, C.S., Baker, G.R. *et al.* (2018). Mechanisms, contexts and points of contention: operationalizing realist-informed research for complex health interventions. *BMC Med Res Methodol,* 18: 178.

Smith, D., & McVie, S. (2003). Theory and method in the Edinburgh study of youth transitions and crime. *British Journal Criminology*, 43, 169-195.

Tasca, G., Cabrera, C., Kristjansson, K., MacNair-Semands, R., Joyce, A. & Ogrodniczuk, J. (2016) The therapeutic factor inventory-8: Using item response theory to create a brief scale for continuous process monitoring for group psychotherapy. *Psychotherapy Research*, 26(2), 131-145.

Yalom, I. & Leszcz, M. (2005). *The Theory and Practice of Group Psychotherapy*. New York: Basic Book

YOUTH ENDOWMENT FUND

Impetus

Home Office

youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

@YouthEndowFund

The Youth Endowment Fund Charitable Trust

Registered Charity Number: 1185413