EVALUATION REPORT

Becoming A Man (BAM)

Pilot study report

Finlay Green, Nick Axford, Cristina Preece, Julie Harris, Kate Allen, Sean Manzi, Julia Mannes, Lynne Callaghan, Vashti Berry, Tim Hobbs, Ediane Santana de Lima and Amy Woodburn

July 2024

DARTINGTON SERVICE DESIGN LAB

UNIVERSITY OF PLYMOUTH

UNIVERSITY OF EXETER | MEDICAL SCHOOL

YOUTH ENDOWMENT FUND

About the Youth Endowment Fund

The Youth Endowment Fund (YEF) is a charity with a mission that matters. We exist to prevent children and young people from becoming involved in violence. We do this by finding out what works and building a movement to put this knowledge into practice.

Children and young people at risk of becoming involved in violence deserve services that give them the best chance of a positive future. To make sure that happens, we'll fund promising projects and then use the very best evaluation to find out what works. Just as we benefit from robust trials in medicine, young people deserve support grounded in the evidence. We'll build that knowledge through our various grant rounds and funding activities.

And just as important is understanding children and young people's lives. Through our Youth Advisory Board and national network of peer researchers, we'll ensure they influence our work and we understand and are addressing their needs. But none of this will make a difference if all we do is produce reports that stay on a shelf.

Together, we need to look at the evidence and agree on what works, then build a movement to make sure that young people get the very best support possible. Our strategy sets out how we'll do it. At its heart, it says that we will fund good work, find what works and work for change. You can read it here.

For more information about the YEF or this report, please contact:

Youth Endowment Fund
C/O Impetus
10 Queen Street Place
London
EC4R 1AG

www.youthendowmentfund.org.uk

hello@youthendowmentfund.org.uk

Registered Charity Number: 1185413

Contents

About the evaluator

This evaluation was conducted by researchers from the Dartington Service Design Lab ('Dartington'), the University of Plymouth ('Plymouth') and the University of Exeter, with support from Black Thrive.

Dartington is an independent research and design charity that is committed to improving outcomes for children, young people and families. The organisation's expertise in evidence generation in the social sciences stems from a 60-year history of leading evaluations of varying designs and complexity, from formative to experimental.

The Community and Primary Care Research Group at the University of Plymouth has a strong track record of health and social care research, including prevention and early intervention to improve child and youth psychosocial outcomes.

The Children and Young People's Mental Health research collaboration at the University of Exeter researches the mental health and well-being of children and young people, with the aim of developing evidence-based policy and practice to improve the lives of children/young people and the communities around them.

Black Thrive Global evolved from the work of the Black Thrive Lambeth partnership, which was established in 2016 to address the inequalities that negatively impact the mental health and well-being of Black people in Lambeth. The Black Thrive Partnership brings together individuals, local communities, statutory agencies and voluntary organisations to address the structural barriers that prevent Black people from thriving.

The time of Kate Allen, Nick Axford, Vashti Berry and Lynne Callaghan is supported by the National Institute for Health and Care Research Applied Research Collaboration South West Peninsula. The views expressed in this publication are those of the authors and not necessarily those of the National Institute for Health and Care Research or the Department of Health and Social Care.

Evaluation team email: info@dartington.org.uk

Executive summary

The project

Becoming a Man (BAM) aims to support 12- to 16-year-old boys to improve their social and emotional skills, self-awareness, relationships and engagement in school. BAM was developed by Youth Guidance (YG), a Chicago-based non-profit, and has demonstrated positive impacts on reducing arrests in robust, US-based evaluations. It is currently being delivered in the UK by the Mental Health Foundation (MHF). The central component of BAM is the BAM Circle, a weekly group session delivered in school to 8–12 young people. These sessions are led by a BAM counsellor. BAM Circles include a range of activities featured in the BAM Manual, including check-ins, role plays, group missions, videos, lectures, and stories. These activities sit across the manual's thirty lessons, which BAM counsellors deliver over the course of fifty BAM Circle sessions over two years. The activities promote the internalisation of BAM's core values, including Integrity, Accountability, and Positive Anger Expression. Alongside BAM Circles, the intervention also facilitates special activities (group activities outside of school property), brief encounters (quick informal check-ins between the BAM counsellor and young person), and one-to-one support (individualised support for those with greater levels of need).

In this project, BAM was targeted at 12–14-year-olds (who became BAM scholars) in three Lambeth secondary schools. Counsellors worked with senior leadership, heads of year and the pastoral teams in each school to identify scholars. The intention was to select a group of children where 15% were largely in line with expected levels of development but exhibited some challenges with social and emotional skills; 70% were to have a balance of strengths and social and emotional challenges; and 15% were intended to be approaching crisis, or in a crisis. This selection approach aimed to prevent negative labelling of the intervention.

YEF previously published a feasibility study of the first year of BAM delivery (in the academic year 2020-21). 95 children in two schools and one Pupil Referral Unit in Lambeth received BAM during this period. The feasibility study established that the quality of delivery was generally successful, and participants had positive perceptions of the programme. It also highlighted challenges including completing the curriculum, and attendance (which was lower than anticipated). However, the first year of programme delivery was shown to be feasible, and so YEF progressed to funding this pilot study. This pilot study aimed to further explore BAM's implementation and ascertain how successfully BAM was implemented, explore whether implementation may be contributing to reported changes in social and emotional development, academic performance and behaviour, and establish whether the intervention may lead to any unintended consequences. The evaluation collected quantitative data including routinely collected programme data on programme participation, participant socio-demographics and the Holistic Student Assessment and the Strengths and Difficulties Questionnaire. Qualitative data collected included interviews (with six counsellors, 14 parents/carers, 11 children, three school staff, a staff member from Youth Guidance and a staff member at the Mental Health Foundation) and focus groups (with counsellors). 97 children participated in the pilot and the pilot phase covered both the academic year 2020-21 (explored by the previous feasibility study), and the academic year 2021-22.

Key conclusions
Some elements of BAM were implemented with success, while others proved more challenging.
Counsellors perceived recruitment of children to be effective and six of the nine BAM groups reached the target size (8 to 12). However, there were some challenges in effectively recruiting children from all three tiers of need.
Children responded well to the curriculum, finding it fun, engaging, relevant and varied. However, it did prove challenging to progress through each of the 30 lessons. Most scholars and carers who were interviewed perceived counsellors to be empathetic, supportive and trustworthy. Counsellors' previous skills and experience, in addition

to BAM training and coaching, supported them to deliver sessions. However, peer supervision and project management support appeared to have less impact.
Interviews with counsellors, scholars, parents/carers, and school staff suggest that BAM contributed to children's social and emotional development and responsible decision making. However, the pilot was not designed to attribute impact on these skills to BAM based on a comparison with a counterfactual group.
Negative labelling of BAM groups within schools did not appear to occur. Some school staff did note that missing school lessons may be a risk for their academic performance.

Interpretation

Some elements of BAM were implemented with success, while others proved more challenging. For instance, children's attendance on the programme was mixed. Children attended an average of 17 BAM Circle sessions in year one and 13 in year two (compared to a minimum annual target of 13, and a total possible 45 sessions). Higher attendance levels were prevented by some scholars joining the programme late and some leaving early (due to leaving the school, being removed due to their behaviour, missing several consecutive sessions, parents withdrawing them to focus on lessons, or being removed due to behaviour). Scheduling the circle sessions also posed a challenge, with only one school arranging the intended 45 sessions. Counsellor sickness, transport and teacher strikes, and school closures at one school prevented the arrangement of all planned sessions. Fewer children than expected received brief encounters and one-to-ones; the target number of brief encounters was met for only one month, in one school.

Counsellors perceived the recruitment process to be effective, and six of the nine BAM groups reached the target size (of 8 to 12 children). However, there were some challenges in effectively recruiting children from all three tiers of need; 56% of children in one of the schools had children in crisis (compared to the target of 15%).

Children responded well to the curriculum, finding it fun, engaging, relevant and varied. However, it did prove challenging to progress through each of the 30 lessons, and no group completed all of the content. Pre-existing factions within the groups may have also undermined group safety and limited curriculum progression. Most scholars and carers who were interviewed perceived counsellors to be empathetic, supportive and trustworthy. Counsellors' previous skills and experience, in addition to BAM training and coaching, supported counsellors to deliver sessions. However, peer supervision and project management support had less impact. Most counsellors could improve at challenging and confronting scholars. BAM was successfully adapted to an English context, with both surface adaptations (to language, symbols and illustrative examples), and deep adaptations (to the core elements of the theory of change) being made.

While data was collected on a range of outcomes, the absence of a counterfactual group means that the pilot was not designed to attribute impact on social-emotional skills, responsible decision-making, attendance, suspension, behaviour, and attainment to BAM.

Interviews with counsellors, scholars, parents and carers, and school staff reveal that BAM was perceived to support some children's social and emotional development and their decision making. Children who engaged most in action, and actively experienced BAM's core values before reflecting on these experiences, were perceived to be most likely to benefit from BAM. Where BAM was not perceived to positively impact children, the evaluator suggests that this may have been due to limited exposure, difficulties in counsellors managing group work, and other challenges in children's lives that might outweigh the potential impact of BAM. Further evaluation is required to establish a robust estimate of impact.

Negative labelling of BAM groups within schools did not appear to occur. While factions were present in groups, these were not groups of friends widely regarded as badly behaved within the wider school, which may have precipitated negative labelling. While counsellors, children, and parents did not identify this as a risk in interviews, some school staff did note that missing school lessons may be a risk to attainment, particularly for those children receiving other interventions during school time.

YEF is currently considering whether to proceed with further evaluation of BAM.

Introduction

Background

Reducing the number of young people involved in violence and offending continues to be a public health priority in the UK (GOV.UK, 2019). In a recent survey of over 7,500 young people aged 13 to 17 in England and Wales, 16% reported having been a victim of violence in the preceding 12 months, and 15% young people were victims (21%) and perpetrators (22%) of violence compared to White young people (16% and 14%, respectively). Involvement in violence is associated with an increased risk of poorer life outcomes, including problems related to both mental health (Miliauskas et al., 2022) and physical health (Wright et al., 2017).

The available academic literature identifies a range of risk factors that may be associated with violence and offending among young people, including those at the level of the young person (e.g. self-esteem, poor self-regulation, social-cognitive deficits), the family (e.g. poor parental attachment), the peer group (e.g. relationships with peers engaged in anti-social behaviour), the community (e.g. living in socially disadvantaged neighbourhoods, disengagement from school) and society (e.g. belonging to a racially minoritised ethnic group) (Early Intervention Foundation and Cordis Bright, 2015; Lammy, 2017).

Recent reviews highlight a range of interventions that have been shown to help address these risk factors and reduce offending behaviour among young people (O'Connor and Waddell, 2015; Hendriks et al., 2018). Becoming A Man (BAM) is one such intervention. It was selected because it has a Level 4 evidence rating in the Early Intervention Foundation Guidebook, meaning that it has "evidence of a long-term positive impact on child outcomes through multiple rigorous evaluations".[1]

BAM is a selective, school-based, social-emotional learning (SEL) intervention for 12-to-16-year-old boys.2 Two randomised controlled trials (RCT) involving 2,740 and 2,064 students, respectively, have shown positive impacts on the 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; 2015). In both studies, all participants were from minority ethnic backgrounds, with around 70% African American and the remainder Hispanic.

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. Several other school-, family- and community-based interventions originating in the US that are concerned with preventing or addressing poor youth psychosocial outcomes (including crime and violence) have struggled to replicate positive effects when trialled in Europe (e.g. Sundell et al., 2008; Skärstrand et al., 2013; Baldus et al., 2016; Humayun et al., 2017; Fonagy et al., 2018; Segrott et al., 2022). There are numerous theories for why this occurs, including the possibilities that (i) incoming interventions inadequately fit the local service systems and culture and (ii) adaptations to the

[1] Early Intervention Foundation Guidebook: https://guidebook.eif.org.uk/programme/becoming-a-man

[2] In this project BAM was implemented with young people in Years 8 to 11 (aged 12–16 years).

intervention for the purposes of ensuring a better fit remove or dilute active ingredients and thereby nullify the theory of change (Burkhart et al., 2019).

Therefore, it is necessary to (i) gather evidence on whether, how, for whom and under what conditions BAM can feasibly be implemented in the UK and (ii) advance our collective understanding of how BAM works, for whom and under what conditions to establish whether and how BAM can be adapted for delivery in the UK. The feasibility phase of the study (2020–22) focused primarily on the first of these aims. It sought to help 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 at the outset that if BAM showed 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 feasibility study report (Green et al., 2023a), this criterion was met. While contextual issues in participating schools contributed to challenges with recruitment, retention and curriculum progression, counsellors were able to deliver enough sessions of sufficient quality and fidelity to justify BAM's progression to the next step. The resulting pilot phase (2021–23) of the evaluation, which is the subject of this report, focused on the second of the aforementioned two aims.

The broader context for this project is the state of SEL in secondary schools in England. SEL is concerned with fostering children's social and emotional skills, alongside their academic skills, in educational settings and can include developing young people's relationships, communication, decision-making, self-esteem and behaviour (Gedikoglu, 2021). There is substantial evidence for the effectiveness of SEL education on youth outcomes such as social and emotional skills, school bonding, conduct problems and academic performance (Durlak et al., 2011; Taylor et al., 2017; Clarke et al., 2021). Although school-based SEL programmes have been promoted in the UK since the late 1990s, Personal, Social, Health and Economic education was a non-statutory subject until 2020. By then, it had become clear that provision was patchy and not being monitored systematically (Feinstein, 2015; Donnelly et al., 2020).

Therefore, in September 2020, Relationships, Sex and Health Education was made compulsory in all secondary schools (Department for Education [DfE], 2019).[3] This meant that it needed regular curriculum time like any other subject. Provision should be tailored, however, so in addition to regular universal provision (i.e. for all students, regardless of need), schools may offer targeted interventions for some pupils according to their needs (such as elevated risk of involvement in youth crime or violence). BAM falls into the category of additional targeted support. Owing to a lack of data about SEL provision in secondary schools in England, it is unknown how prevalent interventions like BAM are. However, other targeted initiatives exist, for instance, those provided by charities such as Football Beyond Borders, Kids Inspire and Khulisa (Granada et al., 2022). These offer a mix of mentoring and therapeutic support in group and one-to-one formats, and provider organisations work with a select number of local authorities or schools at any one time.

[3] Full implementation of this was delayed until September 2021 because of the Covid-19 pandemic.

Intervention

Introduction

BAM 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)[4] and one-to-one support (individualised support for those with greater levels of need to reinforce learning from the BAM Circles and offer emotional support to scholars in a confidential space). Practitioners – known as BAM counsellors – deliver the intervention in schools with groups of eight to 12 participants, called BAM scholars, over two school years. Collectively, these are designed to help young people internalise six core values: integrity, self-determination, positive anger expression, accountability, respect for womanhood and visionary goal-setting. Further details on BAM activities are provided below.

BAM was developed and is supported by Youth Guidance (YG), a non-profit organisation based in Chicago, US. The programme is being delivered in the UK by the Mental Health Foundation (MHF), a London-based charity working in the UK to promote good mental health for all.

Target group

BAM primarily targets adolescent boys living in deprived areas who are experiencing challenges with their social-emotional development. In this evaluation, scholars were eligible if they were boys aged 12–14 years in one of three secondary schools[5] in the South London borough of Lambeth and were experiencing challenges in their social-emotional development. Counsellors worked with senior leadership, heads of year and the pastoral team to identify scholars about whom they had concerns regarding their social-emotional development.[6]

The success of this endeavour was later determined using the Holistic Student Assessment (HSA; Allen et al., 2017). The HSA was administered between January and March 2022, three to five months after the start of the programme. Specifically, scholars were confirmed to be members of the target group if they were identified as experiencing challenges in at least one of the 14 subscales in the HSA (Allen et al., 2017), which are grouped into three categories of life skills (i) resiliencies (internal and external), (ii) relationships and (iii) learning and school engagement (further information on the HSA can be found in the Methods section). Race/ethnicity was not a target group criterion. If scholars were not in the target group, this did not affect their participation.

[4] Brief encounters are any ad-hoc, individualised engagements between young people and counsellors that involve more than a simple greeting but last less than 15 minutes. They usually (but not always) occur in a communal space within the school, such as the playground or the hallway in between classes.

[5] Two of the schools were also involved in the feasibility phase.

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

The HSA is delivered at this later stage because the first two to three months of BAM represent an orientation period, during which membership of and attendance at groups are flexible. Young people from the prospective target group are invited to take part. Some attend and continue to do so beyond orientation. Some attend and decide it is not for them. Others do not accept the invitation at all. The counsellors respond to this fluctuation in collaboration with schools by inviting others who also meet the selection criteria and who enable the balance in groups to be maintained to fill vacancies.

This gives young people a chance to see if BAM really is for them, which supports engagement and gives the counsellors and schools time to adjust group membership depending on who commits and who leaves. The goal is to get to months three or four with a group of eight to ten highly engaged scholars who want to be there and who are experiencing challenges with their social-emotional development. It is only at this stage, once the group is settled, that the HSA is administered.

Counsellors also supported school staff to recruit a range of scholars with different types and levels of need to prevent negative labelling of the intervention as a programme for 'bad' students. Specifically, all groups and school cohorts aimed for approximately a 15%/70%/15% split across the three respective tiers of the HSA. This was an approximation based on the knowledge of school staff of the specific context of each child:

- 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

Only Tier 1 students still deemed to be experiencing at least some challenges in their social-emotional development by school staff were selected to ensure all scholars on the programme could potentially benefit from BAM. Again, the success of this effort was established through the administration of the HSA in months three to five once the groups had settled.

Counsellors also worked with school leadership to ensure that, as far as possible, groups were not made up of smaller factions of young people (i.e., discrete friendship groups).

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

Activities

BAM is made up of four activities:

- BAM Circles (group sessions delivered in a school setting with eight to 12 scholars)
- Special activities (group activities outside of school property or school time)
- Brief encounters (quick, informal check-ins in public school spaces, such as the hallway or playground)
- One-to-one support (individualised intervention for those with greater levels of need used to reinforce learning from BAM Circles and offer emotional support to scholars in a confidential environment)

BAM circles constitute the central element and are delivered over two years (50 one-hour sessions in total, around 25 per year). They are delivered by a male counsellor with Qualifications and Credit Framework (QCF)-6 qualifications.[7] Counsellors are required to have clinical or therapeutic qualifications and experience working with young people. Counsellors are recruited from the communities in which scholars live and share some of their lived experiences to ensure they are relatable to boys on the programme. Sessions typically occur during school hours, substituting for a lesson.[8] Activities in BAM Circles are delivered in line with a 30-lesson curriculum called the BAM Manual (also known as the "BAMual"). One lesson takes one to three sessions to complete, with each lesson covering a range of activities, including check-ins and check-outs to open and close sessions, role plays, group missions, video education, lectures, stories and homework.

The sample for the evaluation constituted all scholars participating at this point, as well as those who joined after to fill spaces left by students who left prematurely or created through the splitting of larger groups into smaller groups. Further details on this are provided in the findings concerning Research Question 1.

Training and support

Counsellors receive 300 hours of programme training; they are recruited, trained, coached and supervised by the MHF, alongside YG, to support their development across a series of core competencies (see Appendix 1).[9] An advisory council is usually established to support 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 in which BAM scholars are based, 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 that 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.

Community partners in Lambeth include Black Thrive, a partnership of Black communities and service providers for Black well-being,[10] and Colourful Minds, a partnership of South London and Maudsley–employed psychiatrists and psychologists working particularly with racially minoritised communities.[11]

Theory of change

[7] Qualifications and Credit Framework: https://www.accreditedqualifications.org.uk/qualifications-and-credit-framework-qcf.html

[8] There were no official restrictions on which lessons BAM scholars could miss in order to attend BAM.

[9] 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.

[10] Black Thrive: www.blackthrive.org

[11] Colourful Minds: www.colourful-minds.org.uk

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 (see Green et al., 2023a), can be found in Appendix 2. It can be summarised as follows.

- In the long term, BAM aims to help scholars 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 scholars 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 scholars from risk while empowering them to take advantage of opportunities and resources in their environment. Whether they do so depends on wider influences in scholars' lives and whether these influences reinforce or limit scholars' efforts to apply the core values.
- BAM helps scholars internalise the core values through successful implementation (i.e. BAM is delivered with quality and fidelity to the intended target population, who attend BAM activities regularly); This is because of the different ways in which scholars actively experience each value before collectively reflecting on these experiences. Whether they engage in action and reflection depends on whether scholars come to BAM sufficiently ready and able to establish and maintain healthy groups – fun, safe spaces in which scholars challenge themselves and each other to be open, honest and vulnerable.

Research questions

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

We sought to establish evidence of promise by gathering evidence on BAM's programme theory, including evidence of unintended consequences, as recommended by YEF in their pilot study guidance (YEF, 2022). A theory of change for BAM was developed in collaboration with YG and MHF at the start of the feasibility study in 2020. This has been revised subsequently in the light of findings from the feasibility phase of the evaluation. BAM's programme theory is made up of four sub-theories: implementation, intermediate outcomes, ultimate outcomes and unintended consequences. These sub-theories gave rise to the research questions that guided 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,[12] for whom, under what circumstances and why?

[12] Subject to finding a suitably acceptable way to measure this

4. To what extent has implementation contributed to unintended consequences, for whom, under what circumstances and why?

Success criteria and/or targets

It was agreed that at the end of the pilot, evidence regarding intervention promise and the necessity, viability and nature of a next-stage evaluation would be interrogated by key stakeholders, including YEF, YG and MHF. The following criteria, which build on the pilot's research questions, were 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 used contribution analysis (Mayne, 2001).

Contribution analysis is a collaborative, systematic approach to establishing the extent to which a particular intervention contributed to changes in a specific set of outcomes. It can support evaluators in collecting and analysing data related to impact in projects that lack comparison groups or large samples. In contribution analysis, rigour is understood in terms of triangulation and the inclusion of multiple perspectives. Different data are collected from 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), it was agreed that the contribution story assembled during the pilot would 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, exposure and fidelity (see Table

1).[13] 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 are considered alongside evidence regarding with whom BAM can be successfully implemented, under what circumstances and why.

Table 1: Success criteria

Implementation area	Target
Recruitment	Each counsellor is delivering the intervention to five groups of eight to 12 young people by January 2022.[14]
	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 et 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.
Exposure	Young people attend a minimum of 13 out of an intended 25 BAM circle sessions on average every school year (26 attended in total across the two school years).
	Every scholar receives at least one brief encounter a month during term time.[15]
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 Circles 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.[16] 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. Further details on the

[13] Retention is an important implementation variable, but this is largely due to its influence on exposure. Therefore, a target for retention was not set in this study. Instead, retention is considered as an influencing factor in relation to exposure in the findings related to Research Question 1 below.

[14] This criterion has been updated. The target date in the pilot protocol is early November 2021. This was revised early, at the beginning of the pilot phase, to reflect YG's standard target date: the end of the calendar year during the first term of delivery.

[15] This criterion has been updated. The pilot protocol described the relevant target as "80% of scholars receive a brief encounter every week during term time." This target was used historically by YG. However, it has been revised in the intervening period because YG found that delivering brief encounters weekly was both unrealistic and difficult to assess and track.

[16] Subject to finding a suitably acceptable way to measure this

step-wise process the evaluation team followed to verify the theory of change and the strength of supporting evidence are provided in the Data Collection and Analysis sections below.

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

The success criteria regarding evaluation design concerned the extent to which outcome measures could be applied successfully (completion rates at baseline and follow-up), including the extent to which outcome measures were acceptable to participants and data collection procedures were efficient and minimally burdensome (from discussion with counsellors/school staff).

Ethical review

The Warren House Group Ethics Committee approved the ethics submission for the feasibility phase of the evaluation (Ref: 19/20-1301, dated 17 December 2020; see Appendix 3). An ethics amendment for the pilot outcomes evaluation phase was approved on 18 February 2021 (see Appendix 4).

Consent from scholars and their parents/carers

Routinely collected data

BAM counsellors sought written, informed consent from scholars and their parents/carers at the point of entry into the intervention (see Appendices 5 and 6) for the MHF to share the following with the evaluation team: (i) data collected by the school on academic attainment, attendance and behaviour and (ii) programme data collected routinely by counsellors. Parents/carers were notified why this was needed (to help with learning more about how BAM works in the UK and how programmes like BAM help scholars) and that parents'/carers' permission to share this would be sought separately. Further information was also sent to families at the start of the second year of the pilot, reminding them of BAM (and the data collection involved in BAM) and giving them an opportunity to opt out of the programme if they wished (see Appendix 7). Participation in BAM was not contingent on participation in the evaluation.

Parents/carers were also given an opportunity to opt their child out of taking the HSA and Strengths and Difficulties Questionnaire (SDQ) surveys at different times via an information sheet sent home to them by MHF. Students were also informed at the time of taking these surveys that they were voluntary.

Primary data collection

Informed consent was sought from participants prior to primary data collection. Interviews were conducted by members of the evaluation team via the telephone, a digital platform (Zoom or Microsoft Teams) or in person, depending on what was convenient for the participant. Potential participants were provided with either hard or digital (emailed) copies of the information sheets and consent forms as well as invitation

sheets for scholars (see Appendix 8) and could either provide written consent (if consent was taken in person) or audio-recorded consent (if consent was taken via telephone, Microsoft Teams or Zoom, which was recorded by the researcher on the consent form).

Contact with parents/carers was made via phone, text message or email. Contact with scholars was made through the BAM counsellor, other MHF staff or parents/carers. No arrangements were made with a young person directly by the evaluation team for data collection. In the case of collecting data from scholars, the information sheet and consent form described above were used, meaning that data could only be collected from a young person if the young person and their parent/carer consented. In the case of collecting data from parents/carers, the parent/carer was asked to provide consent for the evaluation team to interview them. Audio-recorded consent was stored separately from the data.

Ministry of Justice data and YEF data archive

An ethics amendment was obtained to permit (1) the collection of Ministry of Justice (MoJ) data on involvement in crime by BAM participants and (2) the deposit of data on participants' outcomes and participation in BAM in the YEF data archive. Submission of data to the YEF archive is a stipulation of YEF funding and a requirement of all pilot evaluations. The YEF data archive is stored in the Office for National Statistics (ONS) Secure Research Service. Approved researchers are able to use the YEF data archive to look at crime records (held by the MoJ) and education records (held by the DfE) and to connect those with young people's BAM records using a pupil reference number. There is more information about the YEF data archive on the YEF website: https://youthendowmentfund.org.uk/evaluation-data-archive/faqs-the-youth-endowment-fund-data-archive/. In both cases (MoJ data and YEF data archive), it was stipulated that active, informed consent was needed from both the young person and their parent/carer. Young people were given an information sheet and consent form (with separate sections to cover both requests), and a member of the research team visited the schools to meet them, explain the two requests and answer any questions they had. Parents/carers of any young people who consented were sent an information sheet and asked to provide consent to either or both requests.

Consent from counsellors, MHF and YG programme staff, and school staff

These participants were approached via email or telephone at the relevant time to ask if they would be interested in taking part in a semi-structured interview or focus group. Written or audio-recorded informed consent was provided by all participants using the informed consent form (see Appendix 9).[17] Audio-recorded consent was collected separately from the interview or focus group.

[17] For interviews with staff from MHF, YG and schools, it was possible to anonymise findings shared with the YEF, wider stakeholders, the general public and young people and their families. However, it was difficult to guarantee anonymity for findings shared with the project team responsible for implementation (the counsellors and other programme staff from MHF and YG). This is due to the likelihood of deductive disclosure. The informed consent form for wider stakeholders (see Appendix 9) explains this in greater detail.

Data protection

Legal basis

The legal basis for collecting, sharing and processing data for this project was 'Legitimate Interest'. The legitimate interests identified were the broader societal benefits of the study's results through advancing our collective understanding of whether, how, for whom and under what conditions BAM may be feasible to deliver in Lambeth. A Legitimate Interests Balancing Test was carried out that confirmed that programme data routinely collected by MHF on the background and participation of scholars in BAM was necessary for the evaluation to fulfil its aims.

Special category data was processed in accordance with GDPR Article 9(2)(j) (for research and statistical purposes which may be in the public interest), and in accordance with UK law and, in particular, the Data Protection Act 2018 Schedule 1, Part 1, paragraph 4(a), Research etc., the processing is necessary for research purposes, is carried out in accordance with UK GDPR Article 89(1) and is in the public interest.

Data processing roles

Routinely collected programme data was shared by MHF with the main evaluator partner on the basis of a data-sharing agreement in which both parties were data controllers. This programme data, as well as primary qualitative data collected by the main evaluation partner, was then shared with the other two evaluation partners on the basis of two separate data-sharing agreements, in which the main evaluation partner was a data controller, and the other two evaluation partners were data processors. Coaching data collected by YG on counsellors was shared with the main evaluation partner on the basis of a data-sharing agreement in which all parties were data controllers.

De-identifying research participants

For scholars and their parents/carers, qualitative data will be de-identified. No real names or other identifiers or distinguishing features of participants have been or will be used on any reports, presentations or papers.

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

Confidentiality

Each individual participant was assigned a pseudo-ID known only to the evaluation team. The master index linking pseudo-IDs to personal identifiers (first name and family name) was stored on a secure, shared drive in a file separate from the evaluation data. Thus, no names or personal identifiers were attached to the evaluation data. All routinely collected data that the MHF shared with the evaluation team was pseudonymised before it was sent.

Transfer of data

All data was transferred from MHF and YG to the evaluation team via a secure platform (Microsoft OneDrive).

Data quality

Procedures for collecting routine data (outcomes, implementation) as part of service delivery were overseen by the Senior Evaluation Officer at the MHF. In practice, this included (i) providing counsellors with training in how to administer measures to minimise bias and 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) and (ii) analysing data monthly to identify and rectify any issues as quickly as possible, a process that was led by the Senior Evaluation Officer at MHF**.**

Regarding primary data collected by the evaluation team, staff had previous experience of using such methods and received additional training on the specific tools.

Data sharing and storage

Data was shared by the MHF with the Dartington Service Design Lab according to a data-sharing agreement. Primary data was stored in password-protected files on secure servers at the Lab. Data was only accessible to evaluation team members and from password-protected computers. Photos taken of hard copy data (e.g. consent forms) were also uploaded to the Lab's secure servers and subsequently destroyed.

It is a requirement of the funding that limited identifiable data be shared with the DfE for the purposes of data archiving. However, the decision was taken by the evaluation team and YEF at the culmination of the study that the data archiving should not go ahead. This was due to the small number (n = 9) of scholars and parents/carers who consented to the process, which would limit the meaningfulness of follow-up analyses to be conducted by ONS-accredited researchers in the future.

Project team/stakeholders

Delivery team

Jane Caro (associate director of programmes, MHF)

Ntale Eastmond (BAM programme manager, MHF)

Mariyam Farooq (programme advisor, MHF)

Victoria Zamperoni (senior research officer, MHF – until September 2022)

Catherine Negus (senior research officer, MHF – involved from January 2022, leading from September 2022)

Dean Idoniboye-Obu (BAM counsellor)

Hugh Mayers (BAM counsellor)

Kohliah Roberts (BAM counsellor)

Antony Di Vittorio (programme 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, one-to-one, etc.): BAM counsellors

Routine data collection: BAM counsellors

Data management and quality: VZ and CN

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

Evaluation team

Dr Tim Hobbs, Dartington Service Design Lab (until June 2022)

Dr Julie Harris, Dartington Service Design Lab (from July 2022)

Dr Nick Axford, University of Plymouth

Finlay Green, Dartington Service Design Lab

Cristina Preece, Dartington Service Design Lab

Julia Mannes, Dartington Service Design Lab

Ediane Santana De Lima, Dartington Service Design Lab

Dr Sean Manzi, Dartington Service Design Lab

Amy Woodburn, Black Thrive

Dr Lynne Callaghan, University of Plymouth

Dr Kate Allen, University of Exeter

Professor Vashti Berry, University of Exeter

Roles

Co-principal investigators (responsible for oversight of study): TH (until June 2022), JH (from July 2022) and NA

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

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

Recruitment and data collection: FG, CP, JM, KA, LC, NA, ESL and AW

Quantitative analysis: KA, SM and FG

Qualitative analysis: FG, CP, JM, LC, KA, ESL and NA

Report writing and dissemination: FG, NA, KA, SM, JH and VB

Methods

This pilot phase evaluation represents the second part of a staggered two-phase evaluation. It was agreed that should BAM show strong feasibility with the first cohort (n = 95) by the end of the first year of delivery (July 2021), the evaluation of the second cohort (n = 97; September 2021 to July 2023) would advance to a pilot outcomes evaluation. This criterion was met, and a report of the feasibility phase covering BAM's first cohort in the UK (September 2020 to July 2022) was published in 2023 (Green et al., 2023a). The current report is the final report for the pilot phase, covering the entire two-year period of delivery for BAM's second cohort in the UK (September 2021 to July 2023).

Participant selection

There are four sets of participants in the evaluation: young people taking part in BAM (scholars), parents/carers of those young people, BAM counsellors delivering BAM and other stakeholders (the BAM Programme Manager, BAM School Leads and the YG Replication Specialist).

All young people participating in the intervention were eligible to participate in the evaluation. At the point of entry into BAM, all scholars provided active, informed consent for the sharing of routine programme data and data on academic attainment, attendance and grades, and all parents and carers provided active, informed consent for the young people in their care.

Our sampling approach involved two steps: step one involved purposive sampling to identify which stakeholder was deemed best placed to speak to a particular element of our theory of change; step two involved maximum variation sampling (where necessary – essentially for young people and parents/carers) to incorporate as wide a range of perspectives as possible. (The three BAM counsellors, the BAM Programme Manager and the YG Replication Specialist were technically all involved by default, although all were still invited to provide informed consent.) The number of interviews conducted was arrived at through a decision based on data saturation, resource management, research burden on MHF and challenges regarding access.

At the point of entry to BAM, 71 of the scholars' parents/carers consented to MHF sharing their contact information with the evaluation team. Researchers contacted each parent via text message and email. This included texting first and then finding a suitable time to talk about what would be involved in taking part in an interview before arranging a time for the interview itself. For those parents and carers who agreed to be interviewed, consent was then sought to interview their child. Where that consent was gained, informed consent was then sought from the scholar before an interview took place. In the interests of pursuing the most convenient location for either stakeholder group, all interviews with parents and carers were carried out over the phone, while all those with scholars were conducted in person at their respective schools.

Theory of change/logic model development

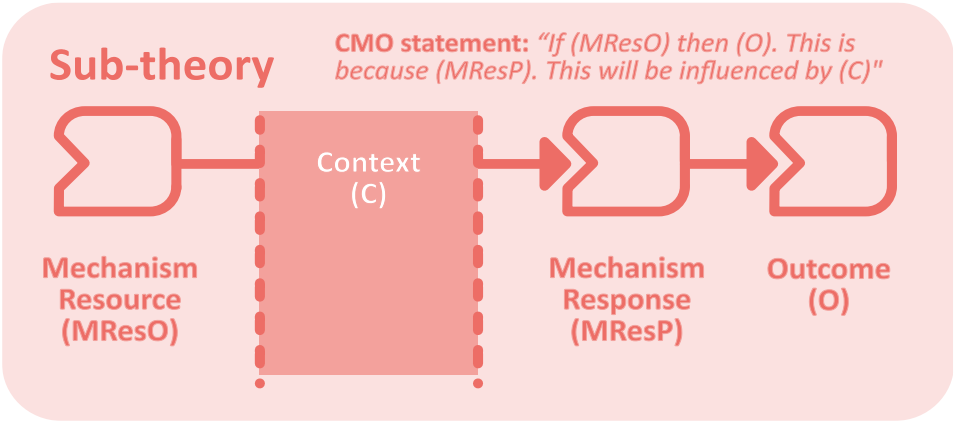
Evaluations based on theories of change have been criticised for failing to develop and test the causal logic of interventions in sufficient depth (Breuer et al., 2016). Realist evaluation is well-placed to address this challenge (Blamey and Mackenzie, 2007; Rolfe, 2019). Realist evaluations begin by generating 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, helping evaluators to establish "what works, for whom, under what circumstances and why." Theories of change can also address some of the shortcomings of realist evaluation. This includes retaining a focus on the whole intervention, providing a visual and more accessible

format for engaging stakeholders in the evaluation and supporting the identification and prioritisation of those elements of the theory deemed most important to test (Blamey and Mackenzie, 2007; Rolfe, 2019). **For these reasons, this study combined theories of change and realist evaluation to develop and test the programme theory underpinning the BAM model.** Combining both approaches during the feasibility study helped us strengthen the reasoning underpinning the theory of change. The development of the theory of change is described in the feasibility phase report (Green et al., 2023a), and the version used in the pilot phase can be found in Appendix 2.

The theory of change is divided into sub-theories – implementation, intermediate outcomes, ultimate outcomes and unintended consequences. The first three sub-theories set the conditions for the one to follow, and all are structured as a realist matrix (Ebenso et al., 2019). Each sub-theory constitutes one candidate theory, structured using the amended CMO framework proposed by Dalkin et al. (2015), which constitutes four components: mechanism resource, context, mechanism response and outcome (see Figure 1– also shared as Appendix 10). The act of separating resource from response, the two defining features of mechanisms within the realist methodology, alongside an explanation of the relationship between the two, helps evaluators avoid conflating programme strategy with mechanisms, a common issue in realist evaluations (Pawson and Manzano-Santaella, 2012).

Figure 1: Amended CMO framework as proposed by Dalkin et al. (2015)



Data collection

Overview

To guide the data collection process and consistent with 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 was broken down into individual elements and entered into a data collection table. This table sets out the attributes of each element in further detail, alongside the associated methods, sources of data, targets and participants, building on the approach proposed by Funnell and Rogers (2011). Table 2 provides a high-level summary. Multiple sources of data were collected for most elements in our programme theory, facilitating a comprehensive degree of triangulation across the theory of change. Triangulation is an important feature of contribution analysis: a systematic, mixed methods, collaborative

approach to measuring impact, which informed the collection, analysis and synthesis of data in this study (discussed further in the Analysis section below).

Quantitative data collection covered two main areas: outcomes for young people and implementation. Data on these were provided by the MHF, since they were collected as part of routine service delivery and monitoring. The MHF also shared socio-demographic data on the backgrounds of students and counsellors, including age (and year group of students) and ethnicity. Qualitative data collection involved the evaluation team conducting interviews (with young people, parents/carers, school staff and other stakeholders) and a focus group (with counsellors). This primary data collection took place in the period June to July 2023. Further details on quantitative and qualitative data collection are provided in the following sections.

Table 2: Methods overview

Theory of change element	Data collection methods	Participants / data sources
Research Question 1: To what extent is BAM being successfully implemented, with whom, under what circumstances and why?		
1A) Backbone support	Qualitative interviews and coaching data	Replication specialist, counsellors, programme managers (MHF) and school staff
1B) Positive learning environment	Focus groups, qualitative interviews and coaching data	Parents and carers, counsellors, replication specialist, programme managers (MHF) and school staff
1C) Successful implementation	Qualitative interviews (Years 1 and 2), Holistic Student Assessment (HAS) data, routinely collected programme data, coaching data and focus groups	Replication specialist, counsellors, school staff, programme managers (MHF), scholars and parents/carers
1 – Context) School context	Qualitative interviews, focus groups and coaching data	Counsellors, school staff, programme managers (MHF), parents/carers, replication specialist and scholars
Research Question 2: To what extent is successful implementation contributing to any reported changes in social-emotional development, for whom, under what circumstances and why?		
2A) Action and reflection	Qualitative interviews	Scholars, counsellors, parents/carers and replication specialist
2B) Internalise the core values	Qualitative interviews, Strengths and Difficulties Questionnaire data and HSA data	Scholars, counsellors and parents/carers
2 – Context) Group conditions	Qualitative interviews	Scholars, counsellors, parents/carers and replication specialist
Research Question 3: To what extent is social-emotional development contributing to any reported changes in academic performance and behaviour and the avoidance of or reduced involvement in crime/anti-social behaviour, for whom, under what circumstances and why?		

3A) Protecting and empowering	Qualitative interviews	Parents/carers, counsellors and scholars
3B) Responsible decision-making	Qualitative interviews and school data on academic attainment, behaviour and attendance	Parents/carers, counsellors and scholars
3 – Context) Wider influences	Qualitative interviews	Parents/carers, counsellors and scholars
Research Question 4: To what extent is implementation contributing to any reported unintended consequences, for whom, under what circumstances and why?		
4A) Recruitment challenges	Qualitative interviews, HSA data and routinely collected data (recruitment, attendance and fidelity)	School staff, parents/carers, counsellors, scholars and replication specialist
4B) Labelling and negotiating	Qualitative interviews	School staff, parents/carers, counsellors and scholars
4C) Irresponsible decision-making	Qualitative interviews and school data on academic attainment, behaviour and attendance	School staff, parents and carers, counsellors, scholars
4 – Context) Wider influences	Qualitative interviews	School staff, parents/carers, counsellors and scholars

Minimisation of bias

To help minimise bias, counsellors received training from both the MHF Senior Evaluation Officer and the PEAR Institute (which developed one of the measures used) 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 were conducted by experienced and trained members of the evaluation team. Participants were told that the information they provided would be confidential and their comments reported anonymously.

Quantitative data: outcomes

Based on the aims of BAM, we sought to collect data on the following outcomes:

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

The MHF collected data for outcomes 1 and 2 from young people participating in BAM at two times with both cohorts as part of their standard monitoring and evaluation procedures. Young people used the IT facilities available in their school to complete the assessments online during a BAM circle session. BAM counsellors facilitated the session with support from the Senior Research Officer at MHF. Baseline data was collected across January and March 2022 with all young people who had been formally recruited following the end of the orientation period (i.e. following the first three orientation lessons of the curriculum), after which groups are considered 'closed'. The second time was during the final month of BAM Circles.

Social-emotional skills/competencies (including behaviour)

The first outcome area cited above was tested via two measures. One was the youth self-report version (11–17 years, without impact supplement) of the SDQ (Goodman, 1997). 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). The first four subscales are summed to create a total difficulties score[18] in the range of 0–40, with higher scores indicating greater difficulties. The other measure used to test the first outcome was the HSA (Allen et al., 2017). Since this was also used to measure the second outcome, it is described in the next section. The SDQ and HSA were collected as part of MHF's routine monitoring and evaluation processes.[19] Students completed the SDQ and HSA online or on paper. Paper copies were then entered into the online platform by MHF. The evaluation team was provided with subscale scores and total difficulties scores for the SDQ and normed scores for each item on the HSA.

Engagement with school and learning

The HSA (Allen et al., 2017) is a validated 61-item self-report assessment tool that measures social-emotional skill development, learning and school engagement. The HSA assesses respondents against 14 subscales grouped into three categories of life skills: (i) resiliencies (internal and external), (ii) relationships and (iii) learning and school engagement. Each of the 14 subscales 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.[20] 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.

We additionally used data routinely collected by schools, including academic attainment, attendance and behaviour (specifically, records of internal and external exclusion, both temporary and permanent). The evaluation team pre-populated an Excel spreadsheet with different column headings for each of these constructs at different times before and during BAM. This was then shared with schools, who completed it for each scholar before returning it to the evaluation team. These data were also used to inform understanding of the extent to which missing school lessons to attend BAM activities had a negative influence on educational performance.

Offending/anti-social behaviour

[18] The sum of four subscales: conduct problems, hyperactivity, emotions and peer relations

[19] Some amendments to YG's standard assessment processes were introduced as part of this evaluation. Normally, the SDQ is only delivered at pre-test and the retrospective pre–post HSA is delivered at post-test. To support the evaluation's requirements, the SDQ and HSA were delivered at pre- and post-test.

[20] This sample is based on a stratified random sample of 9,000 male and female participants aged between nine and 19 years old, drawn from a population of 27,808 HSA respondents in the US (Allen et al., 2017).

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 six/12 months (e.g. criminal damage, stealing, robbery; 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, in turn, could potentially reduce the effectiveness of the programme. There was also a concern that the SRDS could reinforce negative racial stereotypes surrounding Black adolescent boys in Lambeth, an area with longstanding issues regarding the criminalisation of Black people and institutional racism within the police. Therefore, prior to using the SRDS in the evaluation, it was agreed with the YEF that the evaluation team, MHF and Black Thrive (a community partner on the BAM project) would work together to explore options for measuring this outcome in a manner that was both acceptable and rigorous.

Therefore, we convened a workshop in Lambeth involving nine young people who were not involved in BAM but nevertheless shared some key characteristics with BAM participants: they were all male, they were at or very close to the appropriate age (eight were 12 to 16, while one was 11) and they were all from a Black or racially minoritised ethnic background. None of the participants attended schools in which BAM was being delivered. The workshop involved (1) explaining the purpose of the session, including what BAM was, why we were interested in inviting BAM participants to complete surveys relating to offending and anti-social behaviour, and why we were seeking their feedback on the survey, (2) sharing printed versions of the SRDS and the Modified Aggression Scale (Orpinas, 1993), a 22-item alternative assessment that measures aggression and prosocial behaviour across four subscales (fighting, bullying, anger and cooperation/caring) and inviting participants to colour-code items to indicate whether they would feel able to answer the question honestly and comfortably and (3) an open discussion of participants' thoughts about the measures, specifically their acceptability and the likelihood of them answering questions honestly.

Most participants in the workshop expressed discomfort with both surveys. They felt they were being asked the questions because of suspected involvement in criminal behaviour, that this was because of their race and, consequently, felt that they would find it upsetting to complete these surveys. The workshop highlighted the existence of distrust of white people specifically and authority more generally among Black adolescent boys in Lambeth. We concluded that if either measure were used in the evaluation of BAM, (1) there would be doubts concerning the honesty of adolescent boys' responses to the questions, and (2) Black adolescent boys might feel discriminated against through being asked to complete these questions, regardless of measures put in place by the evaluation team to frame or construct the survey in a way that might limit these reactions.

Following the workshop, it was therefore agreed with the YEF that it would be inappropriate on ethical and scientific grounds to use such a measure in the evaluation, and instead, we undertook to explore using official police data on BAM participants' participation in crime. In order to obtain such data, it was necessary to apply formally to the MoJ, specifying the nature of the project; the data required; why it was important; when it was needed; ethical considerations; evidence of approval; risks; how the data would be processed, stored and destroyed in line with GDPR requirements; the time period for which data was needed; the likely number of the number of participants about whom data was required; and the limitations of the research. Subject to MoJ approval, parent consent and student assent, we would need to provide the MoJ with the following information for each participant: a unique project ID, name, date of

birth and postcode. The MoJ would then provide an Excel file with the following information for each participant, as requested: case and offence IDs, offence date, offence group/code, hearing result, caution/conviction date and disposal (all for the period January 2015 to the end of July 2023). This would permit analysis of the profile of participants on entry to BAM and changes to their involvement in crime over the course of their participation in the programme. However, we were unable to secure a large enough number of consenting participants to allow us to obtain the data. Specifically, only 22 young people agreed to us obtaining MoJ data, and of these, parents/carers agreed in only five cases. This was a prohibitively small number and would not have allowed for meaningful analysis of crime data, so we were unable to obtain the data from the MoJ.

Quantitative data: implementation

Implementation outcomes were drawn from existing frameworks, including the Education Endowment Fund's Implementation and Process Evaluation handbook (Humphrey et al., 2019) and others (Carroll et al., 2007) recommended in the latest UK Medical Research Council (MRC) guidance for evaluations of complex interventions (Skivington et al., 2021). We drew on existing YG and MHF systems and measures to track implementation. BAM counsellors collected routine data on programme implementation (recruitment, attendance and fidelity). Data collection tools for this purpose were 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 Research Officer for BAM at MHF. All counsellor-completed records were entered into MHF's data management system. MHF checked these records for errors and inconsistencies, following which minor corrections were made before the records were shared with the evaluation team.

Thus, quantitative implementation data included details on the delivery of each of the four main activities (BAM Circles, brief encounters, one-to-one support and special activities[21]), including when a particular session takes place, who attends and what the session covered (including implementation checklists after each session). More specifically, data was collected on exposure and dose (counsellor-completed records of the number of BAM Circles/brief encounters/one-to-one sessions delivered/attended and the length of sessions), adherence to content (counsellor-completed implementation checklists after each session and formal assessments of counsellors by the replication specialist), recruitment (retention and the proportion of scholars, groups and schools meeting eligibility criteria according to the HSA)[22] and quality (including counsellors' core competencies assessed by the YG replication specialist).[23]

[21] Special activities were not delivered in the first year of the feasibility study owing to COVID-19.

[22] For example, the proportion of participants attending 80% of BAM Circles (the target threshold)

[23] There are five core counsellor competencies: systemic leadership (how counsellors engage school staff and parents), clinical skills, modelling, group work and youth engagement. Counsellors are also assessed on curriculum fidelity and 'men's work'. YG 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 by the replication specialist for each competency against a five-point metric 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.

Regarding the latter, a BAM coach at YG collected data on the development of BAM counsellors. This data covered how BAM counsellors progressed 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 were given one of five grades across each of the competencies: basic, novice, intermediate, advanced and expert. Neither sums nor averages were taken of these grades; instead, areas for improvement were identified and prioritised for the next phase of each counsellor's development. This information was stored on YG's Case Management System.

Qualitative data

Primary qualitative data collection included interviews (with young people, parents/carers, school staff and other stakeholders) and focus groups. In line with standard realist evaluation practice, we adopted a purposive approach to sampling based on which stakeholder was deemed best placed to speak to a particular element of our programme theory (Manzano, 2016). Thus, we conducted 36 interviews with scholars (n = 11), caregivers (n = 14), counsellors (n = 6), school staff (n = 3; the BAM lead in each of the three secondary schools), the replication specialist at YG (n = 1) and the programme manager at MHF (n = 1), as well as one focus group with the three counsellors. All interviews and focus groups were carried out by members of the evaluation team. Following learning from the feasibility phase, interviews with scholars were conducted by two female researchers from racial minority backgrounds.

We built on Manzano's (2016) realist interview principles across all interviews and focus groups by grounding the topic guide for each one in the relevant elements of the theory of change as set out in the data collection table. For counsellors, the replication specialist and the programme manager, this included explicitly orientating them to the theory being evaluated and asking them to speak to that theory based on their experience. Topic guides (Appendix 11) for qualitative interviews and focus groups were informed by those developed by Brand et al. (2019) to aid this process. Interview questions were featured alongside associated theory, which was included as a reference in all topic guides. These topic guides also included open-ended exploratory questions to ensure that the evaluation remained open to alternative and unanticipated contextual factors and unintended consequences. For scholars, caregivers and school staff, a less direct approach was taken, given their unfamiliarity with some of the more complex and abstract elements of BAM's underlying theory.

Interviews with scholars centred on an activity rather than a conversation, drawing from but building on resources developed by others (Macedo, 2022). Scholars were invited to think of their journey over the preceding two years (the period covering their participation in BAM) as a hot-air balloon journey. They were asked how they felt at the start of the journey and at the end, what sources of wind were pushing them along, who was with them in their basket helping them, what stormy weather was getting in the way and what sunny weather was making things easier. Emoji cards and other prompts were provided to stimulate reflection. Their engagement with BAM was discussed at the end.

This approach to data collection with BAM scholars had several benefits, including (i) creating a fun and engaging activity for scholars, (ii) helping us to consider influences other than BAM in their journey (an important part of contribution analysis) and (iii) helping us to simplify our reflections with scholars on mechanisms (the wind) and context (the basket, the stormy weather and the sunny weather).

Each of the 11 scholars we interviewed was treated as an embedded case study (Yin, 2018). This means that they were each treated as sub-cases within the higher-level case under investigation (i.e. the programme). Embedded case studies are recommended tools for those conducting contribution analysis (Delahais and Toulemonde, 2012) due to the way in which they focus data collection and analysis on the journey of each sub-case through the programme. A common pitfall of theory-driven evaluations is that individual elements within programme theory can become 'unconfigured' from one another as the evaluation becomes more focused on the elements themselves and less so on the relationships between them (Pawson and Manzano-Santaella, 2012).

Embedded case studies are also built on the same understanding of rigour as contribution analysis, a conceptualisation that focuses on the degree to which triangulated sources converge or diverge. To maximise the amount of data collected on each sub-case, counsellors, school staff and parents/carers were also asked in their interviews about the journey of each young person we interviewed. Further details on the ways in which their perspectives were triangulated are provided in the Analysis section.

Analysis

The overarching analytical approach for the pilot phase of the study was informed by contribution analysis (Mayne, 2001). While more commonly used in policy and systems change evaluations, contribution analysis has several strengths that justify its use here:

- Its systematic approach to improving confidence about the relative contribution of interventions, associated mechanisms and contextual factors to results (Lemire et al., 2012), which is challenging in both realist evaluations (Marchal et al., 2018) and those based on theories of change (Breuer et al., 2016)
- Its attention to rival influences on impact, which theory-based evaluations sometimes overlook (Breuer et al., 2016)
- Its suitability for evaluations that aim to gather evidence of an intervention's promise but lack comparison groups and large sample sizes (Lemire et al., 2012)

Contribution analysis draws on the analysis and synthesis of qualitative and quantitative data. The details of what this involved in this pilot phase are described below.

Contribution analysis: constructing contribution claims

We applied contribution analysis across each of our sub-theories by constructing four contribution claims – one per sub-theory (implementation, intermediate outcomes, ultimate outcomes and unintended consequences) to answer the associated research question. Contribution claims summarise the degree to which some or all the theory of change was verified, accounting for other key influencing factors (Mayne, 2012). We adapted Delahais and Toulemonde's (2012) approach to constructing contribution claims to account for the four components within each sub-theory (mechanism resource, context, mechanism response and outcome). Each contribution claim was assembled by compiling evidence on the extent to which:

1. The outcome happened …
2. Due to the mechanism response …
3. Which was due to the mechanism resource …
4. Of which the link was mediated by context …
5. While considering rival influences

Qualitative data analysis

The qualitative analysis built upon the approach we used in the feasibility phase of the evaluation, which was based on 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). It was also grounded in framework analysis (Ritchie and Spencer, 1994). We used a form of framework analysis similar to that applied by Brand et al. (2019) during their realist formative process evaluation of an intervention to support prison leavers. We also built on the approach outlined in Gale et al. (2013), particularly their guidance for evaluation teams on best practice for the collective and consistent charting of data into a framework. The NVivo 12 software package was used to build and apply our working analytical framework.[24]

First, we developed a framework matrix. The columns represented the individual elements of the theory of change, which aligned with the data collection table, while the rows represented one type of data for one participant (e.g. one interview with a parent/carer). Every transcript was then coded against this framework, with each cell representing "*relevant evidence from one type of data for one participant in relation to one piece of programme theory"* (Brand et al., 2019, p.156). Information entered in each cell considered 'certainty' (Lemire et al., 2012), i.e. the degree of alignment between the evidence and our programme theory (i.e. the extent to which the evidence did or did not follow the pattern predicted in the relevant part of the theory of change). The framework matrix was reviewed by a second researcher for each transcript coded to ensure consistency.

Quantitative data analysis

Data from pre–post measures (SDQ and HSA) were subjected to descriptive and inferential statistical analysis. Basic descriptive statistics included examining data completeness and the demographics of the sample with and without pre- and post-intervention data. Inferential statistical analyses included paired sample t-tests to measure whether outcomes changed from pre- to post-intervention. These analyses are exploratory, as they do not take into account any potential nesting of students within schools (given the small sample size). We also conducted McNemar tests to examine whether there was a change in participants' level of difficulty (SDQ) or strength/challenge (HSA) from pre- to post-intervention. To do this, we categorised the four-banded solution of the SDQ (Youth*in*mind, 2016) into two groups: 'close to average or slightly raised' versus 'high or very high'. For the HSA, we categorised scores into 'challenge' versus 'average/strength'. This approach to categorisation was used as the McNemar test requires testing the change between two categories pre–post in a 2×2 table. Using more categories would require more tests to be conducted, reducing the sample size and making the test impractical with the available data.

[24] NVivo qualitative data analysis software; QSR International Ply Ltd. Version 12, 2018

The categories for each test were combined to support the balanced identification of change from low to high (a mid-split for the SDQ with four categories and a low to higher split for the HSA).

For the SDQ and HSA analyses, Bonferroni correction was used to account for multiple testing, setting the alpha level to 0.008 for the SDQ analysis and 0.004 for the HSA analysis. Bonferroni correction was suitable for these analyses, as we conducted multiple tests on both the SDQ (Total Difficulties score and subscale scores) and HSA (14 subscales; Bland and Altman, 1995). Applying the Bonferroni correction was necessary because we applied multiple tests to the same individuals and needed to ensure sensitivity due to considerable amounts of missing data and the small sample size, reducing the possibility of a type 1 error.

Summary statistics were used to describe implementation in terms of recruitment, attendance, fidelity and quality, as well as the number and socio-demographic characteristics of the scholars recruited to BAM. Insights from these analyses were added to the framework matrix to enable us to apply an integrated, mixed methods approach during synthesis. In more detail:

- Recruitment: group size and proportion of young people in each HSA tier (1, 2, 3) in each BAM group
- Attendance: the number of BAM Circles held, mean number of BAM Circles attended, percentage of BAM Circles attended, proportion of young people receiving ≥1 brief encounter each month and proportion of young people receiving a one-to-one each month
- Fidelity: progression through the BAM manual (number of lessons reached per group) and adherence to the BAM manual[25].
- Quality: the 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 of experience).[26]

Summary statistics were used to describe the number and socio-demographic characteristics of young people recruited to BAM. Insofar as data permits, these were compared with school and neighbourhood statistics.

School-related behaviour and academic attainment were obtained, as previously described, for the period of the pilot study. Attendance, internal suspensions and external suspensions were used as proxy measures of behaviour, and academic attainment was measured variably across the participating schools using either a teacher-assessed approach or numerical grade classifications. Socio-demographics were used to describe the participants. Mean values and the standard error around the mean were calculated for each school across all available time points.

[25] Quantitative data on adherence was captured and analysed using two sources. First, counsellors used 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 formally assessed counsellors' ability to deliver the curriculum with fidelity. Higher ratings will also indicate greater adherence.

[26] This will be discussed and agreed with the replication specialist.

The use of inferential statistics in the analysis of school data was deemed inappropriate for several reasons. First, there were concerns with data quality, particularly with data on attainment, which was reliant on internal and variable assessments of performance and predicted grades. Second, different measurement approaches were used in each school, reducing the sample size available for analysis and limiting what could be interpreted from inferential statistics. Third, the collection of data at multiple time points across each measure would increase the chances of error in the application of inferential statistics.

Synthesis

According to Delahais and Toulemonde (2012), "*The robustness of a contribution claim depends on the supporting items of evidence that may or may not be strong, convergent, and triangulated"* (p.287). Therefore, to synthesise our analysis, we developed a triangulation assessment. Building on the approach of Campbell et al. (2020), this involved establishing the degree of convergence between individuals (e.g. between caregivers), across stakeholders (e.g. between caregivers and scholars) and across methods (e.g. between routinely collected programme data and interviews). The level of convergence at each level was described consistently using the framework developed by Farmer et al. (2006), which includes four categories: (1) full convergence, (2) partial convergence, (3) silence and (4) divergence.

Preliminary findings were shared with BAM counsellors and programme staff at MHF in October 2023 to identify weaknesses in the contribution story assembled by the evaluation team. Their feedback has been incorporated into this report.

Timeline

Table 3 outlines when each activity in the pilot phase of the study took place.

Table 3: Timeline of the pilot phase

Date	Activity
By end Dec 2021	Set-up, including ethics approval and refinement of the design for the pilot outcome evaluation phase (including the protocol)
Jan–Mar 2022	Baseline data collection of routine outcome data (Holistic Student Assessment [HSA] and Strengths and Difficulties Questionnaire [SDQ]) with young people formally recruited to BAM
Dec 2021/Jan 2022	Recruitment into the evaluation of parents/carers and students formally in the BAM intervention
Jan–May 2023	Qualitative data collection (interviews/focus groups) with BAM counsellors, parents/carers, scholars and other stakeholders and analysis of the data
Jun/Jul 2023	End-point HSA/SDQ
Aug–Nov 2023	Analysis of HSA and SDQ Write-up of the pilot outcome phase, leading to the final report
Nov 2023–Jul 2024	Revisions to the report following peer review comments Dissemination of the final report

Findings

In this section, background information on scholars is provided before the findings related to each of the four research questions are shared. Findings related to BAM's feasibility are discussed in relation to Research Question 1. Findings related to BAM's evidence of promise are discussed in relation to Research Questions 2 to 4.

The findings related to each research question are grounded in the sub-theory in the theory of change related to that question. The contribution claim that summarises the findings for each question uses abbreviations to signal which component of the sub-theory is being referenced in the statement, including abbreviations for mechanism resources (MResO), contextual factors (C), mechanism responses (MResP) and outcomes (O). Rival or alternative explanations are also highlighted (R). This is to ensure that (i) the findings can be clearly linked back to the theory of change and (ii) the findings explicitly advance our understanding of the contexts, mechanisms and outcomes that explain how BAM works. This approach adheres to the RAMESES guidelines for the reporting of realist evaluations (Wong et al., 2016).

As outlined above, references to the strength of evidence are built on the degree to which different sources of evidence converged. 'Strong' evidence is defined as full or fundamental convergence at two or more levels of triangulation, i.e. between individuals (e.g. between caregivers), across stakeholders (e.g. between caregivers and scholars) and across methods (e.g. between routinely collected programme data and interviews).

To ensure transparency while protecting readability, we use verbal markers of quantitative information to capture the number of scholars and parents or carers who shared a particular view in interviews, as is common in the field of qualitative research (Sandelowski, 2001; Maxwell et al., 2010).

- 'Most' is used to describe views shared by a majority of interviewees (i.e. six or more out of 11 scholars interviewed, and eight or more out of 14 parents and carers interviewed).
- 'A minority' is used to describe views shared by less than half of the interviewees (i.e. less than six of 11 scholars interviewed and less than seven out of 14 parents and carers interviewed).

Each of the 11 scholars we interviewed constitutes a case study. They were constructed by triangulating the perspectives of the counsellor, the scholar, their parent/carer and school staff regarding the journey of each scholar over the preceding two years, alongside their implementation, SDQ and HSA data. Three of the 11 case studies are highlighted and shared in depth throughout the findings to illustrate and personalise the insights. They were selected because they captured the variety of experiences contained within the wider pool of 11. Each case study is given a pseudonym: those of the highlighted case studies are Elijah, Sanjay and Mark. Similarly, each of the counsellors is given a pseudonym: Louis at Pine, Marcus at Birch and Grant at Oak.

Participants

Enrolment

Ninety-seven scholars participated in BAM and the evaluation during the pilot phase (see Figure 2). Eighty-two scholars had joined the programme by the end of the orientation period (December 2021), at which point membership of BAM settled. A further 15 joined subsequently, for whom baseline data was not collected. This was due to the research burden that administering the baseline assessment individually with each late starter would place on counsellors. By the end of the programme, at the end of year two (July 2023), 78 remained. Further details on the reasons for attrition are provided in the findings section related to Research Question 1.

Most of the participants – 55 of the 97 scholars – attended Pine. While BAM had previously been delivered in Oak and Birch, this was the first BAM cohort at Pine, which enabled the counsellor – Louis – to recruit a full cohort (five groups) at the start of the year. The counsellors at Oak and Birch – Grant and Marcus – were both delivering four groups in addition to the two represented within the pilot cohort that started in October 2021.

Year group

Most scholars were drawn from Year 9 (52%), followed by Years 8 (37%), 10 (10%) and 11 (1%) (see Table 4). There were fewer scholars of General Certificate of Secondary Education (GCSE) age for this cohort: 11% of scholars started BAM in Years 10 and 11 in the pilot phase, compared with 38% during the feasibility phase. This was the result of a conscious effort made by MHF to reduce the level of disruption that GCSEs imposed on implementation. Competing demands on students' time and increased pressure for them to remain in lessons limited some of the exposure to BAM of scholars in Years 10 and 11 during the feasibility phase.

Ethnicity

Two-thirds of the scholars identified as Black/Black British (66%), followed by Mixed Ethnicity (19%), White/White British (10%) and Other (6%). There was no ethnicity data for 2% of scholars due to challenges faced by counsellors in collecting the information from all young people (see Table 5). The latest demographic estimates for 10–19-year-olds in Lambeth (GLA, 2016) suggest that while those from a Black/Black British background were over-represented among BAM students (66% vs 41%), those from a White/White British background were under-represented (10% vs 32%). However, this comparison should be treated with caution; these data for Lambeth are seven years old and capture demographic information for the borough as a whole rather than the neighbourhoods surrounding each school.

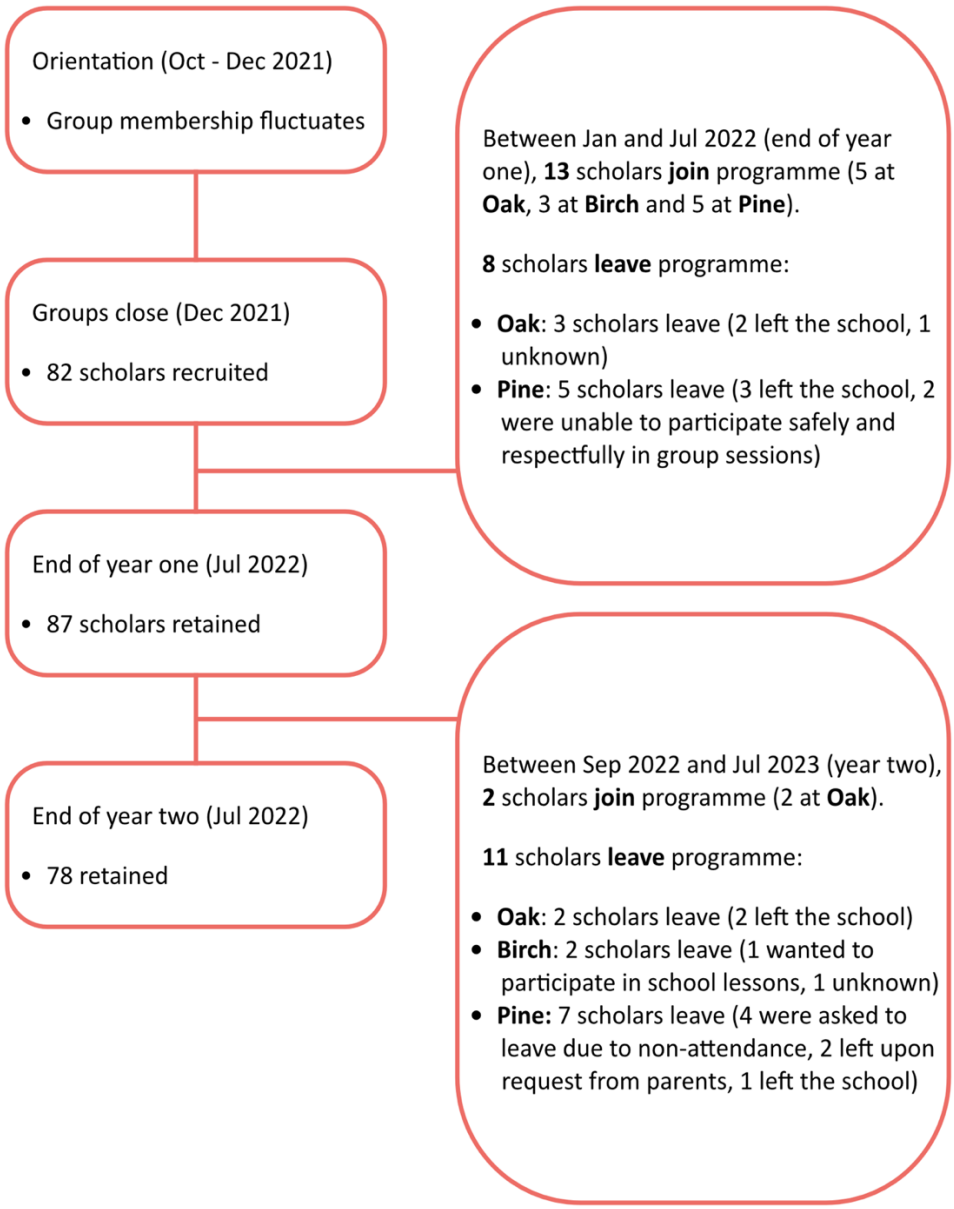


Figure 2: Flowchart depicting the number of scholars recruited, lost through attrition and retained. Data was captured by counsellors on an ongoing basis and shared with the evaluation team following the end of the programme in September 2023

Table 4: Percentage of scholars participating in BAM by year group at the start of the programme

Year	% (Feasibility)	% (Pilot)
8	17	37
9	53	52
10	23	10
11	5	1

Table 5: Percentage of scholars participating in BAM by ethnicity

Ethnicity	% In BAM (Feasibility)	% In BAM (Pilot)	% In Lambeth
Black/Black British	56	66	41
Mixed Ethnicity	18	19	16
White/White British	13	10	32
Other	3	3	9
Missing	7	2	-

School-level data

At a whole school level, all three schools had more pupils eligible for free school meals (FSM) and with English as an additional language (EAL) than the national average, but a similar share of pupils with an Education, Health and Care (EHC) Plan (see Table 6; GOV.UK, 2023). Oak and Birch had a slightly higher share of pupils requiring Special Educational Needs (SEN) support.

Table 6: Percentage of pupils in each school eligible for FSM, with an EHC Plan, requiring SEN support and with EAL

School	% Pupils eligible for FSM	% Pupils with EHC Plan	% Pupils requiring SEN support	% Pupils with EAL
Oak	43.6	4.0	20.0	51.2
Birch	38.1	7.8	17.8	60.4
Pine	52.8	3.6	13.0	51.5
National Average	23.8	4.3	13.0	20.2

Evaluation feasibility: Research Question 1

To what extent is BAM being successfully implemented, with whom, under what circumstances and why?

There is strong evidence that the implementation of BAM varied by implementation area (O)[27]. While adaptation and the quality of delivery were successful overall, recruitment and exposure experienced mixed results. Evidence on the quality of counsellors' adherence to the manual was positive, but their ability to progress through the curriculum was limited by group scheduling, group size and scholar engagement. There is some evidence that counsellors' previous skills and experience supported the delivery effort (R). However, there is strong evidence that some elements of 'backbone support' provided by MHF and YG (MResO) – particularly training and coaching – made an important contribution to the implementation successes observed by creating a positive learning environment for counsellors (MResP). The contributions of peer supervision, project management support and community partners to this environment were limited due to capacity and resource limitations. The extent to which this environment enabled implementation was influenced by engagement from scholars, school context and counsellor absences.

The findings related to this research question are structured as follows. First, the findings related to the mechanism resource are shared. These concern the degree to which the suite of implementation support activities provided by MHF and YG were delivered as intended (these activities were labelled backbone support by the counsellors). These are shared alongside a description of the degree to which each element of this support contributed to the development of a positive learning environment for counsellors (the mechanism response). This is followed by a description of the contribution of this learning environment to each area of implementation (the outcome of this sub-theory), alongside a consideration of the role of school context, counsellor absence and scholar engagement in mediating this contribution.

Backbone support

The evidence was mixed on whether the backbone support offered by MHF and YG helped to create a positive learning environment for counsellors. While training and coaching made important contributions, the roles played by peer supervision, project management support and community partners were more limited.

Training and coaching

At YG, training is more focused on supporting counsellors in delivering the curriculum with fidelity, while coaching is more concerned with supporting the professional development of counsellors.

While training is carried out in person in Chicago, during the first year of the pilot phase, training was conducted mostly online via Zoom due to COVID-19 restrictions. The sessions were shorter (two hours rather than a half or full day) and more frequent (once a week instead of once every four weeks) than the in-person model to prevent fatigue, but otherwise, YG attempted to mimic the BAM Circles as much as possible. In the

[27] See Figure 1 on page 20 for further explanation regarding the use of the abbreviations. This is also shared as Appendix 8.

second half of each session, the three counsellors watched video run-throughs of the next lesson. This was accompanied by advice and an explanation from the replication specialist, following which they had the opportunity to ask questions. In the first half of the following week's session, counsellors would each deliver the session to the group and receive feedback from each other and the replication specialist.

YG's six-step coaching model was delivered largely as intended during the feasibility phase. Across ten visits to the UK (five in each school year), the replication specialist was able to support counsellors through a comprehensive suite of coaching activities, including goal setting, observation, appraisal and feedback.

As captured in the feasibility phase report (Green et al., 2023a, p.41), the counsellors felt that the frequency, intensity and interactivity contributed to a positive learning environment that helped them develop their competencies, including their ability to deliver the curriculum with fidelity. During year two, two of the counsellors felt that the training and coaching contributed more than any other implementation support activity to their learning environment, while the other ranked it second. The coaches and the replication specialist felt that the replication specialist's support was more personalised to their needs and development goals.

- MHF started to deliver BAM in an additional three schools in the London borough of Islington in autumn 2022. This meant that the Lambeth counsellors received fewer group training sessions and more individualised coaching during the 2022/23 school year because they and the new Islington counsellors were progressing through the curriculum at different rates:

  "It was honestly … a lot more intimate and engaged … I could address certain questions that may have stuck out with the counsellor and do a more thorough review of … lessons and feedback." (Replication specialist)

- Counsellors received USBs with online training materials, including video walkthroughs, that had been developed by YG to support counsellors during the pandemic:

  "I can look at the material; I can look at the videos in either my own time or at the time of training, and I feel like … that helped because … I just felt more ahead of myself." (Marcus, counsellor at Birch)

- The replication specialist observed more groups across his five visits in year two due to the lifting of COVID restrictions:

  "He'd come in and sit in and see how I deliver, see how I managed challenging situations. And he [could] give me feedback … it's been such vital and useful information that I've received from those sessions." (Grant, counsellor at Oak)

Peer supervision

Two of the counsellors ranked peer supervision in terms of its contribution to their development during year two. One ranked it first but felt it had deteriorated in year two; his assessment was based on its standard during year one and the superior quality of emotional and practical support it used to offer.

> *"Peer supervision was at that height … And I think … it's definitely not the way it was."* (Marcus, counsellor at Birch)

Scheduling issues, due partly to pressures on the MHF programme manager's time, limited the frequency of peer supervision during year two, which negatively impacted counsellors.

> *"With all these different demands … I don't feel like it's institutionalised like it used to be … I don't know how it's affected delivery because I feel quite experienced, but … I do feel affected emotionally. I do miss that space … I still struggle [in] coming to terms with how it is."* (Louis, counsellor at Pine)

Project management

The role of BAM's programme manager at MHF is an important one. As well as line managing each of the counsellors, the programme manager is responsible for organising training and peer supervision for the counsellors and advocating for them with staff at their schools. This year, project management support was more limited. While some counsellors appreciated the vote of confidence and did feel supported when necessary, they also felt they would have benefited from more timely and proactive support, particularly with schools. The replication specialist and programme manager agreed. Doubling the number of counsellors and schools, challenges with recruiting counsellors in Islington and a reduction in the capacity of the MHF team made it difficult for the programme manager to keep up with the demands on his time.

> *"[There have] been a lot of challenges this year, to the point where I felt that my … line management side of the job has taken a bit of a hit."* (Programme manager)

Black Thrive and Colourful Minds

The level of support from the community partners did not meet the counsellors' expectations. While Black Thrive was able to apply its expertise on anti-racism and mental health to adapt the curriculum and train counsellors in anti-racist practice, these contributions were perceived as rare and insufficient. The ad hoc training sessions provided by Black Thrive offered little in return for the demands they placed on counsellors' already limited capacity.

> *"I don't think the boys have really benefitted … nor have I as a counsellor really benefitted from Black Thrive's involvement."* (Grant, counsellor at Oak)

Adaptation

The positive learning environment within which counsellors operated was also important in adapting BAM to the UK.[28] Adaptation took place through a twin-track process in the pre-implementation period. First, adaptations were made to the teams and individuals that comprised counsellors' backbone support based on perceived differences in context to increase the likelihood of a good intervention–context fit. Counsellors and staff at MHF and YG and the evaluation team systematically identified discrepancies in context by using the BAM theory of change and examining relevant features of context – identified by Craig et al. (2018) and Damschroder et al. (2022) – to design adaptations where necessary.

[28] The process and outcome of the adaptation effort are documented in detail in the evaluation team's report on adaptation submitted to the YEF in November 2020.

For example, BAM has a considerable profile in the US, which, according to YG, motivates schools, communities and parents to support and promote its delivery in their neighbourhoods. No such profile exists in the UK. One of several responses to this discrepancy involved making a series of amendments to the school-level agreement that schools sign with MHF prior to the delivery of BAM to make the language more collaborative, accessible and based on schools' needs and priorities.

The second set of adaptations concerned the curriculum. The team responsible for leading the adaptation process was formed in the pre-implementation phase, prior to the recruitment of scholars. It included a diverse group of adults and practitioners only, including representatives from YG, MHF and Black Thrive. Counsellors were placed at the heart of the effort, given their experience of working with scholars and some of their shared lived experiences.

Adaptations to the BAM curriculum in advance of delivery were classified as either 'surface' or 'deep' by YG. Surface adaptations, which are summarised in Table 7, largely concerned amending language, symbols and illustrative examples to better fit the context of Lambeth. Deep adaptations concerned those that affected core elements of the theory of change. Further detail on the process and outcome of the adaptation effort is documented elsewhere (Green et al., 2023b).

Table 7: Summary of surface adaptations

Nature of adaptation	Example(s)
Superficial changes to language, where the reference or meaning of the word(s) remains intact	'City blocks' became 'streets'.
Changes to cultural references, where both language and reference change, but the purpose and function of the reference remain the same	Basketball references were replaced with football. The '$10 role play' became the '£10 role play'. Films that may resonate more with a US audience were replaced with those deemed better suited to London youth, e.g. Pursuit of Happyness (2006) was substituted for Miracle (2004).
Amendments to the timing and structure of sessions on 'academic integrity'[29] to account for differences in the assessment processes between the two countries	Sessions were aligned with the release of grades throughout the school year.
Adaptation to UK COVID-19 restrictions	Hand sanitiser was used prior to group sessions.

[29] During academic integrity sessions, participants take turns to update the group on their most recent grades. The group then affirms those students who have passed all their classes and challenges those who have failed all or some of their classes.

Replacement of the language used during check-ins with terms more commonly used by London youth (every session should include at least one check-in, where the group takes turns to describe how they are feeling and why)	Certain rituals used during check-ins were changed. For example, the way the group responds to someone checking in by saying 'Asé', a Yoruban word used in the context of circles to affirm the person checking in was replaced in London with 'safe', 'say less', or 'calm', colloquialisms common in London and used, among other things, as greetings.

Surface adaptations

During the feasibility phase, while some pre-emptive surface adaptations worked well, others did not. Both the replication specialist and counsellors were empowered to respond reactively in these situations and tailor the content and framing to scholars' preferences while retaining the overall message or function of the activity (Green et al., 2023a, p.43).

During the pilot phase, the replication specialist and counsellors felt that the learning from year one of the feasibility phase and the pre-emptive adaptation effort at the pre-implementation effort limited the amount of reactive surface-level adaptation required of them during year two of the pilot phase (see Table 7).

> *"I don't feel like there was much adaptation beyond the initial adaptations that we had gone through in the first year, and that happened prior to us even starting to deliver."* (Grant, counsellor at Oak)

Deep adaptations

Deep adaptations to the curriculum covered BAM's use of three concepts: 'tribal societies'; the archetypes of 'savage' and 'warrior', which are covered in the section focused on the core values of Positive Anger Expression; and the archetypes of 'liberator' and 'oppressor', which are covered in the section focused on Respect for Womanhood. Findings regarding 'tribal societies' are considered in the feasibility phase report (Green et al., 2023a, p.44). There is insufficient data on the success (or otherwise) of the adaptation regarding the 'liberator' and 'oppressor' archetypes because most counsellors did not reach the Respect for Womanhood core value in the curriculum with their groups (the reasons for which are outlined below). Therefore, only findings related to 'savage' and 'warrior' are described here.

In the original curriculum, 'savage' is used during sessions on Positive Anger Expression to conjure up images of destructive, uncontrolled anger that creates guilt and shame, while 'warrior' is associated with constructive, controlled anger that brings dignity and integrity.

MHF felt these concepts were problematic. First, both have violent connotations, which could unintentionally suggest that anger equates to violence. Second, 'savage' has been used historically as a derogatory term for indigenous peoples. Third, 'savage' is already used as a colloquialism among scholars in London, often as a compliment to imply strength. As a result, MHF and YG agreed to use the terms 'constructive' and 'destructive' instead.

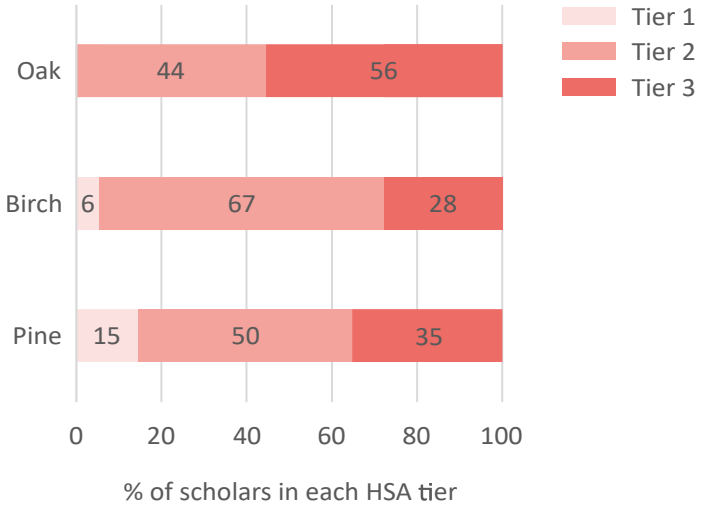
The replication specialist predicted during the feasibility phase that these replacements would lack the same weight with scholars. This is because 'savage' and 'warrior' were archetypes (Jung, 1969) – that is, symbols

with universal meanings that naturally repel or attract scholars. 'Constructive' and 'destructive' are not archetypes.

During the pilot phase, this prediction proved accurate. Counsellors reverted to 'savage' and 'warrior', as these terms resonated more with scholars.

> *"When I was actually going through … that lesson, it just didn't feel right … it felt forced, and it felt like I was like … it didn't feel like it connected … [the counsellor at Pine] said no, I'm still using the savage and the warrior, and I applied the same method, and I was like, you know what, that feels more real, and not even that it just felt real, I felt like the actual students connected a little bit more to it. They could actually visualise it a little bit more, and it didn't hold the same connotations that we were trying to attach to it ... the young people use the word savage. You know, if you do something that's a bit … like, 'That's savage, bro'. So, it's like, it's still language that they use … it's not, sort of, colonial … it's still something that they can identify [with]."* (Marcus, counsellor at Birch)

Figure 3: Percentage of scholars in each HSA tier in each school



Recruitment

Level and spread of need

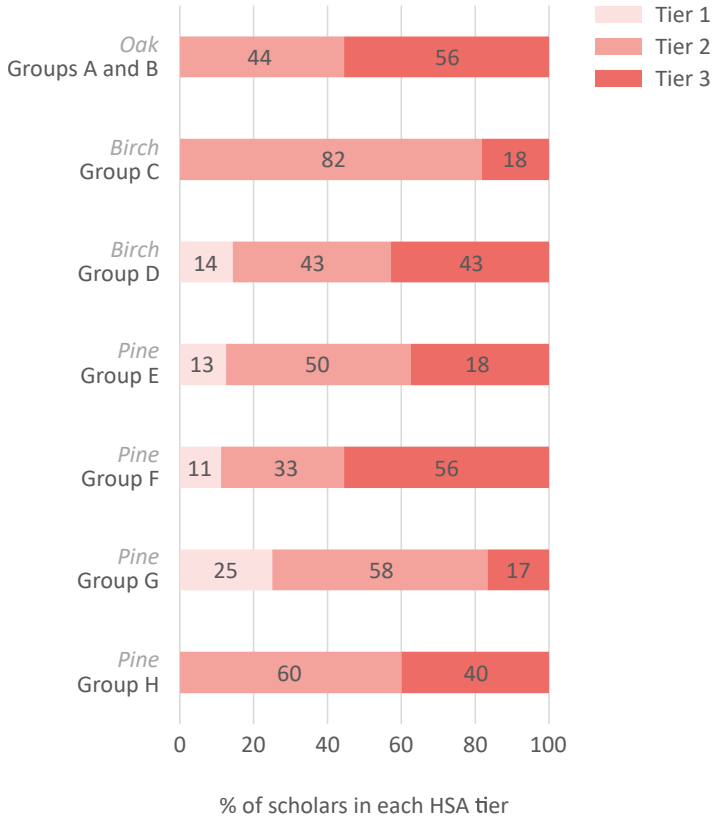
Sixty-one of the 97 participating scholars completed the HSA at baseline. Of those, 55 had at least one challenge related to emotional well-being (the target for all participants).

Birch and Pine showed a spread of need across each of the HSA's three tiers,[30] albeit with more representation from Tier 3 and less from Tier 2 than the targeted 15/70/15 split (see Figure 3). However, the pilot cohort at Oak was skewed heavily towards those with higher needs; there were no Tier 1 scholars at Oak, while 56% of the scholars met the threshold for Tier 3, albeit from a small pool of nine scholars from

[30] Findings regarding the spread of need across the three tiers were calculated using data drawn from all 61 scholars who completed HSA at recruitment.

a possible 11 who completed the baseline HSA at Oak. At a group level, while half had representation from all three tiers, half did not (see Figure 4).[31]

Figure 4: Percentage of scholars in each HSA tier in each group



Group sizes

Six out of nine groups reached the target size (eight to 12 scholars) by the target date (end of December 2021; see Table 8). At Pine, Group I was formed later than the others upon request from the school, which limited its size. At Oak, MHF had intended to form only one group as part of the pilot phase but decided to split the original group into two in November 2021 – hence the smaller number of scholars in Groups A and B. This was due to concerns with the group dynamic. The counsellor observed higher levels of need and more challenging behaviour in the group, which threatened his ability to deliver the programme and progress through the curriculum as intended.

[31] Both groups at Oak are combined due to one of the groups having a response rate of less than five. One group at Pine – Group I – was formed later and did not complete HSA at baseline as a result.

Table 8: Group sizes by December 2021

School	Group	Size by Dec 21
Oak	Group A	7
	Group B	5
Birch	Group C	12
	Group D	8
Pine	Group E	9
	Group F	11
	Group G	12
	Group H	12
	Group I	6

School context

Despite the level of need being higher than intended, overall, the counsellors felt the recruitment process went well during the pilot phase due to experience gained during the feasibility phase. A small number of scholars self-referred (<15) after hearing about BAM from friends or teachers. A decision was then taken between the counsellor and the school on their participation depending on their needs and the existing level and spread of need in and size of each group. However, most scholars were identified and invited to participate pre-emptively and accepted the invitation due to BAM's positive reputation in the school and their positive experience of the orientation sessions (see Figure 5).

Figure 5: Elijah, Sanjay, and Mark's experience of recruitment.

Case Study 1: Elijah

- Elijah self-referred to BAM. According to his counsellor he ***"made the cut"*** because they thought ***"he might get something out of it"***; his counsellor felt he was ***"a little bit below average in some areas."*** Elijah felt ***"nervous and shy because I was just getting into the school."***
- When Elijah's friend told him about BAM, he liked what he heard: ***"My friend said like BAM was about like becoming a man and like trying to better society and your mindset, trying to change it, and I was interested in that."***

Case Study 2: Sanjay

- According to Sanjay's counsellor, he was referred to BAM because he was ***"an anxious young man"***. When BAM started, school was ***"a bit overwhelming"*** for Sanjay: ***"I'd say a bit stressful… Just a lot going on, not just school but in personal [life]."***
- Sanjay's interest was piqued ***"a week before I joined BAM… he gave us a booklet to read about BAM and that's how I got to know about BAM… I thought it was good. A lot of positives in it, and I thought it's something I should do."***

Case Study 3: Mark

- Mark asked to join BAM because he heard about it through his friends. According to his counsellor ***"he was doing okay or a little bit above okay",*** but there ***"was something about him being in his shell,"*** so he and the school agreed he could stay.
- At the start of BAM, Mark was finding things hard. He had found COVID-19 and the transition to secondary school difficult to deal with: ***"I didn't really have like that much confidence and I was very quiet… I didn't really know a lot of people."***

At Oak and Pine, Grant and Louis worked closely with senior leadership, heads of year and pastoral staff to identify potential scholars. At Oak, this built on lessons learned during the feasibility phase, where the over-representation in one group of racially minoritised, higher-need scholars with particularly poor behaviour records may have led to the development of negative labels, both within and outside of the group, that the group was for 'bad Black boys'.

> *"I think that went really well because we learnt from our experience with the year before and how that group and those students created a stigma around the programme … that was … the focal point of our recruitment process."* (Grant, counsellor at Oak)

At Pine, where BAM was being delivered for the first time, the counsellor was clear with school staff on the criteria for BAM.

> *"I did really push and put my stalls out that it's not a programme for naughty boys, so the heads of years, when I first started here, they really supplied me with a lot of young people … in terms of needs, there was a good mixture."* (Louis, counsellor at Pine)

However, at Birch, the counsellor felt the school's misunderstanding of who BAM was for and what young people stood to gain led to a disproportionately high number of scholars with high levels of need joining the programme.

> *"Despite the conversations had around what BAM is, I still feel like within the school they still had this idea that BAM is for … getting students to, sort of, fix up their behaviour … I don't think they really got to know the social and emotional part of it … it was overcast by the … idea of bad students … that was still a driving force … of a lot of the groups."* (Marcus, counsellor at Birch)

Quality

Some alternative influences on counsellors' competencies (full definitions of which can be found in Appendix 1), besides the positive learning environment, were offered, specifically around their previous skills and experience. Counsellors and the replication specialist felt this had given them a head start, particularly with listening, assessing and engaging scholars. Yet, there is strong evidence that the positive learning environment that characterised training and coaching and the experience gained by counsellors during the feasibility phase made an important contribution to counsellors' development.

Table 9 shows the performance ratings[32] given by the replication specialist for each counsellor at the end of the second year (counsellors' pseudonyms are not included to protect their identities). Each rating includes an associated statement by the replication specialist setting out the extent to which their performance was in line with their expected rate of development, given their level of experience. For all competencies on which counsellors were rated 'Advanced', they were deemed to have exceeded expectations. Expectations were reported to have been met for all other competencies except 'Challenging and confronting', where performance was deemed to be 'Below expectation' for two counsellors, and 'Systemic leadership', where the same was true for one counsellor.

Table 9: BAM counsellor skill competency level

Competency	Counsellor 1	Counsellor 2	Counsellor 3
Clinical skills – Assessing	Advanced	Intermediate	Intermediate
Clinical skills – Listening	Advanced	Intermediate	Intermediate
Clinical skills – Challenging and confronting	Intermediate	Novice	Novice
Modelling	Intermediate	Advanced	Advanced
Group work	Advanced	Intermediate	Intermediate
Youth engagement	Advanced	Advanced	Advanced
Systemic leadership	Advanced	Novice	Intermediate
Men's work	Intermediate	Intermediate	Intermediate

Clinical listening and assessing

Clinical listening concerns counsellors' ability to listen actively and attentively to young people. Clinical assessing concerns their ability to conduct here-and-now processing (identifying actions or patterns of behaviour occurring in BAM Circles and using these to stimulate discussion and reflection) (Yalom and Leszcz, 2005).

[32] The assessment uses the following scale: (1) basic knowledge – is familiar with the competency but has no experience in that area, (2) novice – has limited experience in this competency, focus on skill development, (3) intermediate – is experienced to the degree of being proficient, focus on skill enhancement; (4) Advanced – is highly proficient, with no further development needed, focus on nuances of competency and (5) expert – the highest degree of proficiency – is able to coach and lead others in this area.

According to their coaching assessments, counsellors were at either an advanced or intermediate level for listening and assessing. Most scholars and carers agreed that they felt counsellors were empathetic, supportive and trustworthy (see Figure 6).

> *"I feel comfortable with [my counsellor] … I like him … he's just nice, cool and [collected] … he cares a lot."* (Mark, scholar)

> *"[My counsellor's] just open … he's a person you can talk to about anything … he can understand your problems … he gives you a lot of support."* (Sanjay, scholar)

> *"[My counsellor's] very calm … in terms of patience as well … he helps different people in the group, outside and inside the sessions."* (Jacob, scholar)

Figure 6: Elijah, Sanjay and Mark's experiences of counsellors' competencies.

Case Study 1: Elijah ▲

- Elijah felt like his counsellor created a space in BAM that allowed him to be vulnerable. This was partly because of the way in which the counsellor created a relaxed and warm group dynamic: ***"I feel like I can speak honestly and express my emotions. I feel like my BAM counsellor, like, he says whatever and like it makes us feel more comfortable and be able to say things when we see."***

Case Study 2: Sanjay ●

- Sanjay felt he and his peers could speak freely in BAM: ***"you feel more comfortable and you feel more relaxed just to talk about your day and what's been going on in your day… it's a chance to know what everyone's going through."***
- His counsellor felt his ability to model authenticity contributed to this safe space: ***"me being able to sort of demonstrate my own vulnerabilities of what I share, I think it helps create this awareness that young people in that space can do that."***

Case Study 3: Mark ■

- Mark admired his counsellor: ***"He teaches us life skills and stuff like that and, like, the way he manages the session as well, it's really good… He cares a lot… He just like asks us what's going on and stuff like that. Like, if you look upset or something, he asks that and he checks up. Yeah, he makes sure that we're okay, so yeah… I really enjoy like everything he does."***

Brief encounters and one-to-ones played an important role in compounding the unconditional positive regard counsellors conveyed to scholars in the BAM Circles. Brief encounters offered counsellors opportunities to generalise learning from BAM Circles within the wider school context while helping to augment relationships with scholars by reinforcing the consistency and security of their presence in scholars' lives.

> *"Having the counsellor on-site being someone that, you know, if there are any issues … even those very short conversations and interactions can be really positive to [us] ... just a general reminder [of] what they're trying to achieve."* (School staff at Oak)

These relationships were augmented further through one-to-ones, which enabled counsellors and scholars to address particularly sensitive or personal issues in greater depth. They provided unique opportunities for scholars to feel important and listened to.

> *"The one-to-ones are quite unique, you know, and there, you are able to tackle issues that you wouldn't tackle in a group. So, I think the one-to-ones, we've got some very high-profile students who just present many, many, many issues … some medical, some behavioural, and the one-to-ones have really helped them to ... there's one particular child who definitely has benefited from BAM, who has, I don't know, suicidal ideation and all sorts, and he's benefited from it just because ... that's his go-to. He needs to go to BAM. He needs to go to BAM."* (School staff at Birch)

> *"I've had one-to-one sessions with [the counsellor], and we talked a lot about family, football and a lot of things. And I think I feel good about speaking to someone about what's going on in my life … he's the right person to speak to … he's really truthful. He asks you a lot of questions, and he's just really helpful."* (Femi, scholar)

Youth engagement and modelling

Youth Engagement concerns counsellors' ability to be natural and relatable when engaging with scholars. This helps to establish and maintain engagement and foster strong, positive relationships. Modelling concerns their ability to demonstrate authenticity and self-belief, to show how vulnerability can be associated with strength and respect, thereby offering scholars a different version of manhood to aim for – one aligned with BAM's core values.

For youth engagement, the counsellors were all at an advanced level. Along with most BAM scholars and carers, the replication specialist felt counsellors had *"great relationships with the young men"* and that *"you can see youths being happier, you know, being in a space"* with counsellors.

Most counsellors were at an advanced level for modelling the core values. The replication specialist was *"really impressed"* with their development, particularly their ability to be *"vulnerable and transparent"* with scholars and engage in *"healthy self-disclosure.";* Counsellors felt that this helped improve scholar engagement:

> *"These boys hear me say 'Oh, I messed up' all the time, and I think when they see that humility … there is this sense of … 'This person's being authentic; I appreciate that; I'm going to keep coming back to this because I feel as though there's a give and take; it's not just me showing up and being held accountable; this person's doing the work themselves'."* (Marcus, counsellor at Birch)

Group Work, Challenging and Confronting and Men's Work

Group Work concerns counsellors' ability to support scholars in adhering to the three group conditions (fun, safety/respect and challenge). Challenging and Confronting refers to how much counsellors consistently identify and share discrepancies between scholars' stated values and behaviours. Men's Work concerns counsellors' processes of internalising the core values. This requires doing their own ruthless self-examination (deep personal reflection) in order to understand and be able to articulate their own story with authenticity and to mitigate countertransference.

Most scholars we interviewed felt their groups were safe.

> *"I can speak honestly and express my emotions … my BAM counsellor … makes us feel more comfortable."* (Elijah, scholar)

However, a few did not. Tensions or social hierarchies between scholars that existed before and outside of BAM meant that some scholars felt less inclined to be vulnerable in the group.

> *"Some people … I've had problems with before. So I don't like to … be totally honest in BAM."* (Darius, scholar)

Similarly, while the replication specialist felt most counsellors were at an intermediate level for group work, he observed some groups in which participants weren't respecting one another.

> *"There [were] a lot of interruptions during groups … scholars … actually interrupting with unhealthy jokes and different things."* (Replication specialist)

The replication specialist felt this was due to the continued difficulties counsellors faced with challenging and confronting scholars. While improvements were felt by counsellors, most of them were at a novice level in this competency.

> *"It's often hard to challenge young men in certain situations, and so I found that that was the area that I wanted to, kind of, focus on … because of … different groups that I was not seeing progress, you know, in regards to the safety container … I would have liked to see after a second-year group."* (Replication specialist)

The replication specialist felt these difficulties were driven by the lack of opportunity counsellors had to actively experience challenging others and being challenged themselves. While the greater focus on individualised coaching during year two yielded some benefits, it came at the expense of group-level training, which offers a space for counsellors to practice and develop the core values themselves.

The replication specialist felt that this was compounded by the fact that counsellors were yet to attend the Mankind Project Weekend Training Adventure, which all new counsellors are invited to attend when they join BAM, due to challenges with scheduling and capacity. This is a male-only retreat that supports relationship-building between counsellors and challenges them to be open, honest and vulnerable with a group of men supporting them on their journey towards internalising the core values. In other words, it helps them to practice Men's Work and is important in helping counsellors model the core values and challenge scholars to remain in alignment with them. The replication specialist felt this affected the development of the Challenging and Confronting competency as well as *"the man's work that they need to do."*

Systemic leadership

Systemic leadership refers to the ability of counsellors to create the conditions necessary for scholars to benefit from BAM, particularly through their relationship and communication with school staff. One counsellor was at a novice level by the end of year two.

> *"I found out that one of his colleagues [at the school] that he worked with wasn't necessarily really familiar with what BAM did … and the programme itself, and so, you know, I encouraged him to be a little bit more … like, professional development opportunities or inset days so he can have, you know, some form a presentation, just to get the word out a bit more."* (Replication specialist)

One counsellor finished year two at an intermediate level and the other at an advanced level. They both communicated regularly with school staff to establish and maintain an enabling context for delivery.

Exposure

As Table 10 shows, during the pilot phase, all scholars recruited to the BAM programme attended at least one session (i.e. one BAM Circle delivered over the length of one school lesson during school hours). Every scholar who attended at least one session in years one and two attended an average of 17.47 and 13.15 sessions, respectively. The target is 13 per year, set jointly by the evaluation team and YG in light of the average for the first US-based RCT of 13.47. This means that during year one, exposure exceeded the target, while during year two, it was in line with the target.

Attendance was high throughout the pilot phase; scholars attended 80% of sessions or more on average in all three schools.

Table 10: Scholar exposure to BAM (i.e. number of sessions) across both the RCTs, the feasibility phase and the pilot phase

Study	Year	Ever attended	Mean total sessions attended	Mean total sessions | ever attended
Feasibility phase	Year 1	98%	14.67	14.99
	Year 2	68%	9.4	13.74
	Total	98%	24.07	24.59
Pilot phase	Year 1	96%	17.11	17.47
	Year 2	91%	11.63	13.15
	Total	100%	28.66	28.66

Some scholars kept attending BAM because they found it helpful. Some valued the rare opportunity to talk about their feelings, while others learned things that were useful and relevant that they could apply and practice in their everyday lives.

> *"Because [the group sessions were] helpful … you could use it outside of the BAM Circle. So, like, the things that he teaches us, I usually use it when I'm in my lessons, and it really helps me, so that's why I keep coming to BAM."* (Darius, scholar)

However, most scholars we spoke to kept attending BAM because it was fun, active, competitive and informal. BAM offered variety and stimulation, meaningful conversations and high-octane physical activity, and throughout it all, they were encouraged to be themselves.

> *"I feel I keep on going to BAM because it's enjoyable … I feel like sometimes we might sit down in a circle, watch a movie or … go outside, play sports, or next thing they could be, like, telling a story [in] which they're having a debate about what's wrong and stuff like that."* (Elijah, scholar)

The counsellors agreed.

> *"The scholars really value BAM; they enjoy it. And I don't think they just enjoy, oh, it's all about learning, learning about this. Their friends are there; we're probably going to have a fun activity … BAM is associated with fun. It's associated with learning sometimes when it goes serious … all of it was important. But fun is definitely, like, a big part of it. We have a laugh."* (Louis, counsellor at Pine)

Nonetheless, while 13 sessions represented the minimum average attendance deemed acceptable, at the start of the evaluation period, YG and MHF held ambitions for higher numbers of sessions attended – closer to the number reached in the second year of RCT 2 (21.07) – in order to maximise the amount of BAM scholars received. Challenges with exposure were shaped by scholars' start dates, retention and group scheduling.

Start dates and retention

Ninety-seven scholars took part in the pilot phase. Of those, 15 joined late (from January 2022) and 19 left early, which meant that 35% did not complete the full course of BAM (see Table 11).

Table 11: Number of scholars joining and leaving the programme in each school

School	In (Oct 2021–Jan 2022)	In (Jan 2022–Jul 2023)	Out (Jan 2022 –Jul 2023)	Final cohort size (Jul 2023)
Oak	12	7	5	14
Birch	20	3	2	21
Pine	50	5	12	43
Total	**82**	**15**	**19**	**78**

Eight scholars left early because they left the school. Six were asked to leave by counsellors, either because their behaviour was disruptive in the circle (two) or because they failed to attend multiple sessions in a row (four).[33] Three left because they or their carers wanted them to prioritise lessons. For two, the reasons are unknown. At Oak and Pine, some late starters were recruited to fill these spaces. At Oak, some scholars

[33] No minimum level of consistency was established. Decisions regarding group membership in light of inconsistent attendance were made by the counsellor with support from the replication specialist, on a case-by-case basis.

joined late because one group in each school was split during the recruitment period (Oct–Dec 2021), which prompted Grant to recruit more scholars.

Group scheduling

The biggest limiting factor on exposure was the number of groups counsellors were able to schedule. Only three groups, all at Pine, achieved the target of 45 groups held in total (see Table 12).

Table 12: Number of BAM Circles held in each group and year

School	Group	Year 1	Year 2	Total
Oak	A	22	14	36
	B	20	20	40
Birch	C	15	4	19
	D	19	9	28
Pine	E	33	25	58
	F	29	25	54
	G	32	26	58
	H	27	26	53
	I	29	11	41

At Birch, limits on group scheduling were driven by the school context. The school contact in post at the end of the 2022/2023 school year, only assigned to oversee BAM during the last two months of the school year, was an advocate for the programme. He had spent time with Marcus to understand what BAM offered and how it worked, and it resonated with him.

> *"I think if we can engage young people of all ethnicities in mental health and understanding of their own mental health, it's actually a very positive thing. So, I think BAM is good for that side. It's good for getting children to understand how their mental health affects them and how they may be making decisions based on something that they can't actually see, something that's hidden."* (School staff at Birch)

Yet he felt that within the school, "*other people, I don't think they see it as ... as effective as I do,"* which he felt here was grounded in a cultural misalignment between Birch and BAM.

> *"We're punitive here. Definitely. And it sort of goes against BAM's ethos."* (School staff at Birch)

Marcus' original contact at Birch had helped to mitigate the effects of this tension. That contact left Birch at the end of the 2021/2022 school year and was replaced by a less supportive member of staff, who was only replaced by the current, more supportive staff member late in the year.

> *"They just left me to my own devices."* (Marcus, counsellor at Birch)

Marcus and the school contact felt this limited BAM's visibility, allowing this misalignment to fuel misconceptions of BAM; some staff felt BAM undermined the school's more authoritative approach.

> *"There would be a perception that he's trying to be the children's friend when it's not … his job … because here we have to have the uniforms a certain way; they say, 'Oh, the BAM kids, they never have their uniform on'. So there's a ... blame game as well for things that may go wrong in the school."* (School staff at Birch)

The school contact felt that this had shaped perceptions of BAM as *"a loose programme and, you know, kids on it were just having a bit of a jolly."* Consequently, school leadership prioritised scholars' presence in lessons rather than BAM and failed to allocate a dedicated room to the counsellor at Birch.

> *"I couldn't get them out of their core lesson … the school had made changes to how they felt I should, sort of, engage with students, which I didn't experience in the first year … I struggled a bit, man, completely struggled with it."* (Marcus, counsellor at Birch).

The replication specialist felt the situation was exacerbated by the need for BAM *"to build other champions within the school. I just think … we needed to do a better job with that."* The replication specialist, programme manager and Marcus all felt that the programme manager's lack of capacity further limited his ability to support the counsellor and address the problems at the school.

> *"Oftentimes, I would find myself during my visits having certain conversations that I felt, like, you know, could have been done from the Mental Health Foundation so that the counsellors have a bit more support."* (Replication specialist)

The environment at Birch stood in contrast to that at Oak, in particular, but also that of Pine. Both schools had a dedicated BAM room, strong alignment with BAM's approach and a supportive school contact who championed BAM at a leadership level.

> *"There's [a] very strong contrast between myself and [the counsellor at Birch's] experience … I think having senior administration on board … makes for a real difference."* (Grant, counsellor at Oak)

> *"They don't get any support with how to deal with their mental health, so that's a focus for us as a school, and I think that's why BAM works nicely here because [the counsellor at Pine] is very much on board with … the wider vision of the school and making sure these kids feel safe and supported."* (School staff at Pine)

Instead, group scheduling issues at Oak were driven by counsellor sickness, transport and teacher strikes, and school closures. At Pine, the only group not to reach the target of 45 sessions included year 11s, who were often unavailable due to GCSE-related commitments, including mock exams and revision sessions.

Brief encounters and one-to-ones

YG's target is for all scholars to receive at least one brief encounter with a counsellor a month. This target was met for one month at Oak and no months at Birch or Pine (see Table 13).

Table 13: Scholar engagement with brief encounters

School	# Months in which all scholars received at least one brief encounter	Average % of scholars receiving at least one brief encounter a month*
Oak	1	72
Birch	0	9
Pine	0	65

* Does not include August, April or September–December 2021, when BAM groups had not yet closed

YG expects that roughly 10% of scholars will need one-to-one support from counsellors. This is more of a guide than a target, as the needs of each cohort are different. Counsellors decide who receives a one-to-one based on their capacity, requests from scholars and their understanding of which scholars would benefit from one-to-one support, which is based on their judgment rather than a formal assessment (e.g. the tiers of the HSA). For most months in which BAM was being delivered, more than 10% of scholars at Pine received a one-to-one, but not at Oak or Birch (see Table 14).

Table 14: Scholar engagement with one-to-ones

School	% months in which more than 10% of scholars received at least one one-to-one*	Average % scholars receiving at least one one-to-one a month*
Oak	25	6
Birch	25	5
Pine	81	28

* Does not include August, April or September–December 2021, when BAM groups had not yet closed

There were concerns with data quality regarding brief encounters. Given their duration (up to 15 minutes and usually less than five) and their frequency (counsellors can sometimes hold 20–30 brief encounters a day), counsellors had difficulties with data completeness. It is likely they were under-reported. With one-to-ones, exposure was limited by the same issues that impacted counsellors during BAM Circles: at Birch, Marcus struggled to withdraw scholars from lessons for one-to-ones, while at Oak, Grant was intermittently unable to work due to periods of long-term absence. At Pine, Louis' regular participation in playground duty at lunchtime offered opportunities for both brief encounters and one-to-ones, while scholars at Pine also visited the counsellor in the BAM room for ad-hoc one-to-ones as and when they were able or needed to.

Fidelity

With more experience and the replication specialist's support, counsellors adhered to the curriculum well during the pilot phase for the lessons they were able to deliver, building on their performance during the feasibility phase (they were at an advanced or intermediate level for this competency). The scholars we

interviewed responded well to the curriculum: they felt it was fun, engaging, relevant and varied (see Figure 7).

Figure 7: Elijah, Sanjay and Mark's experiences of the curriculum

Case Study 1: Elijah

- Elijah enjoyed how varied the activities were: ***"we might sit down in a circle, watch a movie… go outside, play sports or next thing they could be telling a story."***
- He felt that being active helped him blow off steam: ***"We might go outside… if we're too hyperactive. He might do an activity with us to calm myself down… that helps us. That calms myself down to communicate with other people."***
- His counsellor ranked ***"active learning first"*** among BAM's influences on Elijah.

Case Study 2: Sanjay

- Sanjay's counsellor felt he particularly enjoyed the physical side of the curriculum: ***"he's got such a boyish energy that you could see that he just wants to move. You know, even when he's speaking, sometimes he's leaning in his chair and he's rocking."***
- He felt the activities helped Sanjay to form stronger bonds with his BAM peers: ***"he doesn't sit on the fence when it comes to being active… he likes to bounce off people… to sort of just connect with his peers… Yeah, I think that's what it***

Case Study 3: Mark

- The groups were Mark's favourite activity: ***"we talk about what's going on in our life and sometimes we do a group challenge… we get to bond."***
- His counsellor felt this desire for closeness motivated Mark to engage in BAM and brought the curriculum to life: ***"being taught those core values with his peers, role play, the videos, all the different BAM lessons, I think that they… it helped to crystallise… his peers valuing it and challenging each other and articulating… that brought it home."***

However, they experienced challenges progressing through each of the 30 lessons (see Table 15). Lesson progression was limited by group scheduling, group size and scholar engagement.

Table 15: Lesson reached in the curriculum in each group at the end of year one and year two

School	Group	Year 1 (target = 15)	Year 2 (target = 30)
Oak	A	8	16
	B	9	14
Birch	C	6	12
	D	5	13
Pine	E	15	21
	F	15	23
	G	15	23
	H	13	19

	I	15	22

Group scheduling

As described above, counsellors experienced limitations in their ability to host sessions. These were driven by the school context at Birch and Pine and counsellor sickness, transport and teacher strikes, and school closures at Oak. This impeded their movement through the curriculum, particularly at Birch.

> *"I had to stop seeing groups … working around it … was definitely difficult with the fidelity as well."* (Marcus, counsellor at Birch)

Group sizes

Table 16: Number of scholars joining and leaving each group

School	Group	In	Out	Final
Oak	A	9	2	7
	B	10	3	7
Birch	C	13	0	13
	D	10	2	8
Pine	E	10	2	8
	F	11	3	8
	G	12	1	11
	H	14	4	10
	I	8	2	6

Smaller groups also posed challenges, as one or two scholars failing to attend would render the lesson infeasible.

> *"I might have a group that's got … four students in it because one might be on internal exclusion, and all of a sudden, there's a mission that requires at least eight; can't do the mission, you know, and it sort of drags out, and it delays it…"* (Marcus, counsellor at Birch)

Engagement from scholars

The replication specialist felt that the difficulties counsellors faced with the group work competency impacted curriculum progression.

> *"The development of the group, the stages that they got to … I saw a norming stage, but … it was a lack of getting to that performing stage ... with the progression in the curriculum,*

> *I would have … possibly expected, you know, later on within the curriculum for those groups."* (Replication specialist)

Counsellors felt that some of the challenges they faced with group dynamics stemmed from issues with recruitment. While the HSA indicated cohorts were weighted towards Tiers 2 and 3, and the counsellor at Birch did feel the school referred scholars they felt were '*bad students'* due to their misconceptions of BAM, counsellors did not feel this unduly affected the group dynamic.

Instead, all counsellors felt they faced issues with pre-existing factions in groups, which undermined group safety and limited curriculum progression.

> *"That was a bit of a challenge in the dynamic because they were … yeah, a lot of scholars were friends."* (Louis, counsellor at Pine)

Evidence of promise: Research Question 2

To what extent is implementation contributing to improvements in young people's social-emotional development, for whom, under what circumstances and why?

While the trajectory of scholars' social-emotional development was positive, the change observed through the SDQ was not statistically significant at a Bonferroni-corrected alpha level of 0.008. The change observed through most of the domains in the HSA was not statistically significant at a Bonferroni-corrected alpha level of 0.004. However, a statistically significant positive change was observed for the Empathy domain (O). There is strong evidence from the 11 case studies that BAM was not the most important contributor to scholars' development – scholars generally ranked it second or third, below family, friends, faith or football (R). Nonetheless, there is strong evidence that BAM made an important contribution to the trajectories of the case studies, regardless of their size and direction (MResO). It did so by supporting engagement in action and reflection, which helped scholars manage their emotions, bolstered their self-esteem and offered them catharsis (MResP). There is strong evidence that for most of the case studies, whether and how they engaged in action and reflection was influenced by their readiness and willingness to observe the three group conditions: fun, safety and challenge. Their ability to do so was shaped by wider influences in their lives, including the strength of their support network.

The findings related to this research question are structured as follows. First, the findings related to social-emotional development are shared (the HSA and SDQ results) – the outcome, or O, of this sub-theory. This is followed by an assessment of the contribution of BAM (the mechanism resource, or MResO) to this development through an examination of the 11 case studies. These case studies are split into two clusters, A and B, with the former constituting those who experienced a more positive journey, characterised by larger improvements in social-emotional development, and the latter constituting those who experienced a less positive journey. The findings from each cluster are taken in turn. Each considers how BAM contributed to social-emotional development for those in the cluster by helping them to actively experience the core values before reflecting on those experiences (the mechanism response, or MResP). This information is shared alongside an explanation of the role that context, or C, played in shaping the ability of some scholars in each cluster to engage more in action and reflection than others.

Changes in social-emotional development: HSA and SDQ results

Participants

A total of 97 participants were involved in the pilot phase of the programme. All 97 were part of the evaluation. Table 17 provides an overview of the number of participants with pre- and post-intervention data on the SDQ and HSA. Overall, there were 81 participants who provided at least some form of quantitative data for the pilot phase of the project (either at baseline or post-test). Six participants did not complete the baseline assessment because they joined the programme after the assessments were conducted. According to counsellors, most of the remaining 30 did not complete it due to non-attendance at school on the assessment day, primarily due to illness but also due to temporary internal and external exclusions. There were also some technical difficulties with completion at Oak, including a poor internet connection, which led to some scholars being unable to enter and submit their responses.

Nineteen participants did not complete the post-intervention assessment because they left the programme prematurely (see Figure 2 for more information), while the 10 Year 11 students at Pine were already on study leave by the time of the follow-up assessment. According to counsellors, the one remaining SDQ

assessment and 14 remaining HSA assessments were missing due to non-attendance. The counsellor at Pine was able to follow up with 13 students at a later date to invite them to complete the SDQ (a short assessment) but not the HSA (a much longer assessment). Due to these known systematic reasons for the lack of data completeness, the missing data in this study should be considered 'Missing not at random'. Where inferential statistics are employed, Bonferroni correction has been used to ensure sensitivity, given multiple measurements for a small sample.

Table 17. Number of participants with complete data in Oak, Birch and Pine and in total

Data complete	Oak (n = 19)	Birch (n = 23)	Pine (n = 55)	Total (n = 97)
Data on pre-intervention SDQ – N (%)	9 (47%)	18 (78%)	34 (62%)	61 (63%)
Data on post-intervention SDQ – N (%)	10 (53%)	15 (65%)	42 (76%)	67 (69%)
Data on pre-intervention HSA – N (%)	9 (47%)	18 (78%)	34 (62%)	61 (63%)
Data on post-intervention HSA – N (%)	10 (53%)	15 (65%)	29 (53%)	54 (56%)
Any complete data – N (%)	15 (79%)	22 (96%)	44 (80%)	81 (84%)

The demographic characteristics of the sample who provided some form of outcome data (n = 81) can be seen in Table 18. In the total sample, the mean age of participants who provided some form of data was 13 years, the majority of participants were Black/Black British (70%) and most participants were categorised as having a Tier 2 level of need at pre-intervention (41%).

Table 18. Baseline demographic characteristics of participants who provided at least some form of data

Demographics	Oak (n = 15)	Birch (n = 22)	Pine (n = 44)	Total (n = 81)
Age at enrolment – mean (SD)	13.07 (0.26)	12.73 (0.7)	12.82 (0.79)	12.84 (0.7)
Year group – N (%)				
Year 8	-	9 (41%)	19 (43%)	28 (35%)
Year 9	14 (93%)	13 (59%)	19 (43%)	46 (57%)
Year 10	1 (7%)	-	6 (14%)	7 (9%)
Other	-	-	-	-
Missing	-	-	-	-
Ethnicity – N (%)				
Black/Black British	10 (67%)	18 (82%)	29 (66%)	57 (70%)
Mixed Ethnicity	3 (20%)	3 (14%)	8 (18%)	14 (17%)
White/White British	2 (13%)	-	4 (9%)	6 (7%)
Other	-	1 (5%)	2 (4%)	3 (4%)
Missing	-	-	1 (2%)	1 (1%)
Pre-intervention level of need – N (%)				
Tier 1	-	1 (18%)	5 (11%)	6 (7%)
Tier 2	4 (27%)	12 (55%)	17 (39%)	33 (41%)
Tier 3	5 (33%)	5 (23%)	12 (27%)	22 (27%)
Missing	6 (40%)	4 (18%)	10 (23%)	20 (25%)

Baseline outcome data

Table 19 details pre-intervention SDQ Total Difficulties scores, subscale scores and four-band categorisation of the SDQ Total Difficulties scores indicating whether participants fall into 'close to average', 'slightly raised (slightly lowered)', 'high (low)' or 'very high (very low)' categories based on cut-off points derived from UK population surveys (see Youth*in*mind, 2016 for score ranges). Around half of the participants fell into the 'close to average' category for the SDQ Total Difficulties score and subscale scores. Mean scores were highest for the hyperactivity subscale (indicating greater difficulties), followed by the conduct problems subscale. Considering the four-band categorisation of the SDQ, over 20% of participants fell into the 'very high' range for the Total Difficulties score and over 20% of participants were also classified as having 'very high' levels of difficulties for both hyperactivity and peer problems.

Table 19. Pre-intervention SDQ scores for the sample with complete baseline data (n = 61)

SDQ subscale	Overall Mean (SD^)	Close to average N (%)	Slightly raised N (%)	High N (%)	Very high N (%)
Emotional problems score	3.34 (2.54)	42 (69%)	6 (10%)	4 (7%)	9 (15%)
Conduct problems score	3.75 (1.81)	31 (51%)	9 (15%)	12 (20%)	9 (15%)
Hyperactivity score	5.13 (2.58)	36 (59%)	6 (10%)	6 (10%)	13 (21%)
Peer problems score	2.67 (2.0)	31 (51%)	8 (13%)	9 (15%)	13 (21%)
Prosocial score*	6.44 (1.82)	28 (46%)	12 (20%)	17 (28%)	4 (7%)
Total difficulties score	14.9 (6.1)	29 (48%)	13 (21%)	6 (10%)	13 (21%)

*Categories for prosocial scores are as follows: 'close to average', 'slightly lowered', 'low' and 'very low'.

The mean normed scores and categorisations (i.e. whether the participant falls into the category of 'challenge', 'average', or 'strength' for each HSA item) for baseline HSA can be seen in Table 20. The subscales for HSA are also outlined in Table 20.

- AMSS – academic motivation
- AOSS – action orientation
- ASSS – assertiveness
- CTSS – critical thinking
- ECSS – emotion control
- EMSS – empathy
- LNSS – learning interest
- OPSS – optimism
- PSSS – perseverance
- RFSS – reflection
- ADSS – relationships with adults
- PRSS – relationships with peers
- SBSS – school bonding
- TRSS – trust

Most participants across all subscales fell into the 'average' category on the HSA. Approximately 20–30% of participants were classified under the 'challenge' category and the remainder (<15%) fell into the 'strength' category.

Table 20. Pre-intervention HSA scores for the sample with complete baseline data (n = 59)

HSA subscale	Overall Mean (SD)	Challenge (scores <-1.0) N (%)	Average (scores -1 to 1) N (%)	Strength (scores >1) N (%)
Academic motivation (AMSS)	−0.43 (0.9)	19 (32%)	34 (58%)	6 (10%)
Action orientation (AOSS)	−0.42 (1.11)	18 (30%)	36 (59%)	7 (11%)
Assertiveness (ASSS)	−0.22 (1)	10 (16%)	43 (70%)	8 (13%)
Critical thinking (CTSS)	−0.36 (1.04)	18 (30%)	36 (59%)	7 (11%)
Emotion control (ECSS)	−0.3 (0.95)	17 (28%)	36 (59%)	8 (13%)
Empathy (EMSS)	−0.62 (0.81)	26 (43%)	31 (51%)	4 (7%)
Learning interest (LNSS)	−0.37 (0.97)	20 (33%)	33 (54%)	8 (13%)
Optimism (OPSS)	−0.64 (0.93)	24 (39%)	35 (57%)	2 (3%)
Perseverance (PSSS)	−0.49 (1.17)	22 (36%)	32 (52%)	7 (11%)
Reflection (RFSS)	−0.33 (0.96)	17 (28%)	39 (64%)	5 (8%)
Relationships with adults (ADSS)	−0.43 (0.87)	20 (33%)	39 (64%)	2 (3%)
Relationships with peers (PRSS)	−0.24 (1.06)	14 (23%)	38 (63%)	8 (13%)
School bonding (SBSS)	−0.96 (0.84)	33 (54%)	27 (44%)	1 (2%)
Trust (TRSS)	−0.37 (0.9)	13 (21%)	43 (70%)	5 (8%)

SDQ

There were 44 BAM scholars (45% of all programme participants) with complete SDQ data pre- and post-intervention. Table 21 describes the demographic characteristics of participants with and without complete pre- and post-intervention SDQ data. Using the available data, the characteristics of the two groups and the distribution of the level of need were similar in both groups. However, there were higher levels of missing data on the level of need for participants without pre- and post-intervention SDQ data.

Table 21. Demographic characteristics of participants with and without complete pre- and post-intervention SDQ data

Demographics	Participants with pre- and post-intervention SDQ data (n = 44)		Participants without pre- and post-intervention SDQ data (n = 53)	
Age at enrolment – mean (SD)	12.61 (0.54)		13.13 (0.86)	
Year group – N (%)				
Year 8	20 (45%)		16 (30%)	
Year 9	24 (55%)		26 (49 %)	
Year 10	-		10 (19%)	
Other	-		1 (2%)	
Missing	-		-	
Ethnicity – N (%)				
Black/Black British	30 (68%)		34 (64%)	
Mixed Ethnicity	8 (18%)		10 (19%)	
White/White British	4 (9%)		6 (11%)	
Other	1 (2%)		2 (4%)	
Missing	1 (2%)		1 (2%)	
Level of need – N (%)	Pre	Post	Pre	Post
Tier 1	4 (9%)	14 (32%)	2 (4%)	5 (9%)
Tier 2	23 (52%)	23 (52%)	10 (19%)	14 (26%)
Tier 3	17 (39%)	7 (16%)	5 (9%)	5 (9%)
Missing	-	-	36 (68%)	29 (55%)

Paired sample t-tests on SDQ Total Difficulties and subscale scores (see Table 22) indicated no significant pre- to post-intervention change when the Bonferroni correction was applied to account for multiple testing.

Table 22. Paired t-tests on SDQ Total Difficulties and subscale scores (n = 44)

SDQ subscale	N	Pre-intervention mean (SD)	Post-intervention mean (SD)	Mean difference (95% CI ††)	Effect size†	t-score	p-value (two-tailed)^
Total difficulties score	44	14.73 (6.58)	13 (5.83)	−1.73 (−3.34 to −0.12)	−0.26	−2.16	0.04
Emotional problems score	44	3.05 (2.6)	2.77 (2.06)	−0.27 (−1.03 to 0.48)	−0.1	−0.73	0.47
Conduct problems score	44	3.61 (1.81)	2.93 (1.82)	−0.68 (−1.18 to −0.18)	−0.38	−2.76	0.009
Hyperactivity score	44	5.2 (2.66)	4.48 (2.62)	−0.73 (−1.3 to −0.15)	−0.27	−2.54	0.01
Peer problems score	44	2.86 (2.03)	2.82 (1.83)	−0.04 (−0.59 to 0.5)	−0.02	−0.17	0.87
Prosocial score	44	6.43 (1.9)	6.73 (1.9)	0.3 (−0.26 to 0.85)	0.16	1.07	0.29

^Significance was determined based on a Bonferroni-corrected alpha level of 0.008. †Cohen's D. CI = confidence interval.

Similarly, McNemar tests on the SDQ scores categorised into two groups ('close to average' or 'slightly raised' versus 'high' or 'very high'; see Table 23) displayed no evidence of change in group membership from pre- to post-intervention. This was the case regardless of whether the Bonferroni correction was applied.

Table 23. McNemar tests on the SDQ four-band difficulty division (n = 44)

SDQ subscale	N	Pre-intervention 'high/very high' N (%)	Post intervention 'high / very high' N (%)	Difference: post minus pre (95% CI)	p-value^
Total difficulties score	44	11 (25%)	11 (25%)	−0% (−18% to 18%)	1.0
Emotional problems score	44	9 (20%)	5 (11%)	−9% (−25% to 7%)	0.34
Conduct problems score	44	14 (32%)	12 (27%)	−5% (−21% to 12%)	0.75
Hyperactivity score	44	14 (32%)	10 (23%)	−9% (−22% to 4%)	0.22
Peer problems score	44	17 (39%)	17 (39%)	0% (−18% to 18%)	1.0
Prosocial score†	44	15 (34%)	15 (34%)	0% (−18% to 18%)	1.0

^Significance was determined based on a Bonferroni-corrected alpha level of 0.008. †Pre-intervention and post-intervention N (%) for prosocial score are categorised as 'low/very low'. CI = confidence interval.

Subgroup analyses, including school-level t-tests and McNemar tests, are provided in Appendix 12. There was no significant pre- to post-intervention change in scores or group membership at the school level.

HSA

A total of 32 BAM scholars (33% of all programme participants) provided complete pre- and post-intervention HSA data, and a further two participants provided near complete pre- and post-intervention HSA data (i.e. missing data on one or more subscales). Therefore, there were 34 participants who provided enough HSA data to be included in the analysis. A fuller discussion of the implications of the relatively low response rate is provided in the Limitations section. The characteristics of participants with and without pre- and post-intervention HSA data can be seen in Table 24.

Table 24. Demographic characteristics of participants with and without complete pre- and post-intervention HSA data

Demographics	Participants with pre- and post-intervention HSA data (n = 34^)		Participants without pre- and post-intervention HSA data (n = 51)	
Age at enrolment – mean (SD)	12.65 (0.54)		12.94 (0.79)	
Year group – N (%)				
Year 8	15 (44%)		16 (31%)	
Year 9	19 (56%)		28 (55%)	
Year 10	-		7 (14%)	
Other	-		-	
Missing	-		-	
Ethnicity – N (%)				
Black/Black British	25 (74%)		35 (68%)	
Mixed Ethnicity	7 (44%)		7 (14%)	
White/White British	-		5 (10%)	
Other	2 (3%)		3 (6%)	
Missing	-		1 (2%)	
Level of need – N (%)	Pre	Post	Pre	Post
Tier 1	-	13 (38%)	-	6 (18%)
Tier 2	20 (59%)	14 (41%)	13 (48%)	23 (67%)
Tier 3	11 (32%)	7 (21%)	11 (41%)	5 (15%)
Other	3 (9%)	-	3 (11%)	-
Missing	-	-	-	-

^ Two participants were missing data on one or more HSA subscales but were counted in the complete data sample, as they had both pre- and post-intervention data.

Paired sample t-tests on the 14 HSA subscales (see Table 25) revealed no significant change over time when the Bonferroni correction was applied, except for the Empathy (EMSS) subscale. EMSS scores improved from pre- (mean normed score = −0.73, SD = 0.82) to post-intervention (mean normed score = −0.23, SD = 0.88). The size of this effect can be described as moderate (Cohen's D = 0.64). When we look at just the effect size, no HSA subscales demonstrate a moderate level of change (Cohen's D greater than 0.5). School bonding (SBSS) and Assertiveness (ASSS) are the only two other HSA subscales that have effect sizes of 0.4 or more (0.47 and 0.4, respectively).

Table 25. Paired t-tests on HSA subscales (n = 34)

HSA subscale	N	Pre-intervention mean (SD)	Post-intervention mean (SD)	Mean difference (95% CIs)	Effect size†	t-score	p-value (two-tailed)^
Academic motivation (AMSS)	32	−0.47 (0.79)	−0.48 (0.94)	0.01 (−0.38 to 0.4)	0.01	0.06	0.95
Action orientation (AOSS)	33	−0.41 (1.14)	−0.07 (0.96)	−0.35 (−0.73 to 0.04)	0.32	−1.85	0.07
Assertiveness (ASSS)	33	−0.33 (0.99)	0.15 (0.93)	−0.48 (−0.91 to −0.05)	0.4	−2.29	0.03
Critical thinking (CTSS)	34	−0.29 (1.08)	−0.08 (0.86)	−0.21 (−0.57 to 0.15)	0.21	−1.2	0.24
Emotion control (ECSS)	33	−0.27 (0.77)	−0.08 (0.87)	−0.19 (−0.49 to 0.11)	0.23	−1.3	0.2
Empathy (EMSS)	33	**−0.73 (0.82)**	**−0.23 (0.88)**	**−0.5 (−0.78 to −0.22)**	**0.64**	**−3.69**	**0.001**
Learning interest (LNSS)	34	−0.44 (0.97)	−0.12 (0.88)	−0.32 (−0.69 to 0.05)	0.31	−1.78	0.08
Optimism (OPSS)	34	−0.67 (0.92)	−0.49 (0.83)	−0.18 (−0.59 to 0.23)	0.15	−0.89	0.38
Perseverance (PSSS)	33	−0.39 (1.28)	−0.23 (0.97)	−0.16 (−0.53 to 0.21)	0.16	−0.89	0.38
Reflection (RFSS)	33	−0.49 (0.96)	−0.29 (0.95)	−0.21 (−0.57 to 0.16)	0.2	−1.16	0.26
Relationships with adults (ADSS)	33	−0.42 (0.85)	−0.23 (0.98)	−0.19 (−0.55 to 0.16)	0.19	−1.12	0.27
Relationships with peers (PRSS)	34	−0.29 (1.04)	−0.03 (0.98)	−0.27 (−0.61 to 0.08)	0.27	−1.59	0.12
School bonding (SBSS)	34	−1.03 (0.81)	−0.54 (0.87)	−0.49 (−0.85 to -0.13)	0.47	−2.75	0.01
Trust (TRSS)	33	−0.35 (0.89)	−0.03 (0.97)	−0.31 (−0.74 to 0.11)	0.26	−1.51	0.14

^Significance was determined based on a Bonferroni-corrected alpha level of 0.00357. †Cohen's D. CI = confidence interval.

A significant change in the HSA subscale sore for empathy (EMSS) was observed, with a 0.5 mean difference indicating an increase in empathy overall across the whole sample (see Table 25). Similarly, McNemar tests on the HSA scores categorised into two groups ('challenge' versus 'average/strength'; see Table 26) displayed no evidence of change in group membership over time. This was the case regardless of whether the Bonferroni correction was applied.

Table 26. McNemar tests on the HSA three-band categorisation (n = 34)

HSA subscale	N	Pre-intervention challenge N (%)	Post-intervention challenge N (%)	Test statistic	p-value^
Academic motivation (AMSS)	32	9 (28%)	9 (28%)	0	1
Action orientation (AOSS)	33	10 (30%)	5 (15%)	2.29	0.13
Assertiveness (ASSS)	33	6 (18%)	5 (15%)	0	1
Critical thinking (CTSS)	34	9 (26%)	7 (21%)	0.1	0.75
Emotion control (ECSS)	33	7 (21%)	4 (12%)	0.8	0.37
Empathy (EMSS)	33	17 (52%)	9 (27%)	4.08	0.04
Learning interest (LNSS)	34	10 (29%)	5 (15%)	1.23	0.27
Optimism (OPSS)	34	11 (32%)	10 (29%)	0	1
Perseverance (PSSS)	33	11 (33%)	9 (27%)	0.12	0.72
Reflection (RFSS)	33	12 (36%)	7 (21%)	1.45	0.23
Relationships with adults (ADSS)	33	12 (36%)	9 (27%)	0.36	0.55
Relationships with peers (PRSS)	34	8 (24%)	4 (12%)	0.9	0.34
School bonding (SBSS)	34	18 (53%)	11 (32%)	2.77	0.1
Trust (TRSS)	33	6 (18%)	5 (15%)	0	1

^Significance was determined based on a Bonferroni corrected alpha level of 0.00357.

Subgroup analyses, including school-level t-tests and McNemar tests, are provided in Appendix 13. There was no significant pre- to post-intervention change in HSA scores or HSA group membership at the school level.

Case studies

Each of the 11 scholars we interviewed constitutes a case study. The studies were constructed by triangulating the perspectives of the counsellor, the scholar, their parent/carer and school staff regarding the journey of each scholar over the preceding two years, alongside their implementation, SDQ and HSA data. We compiled the studies to help us understand why some scholars developed more than others and to gauge the role BAM played in that development. Based on their trajectories, two clusters of scholars emerged from these case studies.

Cluster A

Six of these scholars – Cluster A – experienced a more positive journey, characterised by larger improvements in social-emotional development. This cluster includes two of the highlighted case studies, Elijah and Sanjay. These scholars achieved either a positive category change in the SDQ Total Difficulties score (e.g. from 'abnormal' to 'borderline') or experienced more positive than negative category changes across the HSA domains (e.g. from 'average' to 'strength').

For most Cluster A scholars, BAM contributed to their social-emotional development and their internalisation of the core values. Those who did benefit from BAM engaged more in action and reflection. Challenges with curriculum progression meant that the final two core values – respect for womanhood and visionary goal setting – were not delivered in full across each school. Instead, scholars' growth was realised across the first four values: integrity, self-determination, positive anger expression and accountability.

Action

The first core value is integrity – a characteristic of someone who knows who they want to be and acts in alignment with this vision for themselves. Some in Cluster A engaged well with BAM's active learning processes (the 'action' element of the mechanism response within the intermediate sub-theory of the theory of change) – these processes were fun and experiential and helped them to practice the resilience needed to remain in alignment.

> *"It gives us lessons ... in the form of, like, playing and stuff … everyone still remembers their missions."* (Luis, scholar)

> *"So we have to, like, adjust to, like, whatever the task is … so, like, it helps us with our confidence."* (Mark, scholar)

Reflection

For most scholars in Cluster A, the growth in their ability to express anger positively (a core value) was integral to their development. Some were able to learn and apply the two CBT-based skills that constitute the self-determination core value in the curriculum: deep breathing and positive thought replacement.

> *"He really appreciated the idea of deep breathing … that helped him to become grounded."* (Louis, counsellor at Pine)

> *"It really helped me think about what I do when I'm angry and how to, like, remove the anger and replace it with a positive thought … I used it when I was in my class, and I had two warnings; I was angry, but I just thought about no detention … so I just was quiet in my class."* (Femi, scholar)

For some scholars, a lack of self-worth, exacerbated by the difficult transition to secondary school, drove their volatility: their insecurity would manifest as anger in response to minor provocations. Sharing their vulnerabilities in the group and having them accepted helped these scholars to accept themselves and offered catharsis (see Figure 8).

> *"[BAM] allowed him to connect with something that he felt vulnerable about but yet be accepted within that space to make him feel a lot more confident … I think that allowed him to grow within himself."* (Marcus, counsellor at Birch)

> *"I can tell them what I need to say, like, what I need to get off my chest."* (Mark, scholar)

Figure 8: Sanjay's engagement in reflection

Case Study 2: Sanjay

- Sanjay's counsellor felt that at the start of BAM, he would be ***"checking to see if he was going to be judged"*** before speaking, but now, ***"he accepts himself more, so as opposed to needing the group to accept him… he's comfortable, you know, of who he is and I think BAM has allowed him to explore that… to exercise the sense of vulnerability and to know that it's okay."***
- Sanjay agreed: ***"they always try to be supportive of me and you know, they always try to give me positive expectations."***

For some Cluster A scholars, the accountability core value resonated with them the most. BAM offered them the archetype of 'manhood': it provided a framework for realising the mature and respected young men they aspired to be, young men who take responsibility for their actions and treat others the way they want to be treated.

> *"That's his conscious process … 'This thing is telling me about being accountable … and, like, I need to get there somehow … this BAM thing is, kind of, helping me to be that … to be a better man'."* (Louis, counsellor at Pine)

> *"They explain a lot of stuff to us … to show respect and stuff, yeah … Respect people when they speak."* (David, scholar)

Context: the three conditions

Cluster A scholars who engaged in action and reflection did so because they adhered sufficiently to BAM's three conditions: have fun, be safe and challenge yourself and others to be open and honest. They could then share freely and experiment with new skills and identities.

> *"We all get together, we all talk, we all check in … so it's a good space to be in, and I like it ... after a BAM session, I'm normally feeling better."* (Aaron, scholar)

These scholars often had a pre-existing motive that drove them to commit to the conditions. Some were searching for a role model to guide them, and others were searching for a way to tackle racism, their anxiety or low self-esteem (see Figure 9).

> *"Before us even going through any type of curriculum, he had this vision in his head that he wanted to do better … he came into BAM already with that … it allowed [BAM to further] expand what that meant for him."* (Marcus, counsellor at Birch)

Figure 9: Sanjay's engagement in the group conditions

Case Study 2: Sanjay

- His counsellor felt Sanjay engaged so readily in BAM because he already knew ***"he wanted to exceed academically… the active learning and reflection was the… fuel that allowed that to sort of resonate more with him to grow."***
- But his counsellor felt he ***"thrived"*** only ***"when the group became cohesive with respect to safety… when he saw his peers also sharing particular things that were quite vulnerable, I think it allowed him to feel safe to do so himself."***

For some Cluster A scholars, the lack of support elsewhere amplified BAM's impact.

> *"I don't really have anyone to talk to but here."* (Sanjay, scholar)

For others, BAM and their wider support system reinforced one another (see Figure 10).

> *"Support from BAM, from parents, from friends ... all the people close to me always try [to] help me do [the] best … I can."* (Darius, scholar)

Figure 10: Wider influences in Sanjay's life

Case Study 2: Sanjay

- Other influences in Sanjay's life reinforced BAM's contribution to his growth. His mum felt that, in some ways, Sanjay was just ***"growing up… when he started secondary school, he was much younger, so I... also put it down as... being mature and, yeah, I think it's the growth."***
- His counsellor also observed that ***"things felt quite stable within his household,"*** something Sanjay's mum (***"he's great at home"***) and Sanjay (***"I can… talk to them about anything"***) agreed with.

For some in Cluster A, BAM's contribution was less clear. Other supports were more important. Most scholars ranked (in no particular order) family, friends, faith or football as more important sources of support than BAM. For some in Cluster A, the strength of their pre-existing support network limited BAM's impact. One possible contributing factor is BAM's recruitment of young people categorised as Tier 1 on the HSA (i.e. with fewer needs), which may have increased the number of scholars for whom this was the case.

> *"The way that he dresses, the way he carries himself, the way that he speaks, it's clear that he's coming from an environment of social capital ... his values … they aligned with BAM."* (Louis, counsellor at Pine)

Some were able to receive the emotional support and structure they needed from their carers to help them develop their emotion regulation (see Figure 11).

> *"My mum, she's really supportive, she helps me … she's disciplined me a lot, and I think they've helped me to get me to where I am right now."* (Femi, a scholar)

Figure 11: Wider influences in Elijah's life

Case Study 1: Elijah

- Elijah's counsellor felt BAM made a limited contribution to his development: ***"he was in a good place and he maintained that."***
- Elijah agreed. He felt that other supports in his life were ***"a little bit more important. Like my teachers. Because yeah, they do give me good additional feedback, whether that's from exams or… anything like that."***
- He felt that ***"teachers and other people and friends… help me build my confidence… by taking me outside of my comfort zone."***

Others were engaged in other services that they, their carers, school staff and counsellors felt were helping them. One of these services is Football Beyond Borders, a mentoring intervention offered at Oak and Birch, which some scholars felt was as helpful as BAM.

> *"Football Beyond Borders. It helps with … it's kind of like the same thing [as BAM], but, like, after that, we do, like, our football."* (Mark, scholar)

> *"We're getting the same kids, and we're just shoving them all on the same, like, three or four different interventions. And so, it muddies the waters a little bit."* (School staff at Pine)

Many scholars had suspected or diagnosed special educational needs. For some in Cluster A, if they received extra support at school from teachers or from other services, it was hard to gauge BAM's contribution.

> *"The teacher, like, just helped me through a bunch of stuff like anxiety, peer pressure, everything, like, helping me through my, like, exams, telling me to, like, focus and just helping me in general."* (Mohammed, scholar)

> *"Since he started doing all these, you know, all the help that he has been getting, which is fantastic from school and everything … he has improved quite a lot."* (Carer)

For some scholars in Cluster A, they and their carers felt their increased integrity and accountability was largely attributable to their general development.

> *"I just changed … there's nothing really that specific … I grew as a person."* (Luis, scholar)

Cluster B

Three of the 11 case studies – Cluster B – had a less positive journey. This cluster includes one of the highlighted case studies. These scholars:

- Observed either no category change or a negative category change in the SDQ Total Difficulties score (e.g. from 'borderline' to 'abnormal') or
- Experienced the same number of positive and negative category changes or more negative than positive category changes across the HSA domains (e.g. from 'strength to 'average').

For some scholars in Cluster B, implementation challenges limited BAM's impact and their development. Challenges with the school context meant that scholars at Birch attended 4.17 sessions on average in year two. For some in Cluster B, this limited how much they could benefit from BAM.

> *"I feel like he is someone who struggles with anger expression, you know, because it feels very much [like] he's holding on to something, [that] it's not fair; the world is against me … we've not done the deep dive into it because … of timing and what I spoke about earlier [i.e. issues with delivery]."* (Marcus, counsellor at Birch)

For others in Cluster B, the difficulties counsellors experienced with group work and challenging and confronting competencies undermined adherence to BAM's three conditions. These difficulties limited their ability to engage in action and reflection (see Figure 12).

> *"He came out of ... the least vulnerable group. And I might have seen a different side of him had he been in a different group or if there were some individuals that [weren't] in that space."* (Louis, counsellor at Pine)

> *"Let's say, for example, you, like, do something that is wrong and then you go and, like, tell them … some people's reactions could be … angry, furious at the thing that you have done … some people, like, could judge you."* (Mohammed, scholar)

> *"Some people don't always keep it in the circle."* (Anthony, scholar)

Figure 12: Mark's engagement in the group conditions

Case Study 3: Mark

- While Mark's group were respectful, his counsellor found it hard to challenge them to be open and honest with each other: ***"it wasn't a vulnerable circle… certain individuals, it was a tough... that was a tough circle."***
- He felt this limited the depth of Mark's involvement in BAM, as he was more likely to take his cue from more dominant peers in the circle: ***"he was definitely one of the ones that would look around to see if anybody's going to answer first… the leaders… he missed anybody being vulnerable particularly."***

Others in Cluster B did benefit from BAM, according to them, their counsellor and their carer. They did so through the same mechanism as those who benefited from BAM in Cluster A – that is, through engaging in action and reflection and by adhering to the group conditions. However, other challenges in their lives outweighed BAM's impact.

For some scholars, their development was characterised by uncertainty. Some had special educational needs, which limited BAM's contribution (see Figure 13).

Figure 13: Mark's social-emotional development

Case Study 3: Mark

- Mark felt his confidence had grown since the start of BAM: ***"I have more confidence and I feel like I can… make new friends now."***
- He felt BAM and Football Beyond Borders showed him he can overcome adversity: ***"BAM and my other mentoring group they do things to make you… not comfortable… it helps us with our confidence."***
- But his dyslexia still made it hard for him to stay aligned with his goals: ***"I have dyslexia… it's just really challenging and difficult to focus."***

Others experienced different challenges as they grew older.

> *"School is hit or miss, so some good days and bad days. Home, still the same, probably gets frustrated at some times. He does have ADHD … so, that does add to it."* (Carer)

> *"More recently, and I'm assuming it's more of a developmental thing, he has been a lot more rigid and quite quick to flare … I can see that he's having quite a difficult time with, I guess, authority at this point in time."* (Grant, counsellor at Oak)

> *"Sometimes I get detentions for stupid reasons … it's just messing around when you're not meant to, when you're meant to be focusing in class, getting good grades for when you leave secondary school … like, growing up you become less focused."* (Mohammed, scholar)

In some cases, for those in need of extra support, a lack thereof detrimentally affected their well-being (see Figure 14).

> *"I want to push more, but sometimes I get distracted, so I can't really ... sometimes I try to, like, focus. Sometimes, I just can't help it … probably because I have dyslexia. Sometimes it's just really challenging and difficult … [my support worker] left, so I haven't got, like, support, so yeah."* (Mark, scholar)

Figure 14: Limitations on Mark's social-emotional development

Case Study 3: Mark

- Mark's mum felt the improvements she had seen in him were due to a range of factors: ***"It could be a combination of BAM but he also does another mentorship in school called Football Beyond Borders. And I think just maturing as an adolescent."***
- However, support for Mark's dyslexia had been withdrawn - school staff felt Mark was a ***"little bit neglected"*** within the school.

Evidence of promise: Research Question 3

To what extent is social-emotional development contributing to any reported changes in academic performance and behaviour and the avoidance of or reduced involvement in crime/anti-social behaviour, for whom, under what circumstances and why?

There is strong evidence that, as with social-emotional development, responsible decision-making at school remained stable overall. Across attendance, suspensions and academic attainment, performance for most schools and scholars remained unchanged, with some exceptions; while attendance fell and external suspensions increased at Pine, internal suspensions fell, and academic attainment increased at Oak (O). There is strong evidence from the 11 case studies that, while their wider support networks helped (R), social-emotional development made an important contribution to responsible decision-making for most of them, albeit with differences in the extent of this contribution (MResO); This was the case for most scholars regardless of the trajectories of their decision-making over the implementation period. Their development contributed to responsible decision-making by empowering and protecting them (MResP), the extent of which depended on wider influences in their lives (C). These results suggest that BAM's focus on social-emotional development as an intermediate outcome was a legitimate one, given evidence of BAM's contribution to that development and the contribution of that development to responsible decision-making for some scholars.

The findings related to this research question are structured as follows. First, the findings related to academic performance and behaviour are shared – the outcome, or O, of this sub-theory. This is followed by an examination of the 11 case studies, including findings on the contribution of social-emotional development (the mechanism resource, or MResO) to responsible decision-making. It also includes findings on the degree to which this contribution can be explained by the way in which this development empowers scholars to make pro-social choices and protects them from risks (the mechanism response, or MResP). This is followed by an examination of the role of context in enabling or limiting this empowering and protecting effect for different scholars, particularly the presence (or absence) of wider support in their lives.

Changes in academic performance and behaviour: school data results

Analysis

Attendance, behaviour and academic attainment have been analysed for the students participating in the BAM pilot across the three schools. Attendance was recorded as the percentage of days of attendance. Behaviour was measured by the number of days spent in internal and external suspensions. Academic attainment was measured using differing approaches across the three schools. The schools used both grade-based measures and expectation-based categorical measures. Further details are provided below in relation to the relevant schools.

The use of inferential statistics was deemed inappropriate for several reasons. First, there were concerns with data quality, particularly with data on attainment, which was reliant on internal and variable assessments of performance and predicted grades. Second, different measurement approaches were used in each school, reducing the sample size available for analysis and limiting what could be interpreted from inferential statistics. Third, the collection of data at multiple time points across each measure would increase the chances of error in the application of inferential statistics. The results below describe the trends in the data at the different measurement points. The results are also presented individually for each school and not as a whole sample to account for the variation in measurement approaches. Error bars have been

included on all figures to aid in the determination of where larger changes than normal variation might be occurring.

Demographics

Across the schools, young people for whom school data was available were, on average, 13 years of age, mostly from Year 9 and predominantly of Black/Black British ethnicity (see Table 27).

Table 27: Demographic description of the young people included in the academic performance and behaviour analyses.

		Overall N = 79	Pine N = 47	Oak N = 11	Birch N = 21
Mean age (SD)		12.85 (0.75)	12.85 (0.86)	13 (0)	12.76 (0.7)
Year group	Year 8 N (%)	29 (36.7%)	21 (44.7%)	-	8 (38.1%)
	Year 9 N (%)	43 (54.4%)	19 (40.4%)	11 (100%)	13 (61.9%)
	Year 10 N (%)	7 (8.9%)	7 (14.9%)	-	-
Ethnicity	Black/Black British N (%)	54 (68.4%)	31 (66%)	6 (54.5%)	17 (81%)
	Mixed ethnicity N (%)	14 (17.7%)	8 (17%)	-	-
	White/White British N (%)	7 (8.9%)	5 (10.6%)	-	-
	Other N (%)	4 (5%)	3 (6.4%)	5 (45.6%)	4 (19%)

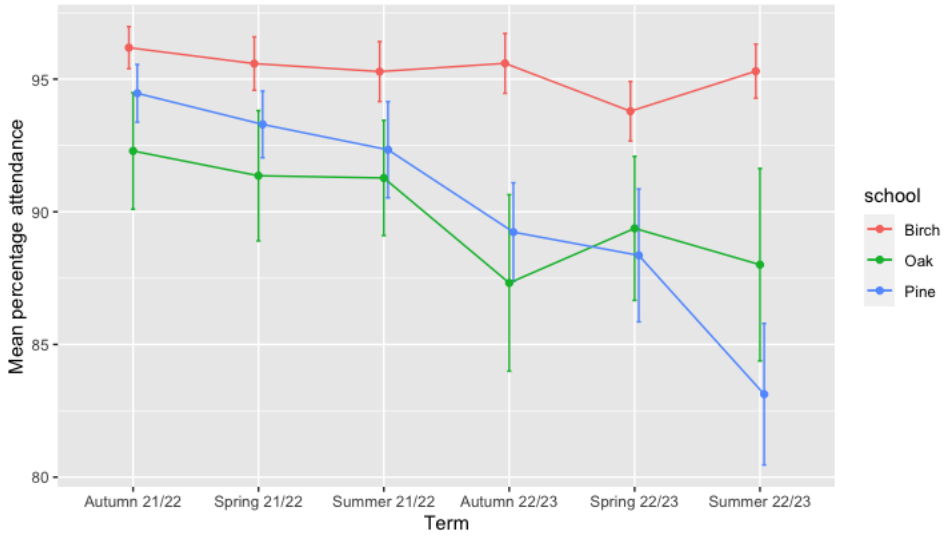
Attendance

The participants in the BAM pilot across the three schools show differing trends in attendance, as seen in Figure 15. Attendance at Birch was stable over all six school terms. Oak had the lowest starting mean attendance. This was stable through the first school year, then dropped slightly in the second year. Young people's attendance at Pine steadily declined over time, with a considerable difference of approximately 10% between the first and last terms of the pilot (see Table 28). The decline in attendance between autumn 2021/2022 and summer 2022/2023 is likely significant due to the lack of overlap between the error bars.

Table 28: Mean percentage attendance by school term between 2021 and 2023 by school

School	Term	Mean (%)	Standard error
Oak	Autumn 2021/22	92.29	2.19
	Spring 2021/22	91.36	2.45
	Summer 2021/22	91.28	2.17
	Autumn 2022/23	87.32	3.33
	Spring 2022/23	89.38	2.71
	Summer 2022/23	88.01	3.63
Pine	Autumn 2021/22	94.47	1.09
	Spring 2021/22	93.29	1.26
	Summer 2021/22	92.34	1.81
	Autumn 2022/23	89.23	1.86
	Spring 2022/23	88.36	2.50
	Summer 2022/23	83.12	2.67
Birch	Autumn 2021/22	96.19	0.79
	Spring 2021/22	95.59	1.01
	Summer 2021/22	95.29	1.13
	Autumn 2022/23	95.60	1.13
	Spring 2022/23	93.79	1.12
	Summer 2022/23	95.30	1.02

Figure 15: Mean percentage attendance with standard error by school term between 2021 and 2023 by school



Suspensions

Internal suspensions

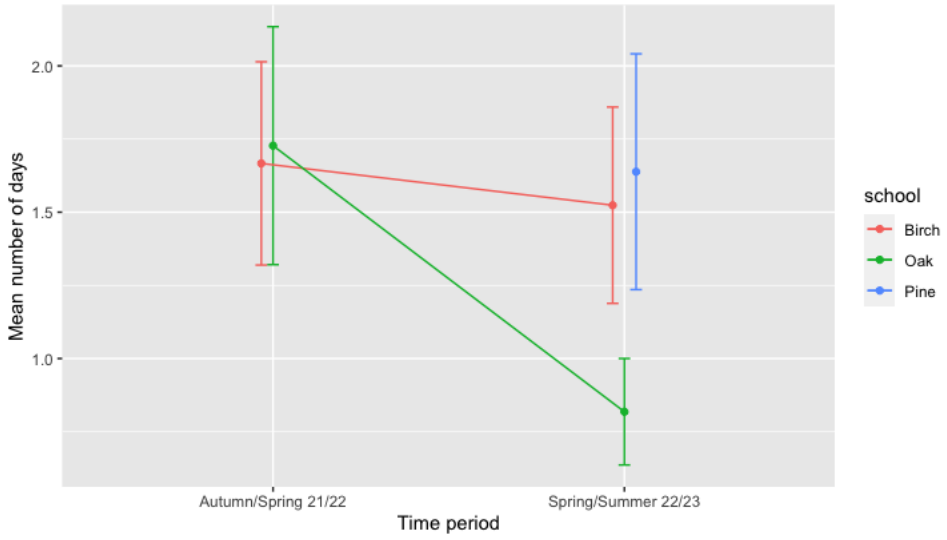
Data on internal suspensions at Pine was only available in the spring/summer time period but were comparable to both schools in autumn/spring and Birch in spring/summer (see Table 29). Figure 16 shows that internal suspensions at Birch were stable across both time periods. Oak had a decline in internal

suspensions from autumn/spring to spring/summer. Please see the Analysis section above for an explanation of why the further use of inferential statistics was deemed inappropriate.

Table 29: Mean number of days of internal suspensions by school over autumn/spring 2021/2022 and spring/summer 2022/2023

School	Term	Mean	Standard error
Oak	Autumn/spring	1.73	0.41
	Spring/summer	0.82	0.18
Pine	Autumn/spring	NA	NA
	Spring/summer	1.64	0.40
Birch	Autumn/spring	1.67	0.35
	Spring/summer	1.52	0.34

Figure 16: Mean number of days of internal suspension with standard error by school and over time



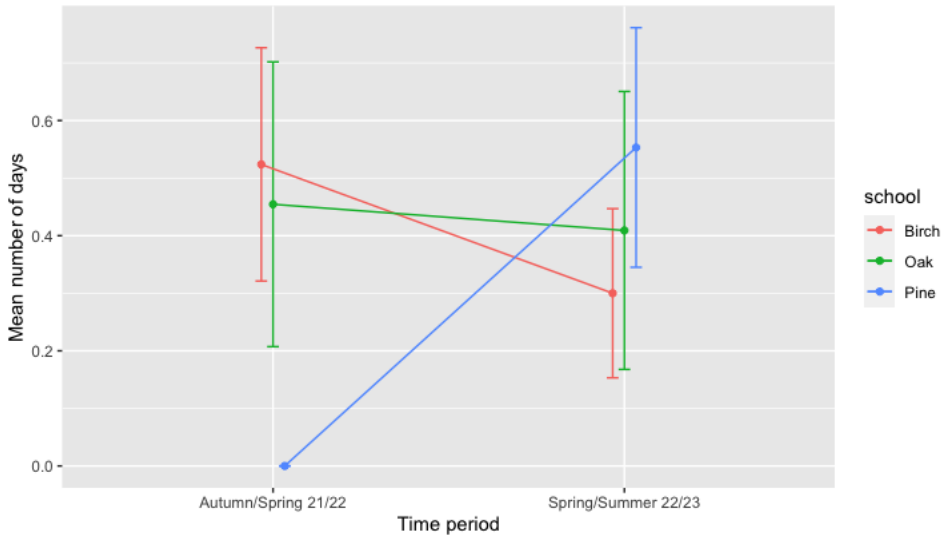
External suspensions

The average level of external suspensions remained consistent across both time periods for Oak and Birch (see Figure 17). No external suspensions were recorded for Pine in the autumn/spring (see Table 30). In spring/summer, external suspensions became comparable to the other two schools.

Table 30: Mean number of days of external suspension by school and over time autumn/spring 2021/2022 and spring/summer 2022/2023

School	Term	Mean	Standard error
Oak	Autumn/spring	0.45	0.25
	Spring/summer	0.41	0.24
Pine	Autumn/spring	0.00	0.00
	Spring/summer	0.55	0.21
Birch	Autumn/spring	0.52	0.20
	Spring/summer	0.30	0.15

Figure 17: Mean number of days of external suspension, including a standard error, by school and over autumn/spring 2021/2022 and spring/summer 2022/2023



Academic attainment

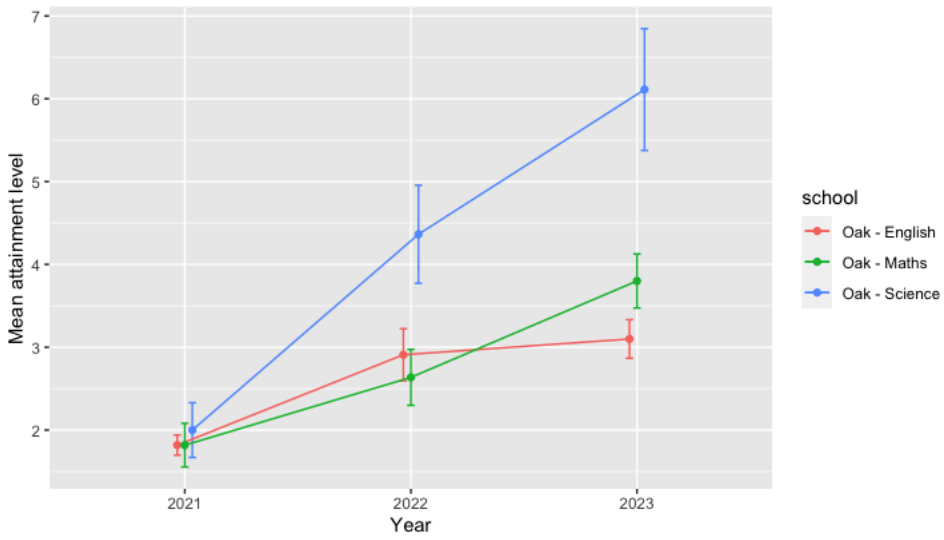
Oak

Academic attainment increased across English, Math and Science at Oak between 2021 and 2023 (see Table 31 and Figure 18). This was recorded as the estimated GCSE grade level (1 to 9, with 1 representing the lowest grade, 4 representing a passing grade and 9 representing the highest possible grade). While grades increased across all three subjects, the mean grade for English and Math did not surpass a passing grade of 4. The mean grade for Science rose above a passing grade, reaching 6.

Table 31: Mean grade categorisation by school for the years 2021, 2022 and 2023

Subject	Term	Mean	Standard error
English	2021	1.82	0.12
	2022	2.91	0.31
	2023	3.10	0.23
Maths	2021	1.82	0.26
	2022	2.64	0.34
	2023	3.80	0.33
Science	2021	2.00	0.33
	2022	4.36	0.59
	2023	6.11	0.73

Figure 18: Mean grade categorisation, including a standard error, for Oak by subject for the years 2021, 2022 and 2023



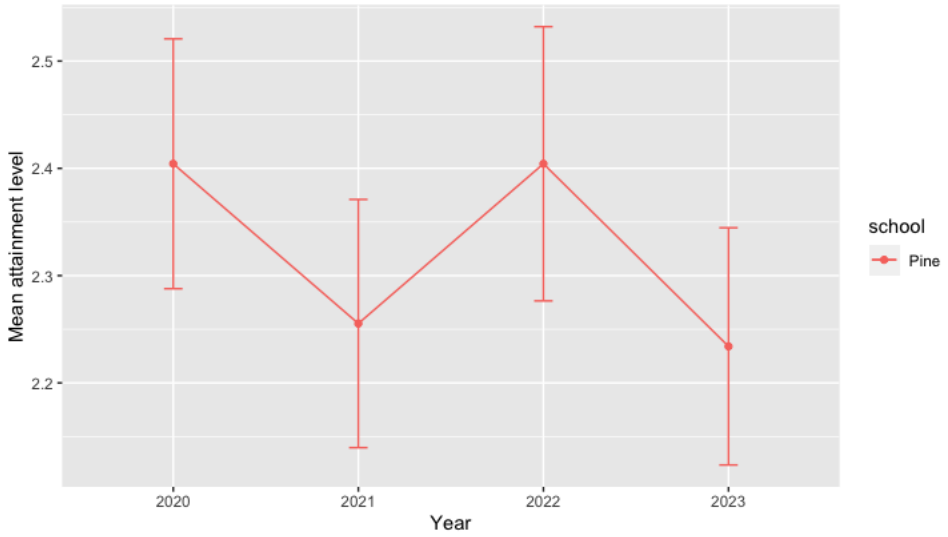
Pine

At Pine, academic attainment was recorded categorically – above or below expectations (see Table 32 and Figure 19). These levels remained consistent between 2020 and 2023 between below expected (2) and expected (3).

Table 32: Mean teacher-assessed academic attainment categorisation for Pine for the years 2020, 2021, 2022 and 2023

Term	Mean	Standard error
2020	2.40	0.12
2021	2.26	0.12
2022	2.40	0.13
2023	2.23	0.11

Figure 19: Mean teacher-assessed academic attainment categorisation, including a standard error, for Pine for the years 2020, 2021, 2022 and 2023



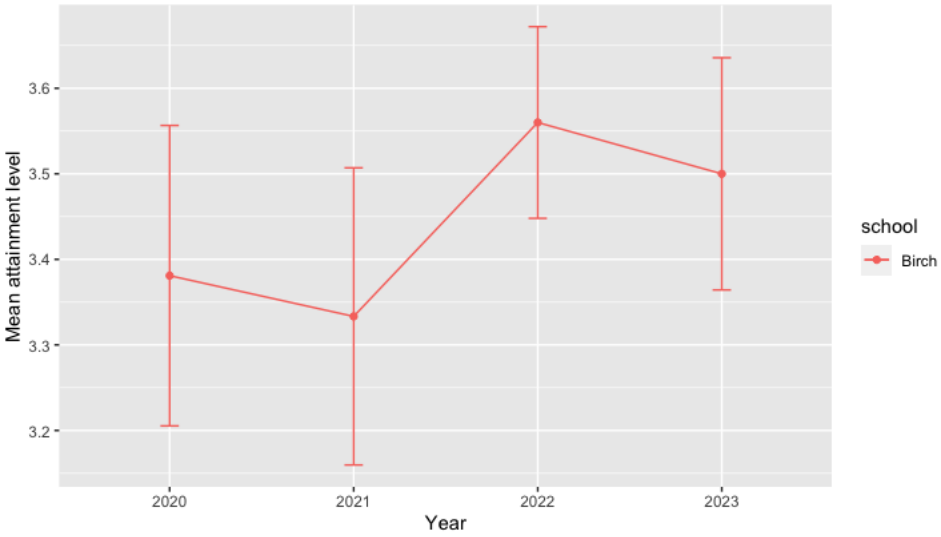
Birch

There was a small increase in the average academic attainment level at Birch between 2021 and 2022. Overall academic attainment levels remained stable (see Table 33 and Figure 20).

Table 33: Mean grade categorisation for Birch for the years 2020, 2021, 2022 and 2023

Term	Mean	Standard error
2020	3.38	0.18
2021	3.33	0.17
2022	3.56	0.11
2023	3.50	0.14

Figure 20: Mean grade categorisation, including a standard error, for Birch for the years 2020, 2021, 2022 and 2023



Case studies

For most of the 11 case studies, findings regarding the extent of change in their academic performance and behaviour are inconclusive, mirroring the implications of the school data for the cohort in general.

Regardless, as with the relationship between implementation and social-emotional development, the direction and size of the change observed are limited in terms of the evidence it provides regarding the contribution of social-emotional development to responsible decision-making. Whether positive or negative, large or small, any changes in scholars' behaviours may or may not be explained by the extent of their internalisation of the core values.

By interrogating the evidence from the 11 case studies, we can better understand this relationship. There is strong evidence from interviews with the scholars, their parents, counsellors and school staff that social-emotional development made an important contribution to the decision-making of most of them, albeit with differences in the extent of this contribution. This was the case for most scholars regardless of the trajectories of their decision-making over the implementation period. Their development contributed by empowering and protecting them, the extent of which depended on their wider context.

Empowering

The core values helped some scholars build deeper relationships with teachers, friends and family and take advantage of other opportunities, including their school education (see Figure 21).

> *"He … wasn't necessarily sure of his place within the school … But certainly, I think ... his interactions with staff … his confidence in the way in which you see him interact with his peers, that he's … in a more positive place around coming to school."* (School staff at Oak)

Figure 21: Elijah's engagement in responsible decision-making

Case Study 1: Elijah

- Elijah's mum did not feel he had changed at home. If anything, she felt things had gotten worse, because he had entered ***"into that typical moody teenage phase."***
- But at school, Elijah noticed some improvement. While Elijah used to be ***"really late"*** for school at the start of BAM, his newfound confidence and positivity meant he would now be late ***"once a month, I think, and like school's doing good. I've learnt new things and I've like had BAM like workshops to help me, but I love my education and like I'll behave in class… right now I think school is amazing."***
- His counsellor noticed some positive changes too: ***"towards the end of the year school year, I noticed that he did make some change... improvements in literature… I did speak to one of his teachers and she did share that with me… that ties to not being afraid, managing self-defeating thoughts."***

This was particularly the case for scholars who had internalised the core value of integrity. They had a clearer sense of who they were and what they wanted and a greater confidence in their ability to overcome obstacles in pursuit of their goals.

> *"I can focus on what I really like and what I enjoy … my acting or my dancing because I wanted to, like, leave this year, but I don't want to anymore because, like, I feel like I'm going to achieve more staying."* (Mark, scholar)

Protecting

The core values also helped to protect scholars from risks. Scholars were better able to manage their emotions in situations that might have deteriorated before and resist pressure from others, including peers, to make irresponsible decisions (see Figure 22).

Figure 22: Sanjay's engagement in responsible decision-making

Case Study 2: Sanjay

- According to Sanjay's mum, his burgeoning self-worth was making him less susceptible to provocation: ***"BAM has helped him to wait a second or more to deal with a situation differently… his behaviour has improved more where he's become... more settled in school."***
- She also felt he was better able empathise with others and take accountability: ***"he was saying to me that before he thought that oh, like, sometimes at school like he feels… the teachers were... against him rather than for him, but… he's realised that they're like, looking out for his best interest."***
- Sanjay attributes his improvement to BAM: ***"I didn't really have a good mentality, I wasn't really like - like it didn't really help because I used to think about negative stuff and that would just hold me back. So, like now that like… now I've got like my BAM counsellor and I get positive expectations from him, he kind of supports me and helps me to, you know, do better."***

> *"I started with anxiety, like, coming to a new school, new friends … they could peer pressure you to do something that you don't want to do at all, but then when I joined BAM, I felt welcomed … It just made me make the right choices."* (Mohammed, scholar)

The protective influence of the core values was reinforced directly by counsellors at critical moments during school. The BAM counsellor became a tool that both teachers and scholars used to de-escalate tensions with peers and school staff, whether through scholars going to see the counsellors to calm down or teachers reinforcing BAM messaging in the classroom.

> *"Another student was squaring up to him last year, and he went and said, 'I want to find [the counsellor],' and he went and found him and regulated himself there."* (School staff at Pine)

> *"An English teacher … was struggling with him in the lessons, where sometimes he would just misbehave, or he [would] challenge her. And then she used me … she said to him, 'I want you to think … whenever you're doing something, I want you to think to yourself, what would [the counsellor] do?'"* (Marcus, counsellor at Birch)

Wider influences

As with Research Question 2, scholars' wider support network (or lack thereof) helped to reinforce (or amplify) the contribution of social-emotional development to responsible decision-making. Similarly, those for whom internalisation of the core values made a larger contribution shared a pre-existing and long-standing clarity on goals or motives.

> *"I want to be a footballer in the future, so that's what's mostly pushing me … it's been a childhood dream, to be honest."* (Aaron, scholar)

However, for some scholars, the size and strength of other challenges in their lives limited the contribution of social-emotional development to responsible decision-making (see Figure 23).

Figure 23: Wider influences in Mark's life

Case Study 3: Mark

- The lack of support for his dyslexia continued to make things difficult at school ***"I get distracted when they're talking… I get in trouble because of that."***
- School staff also acknowledged the challenges Mark continued to experience, both with his engagement (***"[he is] very, very quiet in the classroom"***) and his attendance: ***"His attendance is really poor… we can't get him into school, and when we do get him into school, he's very late."***
- But Mark's confidence was also empowering him to make the most of the opportunities available to him: ***"I feel like [BAM's] still like part of me… it's pushing me because it teaches you about life skills and stuff like that, so I know what to do… it teaches you not to give up."***
- This resilience was helping to protect him from risks; school staff had noticed some improvements in his behaviour: ***"he hasn't had an instance in a very long time."***

The lingering effects of the pandemic, combined with the cost-of-living crisis and high levels of staff turnover, have placed schools' resources under considerable pressure and forced them to make difficult decisions. According to the school contact at Pine, vacant posts and long-term absences accounted for nearly 40% of faculty positions, forcing the school to prioritise. Permanent staff were directed towards Key Stage 4 to support students with their GCSEs, with the majority of temporary staff, characterised by high turnover, taking up posts for Years 7 to 9. They felt this impacted students' school engagement during those years. Counsellors agreed.

> *"I don't think a lot of the boys take the in-year term time, like pop quiz assessments, as seriously … as they do their GCSEs."* (Grant, counsellor at Oak).

Conclusion

The evidence gathered as part of this evaluation supports the causal pathway articulated in the theory of change. For those scholars among the 11 case studies for whom BAM did make a contribution, it did so by helping them to actively experience the core values before reflecting on those experiences. This drove their social-emotional development, which empowered scholars to make responsible decisions and protected them from risks. However, this causal chain did not exist in isolation. The causal links between BAM and social-emotional development and between this development and responsible decision-making were fraught with alternative influences that served to undermine, rival and reinforce BAM's contribution, from the supporting implementation infrastructure and the context of partnering schools to whether BAM was the right programme at the right time for scholars and how it interacted with wider influences in their lives. The interaction between both the programme and scholars and these wider influences generated important differences in experiences for participants, with contributions to outcome change that were similarly wide-ranging.

Evidence of promise: Research Question 4

To what extent has implementation contributed to unintended consequences, for whom, under what circumstances and why?

There is mixed evidence on whether BAM inadvertently supported scholars in engaging in irresponsible decision-making (O). There is strong evidence that negative labels did not develop in BAM groups (MResP). While factions were present in groups, these were not groups of friends widely regarded as badly behaved within the wider school, which may have precipitated negative labelling (MResO). There is mixed evidence on whether missing lessons had a negative impact. None of the counsellors, scholars or parents/carers interviewed shared that they felt missing lessons negatively impacted scholars' academic attainment. However, some school staff did feel this was a risk. This was particularly the case for scholars in receipt of other interventions during school time and for those in settings where school staff and counsellors faced challenges in supporting the application of learning from BAM in school (C).

Irresponsible decision-making

As part of the evaluation, we sought to identify and explore possible unintended adverse effects. We were particularly alert to the impacts of negative group labelling, fractious BAM groups and time out of lessons on school engagement and performance. These issues were prioritised based largely on the experiences of the intervention developer, YG, and its insights into the unintended consequences that occurred most frequently during its 20 years of delivering BAM in the US. These possible adverse effects formed columns on the framework matrix that was used to analyse qualitative data, with any evidence associated with these phenomena recorded against the matrix. We remained open to and actively explored other possible issues during qualitative data collection while also maintaining open channels of communication with both MHF and YG through regular meetings as a structure for raising any serious adverse effects that would have required immediate attention. However, none were observed.

Negative labelling and group factions

Based on their previous experience, YG feared that scholars might develop a perception that BAM is for 'bad kids' due to challenges with recruitment. This could inadvertently support further engagement by BAM scholars in irresponsible decision-making and anti-social behaviour due to the further internalisation of these negative labels and their impact on group dynamics within BAM. This phenomenon has been observed in other UK school- and group-based SEL interventions (Evans et al., 2015). There was also evidence during the feasibility phase that recruitment challenges – specifically groups weighted predominantly towards young people with higher levels of need – precipitated the development of negative labels among scholars in the pupil referral unit and in one of the groups at Oak.

During the pilot phase, largely due to the lessons learned during the feasibility phase, this was less of a concern. In Oak and Pine, Grant and Louis worked with school staff to ensure that despite the relatively high levels of need, groups were not weighted towards students regarded widely among their peers and teachers as being those most associated with poor behaviour. While they were present in the cohort at Birch, this did not unduly affect the perception of the group internally or externally due to BAM's pre-existing reputation. Indeed, all of those who spoke on the matter felt that BAM was something to be proud of and that other students were envious of their participation.

> *"BAM scholars in the school have got clout … BAM has got status … 'He's a BAM scholar', you know … 'He's one of the guys that gets up and leaves. Where are they going? Sir, can I join BAM?' … they're all bopping out, or they're going on a trip, or they ... this American guy's come to talk to them and stuff like that ... so I think that that's got its role as well."* (Louis, counsellor at Pine)

Counsellors did experience challenges with factions in their groups during the pilot phase, but these were not groups of friends widely regarded as badly behaved within the wider school. Rather than leading to irresponsible decision-making, these cliques impeded implementation by challenging counsellors' authority, undermining the group conditions and slowing down progress through the curriculum.

> *"[There were] some subgroups within the wider group initially … maybe three students who were, sort of, trying to be … the joker, you know … not that it had a massive impact, but it was still an impact I recognised that I had to sometimes stop going through the lesson to address particular issues that the boys were, sort of, creating."* (Marcus, counsellor at Birch)

Missing lessons

None of the counsellors, scholars or parents/carers interviewed shared that they felt missing lessons negatively impacted scholars' academic attainment. As reflected on in the findings regarding Research Question 3, BAM's contribution to academic attainment was either positive or inconclusive for the 11 case studies but not detrimental.

Counsellors pursued a range of measures with schools' and MHF's support to limit the degree to which missing lessons impacted students, including regularly rotating the lesson during which BAM was held (at least every term), not holding BAM during core subjects (English, Math and Science), using 'academic integrity' check-ins to talk through school work and grades to allow the group to hold each other accountable for doing the work and achieving the level of performance they each committed to and making it clear that it was the responsibility of students to catch up on any lessons, homework or coursework missed.

> *"There was always that thing about academic integrity, so they knew that they had to make up the lessons, the work in their own lunchtime, home times, after school club, whatever … but it was known that if the teachers have any issues that they could pull rank … that was supposed to keep them on their toes, and I believe that it did."* (Louis, counsellor at Pine)

However, some school staff had reservations. While the school contact at Birch felt that it was too difficult to say, given all the other drivers of school performance a scholar may experience, the contact at Pine was clear: missing lessons could only ever have a detrimental impact, particularly when a student was engaging in multiple interventions during school hours, but this impact was offset by the social-emotional and behavioural benefits they enjoyed.

> *"It's just the impact on their education, being out of lessons. That is something that concerns me quite a lot because I think … it's all well and good being, like, it's only a half hour, it's only an hour, but particularly if they were a student who was on multiple interventions, they might be missing three or four hours a week … in terms of adverse effects, I think we do have to acknowledge that … the benefit of BAM is offset by the impact on their educational outcome … but having looked at those kids, I think you can see … [there are] very significant improvements. So I would argue that … it seems to be a worthwhile endeavour."* (School staff at Pine)

For the school contact at Oak, the issue was less the content missed by not attending lessons and more the risk that substituting lessons for BAM communicated to students that BAM was more important and that students should de-prioritise school. However, they reflected that this had been a more significant issue during the feasibility phase due to the greater alignment between BAM and the school that was evident in the pilot phase.

> *"I think they're talking very, very, kind of, small numbers, but … that was part of the problem with the first group that we put together is [that] I think they were … not necessarily using BAM as an excuse, but then … that then becomes not taking the lessons as seriously or coming out of lessons. Now I've got BAM, do you know what I mean? And starting to use that ... as a reason to not engage in the rest of the school rather than what it's actually designed to be, which is the complete opposite of course ... that doesn't happen with the large majority of the boys that are ... part of the programme because of the way that BAM is embedded into the school"* (School staff at Oak)

Readiness for trial

This pilot study did not explicitly set out to test the readiness of BAM for a trial in the UK. Rather, it sought to establish whether BAM has promise and to test aspects of evaluation design to inform a next-stage evaluation. The results indicate some limited evidence of promise, notably through the contribution of BAM to aspects of scholars' social-emotional development, notably empathy, although there is strong evidence that BAM was not the most important contributor to scholars' development. There is no strong evidence of adverse unintended consequences. There were challenges with outcome measurement, however (explored further in the Limitations section below), and these would need to be attended to in any future evaluation of the impact of BAM.

The findings also highlight aspects of the evaluation design that are in need of attention. These include the need to find an acceptable way to measure participation in crime and anti-social behaviours and concerns raised by MHF regarding the complexities of the consenting processes for this study. It took significant time to prepare, distribute and track these processes and repeatedly encourage families to complete them. Some at MHF felt these procedures formed an additional barrier to time usage, data collection and the capacity of the research members of the team to contribute to active organisational learning during the pilot.

An RCT of BAM – preceded by an internal pilot RCT – could be conducted, and this would arguably provide stronger evidence than could be provided here on the effectiveness of BAM in the UK. Specifically, it would allow a comparison of the progress of a cohort of young people who are offered BAM with a similar cohort of young people who are not offered BAM. As such, it would allow the calculation of the impact of BAM on primary and secondary outcomes, especially the size of any such effects on education, crime and anti-social

behaviour. However, there are important questions regarding the necessity, practicality and suitability of such a path.

There is evidence of a positive effect on education, crime and anti-social behaviour in two trials of BAM in the US. In the UK, we now have some limited evidence of BAM making a positive contribution to aspects of young people's social-emotional development, notably empathy, especially where implementation happens as intended and young people engage, and of the theory of change manifesting as intended (i.e. key theorised mechanisms are firing). Given this evidence, it might be more worthwhile to attend to some of the implementation issues identified, for example, the importance of creating a supportive environment for BAM in schools, and to test the success of any changes made. Similarly, attention is needed to what might work to mitigate the potentially harmful effects on scholars' education of missing regular lessons to attend BAM.

Given the anticipated challenge of randomising scholars within schools, owing to issues with teacher and student acceptability and the likely contamination of the control group (e.g. through contact with BAM counsellors), any trial of BAM would require a cluster randomised trial design. This would likely require a large sample of secondary schools; other definitive trials of school-based interventions typically involve 40 to 80 schools (inclusive of control schools) and depending on the level of clustering involved and the minimum detectable effect size of interest, an even larger sample size is sometimes warranted (Demack, 2019). In turn, this would require a very significant, costly and rapid scale-up of the BAM delivery support infrastructure in the UK. While this may be possible, it would place significant pressure on the MHF and potentially impair the amount and quality of delivery support offered.

Given the likelihood that BAM has a positive influence on some young people, at least when implemented well and where scholars engage, with a reasonably low risk of harmful effects, we argue that rather than conducting a trial of BAM, it would be better to invest scarce evaluation resources in trials of interventions that are genuinely innovative and for which there is real need. Future evaluations of BAM would more usefully focus on the implementation and engagement issues identified above, particularly since this study only involved three schools, and further explore what works for whom, when and why, drawing on the theory of change.

Conclusion

Table 34: Summary of pilot study findings

Research question	Finding
1. To what extent is BAM being successfully implemented, with whom, under what circumstances and why?	There is strong evidence that the implementation of BAM varied by implementation area. While adaptation and the quality of delivery were successful overall, recruitment and exposure experienced mixed results. There is strong evidence that the quality of counsellors' adherence to the manual was positive, but their ability to progress through the curriculum was limited by group scheduling, group size and scholar engagement. There is some evidence that counsellors' previous skills and experience in youth work and group work supported the delivery effort. However, there is strong evidence that some elements of 'backbone support' provided by the MHF and YG – particularly training and coaching – made an important contribution to the implementation successes observed by creating a positive learning environment for counsellors. The contributions of peer supervision, project management support and community partners to this environment were limited due to capacity and resource limitations. The extent to which this environment enabled implementation was influenced by engagement from scholars, school context and counsellor absences.
2. To what extent is implementation contributing to any reported changes in social-emotional development, for whom, under what circumstances and why?	While the trajectory of scholars' social-emotional development was positive (post-test scores were higher than pre-test scores on average), the change observed through the SDQ was not statistically significant at a Bonferroni-corrected alpha level of 0.008. The change observed through most of the domains in the HSA was not statistically significant at the Bonferroni-corrected alpha level of 0.004. However, a statistically significant positive change was observed in the Empathy domain in the HSA. There is strong evidence from the 11 case studies that BAM was not the most important contributor to scholars' development – scholars generally ranked it second or third, below family, friends, faith or football. Nonetheless, there is strong evidence from the qualitative data that BAM made an important contribution to the trajectories of the case studies, regardless of their size and direction. It did so by supporting engagement in action and reflection, which helped scholars to manage their emotions, bolstered their self-esteem and offered them catharsis. There is strong evidence that for most of the case studies, whether and how they engaged in action and reflection was influenced by their readiness and willingness to observe the three group conditions: fun, safety and challenge. Their ability to do so was shaped by wider influences in their lives, including the strength of their support network.

3. To what extent is social-emotional development contributing to any reported changes in academic performance and behaviour and the avoidance of or reduced involvement in crime/anti-social behaviour, for whom, under what circumstances and why?	There is strong evidence that, as with social-emotional development, responsible decision-making at school remained stable overall. Across attendance, suspensions and attainment, performance for most schools and scholars remained unchanged, with some exceptions; while attendance fell and external suspensions increased at Pine, internal suspensions fell and academic attainment increased at Oak. There is strong evidence from the 11 case studies that while their wider support networks helped, social-emotional development made an important contribution to responsible decision-making for most of them, albeit with differences in the extent of this contribution. This was the case for most scholars regardless of the trajectories of their decision-making over the implementation period. Their development contributed by empowering and protecting them (MResP), the extent of which depended on wider influences in their lives. This suggests that BAM's focus on social-emotional development as an intermediate outcome is a legitimate one, given evidence of BAM's contribution to that development and the contribution of that development to responsible decision-making for some scholars.
4. To what extent is implementation contributing to any reported unintended consequences, for whom, under what circumstances and why?	There is mixed evidence on whether BAM inadvertently supported scholars in engaging in irresponsible decision-making. There is strong evidence that negative labels did not develop in BAM groups. While factions were present in groups, these were not groups of friends widely regarded as badly behaved within the wider school, which may have precipitated negative labelling. There is mixed evidence on whether missing lessons had a negative impact. None of the counsellors, scholars or parents/carers interviewed shared that they felt missing lessons negatively impacted scholars' academic attainment. However, some school staff did feel this was a risk. This was particularly the case for scholars in receipt of other interventions during school time and for those in settings where school staff and counsellors faced challenges in supporting the application of learning from BAM in school.

Evaluator judgement of intervention and evaluation feasibility

The evidence presented here on the feasibility of BAM mirrors the findings from the feasibility phase. One target, mean sessions attended, was successfully achieved. Others, such as group size, level and spread of need, were nearly met. However, targets for sessions scheduled and curriculum progression were not achieved. The evaluation of areas without set targets, such as adaptation and quality, indicates effective implementation. Overall, participants responded positively, with strong evidence showing movement through core elements of the theory of change. The 11 case studies suggested that scholars were engaging in action and reflection, which supported aspects of social-emotional development through the internalisation of the core values and contributed to improved decision-making.

The success of the work undertaken by the MHF, working with YG, to adopt and adapt a complex intervention like BAM in a new context should not be underestimated. This required a very considerable effort from all involved in often very challenging circumstances, not least the Covid-19 pandemic and

associated lockdowns and aftereffects. While there are positive signs for BAM's feasibility in London and the UK more generally, due consideration must be given to the challenges that prevented MHF from meeting certain targets during the pilot phase. These include the challenges they faced with schools, particularly Birch, which were exacerbated by capacity limitations at MHF.

In response, MHF has begun a process to strengthen the staffing infrastructure underpinning its implementation support efforts so that it can withstand BAM's expansion. However, this process must be understood within the wider context surrounding the third sector, which has been directly impacted by the current cost-of-living crisis and the subsequent reductions in charitable giving. This continues to affect MHF and its ability to expand its team. Whether and how this context (and MHF's response) evolves will be crucial in determining BAM's long-term trajectory in the UK.

Interpretation

Discussion

Numerous US interventions aimed at enhancing child and youth psychosocial outcomes and educational performance and reducing crime and violence have not replicated positive effects in European contexts (Sundell et al., 2008; Skärstrand et al., 2013; Baldus et al., 2016; Humayun et al., 2017; Fonagy et al., 2018; Segrott et al., 2022). However, the findings from both the feasibility phase and the pilot phase suggest there is evidence of promise for BAM in the UK.

For Research Question 1, while challenges remain, there were several signs of implementation success during both the feasibility phase and pilot phase. As noted in the feasibility phase report, BAM retains several assets that are important in driving implementation (Damschroder et al., 2022) and effective adaptation (Movsisyan et al., 2021; Green et al., 2023a, p.77). These include significant resources; the support of two mature, long-standing third sector organisations in YG and MHF that have established and maintained an effective, collaborative and mutually beneficial relationship; 20 years of iteration and development supported by YG's culture of learning and improvement; and a well-grounded, well-tested programme theory that draws upon evidence, formal scientific theory, practitioner insight and young people's experiences. These factors place YG and MHF in good stead as they work together to address the challenges they are already facing and will continue to face as they scale BAM.

The findings from the 11 case studies regarding Research Questions 2, 3 and 4 largely reinforce the evidence of promise that emerged from Research Question 1. While little can be said about the extent of BAM's contribution, the presence of its contribution was evident: scholars, parents/carers, counsellors and school staff felt that BAM contributed to social-emotional development (most notably empathy) for most scholars to some degree, with due consideration having been given to alternative explanations, wider contextual factors and potential adverse effects.

The strength of the evidence for BAM's promise is further reinforced by the alignment observed during both the feasibility and pilot phases with BAM's theory of change. Considerable time and effort were spent capturing, consolidating, testing and refining the theory of change during the feasibility phase. These efforts bore fruit: no changes were made to the theory of change at the end of the pilot phase, mirroring the outcome of the feasibility phase.

It is largely for this reason that the theory of change and the associated methods and measures for testing it are likely to be useful to future evaluations of BAM, regardless of their aims. There is strong evidence from

both the feasibility phase and the pilot phase that when BAM works, it does so by providing a positive learning environment for counsellors, engaging scholars in action and reflection, supporting their internalisation of the core values and helping them make better decisions by empowering and protecting them. Any evaluators looking to understand why, how and for whom a future iteration of BAM is or is not working might consider studying adherence to this model as a starting point alongside the tools developed for testing it.

Limitations

The pilot phase of the evaluation had several limitations. First, standardised measures were collected by the MHF and not independently by the evaluation team. Scholars completed the measures independently and followed a considered and careful process designed by MHF, with support from the evaluation team. However, the fact that completion of measures is higher for those who complete the intervention in evaluations that rely on programme data can lead to bias (those who complete measures are more likely to have engaged and done well). We recognise the importance and value of independently collected data in evaluations, but we support (and would advocate for) the use of existing programme data to evaluate intervention feasibility and early indications of promise. More should be done by funders and researchers to leverage the use of routine programme data to evaluate interventions in this field. This is partly because the value of independent data collection needs to be balanced against the data burden on participants, and in that respect, it made sense when the evaluation started to use data routinely collected as part of service delivery, given that it mapped onto key constructs of interest. A further consideration is that the SDQ is a core measure of the YEF and is included in most YEF-funded evaluations; it did not make sense – and indeed would have been unethical – to duplicate the collection of the SDQ, since it was collected by the MHF as part of programme delivery (i.e. apart from evaluation).

That said, this needs to be balanced against the risk of this approach yielding high levels of missing data, which can make it harder to answer the research questions. This risk was realised in this evaluation, in which the proportion of participants completing the SDQ and HSA pre- and post-intervention was 45% and 33%, respectively. It would also be possible to collect SDQ data from different sources (teachers, parents/carers), which has the advantage of generating different perspectives on scholars' behaviour and emotional well-being, including in varied contexts.

Second, both MHF and YG raised concerns about the interpretation of the SDQ and HSA using the pre–post method. As part of their standard assessment procedures, YG prefers the HSA retrospective pre–post (HSA-RPP) for measuring changes in the social-emotional development of scholars. The HSA-RPP assesses social-emotional development by asking children and young people to retrospectively self-report on their perceived progress pre-programme compared to post-programme at the endpoint of the programme. While they acknowledge the increased risk of recall bias, YG's preference is largely because of concerns over response shift bias. YG and MHF felt this was a particular issue for a programme that is deliberately designed to increase self-awareness among participants. Students may be more likely to retrospectively acknowledge pre-programme weaknesses in themselves upon completion of the programme due to improvements in their ability to understand themselves. Given YFF's requirement that the SDQ be administered both pre- and

post-test, concerns about the subsequent data burden limited the use of the retrospective pre–post HSA as part of this study. Work is ongoing by the HSA developers to explore the relative benefits and weaknesses of both approaches. Future research on BAM should factor in any learning that emerges.

Third, there was no counterfactual in this study design, making it hard to interpret outcome trajectories. The study was also not powered to detect change, although we did correct for multiple comparisons (using Bonferroni) to reduce the likelihood of type 1 errors. It is possible, for instance, that little or no change in the outcomes measured quantitatively suggests that the programme does not have measurable effects. As part of the qualitative data collection, we inquired whether positive changes observed by participants might be attributed to other developments in the young person's life besides the programme. While this was not always easy for participants to assess, we contend that it is a helpful corrective to assuming uncritically that positive change is caused by the programme.

Fourth, we were unable to obtain quantitative data on the outcome of crime and anti-social behaviour. For ethical reasons, it was deemed to be inappropriate to ask young people to complete a standardised measure such as the SRDS, and owing to the low number of consents received, it was not possible to obtain MoJ data on BAM participants' involvement in crime before and during their participation BAM. Further work is needed to find an acceptable way of measuring this outcome for young people from Black and minority ethnic communities.[34]

Fifth, there were challenges associated with collecting the data on academic attainment from schools and ensuring its comparability across schools. Moreover, that data is based on teacher assessments and is therefore potentially biased, particularly in terms of gender and socio-economic disadvantage (Office of Qualifications and Examinations Regulation, 2018; Lee and Walter, 2020). That said, research in England demonstrates that teacher-assessed grades reliably predict adolescents' final GCSE grades (Rimfeld et al., 2019). Future work could make better use of the potential for data linkage with the National Pupil Database.

Future research and publications

The pilot study suggests that when the intervention is implemented well and young people engage with it, it contributes to positive changes in young people's social-emotional development. There is also evidence that this happens because the mechanisms posited in the theory of change are working as intended. We may conclude from this that BAM replicates well in the UK when careful adaptations are made to both curriculum and implementation.

However, the pilot phase leaves several questions unanswered. First, what impact does the programme have on youth participation in crime and anti-social behaviour? Further empirical investigation of this would require finding a way of measuring this that is both rigorous and acceptable to young people, especially those from Black and minority ethnic communities. It may also be possible to model the effect of BAM on these outcomes based on its known effect on social-emotional development and education.

Second, how can implementation issues be addressed successfully to further support effective implementation and the engagement of all participants? Investigating this will involve working with schools, counsellors and support organisations (MHF and YG) to co-produce and test changes to existing delivery

[34] See the YEF outcomes framework and database: https://youthendowmentfund.org.uk/outcomes/

methods, focusing on whether they improve delivery and student engagement and are acceptable to those concerned. There is likely to be a role for rapid cycle testing (or similar) to achieve this.

Third, what is the extent of the adverse effect on their education of scholars missing lessons in order to participate in BAM? The quantitative school data on this issue was inconclusive, and school staff were the only interviewee group to express a concern about this. Further mixed methods investigation of this issue would help determine whether it really is an issue or if the educational benefits of participating in BAM outweigh the potentially detrimental effect of missing lessons.

In regard to publications, a journal article focusing on the process of pre-implementation has already been published (Green et al., 2023), and we will seek to publish articles on both the evaluation results and method. There is likely to be particular interest in the evaluation field in terms of the practicality and relative merits of combining contribution analysis, realist evaluation and theory of change, as we have done here.

References

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

Baldus, C., Thomsen, M., Sack, P.-M., Bröning, S., Arnaud, N., Daubmann, A. and Thomasius, R. (2016). Evaluation of a German Version of the Strengthening Families Programme 10–14: a Randomised Controlled Trial. *The European Journal of Public Health* [online] 26(6), pp.953–959. Doi: https://doi.org/10.1093/eurpub/ckw082.

Blamey, A. and Mackenzie, M. (2007). Theories of Change and Realistic Evaluation. *Evaluation* [online] 13(4), pp.439–455. Doi: https://doi.org/10.1177/1356389007082129.

Bland, J.M. and Altman, D.G. (1995). Statistics Notes: Multiple Significance Tests: the Bonferroni Method. *British Medical Journal* [online] 310(6973), pp.170–170. Doi: https://doi.org/10.1136/bmj.310.6973.170.

Brand, S.L., Quinn, C., Pearson, M., Lennox, C., Owens, C., Kirkpatrick, T., Callaghan, L., Stirzaker, A., Michie, S., Maguire, M., Shaw, J. and Byng, R. (2018). Building Programme Theory to Develop More Adaptable and Scalable Complex Interventions: Realist Formative Process Evaluation Prior to Full Trial. *Evaluation* [online] 25(2), pp.149–170. Doi: https://doi.org/10.1177/1356389018802134.

Breuer, E., Lee, L., De Silva, M. and Lund, C. (2015). Using Theory of Change to Design and Evaluate Public Health Interventions: a Systematic Review. *Implementation Science* [online] 11(1). Doi: https://doi.org/10.1186/s13012-016-0422-6.

Burkhart, G., Axford, N., Sonthalia, S., Foxcroft, D., Faggiano, F. and De Kock, C. (2019). Why Do Flagship Evidence-based Programmes from the US Run Aground in Europe, and How Should Online Repositories of Programmes Deal With This? In: *European Society of Prevention Research.* EUSPR Conference.

Campbell, R., Goodman-Williams, R., Feeney, H. and Fehler-Cabral, G. (2018). Assessing Triangulation across Methodologies, Methods, and Stakeholder Groups: the Joys, Woes, and Politics of Interpreting Convergent and Divergent Data. *American Journal of Evaluation* [online] 41(1), pp.125–144. Doi: https://doi.org/10.1177/1098214018804195.

Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J. and Balain, S. (2007). A Conceptual Framework for Implementation Fidelity. *Implementation Science* [online] 2(1). Doi: https://doi.org/10.1186/1748-5908-2-40.

Clarke, A., Sorgenfrei, M., Mulcahy, J., Davie, P., Friedrich, C. and McBride, T. (2021) *Adolescent Mental Health: a Systematic Review on the Effectiveness of School-based Interventions*. London: Early Intervention Foundation.

Craig, P., Ruggiero, E.D., Frohlich, K.L., Mykhalovskiy, E., White, M. (2018). *Taking Account of Context in Population Health Intervention Research: Guidance for Producers, Users and Funders of Research* [online] *PubMed*. NIHR Journals Library. Available at: https://www.ncbi.nlm.nih.gov/books/NBK498645/ [Accessed 16 Nov. 2023].

Dalkin, S.M., Greenhalgh, J., Jones, D., Cunningham, B. and Lhussier, M. (2015). What's in a Mechanism? Development of a Key Concept in Realist Evaluation. *Implementation Science* [online] 10(1). Doi: https://doi.org/10.1186/s13012-015-0237-x.

Damschroder, L.J., Reardon, C.M., Widerquist, M.A.O. and Lowery, J. (2022). The Updated Consolidated Framework for Implementation Research Based on User Feedback. *Implementation Science* [online] 17(1). Doi: https://doi.org/10.1186/s13012-022-01245-0.

Delahais, T. and Toulemonde, J. (2012). Applying Contribution Analysis: Lessons from Five Years of Practice. *Evaluation* [online] 18(3), pp.281–293. Doi: https://doi.org/10.1177/1356389012450810.

Demack, S. (2019). *Does the Classroom Level Matter in the Design of Educational Trials? A Theoretical and Empirical Review*. EEF Research Paper 003. London: Education Endowment Foundation.

DfE (Department for Education) (2019) *Relationships Education, Relationships and Sex Education (RSE) and Health Education: Statutory Guidance for Governing Bodies, Proprietors, Head Teachers, Principals, Senior Leadership Teams, Teachers*. London: DfE.

Donnelly, M., Brown, C., Batlle, I. C. and Sandoval-Hernández, A. (2020) *Developing Social and Emotional Skills: Education Policy and Practice in the UK Home Nations*. London: NESTA.

Durlak, J.A., Weissberg, R.P., Dymnicki, A.B., Taylor, R.D. and Schellinger, K.B. (2011) The Impact of Enhancing Students' Social and Emotional Learning: a Meta-analysis of School-based Universal Interventions. *Child Development*, 82(1), 405-432. Doi: https://pubmed.ncbi.nlm.nih.gov/21291449/

Dybdal, L., Nielsen, S.B. and Lemire, S. (2010) Contribution Analysis Applied: Reflections on Scope and Methodology. *Canadian Journal of Program Evaluation* [online] 25(2), 29-57. Doi: https://www.researchgate.net/publication/281688452_Contribution_analysis_applied_Reflections_on_scope_and_methodology (Accessed 15/07/24)

Ebenso, B., Manzano, A., Uzochukwu, B., Etiaba, E., Huss, R., Ensor, T., Newell, J., Onwujekwe, O., Ezumah, N., Hicks, J. and Mirzoev, T. (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* [online] 73, pp.97–110. Doi: https://doi.org/10.1016/j.evalprogplan.2018.12.002.

Early Intervention Foundation and Cordis Bright (2018). *Preventing Gang and Youth Violence: a Review of Risk and Protective Factors* [online] Early Intervention Foundation. Available at: https://www.eif.org.uk/report/preventing-gang-and-youth-violence-a-review-of-risk-and-protective-factors [Accessed 16 Nov. 2023].

Evans, R., Scourfield, J. and Murphy, S. (2014). The Unintended Consequences of Targeting: Young People's Lived Experiences of Social and Emotional Learning Interventions. *British Educational Research Journal* [online] 41(3), pp.381–397. Doi: https://doi.org/10.1002/berj.3155.

Farmer, T., Robinson, K., Elliott, S.J. and Eyles, J. (2006). Developing and Implementing a Triangulation Protocol for Qualitative Health Research. *Qualitative Health Research* [online] 16(3), pp.377–394. Doi: https://doi.org/10.1177/1049732305285708.

Feinstein, L. (2015) *Social and Emotional Learning: Skills for Life and Work*. London: Early Intervention Foundation, Cabinet Office and Social Mobility & Child Poverty Commission.

Fonagy, P., Butler, S., Cottrell, D., Scott, S., Pilling, S., Eisler, I., Fuggle, P., Kraam, A., Byford, S., Wason, J., Ellison, R., Simes, E., Ganguli, P., Allison, E. and Goodyer, I.M. (2018). Multisystemic Therapy versus Management as Usual in the Treatment of Adolescent Antisocial Behaviour (START): a Pragmatic, Randomised Controlled, Superiority Trial. *The Lancet Psychiatry* [online] 5(2), pp.119–133. Doi: https://doi.org/10.1016/s2215-0366(18)30001-4.

Funnell, S. and Rogers, P. (2011). *Purposeful Program Theory: Effective Use of Theories of Change and Logic Models* [online] *Purposeful Program Theory* Available at: https://www.wiley.com/en-gb/Purposeful+Program+Theory%3A+Effective+Use+of+Theories+of+Change+and+Logic+Models-p-9780470478578 [Accessed 16 Nov. 2023].

Gale, N.K., Heath, G., Cameron, E., Rashid, S. and Redwood, S. (2013). Using the Framework Method for the Analysis of Qualitative Data in Multi-disciplinary Health Research. *BMC Medical Research Methodology* [online] 13(1), pp.1–8. Doi: https://doi.org/10.1186/1471-2288-13-117.

Gedikoglu, M. (2021) *Social and Emotional Learning: an Evidence Review and Synthesis of Key Issues*. London: Education Policy Institute.

Goodman, R. (1997). The Strengths and Difficulties Questionnaire: a Research Note. *Journal of Child Psychology and Psychiatry* [online] 38(5), pp.581–586. Doi: https://doi.org/10.1111/j.1469-7610.1997.tb01545.x.

Gov.uk (2019). *Charity Chosen to Deliver £200m Youth Endowment Fund to Tackle Violence* [online] GOV.UK. Available at: https://www.gov.uk/government/news/charity-chosen-to-deliver-200m-youth-endowment-fund-to-tackle-violence [Accessed 16 Nov. 2023].

Gov.uk (2023). *Special Educational Needs in England* [online] GOV.UK. Available at: https://explore-education-statistics.service.gov.uk/find-statistics/special-educational-needs-in-england [Accessed 11 Mar. 2024].

Granada, A., Hallgarten, J. and Hasset, A. (2022) *Catalysing Social and Emotional Learning in Schools in England: a Policy and Practice Review*. London: The Centre for Education and Youth.

Green, F., Axford, N., Harris, J., Preece, C., Mannes, J., Allen, K., Callaghan, L., Berry, V., Santana de Lima, E. and Woodburn, A. (2023a) *Becoming a Man (BAM): Feasibility Study Report* [online] London: Youth Endowment Fund. Available from: https://arc-swp.nihr.ac.uk/publications/becoming-a-man-bam-feasibility-study-report/ (Accessed 15/07/24)

Green, F., Axford, N., Eastmond, N., Berry, V., Mannes, J., Allen, K., Callaghan, L. and Hobbs, T. (2023b). Transporting an evidence-based Program to a New Country: A Narrative Description and Analysis of Pre-

implementation Adaptation. *Journal of Prevention* [online] 44(6). Doi: https://doi.org/10.1007/s10935-023-00742-2.

Greenhalgh, T., Pawson, R., Wong, G., Westhorp, G., Greenhalgh, J., Manzano, A. and Jagosh, J. (2017). *Retroduction in Realist Evaluation*. *The RAMESES II Project* [online]. Available at: https://www.ramesesproject.org/media/RAMESES_II_Retroduction.pdf (Accessed 15/7/24).

Hein, I. M., De Vries, M. C., Troost, P. W., Meynen, G., Van Goudoever, J. B. and Lindauer, R. J. (2015). Informed Consent Instead of Assent is Appropriate in Children From the Age of Twelve: Policy Implications of new Findings on Children's Competence to Consent to Clinical Research. *BMC Medical Ethics*, *16*(1), p.76. Doi: https://doi.org/10.1186/s12910-015-0067-z.

Heller, S., Pollack, H.A., Ander, R. and Ludwig, J. (2013). *Preventing Youth Violence and Dropout: a Randomized Field Experiment* [online] National Bureau of Economic Research Working Paper Series. Available at: https://www.nber.org/papers/w19014 [Accessed 16 Nov. 2023].

Heller, S.B., Shah, A.K., Guryan, J., Ludwig, J., Mullainathan, S. and Pollack, H.A. (2015). *Thinking, Fast and Slow? Some Field Experiments to Reduce Crime and Dropout in Chicago* [online] National Bureau of Economic Research. Available at: https://www.nber.org/papers/w21178 [Accessed 16 Nov. 2023].

Hendriks, A.M., Bartels, M., Colins, O.F. and Finkenauer, C. (2018). Childhood Aggression: a Synthesis of Reviews and Meta-analyses to Reveal Patterns and Opportunities for Prevention and Intervention Strategies. *Neuroscience & Biobehavioral Reviews* [online] 91, pp.278–291. Doi: https://doi.org/10.1016/j.neubiorev.2018.03.021.

Humayun, S., Herlitz, L., Chesnokov, M., Doolan, M., Landau, S. and Scott, S. (2017). Randomized Controlled Trial of Functional Family Therapy for Offending and Antisocial Behavior in UK Youth. *Journal of Child Psychology and Psychiatry* [online] 58(9), pp.1023–1032. Doi: https://doi.org/10.1111/jcpp.12743.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R. and Kerr, K. (2016). Implementation and Process Evaluations (IPE) for Interventions in Education settings: an Introductory Handbook. *The University of Manchester* [online] Available at: https://research.manchester.ac.uk/en/publications/implementation-and-process-evaluations-ipe-for-interventions-in-e [Accessed 16 Nov. 2023].

Jagosh, J. (2020). Retroductive Theorizing in Pawson and Tilley's Applied Scientific Realism. *Journal of Critical Realism* [online] 19(2), pp.121–130. Doi: https://doi.org/10.1080/14767430.2020.1723301.

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.

Lammy, D. (2017). *The Lammy Review: an Independent Review into the Treatment of, and Outcomes for, Black, Asian and Minority Ethnic Individuals in the Criminal Justice System* [online] *GOV.UK*. Available at:

https://assets.publishing.service.gov.uk/media/5a82009040f0b62305b91f49/lammy-review-final-report.pdf [Accessed 16 Nov. 2023].

Lee, M.W. and Walter, M. (2020). *Equality Impact Assessment: Literature Review* (From Ofqual's strategy Risk and Research Directorate /20/6611). Ofqual. Doi: https://assets.publishing.service.gov.uk/media/5e971f1de90e071a145ec51f/Equality_impact_assessment_literature_review_15_April_2020.pdf

Lemire, S.T., Nielsen, S.B. and Dybdal, L. (2012). Making Contribution Analysis Work: a Practical Framework for Handling Influencing Factors and Alternative Explanations. *Evaluation* [online] 18(3), pp.294–309. Doi: https://doi.org/10.1177/1356389012450654.

Macedo, T., 2022. *Hot Air Balloon.* Hyperbox Island Toolbox [online] Available at: https://toolbox.hyperisland.com/hot-air-balloon (Accessed 15/07/24)

Manzano, A. (2016). The Craft of Interviewing in Realist Evaluation. *Evaluation* [online] 22(3), pp.342–360. Doi: https://doi.org/10.1177/1356389016638615.ht.

Marchal, B., Kegels, G. and Belle, S.V. (2018). *Doing Realist Research* [online] SAGE Publications Ltd. Available at: https://methods.sagepub.com/book/doing-realist-research/i349.xml [Accessed 16 Nov. 2023].

Maxwell, J.A. (2010). Using Numbers in Qualitative Research. *Qualitative Inquiry* [online] 16(6), pp.475–482. Doi: https://doi.org/10.1177/1077800410364740.

Mayne, J. (2001). Addressing Attribution Through Contribution Analysis: Using Performance Measures Sensibly. *Canadian Journal of Program Evaluation* [online] 16(1), pp.1–24. Doi: https://doi.org/10.3138/cjpe.016.001.

Mayne, J. (2012). Contribution Analysis: Coming of age? *Evaluation* [online] 18(3), pp.270–280. Doi: https://doi.org/10.1177/1356389012451663.

Miliauskas, C.R., Faus, D.P., da Cruz, V.L., do Nascimento Vallaperde, J.G.R., Junger, W. and Lopes, C.S. (2022). Community Violence and Internalizing Mental Health Symptoms in Adolescents: a Systematic review. *BMC Psychiatry* [online] 22(1). doi:https://doi.org/10.1186/s12888-022-03873-8.

Movsisyan, A., Arnold, L., Copeland, L., Evans, R., Littlecott, H., Moore, G., O'Cathain, A., Pfadenhauer, L., Segrott, J. and Rehfuess, E. (2021). Adapting Evidence-informed Population Health Interventions for New Contexts: a Scoping Review of Current Practice. *Health Research Policy and Systems* [online] 19(1). Doi: https://doi.org/10.1186/s12961-020-00668-9.

O'Connor, R. and Waddell, S. (2015). *What Works to Prevent Gang involvement, Youth Violence and Crime: a Rapid Review of Interventions Delivered in the UK and Abroad* [online] Early Intervention Foundation. Available at: https://www.eif.org.uk/report/what-works-to-prevent-gang-involvement-youth-violence-and-crime-a-rapid-review-of-interventions-delivered-in-the-uk-and-abroad [Accessed 17 Nov. 2023].

Office of Qualifications and Examinations Regulation (2018, March). *GCSE 9 to 1 Grades: a Brief Guide for Parents* [online] The Ofqual Blog. Available at: https://ofqual.blog.gov.uk/2018/03/02/gcse-9-to-1-grades-a-brief-guide-for-parents/. (Accessed 15/07/24)

Orpinas P. (1993). *Skills Training and Social Influences for Violence Prevention in Middle Schools. A Curriculum Evaluation*. Doctoral Dissertation. Houston, TX: University of Texas Health Science Center at Houston, School of Public Health.

Pawson, R. and Tilley, N. (1997). *Realistic Evaluation*. London: SAGE Publications Ltd.

Pawson, R. and Manzano-Santaella, A. (2012). A Realist Diagnostic Workshop. *Evaluation* [online] 18(2), pp.176–191. Doi: https://doi.org/10.1177/1356389012440912.

Rimfeld, K., Malanchini, M., Hannigan, L. J., Dale, P. S., Allen, R., Hart, S. A. and Plomin, R. (2019). Teacher Assessments During Compulsory Education are as Reliable, Stable and Heritable as Standardized Test Scores. *Journal of Child Psychology and Psychiatry*, *60*(12), 1278-1288. Doi: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6848749/

Ritchie, J. and Spencer, L. (1994). Qualitative Data Analysis for Applied Policy Research. In: *Analyzing* Qualitative Data. London: Routledge.

Rolfe, S. (2019). Combining Theories of Change and Realist Evaluation in Practice: Lessons from a Research on Evaluation Study. *Evaluation* [online] 25(3), p.135638901983522. Doi: https://doi.org/10.1177/1356389019835229.\.

Sandelowski, M. (2001). Real Qualitative Researchers Do Not Count: the Use of Numbers in Qualitative Research. *Research in Nursing & Health* [online] 24(3), pp.230–240. Doi: https://doi.org/10.1002/nur.1025.

Segrott, J., Gillespie, D., Lau, M., Holliday, J., Murphy, S., Foxcroft, D., Hood, K., Scourfield, J., Phillips, C., Roberts, Z., Rothwell, H., Hurlow, C. and Moore, L. (2022). Effectiveness of the Strengthening Families Programme in the UK at Preventing Substance Misuse in 10–14-year-olds: a Pragmatic Randomised Controlled Trial. *BMJ Open* [online] 12(2), p.e049647. Doi: https://doi.org/10.1136/bmjopen-2021-049647.

Skärstrand, E., Sundell, K. and Andreasson, S. (2013). Evaluation of a Swedish Version of the Strengthening Families Programme. *The European Journal of Public Health* [online] 24(4), pp.578–584. Doi: https://doi.org/10.1093/eurpub/ckt146.

Skivington, K., Matthews, L., Simpson, S.A., Craig, P., Baird, J., Blazeby, J.M., Boyd, K.A., Craig, N., French, D.P., McIntosh, E., Petticrew, M., Rycroft-Malone, J., White, M. and Moore, L. (2021). A New Framework for Developing and Evaluating Complex Interventions: Update of Medical Research Council Guidance. *BMJ* [online] 374(1), p.n2061. Doi: https://doi.org/10.1136/bmj.n2061.

Smith, D.J. and McVie, S. (2003). Theory and Method in the Edinburgh Study of Youth Transitions and Crime. *British Journal of Criminology* [online] 43(1), pp.169–195. Doi: https://doi.org/10.1093/bjc/43.1.169.

Sundell, K., Hansson, K., Löfholm, C.A., Olsson, T., Gustle, L.-H. and Kadesjö, C. (2008). The Transportability of Multisystemic Therapy to Sweden: Short-term Results from a Randomized Trial of Conduct-disordered Youths. *Journal of Family Psychology* [online] 22(4), pp.550–560. Doi: https://doi.org/10.1037/a0012790.

Taylor, R.D., Oberle, E., Durlak, J.A., and Weissberg, R.P. (2017) Promoting Positive Youth Development Through School-based Social and Emotional Learning Interventions: a Meta-analysis of Follow-up Effects. *Child Development*, 88(4), 1156-1171. Doi: https://artsedsel.org/wp-content/uploads/Promoting-Positive-Youth-Development-Through-School-based-Social-and-Emotional-Learning-Interventions_-A-Meta-analysis-of-Follow-up-Effects.pdf

Wong, G., Westhorp, G., Manzano, A., Greenhalgh, J., Jagosh, J. and Greenhalgh, T. (2016). RAMESES II Reporting Standards for Realist Evaluations. *BMC Medicine* [online] 14(1). Doi: https://doi.org/10.1186/s12916-016-0643-1.

Wright, A.W., Austin, M., Booth, C. and Kliewer, W. (2016). Exposure to Community Violence and Physical Health Outcomes in Youth: a Systematic Review. *Journal of Pediatric Psychology* [online] 42(4), p.jsw088. Doi: https://doi.org/10.1093/jpepsy/jsw088.

Yalom, I.D. and Leszcz, M. (2005). *The Theory and Practice of Group Psychotherapy*. 5th ed. New York: Basic Books.

Yin, R.K. (2018). *Case Study Research and Applications: Design and Methods*. 6th ed. Thousand Oaks, California: Sage Publications.

Youth Endowment Fund (2023). *Children, Violence and Vulnerability: the Second Annual Youth Endowment Fund Report into Young People's Experience of Violence*. [online] Youth Endowment Fund. Available at: https://youthendowmentfund.org.uk/wp-content/uploads/2023/11/YEF_Children_violence_and_vulnerability_2023_FINAL.pdf [Accessed 19 December 2023].

Youth Endowment Fund (2022). *Pilot Study Guidance*. [online] Youth Endowment Fund. Available at: https://youthendowmentfund.org.uk/wp-content/uploads/2022/03/3.-YEF-Evaluations-Guidance-Pilot-studies-March-2022.pdf [Accessed 21 February 2024].

Youth*in*mind (2016). *Scoring the Strengths & Difficulties Questionnaire for Age 4-17 or 18+* [online] SDQInfo. Available at: https://www.sdqinfo.org/py/sdqinfo/c0.py [Accessed 30 Nov. 2023].