EVALUATION REPORT

Face It

Pilot report

Alice Worsley, Lilli Wagstaff, Tess Moseley-Roberts, Jack Martin, Tom McBride

October 2025

THE LAB | ENDING YOUTH VIOLENCE

YOUTH ENDOWMENT FUND

Contents

Project title	Face It Evaluation – Pilot study
Developer (Institution)	Khulisa
Evaluator (Institution)	Ending Youth Violence Lab
Principal investigator(s)	Tom McBride
Evaluation plan author(s)	Alice Worsley, Lilli Wagstaff, Tess Moseley-Roberts, Jack Martin, Tom McBride
Evaluation setting	Secondary schools in London
Target group	13–15-year-olds living in London at risk of school exclusion
Planned number of participants	Five schools, approximately 160 young people
Funding source and declaration of interests	Funding was received for this work from the Youth Endowment Fund and from philanthropist Stuart Roden. There are no known conflicts of interest associated with this publication, and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgements

We are grateful to colleagues from Khulisa (including Iman Haji, Derellanne Knowles and Porsha Robinson) for supporting the evaluation and the delivery underpinning it.

We are also grateful to colleagues from the Behavioural Insights Team (particularly Dr Patrick Taylor and Leonie Nicks), the Youth Endowment Fund (Dr Mollie Bourne, Dr Daniel Acquah, Dr Amy Wells and Fardawza Ahmed) and Dr Nick Axford (University of Plymouth) for their helpful advice and feedback.

1. Executive Summary

The project

Face It is a school-based social and emotional skills development programme designed for children aged 11-16 at risk of offending, exploitation and school exclusion. It aims to develop a range of social and emotional skills, including emotional regulation and self-awareness, and encourages pupils to reflect on the root causes of and triggers for their disruptive or challenging behaviour. The intervention is delivered by Khulisa, a youth mental health charity, over six weeks. Participants in this project were aged 13-15 (in Years 9 and 10) and were referred by teachers due to displaying externalising behaviours and elevated risk of involvement in violence. The six-week programme includes an initial one-to-one session and group session, where facilitators aim to build a sense of safety and empower children to decide whether they want to proceed. This is followed by an intensive five-day programme, where facilitators use a variety of activities (including art and storytelling) to identify the children's habits and triggers, practise new skills and build their self-esteem. The programme ends with another group session and a final one-to-one session to reflect on the child's progress. Each programme is tailored to the participant's needs and delivered by therapeutically trained facilitators (typically in art or drama therapy), who also receive dedicated training and ongoing clinical supervision.

The YEF and Stuart Roden funded the Ending Youth Violence Lab to conduct a pilot evaluation of Face It. It aimed to assess the deliverability, acceptability and evaluability of the programme, in addition to exploring whether it showed any early promise of positive outcomes for children. The evaluation piloted a two-armed randomised controlled trial (RCT). The evaluation aimed to recruit 160 children across five schools but was ultimately only able to recruit two schools. Although 98 children were initially referred, only 67 children ultimately participated from the two schools. Of these, 36 were randomised to receive the programme, and 31 received support as usual. Among participants, 62% were from a Black background, 17% from a Mixed/Multiple ethnicity background and the remaining were White, Asian or from other minority groups (numbers too small to report for anonymity reasons). Around 46% of the total participants received free school meals.

The evaluation collected quantitative data, including programme data on participation, facilitator fidelity reports, satisfaction surveys and a range of outcome measures, including the self-reported Strengths and Difficulties Questionnaire and the Self-Report Delinquency Scale, as well as parent/caregiver survey, attendance and exclusion data. Outcome measures were collected at three points: at baseline, at the end of the programme and three months post the end of the programme. Qualitative data were also collected through participatory groups of children from a school where Khulisa had previously delivered Face It and semi-structured interviews with 13 young people who received the programme, four control group pupils, four Khulisa facilitators, two Khulisa managers and two school staff members (one per school). The study took place between April 2023 and July 2025.

Key conclusions
There were significant difficulties in recruiting schools, with only two schools successfully recruited (40% of the target). Of the two schools recruited, one dropped out before delivery could begin with the second cohort. Therefore, only one school was retained throughout the study.
Despite recruitment challenges, Khulisa facilitators reported that activities were delivered with high fidelity (with over 75% of key activities conducted and 88% of children completing the core five-day programme).
Evaluation feasibility was mixed. Once schools were on board, it was possible to recruit children and carry out randomisation. However, attrition was high (36% post-programme, rising to 46% at three-month follow-up), some evaluation methods (e.g. self-report surveys) were unpopular and parent/caregiver engagement was very limited.

The programme was found to be acceptable and valued by children and school staff. The small number of children and staff who were interviewed and completed surveys perceived that Face It could improve emotional regulation, behaviour, social skills and well-being.
Given the challenges faced by the evaluation (particularly in recruiting schools), Face It is not yet ready for a full-scale RCT. An RCT may be feasible if fundamental changes are made to the way schools are recruited to the programme.

Interpretation

The project aimed to recruit five schools (with two cohorts each). Despite a broad recruitment strategy – including cold outreach to 60 schools, newsletters and meetings with senior leaders – only two schools were recruited (40% of the target). One of these schools later withdrew before its second cohort started the programme, while the other delivered an additional unplanned third cohort, enabling Face It to run with four cohorts in total. Schools that declined cited concerns about the survey's sensitive questions, the programme's intensity and duration, challenges in generating referrals, and issues with data sharing and opt-out consent.

Qualitative insights from Khulisa facilitators found that Face It was delivered largely as intended, with a high level of fidelity. Facilitators reported completing the majority of key activities, and adaptations made during delivery, such as tailored activities to the age or response of the group, were seen as consistent with the programme manual. This flexibility was considered important for engaging children and building trust, as reflected in the high overall level of programme engagement (88%). Participating children generally felt positive about the programme. Quantitative and qualitative insights suggested that facilitator relationships mattered more to children than the specific activities, with some even suggesting that the programme was too short. School staff were also positive, highlighting a safe, calm space. Although they felt the weeklong timetable was challenging, the prevailing view was that the benefits outweighed disruptions. Most concerns raised by school staff were related to the evaluation rather than the programme.

There were challenges in retaining children in the evaluation. Of the 99 children referred, 69 consented and completed the baseline, 67 were randomised and 43 completed the post-test (36% attrition, rising to 46% at the three-month follow-up). Parental responses were also minimal. Randomisation was workable, but some evaluation features, such as sensitive survey items, referral burden and group survey sessions, raised concerns. Overall, an RCT could be feasible, but the recruitment of schools and retention are key barriers. Future trials should extend data collection windows and strengthen pre-evaluation communications.

Given that the pilot trial was not designed to be large enough to detect impact, we rely on qualitative insights to assess the perceived benefits of Face It. This qualitative research generated positive views, with children describing improvements in their emotional regulation, behaviour, relationships, confidence and well-being; however, perceived impacts on attendance and attitude to school were mixed. School staff, while cautious about whether the programme was directly responsible, reported happier pupils, fewer fights and stronger relationships with pupils. Teachers cited trusted facilitators, a safe and calm space, engaging activities and mixed groups that supported peer learning as factors that led to these impacts.

Given the challenges faced in this pilot trial, the YEF has no current plans to proceed with further evaluation. The evaluator judges that Face It is not ready for a full-scale trial, as recruiting a sufficiently large number of schools to enable a well-powered efficacy trial in this context would not currently be possible within any reasonable recruitment window. Fundamental changes would be required before a robust evaluation could be undertaken. These changes include casting the school recruitment net wider from the outset, allowing for longer lead-in times for recruitment, shortening the programme to ease school concerns and offering the programme to a younger cohort (as schools may be more amenable to younger children missing lessons).

2. Study rationale and background

2.1. About the Ending Youth Violence Lab

The Ending Youth Violence Lab ('the Lab') was founded in the summer of 2022, bringing together expertise in intervention, evaluation and youth violence. It was funded by Stuart Roden and the Youth Endowment Fund (YEF), and it is being incubated at the Behavioural Insights Team.

The Lab's mission is to catalyse a step change in understanding and tackling violence. To do this, we do three things: Firstly, we identify promising projects which seek to address youth violence. Secondly, we fund the development and delivery of those projects. Thirdly, we conduct research to assess how delivery has gone, how to make it even better and what the potential is for further evaluation (with a focus on early-stage testing to support the work of the YEF).

We prioritise three strands of activity:

1. **Supporting the importation, adaptation and testing of well-evidenced interventions from overseas:** We will identify approaches with strong evidence of improving youth violence outcomes or related upstream factors in other countries, adapt these to the UK context and deliver early-stage testing.
2. **Working with UK organisations to develop strong ideas into evaluable interventions:** We will work with the sector to find interventions that have strong theoretical underpinnings and are committed to rigorous evaluation and we oversee the development and early-stage testing needed to get them trial-ready.
3. **Working with developers, researchers, practitioners and service users to co-design new and innovative approaches:** We will build partnerships and fund the development of novel approaches to tackling youth violence, with a focus on addressing underserved populations and unmet needs.

The project described in this report forms part of Strand 2 of the Lab's approach.

2.2. Rationale for the evaluation

We identified Face It as a promising programme that would benefit from a pilot evaluation (to establish whether it could progress to a full-scale randomised controlled trial (RCT) on the basis that:

- **Face It's intended outcomes are relevant to the Lab's mission:** The Lab is focused on developing promising programmes to support reductions in youth crime and violence. In the short term, Face It aims to develop young people's emotional self-regulation, coping skills and resilience. There is strong evidence that social and emotional skills are important for the positive development of children and young people and contribute to a range of important long-term educational, economic, health, social and criminal justice outcomes.[1] [2]
- **Face It belongs to a promising area of intervention:** There is good evidence that well-designed school-based social and emotional programmes can be effective, that they are being successfully implemented in UK schools and that they can have a positive impact on students' social and emotional competencies and educational outcomes.[3]

[1] Goodman, A., Joshi, H., Nasim, B., & Tyler, C. (2015). Social and emotional skills in childhood and their long-term effects on adult life.
[2] OECD. (2017). Social and emotional skills: Well-being, connectedness and success.
[3] Clarke, A. M., Morreale, S., Field, C. A., Hussein, Y., & Barry, M. M. (2015). What works in enhancing social and emotional skills development during childhood and adolescence. A review of the evidence on the effectiveness of school-based and out-of-school programmes in the UK. A report produced by the World Health Organization Collaborating Centre for Health Promotion Research, National University of Ireland Galway.

- **Khulisa has shown organisational commitment to developing an evidence base and has generated promising initial evidence for Face It:** Face It indicates early promise as an intervention to improve social and emotional skills. The intervention has demonstrated that it can recruit and retain participants, and qualitative work indicates that the programme is well regarded by participants. An internal evaluation of the programme conducted by Nesta indicated that Khulisa's theory of change highlights relevant outcomes, and both quantitative and qualitative insights suggested largely positive changes in the outcomes examined.[4] A quasi-experimental study exploring outcomes for nine cohorts of Face It participants from 2020/21 to 2021/22 found increased pupil well-being, resilience and emotional regulation outcomes over the course of the programme.[5] These studies collectively show that the intervention shows promise.
- **Face It is in its early stages in terms of evidence and evaluation, and a pilot trial would add value:** There is not yet a robust RCT demonstrating that it is effective at improving outcomes for young people. Before any future full-scale RCT, it is important that a small-scale pilot trial is conducted to support and inform this work – to test and improve evaluation procedures, such as randomisation and data collection, and to generate useful information around sample size determination.

[4] Barber, A. (2020). An internal evaluation of Khulisa's 'Face It' well-being programme in schools. A report produced by Khulisa. Available at: https://www.khulisa.co.uk/wp-content/uploads/images/NESTA-Evaluation-Report-2020.pdf.
[5] ImpactEd. (2024). Khulisa: Face It programme impact evaluation report (Final). A report produced by ImpactEd.

3. The intervention

3.1. Intervention overview

Face It is a school-based intervention designed and implemented by Khulisa. Face It is specifically designed for young people at risk of offending, exploitation and school exclusion. The programme aims to build self-awareness and encourages pupils to reflect on the root causes of and triggers for their disruptive or challenging behaviour. Khulisa believes that early intervention breaks the school-to-prison pipeline, which is exacerbated by exclusion, enabling young people to choose a safe and crime-free future.

The intervention is delivered over six weeks, including an intensive five-day programme of activities, a pre-programme and post-programme group, and one-to-one sessions (delivered prior to and after the five-day programme). Each programme is tailored to participants' needs and uses art, storytelling, and one-to-one and group experiential techniques delivered by trained drama therapists. Programme facilitators are assigned to groups of young people, ensuring that at least one person on each facilitation team has a therapeutic qualification. This is typically a drama therapy qualification, but it may also be relational therapy. A detailed description of Face It using the TIDieR (template for intervention description and replication) framework can be found in Annex A.

Face It aims to help young people learn new coping mechanisms and skills to identify individual triggers (e.g. anger or conflict) and alternative responses that can help improve capacity for concentration and focus during lessons. The programme has a focus on developing emotional regulation skills (planning, decision making and communication), which can have a positive impact on behaviour, sense of self-worth and confidence.

3.2. Theory of change

Figure 1 details a high-level theory of change for the intervention. We have conducted an initial review of the theory of change, discussed this in depth with Khulisa and concluded that the outcomes it specifies and the underlying pathways to achieve them are sufficiently well considered and plausible to justify further evaluation.

Please see Khulisa's most recent version of its theory of change in Annex B.

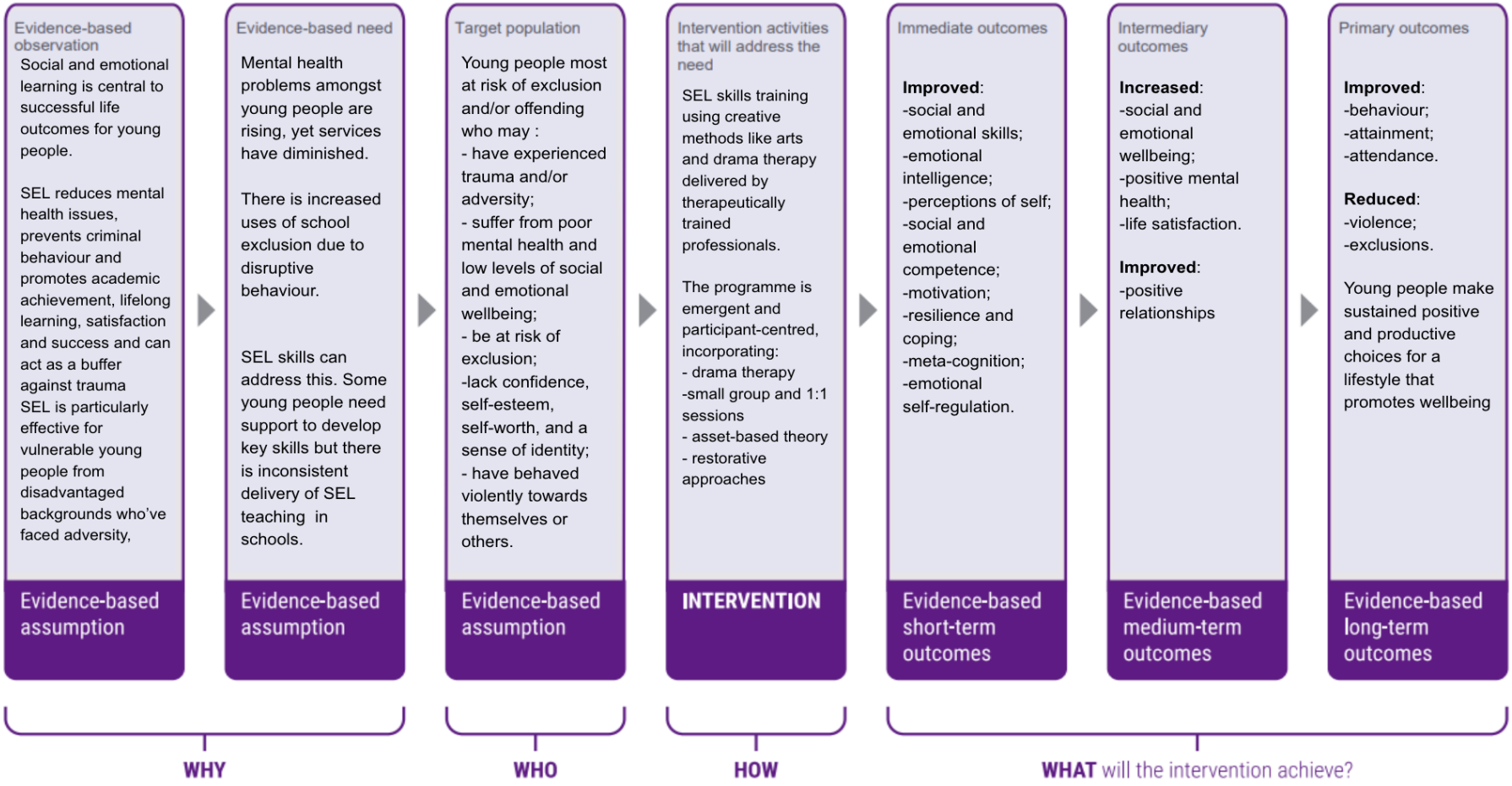


Figure 1 – Face It theory of change

3.3. How the intervention compares to other services and business-as-usual

Face It can be distinguished from many other school-based social and emotional programmes for similar cohorts on the basis that it is delivered by therapeutically qualified practitioners and combines one-to-one sessions with group sessions. In particular:

1. **It takes a therapeutic and trauma-informed approach**: Khulisa's programmes are led by therapists and follow trauma-informed methodologies, including Perry's neurosequential model.[6] [7]
2. **Its activities are creative and experiential:** It is theorised that children with limited emotional literacy respond more effectively to non-verbal activities. Khulisa uses art, drama therapy, role play and games to help participants observe themselves through the experiences of others, which is intended to help them understand the behaviours that benefit or hold them back.
3. **The programme combines group and one-to-one support to embed learning:** Face It uses a multilayered approach to learning. One-to-one sessions give facilitators an opportunity to gain insight into participants' levels of self-awareness and shed light on how to help these participants build confidence, self-worth, and more positive and healthy beliefs. Group sessions aim to enable participants to build relationships, develop empathy and learn to respect other perspectives through group discussion and other activities.

[6] Perry, B. D. (2006). 'Applying principles of neurodevelopment to clinical work with maltreated and traumatized children: the Neurosequential Model of Therapeutics' in Webb, N. B. (Eds.) *Traumatized Youth in Child Welfare*, Guildford Press, New York, 27–51.

[7] Perry, B. D. (2009). Examining child maltreatment through a neurodevelopmental lens: clinical applications of the Neurosequential Model of Therapeutics, *Journal of Loss and Trauma*, 14, 240–255.

4. Research questions

The pilot trial was designed to test the following questions:

- **Deliverability:** Can Face It recruit and retain sufficient numbers of young people from the target population and be delivered with fidelity?
- **Acceptability:** Do young people value the programme and feel that the content and delivery are appropriate to their needs? Do teachers and caregivers find programme delivery acceptable?
- **Evaluability:** Do we have enough confidence in the feasibility of an RCT, particularly in terms of recruitment into evaluation, randomisation and outcome data collection, to justify a future full-scale efficacy trial? Is it possible to collect data which would allow for the testing of mediators, moderators and differential impact as part of any subsequent evaluation?
- **Outcomes:** Does Face It show sufficient promise in terms of improvements in key outcomes (resilience, emotional regulation, social and emotional well-being, school attendance, and school exclusions) to justify a subsequent efficacy trial?[8]

More details on how we addressed these questions are provided in *Table 1*.

Table 1: Research questions

	Overarching question	Sub-question
Feasibility	Can Face It recruit and retain sufficient numbers of young people from the target population and be delivered with fidelity?	**Reach and recruitment** ● Can Khulisa recruit a sufficient number of schools to participate in the programme? ● Can Khulisa recruit and retain a sufficient number of young people at risk of youth violence? ● What factors affect recruitment and retention? ● Do recruitment and retention vary by ethnicity, gender or other baseline characteristics? ● How should the programme be adapted to improve recruitment and retention? **Fidelity** ● Can Face It be delivered with fidelity? ● What factors affect fidelity? ● What variations in delivery are appropriate for effective implementation?
Acceptability	Do young people value the programme and feel the content and delivery are appropriate to their needs? Do teachers and caregivers find programme delivery acceptable?	● Is Face It acceptable to young people at risk of youth violence? ● Does participation in a randomised evaluation affect participants' views of the intervention? If so, how? ● Does acceptability vary by ethnicity and race? ● Is the programme acceptable to schools? ● What factors affect acceptability?

[8] Note that the sample size in this pilot was considered unlikely to be big enough to make strong causal claims about effectiveness. However, our aim was to produce descriptive statistics on key outcome variables and make an estimate of effectiveness with low levels of confidence.

		• How could the programme be adapted to increase acceptability?

Evaluability	Do we have enough confidence in the feasibility of a randomised controlled trial (RCT), particularly in terms of recruitment into evaluation, randomisation and outcome data collection, to justify a continuation of the efficacy trial?	• Is randomisation feasible and adhered to?
		• Is data collection possible and sufficient?
		• What factors affect the ease of data collection and the completeness and quality of the data?
		• How could the approach to data collection be improved to increase ease, completeness and quality?
		• Is it possible to retain schools and young people in the control group?
		• What factors affect retention in the control group? What are the best methods of achieving retention in the control group?
		• Do practitioners or young people feel that the evaluation impacts intervention delivery? If so, what elements of the evaluation affect delivery quality?
		• Is randomisation acceptable to practitioners and teachers?
		• What factors affect acceptability?
		• How could the approach to randomisation be adapted to increase acceptability?
		• Is it possible to gather data to examine which part of Face It leads to improved outcomes?
		• Is it possible to gather data and permissions so that a future RCT could assess whether outcomes vary by:
		  ○ School type or location
		  ○ Gender
		  ○ Age
		  ○ Socioeconomic status/free school meal status.
		  ○ Race and ethnicity

Outcomes	Does Face It show sufficient promise in terms of improvements in key outcomes to justify a subsequent efficacy trial?	• Is there preliminary evidence that Face It improves:
		  ○ Resilience
		  ○ Emotional regulation
		  ○ Social and emotional well-being
		  ○ Behavioural difficulties
		  ○ Attendance
		  ○ Exclusions
		  ○ Self-reported delinquency
		• Will it be possible to link to offending data via the Metropolitan Police?

5. Design and methodology

5.1. Design

This pilot trial was a parallel two-armed superiority RCT of Face It compared to business-as-usual services for young people at risk of exploitation, offending and school exclusion.

5.2. Process

For all participating schools, we aimed to deliver the programme twice – once to one group of Year 9s (typically 13- to 14-year-olds) and once to one group of Year 10s (typically 14- to 15-year-olds). For each school and cohort, the following process was followed:

1) **School recruitment:** Khulisa reached out to schools to explain the project and the offer of participation. Interested schools were then encouraged to sign a service-level agreement, and head teachers or other members of the senior leadership team then provided written consent for pupils within the school to participate in the evaluation and receive the programme.
2) **Young person referral and eligibility assessment:** Referrals were made within schools, typically by teachers with awareness of pupils' behaviour, such as a head of year or form tutor, based on recruitment criteria and guidance shared with schools by Khulisa. School staff were prompted to provide details about young people to Khulisa via an online form. The form accepted submissions for young people with the correct eligibility characteristics and rejected submissions from those who were ineligible. We attempted to recruit an overall sample with an approximately 80:20 split of participants displaying primarily externalising and internalising behaviours, respectively, as Khulisa views this balance as an essential component of building an optimal group dynamic for the programme. We also attempted to recruit an overall sample that was well balanced on gender.
3) **Parental (opt-out) consent:** Parents/caregivers of young people were given an information sheet explaining the evaluation purpose and process, and they were given the opportunity to withdraw their child from the evaluation.
4) **Young person consent and baseline data collection:** Eligible young people were invited to a group data collection session held in their schools. At these sessions, Lab researchers talked young people through the project and the information sheet. Young people were given an opportunity to provide their consent to be involved in the project, and they were then asked to complete the baseline data collection survey on a school computer. Lab researchers were present to answer any questions the young people may have had as they completed the survey. Young people who did not consent to participation in the evaluation or did not provide baseline data were not included in the project and were not randomised.
5) **Randomisation:** Randomisation occurred in batches for each school after students were referred to the programme and assessed for eligibility and after consent was obtained and baseline assessment had taken place. Khulisa and the schools were then informed which young people were in the treatment group and the control group. Control group participants were not permitted to receive Face It, but they could be referred to pre-existing 'business-as-usual' programmes and services available within the school (treatment group children could also still access these other services).
6) **Post-test data collection:** After the treatment group completed the Face It programme, an equivalent group data collection session was held in the young people's schools. They were asked to complete the same survey again using school computers. We attempted to collect outcome data from all young people, regardless of whether they were randomly assigned to the treatment or control groups and regardless of how much of the programme they had completed (an intention-to-treat design).

7) **Follow-up data collection:** Three months after the post-test survey, we attempted to collect the same set of outcome data. Instead of collecting this in Lab-facilitated data collection sessions held within the schools, we asked the schools to run and facilitate these sessions, providing guidance on how to do this. This was partly done to test the feasibility and acceptability of this alternative approach to data collection (which, if successful, could be a lower-cost and easier-to-scale option for future evaluators).

5.3. Randomisation and control group

- **Design:** This pilot trial was a parallel two-armed superiority RCT of Face It compared to SAU for young people at risk of exploitation, offending and school exclusion.
- **Unit of randomisation:** The pilot trial used a within-schools randomisation design, with randomisation taking place at the level of the individual student.
- **Randomisation approach:** Random assignment to the intervention or control groups was implemented by the Lab using a random number generator in Stata. Stratified randomisation was used to ensure that the overall sample ratio of young people presenting with externalising or internalising behaviours was maintained in both the treatment and control groups. In addition to this, we stratified by gender to ensure that the treatment and control groups were balanced on this characteristic and to promote mixed-gender intervention groups, which the programme developers argue improves group dynamics.
- **Control group:** Individuals allocated to the control group received SAU. We investigated what SAU included as part of post-test data collection.

5.4. Participants

5.4.1. School characteristics

We aimed to recruit schools with a high proportion of disadvantaged pupils (more pupils in receipt of the pupil premium than the national average) and that are ethnically diverse (at least 30% of pupils not of White British heritage). Participating schools were identified by Khulisa on the basis of meeting these criteria and not having used the Face It programme previously.

We initially aimed to recruit five schools – three for delivery in the autumn term and then a further two for delivery in the spring term. The target of five schools was selected because it was at the upper boundary of what Khulisa believed was feasible within the timelines of the pilot trial, it was believed to enable a better understanding of the demand for interventions such as Face It within schools and it was judged to be broadly half of the number of schools that would be required for a full-scale efficacy trial, making it an appropriate target for a smaller-scale pilot trial.

5.4.2. Children and young people's characteristics

Participants were young people aged 13–15 in Years 9 and 10 who were referred to Face It by schools and who met programme eligibility criteria.

Our aim was that participants would predominantly be pupils displaying externalising behaviours and considered at elevated risk of involvement in youth violence (proxied by risk of exclusion, history of verbal and physical violence, and/or bullying perpetration). Pupils were identified as at risk using internal school data and teachers' judgement.

5.4.3. Sample size

As this was a pilot trial, we did not aim to test the impact of Face It but rather to establish its evaluability and to generate information that would enable a full-scale impact evaluation in the future. Consequently,

we did not select our sample size to detect any particular effect size but rather based our sample size targets on what was felt to be feasible within the time frame and to enable a suitable test of these key evaluability questions.

Khulisa's view, which was based on its experience of delivering the programme, was that Face It groups ideally include between eight and 10 pupils. Given that we aimed to deliver within five schools and to two cohorts within each, this meant that we anticipated delivering to eight pupils within each cohort and 16 pupils within each school, that is, to 80 young people across all five schools. This would then require an overall sample of 160 pupils, including 80 young people in the control group.

Using pupil-level randomisation, a sample size of 160, a power of 0.8 and estimates for standard deviation on the Strengths and Difficulties Questionnaire (SDQ) from previous papers,[9] we anticipated that we would be able to detect a Cohen's D of 0.5 (or a mean difference of 3 on the SDQ).

5.5. Data collection and analysis

We used a mixed methods approach to evaluate the deliverability, acceptability and evaluability of the Face It programme. In the following section, we provide a high-level overview of the research activities we completed, as well as any deviations from the methodology plans we outlined in the study protocol. For a fuller description of the methods used, please see the relevant annexes of this report, the evaluation protocol and the statistical analysis plan.

5.5.1. Quantitative data collection

Outcome survey

We conducted surveys with young people at baseline (pre-randomisation), at the end of the programme and at three months post-programme. As discussed above, the first two data collection points were conducted via Lab-facilitated survey sessions in schools themselves, with young people using computers to enter their responses in an online survey. For the final data collection point, the online survey sessions were facilitated by schools.

We also conducted surveys with parents/caregivers of young people at the same time points (focused on children's behaviours and parents' well-being). These surveys were completed online only, with parents/caregivers being emailed a link to an online survey to complete in their own time.

Our pre-specified primary outcome was behavioural difficulties as assessed by the SDQ (the total difficulties score, self-reported, at the post-programme time point). We also measured self-reported offending, victimisation, resilience and emotional regulation using a variety of validated surveys. Overall, we focused on the subset of Face It's intended outcomes that relate to the YEF's focus on youth violence and are predictive of getting caught up in crime and violence. Please see Annex C for more details on each of the survey measures used.

[9] Two papers found that the standard deviation for the Strengths and Difficulties Questionnaire was between 5 and 7, so we ran our power calculations assuming 6.5. The papers are cited as follows: Cortina, M., & Fazel, M. (2014). The art room: an evaluation of a targeted school-based group intervention for students with emotional and behavioural difficulties. *The Arts in Psychotherapy*, 11, 10.1016/j.aip.2014.12.003; Vugteveen, J., de Bildt, A., & Timmerman, M. E. (2022). Normative data for the self-reported and parent-reported Strengths and Difficulties Questionnaire (SDQ) for ages 12–17. *Child and Adolescent Psychiatry and Mental Health* 16, 5, https://doi.org/10.1186/s13034-021-00437-8.

ImpactEd data
In addition to the outcomes specified above (measured using surveys), we aimed to use the ImpactEd platform to collect data on school attendance and exclusions. We attempted to collect these data at baseline and three months post-programme.

Satisfaction survey
As part of the outcome survey, we also included some questions aimed at gauging young people's satisfaction with receiving the programme. These questions used 5-point Likert scales, asking young people to self-report their:

- *Satisfaction with the practitioners who delivered Face It to them*
- *Satisfaction with receiving Face It in a group format alongside other young people*
- *Satisfaction with Face It's informational taught content*
- *Satisfaction with Face It's activities*
- *Perception that the programme met their needs*
- *Overall satisfaction with the programme*

Administrative data
Khulisa shared data from referral forms and session attendance lists with the Lab. These administrative data also included:

- The number of young people referred
- Complaints submitted to Khulisa and school staff

Fidelity reports
Programme facilitators self-reported the activities they conducted with young people during the five-day programme. This allowed us to calculate the proportion of the intended activities that were actually delivered to young people.

5.5.2. Quantitative data analysis

Feasibility, acceptability and evaluability analysis
The primary objective of the pilot trial was to establish the feasibility, acceptability and evaluability (i.e. the extent to which conducting a full-scale evaluation of Face It is possible) of the programme. These analyses were conducted using descriptive analyses of the quantitative data specified above.

Effectiveness analysis
We also conducted analyses on the outcomes of the pilot study using the following approach. This pilot trial was not primarily designed to estimate effect sizes or evaluate the impact of the intervention in depth. Instead, analysing our outcome data was intended to support building an understanding of whether there is sufficient evidence of impact to justify a larger and more robust efficacy trial.

All outcome data were analysed using an intention-to-treat analysis and linear (or logistic, where relevant) regressions.

We collected pre-intervention outcomes for all students to increase power and to adjust for regression to the mean. Our control vector included gender, ethnicity, free school meal status, pre-intervention outcomes and allocation reason.[10]

For more details, please see the evaluation protocol (pages 32–33) and the statistical analysis plan (pages 15–17 for more details on the statistical analyses of outcome data).

5.5.3. Qualitative data collection

Below, we summarise the key groups we interviewed for this project. For more details on interview topics and questions, please see the evaluation protocol (pages 34–40).

Participatory groups

Prior to the pilot trial, we conducted a young person participatory group with a number of young people in a school where Khulisa had previously delivered Face It (not a school part of the sample for this evaluation).

We involved a small group of six young people who had participated in the Face It programme within the last two years to help co-construct elements of our research. This session took two hours, with the key aims of asking for advice and input into the evaluation design and of sense-checking evaluation materials, including interview information sheets and consent forms, and asking for advice and input into improvements. Advice from young people resulted in briefer evaluation materials that focused more heavily on the aspects of the work young people felt were most relevant to them.

A second session with four of the same young people was conducted at the conclusion of the trial. The aim of the session was to have young people contribute to our understanding of the pilot evaluation's qualitative research findings. We focused the session on areas of divergence and key thematic areas coming out of interviews that took place with programme staff, school staff and students during the pilot evaluation of the Face It programme. We report on the views of the participatory group throughout this report.

Interviews with young people

To explore evaluability, acceptability and self-perceived impact from young people's perspectives, we aimed to conduct in-person interviews with 4–6 pupils per case study school (12–18 pupils within three schools) six months after the programme had concluded. Young people were identified by school staff on the basis of their availability and willingness to participate. We attempted to purposively sample young people to achieve diversity across the following characteristics: school, gender, year group, ethnicity, free school meal status, and whether the young person presented with primarily internalising or externalising behavioural issues. Overall, we successfully interviewed 13 young people who received Face It from two schools.

[10] This is defined as whether they fall into the 80% of externalising behaviour or the 20% of internalising behaviour.

In addition, over the course of conducting the study, we realised that we would only get a partial view of the evaluability questions (particularly those relating to the acceptability of randomisation among young people) if we only interviewed young people who had been randomised to receive the programme. Consequently, we decided to interview young people who had been assigned to the control group to explore their experiences and feelings around not receiving the programme (and their understanding of randomisation and the reasons they did not end up participating in Face It). We successfully interviewed four young people allocated to the control group in one school.

Interviews with parents/caregivers

To explore evaluability, acceptability and perceived impact from parents'/caregivers' perspectives, we aimed to interview a number of parents/caregivers from one school using a 'drop-in' style at a parents' evening or similar event (and so taking a convenience sample). Unfortunately, we were not able to complete this work, as schools deemed it impractical.

Interviews with Khulisa facilitators

To explore deliverability, acceptability, evaluability and perceived impact from Khulisa's perspective, we aimed to interview three to six programme facilitators close to the end of the programmes they were delivering and to interview two managers at Khulisa towards the end of the project. We were able to interview four Khulisa facilitators and two Khulisa managers.

Interviews with school staff

To explore how schools perceived Face It and to explore deliverability, acceptability, evaluability and perceived impact from schools' perspectives, we aimed to interview three school staff members (one from each of the three case study schools). Overall, we successfully interviewed two school staff members (across two schools).

5.5.4. Qualitative data analysis

We adopted a framework approach to analysing qualitative data. The key themes emerging from both the research questions and the data were incorporated into an analytical framework, with each column of the framework representing a subtheme and each row a participant.

The interview data were then charted (summarised) into that framework, which was used to carry out the analysis. This involved looking at each theme in turn and exploring the range of views held under that theme, with a view to developing categories. Views were grouped into clusters and iteratively developed into clear categories.

The primary aim of this work was to clearly map out the range and diversity of views that exist within each participant population. However, where possible, we proceeded to a higher level of analysis and aimed to look for patterns and linkages in the data, as well as explanations.

5.6. Racial diversity and inclusion

The Lab is committed to conducting research in which equality, diversity and inclusion principles are firmly embedded across all stages of this evaluation, including design, recruitment, data collection and analysis. In the conduct of this work, we focused on:

- **Inclusivity in the design of the evaluation:** As noted above, we included a small group of young people in a participatory group to advise on the design and conduct of the research. We asked the school to suggest a group of young people with a balance of characteristics across the following categories: externalising/internalising behaviours, gender, age and ethnicity.
- **Inclusivity in recruitment to the programme and sampling for the evaluation:**
    - Prior to eliciting student referrals from schools, Khulisa conducted work against unconscious bias with relevant school staff members, using material developed with UNLRN (educating staff about unconscious bias and sharing strategies to minimise this when considering which students to refer to the programme).
    - Our sampling approaches to qualitative work aimed for diversity across a range of characteristics, including ethnicity and gender.
- **Inclusivity during data collection:** We adopted an approach which we hope minimised power imbalances and mitigated access issues. For more details on our approach, please see the evaluation protocol (pages 47–49).
- **Inclusivity in data analysis:** We engaged our young person participatory group in cross-checking our analyses and emerging themes.

6. Findings: Deliverability

6.1. Introduction

As part of the pilot study, we set out to understand:

- **Recruitment and retention:** Can Khulisa recruit and retain sufficient numbers of schools and young people at risk of youth violence?
- **Fidelity:** Can the Face It programme be delivered with fidelity?

6.2. Recruitment and retention relating to the programme

This pilot trial involved recruiting schools and young people into a project which included both participation in an RCT and potential participation in the Face It programme. The section that follows breaks down findings in terms of recruiting schools and in terms of recruiting young people themselves. In this section, we discuss the broader school recruitment strategy and issues with recruitment that pertain to the programme itself (rather than the RCT – these are detailed in the evaluability section of this report).

Findings on recruitment and retention primarily come from administrative data, feedback from schools and interviews with key stakeholders.

6.2.1. Schools

What happened?

Overall, we intended to recruit **five schools** into the project (recruiting three in the autumn term of 2023, followed by two in the spring term of 2024). Each school was intended to deliver Face It to two separate cohorts of young people (a total of 10 cohorts).

Khulisa experienced significant challenges in recruiting schools, and unfortunately, only **two schools** were successfully recruited (**40%** of the target number of schools and **3%** of those approached). Of the two schools that were successfully *recruited*, one dropped out of the project before delivery could begin with the second cohort. Therefore, we only **retained one** school throughout the project. It is worth noting that the school that was retained in the project was keen to continue with delivery, so Face It was delivered to an initially unplanned third cohort at this school. Therefore, Face It was delivered to four cohorts in total.[11]

Khulisa invested significant time and energy into recruiting schools. A multipronged recruitment strategy was developed and refined during the recruitment period, as Khulisa staff were monitoring progress and getting feedback from schools. A '*massive cold call*' of heads

[11] We use the following labels throughout the report:
- School 1 - Cohort 1
- School 1 - Cohort 2
- School 1 - Cohort 3
- School 2 - Cohort 1

of Years 8, 9 and 10 in schools with over 50% of students eligible for the pupil premium was complemented with:

- Extensive advertising in relevant newsletters
- Meetings with directors of education to seek their help in promoting the programme to head teachers
- Promotional materials relating to Khulisa and Face It being dropped at schools
- Existing contacts Khulisa had with academies, individual schools and local authorities being identified

We summarise the number of schools approached and their progress in terms of being recruited into the study in the table below, in terms of five stages:

1) **Prospecting:** Initial contact and attempt to acquire school interest
2) **Cultivating:** Further communication with schools about the project following their expression of interest
3) **Clarifying:** Further discussions around the project, often involving a meeting with the Lab, specifically relating to the evaluation
4) **Committing:** Schools signing service-level agreements and providing school-level consent to participate in the project and evaluation
5) **Delivering:** Proceeding to data collection, randomisation and programme delivery.

Table 2: Summary of school recruitment

Stage of recruitment	Number of schools	Reasons for drop-out
1. Prospecting: Initial contact and attempt to acquire school interest	60	In total, 60 schools were approached to be recruited for this project across the two terms.
2. Cultivating: Further communication with schools about the project following their expression of interest	10 schools retained (50 schools lost from the previous stage)	10 schools responded to Khulisa and expressed interest in exploring their involvement in the project. 50 of the 60 approached schools either did not respond to Khulisa's outreach emails or did respond but subsequently ceased communications with no reason given.
3. Clarifying: Further discussions around the project, often involving a meeting with the Lab, specifically relating to the evaluation	5 (5 schools lost from the previous stage)	5 schools engaged in further discussions about the project and the evaluation. Of the 10 schools that expressed interest in exploring their involvement in the project, 5 schools dropped out following further discussion: ● 2 had concerns about the evaluation survey (i.e. pupils being asked about criminal/antisocial behaviours) ● 2 schools had concerns about the intensity/duration of the programme

		● 1 school was concerned about both
4. Committing: Schools signing service-level agreements and providing school-level consent to participate in the project and evaluation	3 (2 schools lost from the previous stage)	3 schools provided Khulisa with a commitment to participate in the project. Of the 5 schools that engaged in further discussions about the project and evaluation, 2 schools dropped out: ● 1 dropped out because of concerns about the intensity/duration of the programme ● 1 dropped out due to being unable to generate sufficient referrals of eligible young people
5. Delivering: Proceeding to data collection, randomisation and programme delivery	2 (1 school lost from the previous stage)	2 schools went on to participate in the project and randomised controlled trial. The school that dropped out after committing to participate in the project did so when the school's senior management team raised some last-minute concerns around i) parental opt-out consent, ii) data sharing and iii) achieving sufficient referrals.

Key challenges and recommendations

Getting schools to agree to a meeting to introduce the programme was seen as critical by Khulisa staff. However, getting schools to engage in these initial exploratory discussions was very challenging. Khulisa staff reported that they typically work with schools with which they have an established relationship. However, for the trial, the staff had to cast the net wider, to schools that knew little or nothing about Khulisa or the Face It programme. In digging deeper into these recruitment issues, the research team set out four types of potential/hypothetical explanations for these issues with recruiting these new schools:

- **Characteristics of the recruitment approach itself had a negative impact on school recruitment:** It's possible that efforts at reaching out to schools could have been improved in key ways (e.g. clearer communication, more clarity on what involvement required and more emphasis on the attractive aspects of being involved in the project and of the Face It programme).
- **Characteristics of the schools themselves had a negative impact on school recruitment:** Schools may not have been receptive to the recruitment approach if they lacked bandwidth and capacity to engage and were hesitant or unable to commit staff time to setting up and supporting the programme.
- **Characteristics of the programme had a negative impact on school recruitment:** There may have been aspects of the Face It programme itself that were unappealing to schools. For example, the Face It programme is quite intensive and requires a lot of pupil time, with five full days out of the school timetable (with pre- and post-programme sessions in addition to that).
- **Characteristics of the evaluation had a negative impact on school recruitment:** Aspects of the evaluation may have been unappealing to schools. For example, the

large number of referrals required by an evaluation (vs delivery without an evaluation) may have been off-putting to schools. The fact that some young people are assigned to a control group and don't get additional support may have been unattractive to schools, and the nature of the outcome survey and the topics and behaviours it addresses may have been concerning to schools. We discuss this fourth category in greater detail in the evaluability section of this report.

The data we have to determine which of these possible explanations contributed to our challenges with recruitment are limited by the fact that many schools did not provide us with a reason for not taking up the offer of participation. Fifty schools (83% of those contacted) did not go on to participate in the project, and **we do not know why,** as they did not respond to initial outreach or stopped responding to emails from Khulisa. After some time had passed, the research team attempted to get in touch with each of these schools to verify why they did not wish to proceed with the project, and unfortunately, we did not receive responses.

Overall:

- **There is limited evidence from schools that characteristics of the recruitment approach itself had a negative impact on school recruitment.** However, our experience of recruiting for this trial identifies a number of challenges and suggests a number of improvements that could have been made:
    - *Focusing on too few schools initially:* In retrospect, the partnership focused on too few schools at the outset of the recruitment period, only approaching new schools when a dead end was reached with the current schools of focus.
        - *Recommendation 1 – Cast the recruitment net wider from the outset:* Given the low conversion rate (3% of contacted schools participated in the programme), it would be necessary to approach many more schools when recruiting for any future trial. The figures from this trial suggest that we would have had to approach approximately 170 schools (rather than 60) to achieve our target of recruiting five schools. This would, however, introduce other difficulties in terms of the time and resources required to reach out to many more schools and manage communications arising from that.
    - *Tight timelines and insufficient opportunity to build visibility:* Khulisa staff reported that the timetable for recruiting schools (i.e. June–July for a September start) was considered very tight.
        - *Recommendation 2 – Have longer lead-in times and more opportunity to lay the groundwork:* A longer recruitment window would have been more straightforward to manage and would also have given more opportunity for Khulisa to raise the profile of its organisation and programme among a broader network of schools. One of the Khulisa staff members believed that, while Khulisa is effective in engaging schools with which it has established a trusted relationship, to reach new schools, Khulisa's visibility would need to be increased (e.g. by participating in community events and school fairs and improving its marketing and social media presence), which requires time.
    - *Inconveniently timed recruitment windows for schools:* Khulisa staff reported that the timetable for recruiting schools (i.e. June–July for a September start)

was suboptimal, given that July is a difficult time to engage schools due to a combination of exams and the start of the summer holiday. However, it is worth noting that school recruitment continued through August into March of the following year and did not yield additional recruited schools. Given this, we make no strong recommendation against this challenge.

- **There is limited evidence from schools that a lack of bandwidth/capacity among school staff had a negative impact on school recruitment.** While it is almost certainly the case that this played a role, it would not explain why recruitment for Face It in particular was so challenging (as this is likely to affect the recruitment of any other school-based, socio-emotional, skills-focused programme).
    - *One identified challenge was that delivery windows were inconveniently timed for schools:* Khulisa staff reported that the timing of programme delivery did not take into consideration when it would be easiest for schools to fit in the programme. For example, in September, young people are just settling into their new year groups, and schools may be reluctant to offer a programme that may disrupt the settling process.
        - *Recommendation 3 – Align the timing of the programme to when it is most convenient for schools and students:* Given the difficulties reported around expecting strong engagement from schools in the autumn and summer terms, it would be better for a future trial to aim to deliver in the spring term (avoiding the onboarding/settling in of the autumn term and the exams of the summer term).

- **There is some evidence from schools themselves that characteristics of the programme had a negative impact on school recruitment.** Four schools (7% of those contacted) withdrew from the pilot trial, citing concerns about the intensity and duration of the programme.
    - *One identified challenge was that the Face It programme demands too much time out of the school timetable:* It requires five full days out of the school timetable (with pre- and post-programme sessions in addition to that). According to Khulisa staff, the exploratory discussions they had with a few schools revealed that school staff and parents had concerns about students being out of lessons for a week. This issue may be exacerbated by Years 9 and 10 being seen as critical periods by schools, as pupils make their subject choices and transition into preparation for GCSEs. While school staff (and some young people) eventually perceived the length of the programme to be worth it once Face It had been delivered in their school, it is a barrier to engaging with the programme in the first place.
        - *Recommendation 4 – Shorten the programme to ease schools' concerns:* We understand that there is a less intensive three-day version of Face It offered by Khulisa. This may be more appealing to schools that are concerned about taking pupils who are struggling out of class for long periods of time. This would need to be balanced against the potential downsides of shortening the programme (i.e. losing possibly important/efficacious programme content).
        - *Recommendation 5 – Offer the programme to a younger cohort:* Schools may be more amenable to the intensity and duration of the

Face It programme if it is being offered to younger groups, where missing lessons may be perceived to have less of a negative impact. However, whether younger students would need or potentially benefit from the programme to the same extent would need to be carefully considered.

    - *It was also identified that the length of the programme introduces logistical challenges:* Khulisa staff reported that some schools shared concerns about finding a room available for a week to deliver the programme. This emerged as a difficulty in a participating school involved in the qualitative interviews.
        - *Recommendation 6 – Offer flexibility in venues:* Khulisa staff recommended offering schools an alternative external venue if they struggle to find a classroom for the duration of the programme.

- **There is some evidence from schools themselves that characteristics of the evaluation had a negative impact on school recruitment.** Three schools (5% of those contacted) pulled out, citing **concerns relating to the evaluation outcome surveys,** and two schools (3% of those contacted) pulled out, citing concerns about being able to **generate sufficient numbers of eligible referrals.** Both of these points are discussed in greater detail in the evaluability section of this report.

Overall, then, we find evidence that aspects of the programme and the evaluation had a negative impact on school recruitment.

6.2.2. Children and young people

Overall, we intended to recruit **160** young people into the project. Each school was intended to provide 32 young people for the evaluation (two cohorts of 16), with 16 young people receiving Face It (two cohorts of eight) within each school.

What happened?

In terms of **referrals**, we received a reasonable number across all cohorts in the two recruited schools. Our target number of referrals for each cohort was at least 24 to accommodate the fact that not all referred pupils would consent to be involved in the project and to promote an optimal Face It group size post-randomisation. This target was exceeded in three of our four cohorts and was not met in one.

Table 3: Referrals per cohort

Cohort	Number of referrals
School 1: Cohort 1	25
School 1: Cohort 2	28
School 1: Cohort 3	27
School 2: Cohort 1	18
Overall	98

In terms of ethnicity, the referrals we received broke down as follows across both schools and all cohorts, indicating that at the outset, the project successfully engaged young people from a range of backgrounds and reflecting our ambition to recruit schools with high levels of diversity:

Table 4: Characteristics (ethnicity) of young people referred

Ethnicity	Number of referrals
Black	57
Asian	9
White	9
Mixed/Multiple ethnicity	14
Other minority ethnicities	Under 5

In terms of eligibility and profile of risk, all referrals were eligible for the project, insofar as it was not possible to submit an ineligible referral to Khulisa's form. Across all four cohorts, 59% of referred pupils were considered to exhibit externalising behaviour problems, 18% were considered at risk of involvement in criminal activity and 14% were considered at risk of exclusion.

In terms of **programme take-up** (i.e. successfully getting those pupils randomised to the programme to participate), the vast majority of pupils randomised to the treatment group attended at least one Face It session (including the 'pre' one-to-one and group sessions as well as the five main programme days). In the few instances where a pupil was randomised to Face It but did not go on to attend any sessions, we believe the pupils themselves decided not to proceed with the programme.

Table 5: Take-up per cohort

Cohort	Percentage of children and young people in the treatment group attending one or more sessions of the programme
School 1: Cohort 1	88% (7/8)
School 1: Cohort 2	90% (9/10)
School 1: Cohort 3	100% (8/8)
School 2: Cohort 1	86% (6/7)
Overall	91% (31/34)

As part of our analysis plan, we committed to exploring whether key metrics such as take-up differed by characteristics, including ethnic background. Overall, programme take-up did not appear to differ substantively in terms of ethnicity, though it is worth noting that sample sizes for some groups were very small, and therefore drawing strong conclusions or interpretations from these data is challenging.

Ongoing engagement and attendance were relatively high (at 88% overall). Given that the programme is school-based and scheduled during school time, we expected most pupils to attend the whole programme. We did not systematically collect reasons for non-attendance. However, it is likely to include i) **typical reasons for missing school in general** (e.g. illness, medical appointments and unauthorised absence) and ii) **competing priorities within school** (e.g. school-approved activities such as trips, important lessons that young people did not want to miss or teachers who did not want them to miss class).

Table 6: Attendance per cohort

Cohort	Percentage of pupils who attended all five days of the core programme
School 1: Cohort 1	88% (7/8)
School 1: Cohort 2	90% (9/10)
School 1: Cohort 3	100% (8/8)
School 2: Cohort 1	86% (6/7)
Overall	88% (30/34)

As part of our analysis plan, we committed to exploring whether key metrics such as attendance differed by characteristics, including ethnic background. Overall, programme attendance did not appear to differ substantively in terms of ethnicity, though it is worth noting that sample sizes for some groups were very small, and therefore drawing strong conclusions or interpretations from these data is challenging.

Key challenges and recommendations

In terms of **referrals**, although the schools recruited into the project generally provided us with sufficient numbers of referrals, some key challenges identified by school staff and Khulisa facilitators include:

- *Insufficient time:* Some school staff felt they were not given sufficient time for recruitment (this was compounded by the need to generate sufficient referrals for a control group, which is discussed in greater detail in the evaluability section of this report).
- *Referral form indicating a higher level of need:* Some school staff found the referral form confusing, as it included questions that were not applicable to the students they were referring (e.g. self-harm, permanent exclusion). While the students referred were facing some challenges, they were not at the level of intensity some of the questions

suggested, which seemed to be for young people who had gone through extensive trauma.

- *Competition from other programmes:* Facilitators reported that in one school, it was more difficult to recruit because it already had a programme for young people with 'big problems/category 1'.

In terms of **eligibility**:

- *Internalising pupils are harder to identify:* As noted previously, schools were asked to identify young people with externalising or internalising behaviours (with the intention that there would be an 80:20 split in our sample of young people, with 80% displaying predominantly externalising behaviours and the remainder displaying internalising behaviours). One school noted that it had difficulties in identifying eligible young people with internalised behaviours, as they are less likely to be flagged through the behaviour monitoring system, and there is less information about them to feed into the referral process.
- *Desired split of externalising/internalising was not achieved overall:* We note that the proportion of referred pupils considered to exhibit externalising behaviour problems (59%) was less than the 80% we were aiming for. This suggests either a problem in the identification of these pupils or in the communication of this expectation to the schools.
    - *Recommendation 7 – Provide clearer guidance to schools:* It would be useful to further emphasise the requirement to have a split of pupils exhibiting externalising/internalising problems in communications with schools during the onboarding process and the reasons it is important. More advice and guidance around how to identify pupils with these characteristics using existing school data and teacher judgement are likely to be helpful.

In terms of **take-up** (i.e. successfully getting those pupils randomised to the programme to participate), the young people interviewed were generally enthusiastic about joining Face It. The prevailing motivator identified through qualitative work with young people was the ability to miss lessons and to do something different. Feedback from the young person participatory group also reflected these views about missing school lessons being an important motivator. Although it is not clear to what extent this issue limited take-up, the following challenge was identified:

- *Lack of clarity around what Face It involves:* Young people seemed to have little understanding at the outset of what the programme would involve.
    - *Recommendation 8 – Share more information about the programme in advance:* Both facilitators and school staff said young people who are offered a place should have a better understanding of what the programme involves. A participant suggested that it would be helpful for Khulisa to run an assembly to explain what Face It involves and to motivate students to take part. We agree that these measures could be helpful, although we would caution that in the context of an RCT, raising awareness of the programme should only be conducted with those assigned to the treatment group prior to formally beginning programme sessions (to mitigate risks around causing disappointment, resentful demoralisation and differential attrition in the control group).

Ongoing engagement and attendance, as noted, were relatively high. Through the qualitative work, school staff and facilitators reported high levels of engagement with the programme, which was supported by the evidence from young people. The facilitators reported that the programme normally has good retention, and this was the experience of delivering it in the two schools in the evaluation, with one school reported to have had higher retention than normal. Nonetheless, our work identified the following challenges:

- *Withdrawals:* While parental withdrawals were limited in number, there were two across the project. Four young people formally withdrew from the project. In one case, the Lab's and Khulisa's association with violence/crime reduction raised some concerns, with one parent withdrawing their child from the programme after checking Khulisa's website (under the belief that their child was offered a place because they were suspected of being a criminal).
    - *Recommendation 9 – Carefully communicate about the aims of the project to parents:* In our communication with young people, we took great care to simultaneously i) be transparent about the overall aims of the project (which included exploring the impact of a school-based programme on offending behaviours), ii) make it clear that Face It is also intended to impact a range of other outcomes beneficial for young people in general and iii) emphasise that young people were not selected because they were suspected to have gotten into trouble. Future work should ensure that this messaging gets through to parents who, understandably, may have concerns without this context.
- *Competing priorities:* It was reported across interviewees that competing priorities may have introduced some challenges for attendance and engagement. For example, two young people were reported to have dropped out of one school because they did not want to miss football training. Similarly, while some young people reported that, initially, missing a week of school was what attracted them to the programme, once it started, they began to wonder about the possible consequences of the time off lessons (e.g. the need to catch up). Feedback from the young person's participatory group also reflected these views about having to face the possible consequences of missing lesson time, which may be more negative at certain times of the year (e.g. mock and exam periods and at the start of terms, when teachers tend to go over the basic/core concepts) and for certain year groups.
    - *Recommendation 10 – Make the time commitment clear up front:* It is clear that for some young people, the fact that attending the programme would imply missing other activities and catching up with academic work only became clear partway through their involvement with the programme. To reduce non-attendance or programme drop-out, the implications of programme attendance should be made clear to pupils prior to being offered a place on Face It.
    - *Recommendation 11 – Reduce the time commitment to ease young people's concerns:* As noted previously, there is evidence that schools are concerned about the amount of time Face It demands out of the school timetable. There is some evidence that this commitment can be concerning for young people, too. Shortening the programme may alleviate these concerns. We understand that there is a less intensive three-day version of Face It offered by Khulisa, which may be more acceptable to young people.

6.3. Fidelity

In this section, we describe the extent to which the core elements of the programme were delivered as intended, as well as barriers and enablers to delivery and adaptations that were made. Findings on fidelity come from interviews with Khulisa staff and facilitator self-completed fidelity checklists.

What happened?

Overall, Face It was delivered with a high level of fidelity. Fidelity was assessed using facilitator self-reports of key Face It activities specifically conducted across the five-day programme itself. Generally speaking, fidelity was high, with over 80% of key activities being conducted. Fidelity was lower (and in the medium range) for one cohort (School 1: Cohort 2).

Table 7: Fidelity ratings per cohort

Stage of recruitment	Percentage of programme milestones met
School 1: Cohort 1	81%
School 1: Cohort 2	67%
School 1: Cohort 3	88%
School 2: Cohort 1	88%
Overall	81%

Facilitators seemed confident they had delivered all the key programme components as planned. Variations in the programme delivery (e.g. adapting activities depending on the year group) were reported to be in line with the programme protocols, which, for example, outline core activities and optional activities that facilitators can select based on a '*feel for the group'.* This flexibility to vary the programme based on students' feedback was believed to be important to establish trusted relationships, and it was appreciated by young people. As one facilitator explained:

'I think actually it's pretty okay to change the way that you do deliver depending on how the young people are responding to the content because it's for them. You want them to get the most out of it. As long as it's the same content running through, you've got the same aim, it's just that the way you might do it will be slightly different'.

Key challenges and recommendations

No major challenges were identified, as the majority of programmes were delivered with fidelity (albeit with evidence of flex in content based on group dynamics). However, it is worth noting that this was achieved on a relatively small scale. In order to deliver a full-scale efficacy trial, Khulisa will need to recruit and train a large number of additional facilitators. It will need to ensure that the current quality of training is maintained to ensure fidelity at a larger scale.

6.4. Cost

Summary

Below, we report the cost of delivering the intervention, following YEF costing guidance. We:

- Use a bottom-up costing approach and categorise costs into the following categories: prerequisites, set-up costs and recurring costs.
- Report the total cost for a typical single cohort receiving the intervention for one round of delivery and the costs per participant for one round of delivery, assuming full compliance.
- We report cost data for the delivery cohort for which the most complete data were provided, where eight young people were randomised to receive Face It.
- We follow the YEF's guidance and aim to provide costs that reflect the **input cost** of the intervention (i.e. the cost of the resources used in delivering the intervention). These figures **do not** represent the 'market cost' of the programme or any cost of the programme to schools.
- We also aim to provide costs that are broadly reflective of business-as-usual delivery of the programme (excluding costs that are specifically related to the evaluation, making them specific to the context of this project).

Table 8: Set-up and recurring costs per group and per participant

Cost category	Set-up/recurring costs	Costs per group	Costs per participant
Staff costs[12]	Set-up costs	£58.45	£7.31
	Recurring costs	£10,946.68	£1,368.33
Materials and equipment costs	Set-up costs	£306	£38.25
	Recurring costs	£100	£12.50
Programme procurement costs	Set-up costs	£33	£4.13
	Recurring costs	£0	0
Total costs	-	**£11,444.13**	**£1,430.52**

[12] In terms of staff costs, the time estimates we received included a substantial amount of management time. We believe these figures largely reflect the additional time spent meeting the requirements of this specific project and its evaluation rather than business-as-usual time spent running the programme. To avoid overstating costs, we have therefore applied a simplified assumption that everyday management time is proportional to frontline delivery, at a ratio of 1:1. This provides a conservative but reasonable estimate of business-as-usual management input.

Overall, the majority of costs fall into staff costs and equipment/materials costs categories.

Prerequisite costs

Prerequisite costs include what is already expected to be in place before a programme is implemented. It is expected that prior to delivery, schools will have appropriate staff to liaise with Khulisa staff and to facilitate delivery. It is expected that there will be access to suitable spaces to deliver the intervention.

Set-up costs

Set-up costs are one-off costs incurred at the start of a programme. Here, these were mainly accounted for by:

- *Staff costs:* The cost of practitioner time in terms of attending programme training and preparing for delivery for the two Khulisa facilitators responsible for frontline delivery.
- *Materials and equipment:* The cost of laptops to support the delivery of the intervention (prorated in line with their relative use for programme delivery).

Recurring costs

Recurring costs are those required each time a programme is implemented. Here, these were mainly accounted for by:

- *Staff costs:* This includes the cost of practitioner time in terms of delivering the programme (the pre- and post-programme one-to-one meetings, the pre- and post-programme group meetings and the five-day programme) for the two Khulisa facilitators responsible for frontline delivery. It also includes the cost of management and administrative time from 'Head of' level staff at Khulisa (programmes, partnerships, finance and human resources) and an (internal) monitoring, evaluation and learning officer.
- *Materials and equipment:* This includes the cost of snacks for young people in programme sessions, as well as the masks used in the programme and art materials.

6.5. Conclusions

Overall, we conclude that it is possible to recruit and retain young people in the Face It programme and to deliver the programme to them with fidelity. The main deliverability barrier observed was in the recruitment of schools, which was hugely challenging. Overcoming this substantial challenge in future evaluations would likely require:

- *Casting the recruitment net wider from the outset and targeting many more schools than required*
- *Having longer lead-in times and more opportunity to lay the groundwork*
- *Ensuring better-timed recruitment windows*
- *Aligning the timing of the programme with when it is most convenient for schools and students*

- *Shortening the programme to ease schools' concerns*
- *Offering flexibility in venues*

7. Findings: Acceptability

7.1. Introduction

As part of the pilot study, we set out to understand acceptability in terms of:

- **Young people:** Is the Face It programme seen as acceptable and valuable by young people?
- **School staff:** Is the Face It programme seen as acceptable and valuable by school staff?
- **Parents:** Is the Face It programme seen as acceptable and valuable by parents?

7.2. Young people

Young people were asked a series of multiple-choice Likert-style questions about their satisfaction with aspects of the programme (this was included in the post-programme outcome survey). Of the 21 young people assigned to the treatment group who completed the post-programme survey, 18 of them answered these questions. Scores of 4 and 5 indicated satisfaction with the programme (either meaning the young person found a particular aspect of the programme helpful or very helpful or they were happy or very happy with an aspect of the programme).

On average, young people were happy with the facilitators who delivered the programme to them, were happy with receiving the Face It programme in a group format and found the programme helpful in terms of meeting their needs. Overall, they were happy with the experience of the programme. This is reflected in our qualitative findings:

- *Facilitators:* Young people were very positive about the facilitators, who were described as warm and caring, funny and entertaining, always smiling, and non-judgemental. Some young people reported that they were able to establish a trusted relationship with the facilitators, which meant they could talk freely and remain calm even when discussing difficult situations. Young people also appreciated not being shouted at and being given the time to calm down if they got angry or upset. Relatedly, some young people drew attention to:
    - *One-to-one sessions with facilitators:* They enjoyed the one-to-one sessions with the facilitators, the ‘check-ins’. These provided a nice and calm environment where young people could talk freely and honestly with facilitators, who provided comprehensive explanations and answers to questions, supported them in expressing themselves and made them feel listened to and valued.
    - *Flexibility of facilitators:* They enjoyed the flexibility to adapt the activities and the pace of work to their needs, and they never felt rushed. They liked the freedom to choose some activities, work in pairs or in a group, chat and listen to music during the activities.

- *Group format:* Some young people reported that the group sessions instilled a sense of responsibility and respect for one another. Young people enjoyed making the rules

together, working as a team, learning with peers and having a safe space where everybody could express their views and experiences without being judged.

Interestingly, the satisfaction survey suggests that, on average, young people were less positive about the specific activities involved in Face It and the taught content involved in Face It – falling somewhere between finding them neither helpful nor unhelpful and finding them helpful. This may indicate that the overall experience of participating in Face It and the interactions with the facilitators are more important to young people than the specific activities.

While our qualitative work indicates that some young people did not value some specific games/activities, broadly, the responses indicated that young people enjoyed the range of creative activities and the opportunity to learn while having fun. The topics covered were seen as relevant (e.g. learning to value yourself, the triggers of anger, how to control your emotions and how to establish positive relationships).

Table 9: Satisfaction survey for young people

Measure	Mean (standard deviation)
Satisfaction with practitioners	4.83 (0.38)
Satisfaction with group format	4.28 (1.07)
Satisfaction with taught content	3.61 (0.85)
Satisfaction with programme activities	3.89 (1.08)
Perception of the programme meeting their needs	4.00 (0.91)
Overall satisfaction with the programme	4.67 (0.59)

Our satisfaction survey also asked how satisfied young people were with the duration of the programme. Interestingly, on average, young people responding to the survey thought that the programme was a bit too short. Our qualitative work also reflects this. While some young people became concerned about missing lessons during the course of the programme, they still felt that the programme was worth it and suggested lengthening the programme. However, it is worth noting that we did not receive completed satisfaction surveys from all young people, and this sample (and our qualitative sample) may include young people who persisted with the programme (and not include those who may have dropped out owing to concerns about missing school).

As part of our analysis plan, we committed to exploring whether key metrics, such as satisfaction, differed by characteristics, including ethnic background. Overall, satisfaction with the programme did not appear to differ substantively in terms of ethnicity, though it is worth noting that sample sizes for some groups were very small, and, therefore, drawing strong conclusions or interpretations from these data is challenging.

7.3. School staff

While we did not ask school staff to complete a survey, we interviewed two school staff members about their perceptions of the programme.

Overall, school staff were very positive about the programme, as they believed it created a safe space and a calm environment for young people to explore their emotions and feelings through a range of creative activities, which helped to support engagement. They highlighted the skills and experience of the facilitators as key to the programme's success, as young people quickly established good relationships with them. It is worth noting that School 1 was keen for Khulisa to deliver to a third, unplanned cohort and to extend its participation. Khulisa facilitators also reported that the programme is usually very popular with schools, and they believed this was the case in the two schools in the evaluation, where they felt the programme was well received by staff.

As noted previously, some staff reported that the week-long programme was challenging in terms of time off the curriculum. However, they indicated that the problems it created were not insurmountable, and on balance, they felt the benefits of the programme justified its length. Overall, most issues highlighted by school staff in interviews related to the evaluation rather than the programme (please see the evaluability section of this report).

7.4. Parents

Alongside the outcome survey that we asked young people to complete before and after Face It was delivered, we attempted to ask parents to complete a similar survey, which asked questions about parents' perceptions of the programme. We received an extremely limited response to this survey (two responses to the post-programme survey), so we do not draw any conclusions from these results.

We had also planned to interview several parents about their perceptions of the programme in a school-based drop-in session. Unfortunately, we were not able to complete this session, as the two recruited schools were hesitant to commit to this (believing it to be too complex or difficult).

Therefore, we have very limited data on the acceptability of the programme to parents. We are aware that there was a slightly increased number of parental withdrawals from the programme (compared to standard delivery of Face It), but in general, this was low.

7.5. Conclusions

Overall, we conclude that the Face It programme is acceptable to and valued by young people and school staff, having received very positive feedback through the satisfaction survey (noting the limitations of this in that only some young people completed the survey and that this may largely reflect young people who enjoyed the programme and persisted with it) and qualitative interviews. There are mixed findings about the intensity of the

programme, though these tend to be overcome once a school commits to the programme, it is delivered, and young people and teachers perceive the benefits.

8. Findings: Evaluability

8.1. Introduction

As part of the pilot study, we set out to understand:

- **Recruitment into the study:** Can sufficient numbers of schools and young people be recruited into the RCT?
- **Randomisation:** Is randomisation feasible and adhered to? Is it acceptable to young people and schools?
- **Data collection:** Is data collection feasible, and is it possible to retain young people in the evaluation? Is data collection acceptable to young people and schools?

8.2. Recruitment into the study

As noted previously, this pilot trial involved recruiting schools and young people into a project which included both participation in an RCT and potential participation in the Face It programme. In the previous (deliverability) section of the report, we focused on programme-related issues around school recruitment and young people's uptake of the *programme* specifically (once randomised to receive it). Here, we focus on evaluation-related issues around school recruitment and young people's recruitment into the *evaluation*.

Findings primarily come from our survey data, administrative data on referrals, feedback from schools and interviews with key stakeholders.

8.2.1. Schools

What happened?

As discussed above in greater detail, Khulisa experienced significant challenges in recruiting schools. Unfortunately, only **two schools** were successfully recruited (**40%** of the target number of schools and **3%** of those approached). Of the two schools that were successfully *recruited,* one dropped out of the project before delivery could begin with the second cohort. Therefore, we only **retained one** school throughout the project.

Key challenges and recommendations

In addition to the programme-based challenges to school recruitment:

- **There is some evidence from schools themselves that characteristics of the evaluation had a negative impact on school recruitment.**
    - *One identified challenge was around* ***generating sufficient numbers of referrals required for the evaluation.*** While schools generally did not struggle to find enough young people who were eligible for the programme to sustain delivering a programme group (approximately 10 young people), some schools found finding enough referrals to sustain a programme group *and* a control group of a similar size difficult (in addition to overrecruiting to pre-empt drop-outs). Two schools (3% of those contacted) withdrew from the

project, citing concerns about generating sufficient numbers of eligible referrals. It is unclear exactly why this is the case. Khulisa staff believed that the evaluation was seen by schools they were trying to recruit as a considerable burden (and, presumably, generating referrals was the main driver of this, given it was the most intensive part of their role). It is worth noting that this issue did not emerge from the discussions with *participating* schools, although this does not mean it wasn't a barrier for those we failed to recruit.

- Overall, we make **no strong additional recommendation against this challenge**. Given the parameters of i) the optimal group size for Face It and ii) the need to keep treatment and control groups of a broadly similar size, there is little that can be done to mitigate this challenge beyond extending recruitment windows and timing them during relatively quiet school periods to ease the burden on schools (Recommendations 2 and 3 described in the deliverability section of this report).

- *Another identified challenge was around schools'* ***concerns with the evaluation outcome surveys.*** As this project was funded by the YEF to lay the groundwork for a full-scale RCT investigating the impact on offending, our survey included questions about sensitive topics and behaviours. The best example of this is the Self-Report Delinquency Scale (SRDS), which asks young people to self-report whether they have been involved in a series of antisocial and criminal behaviours. Three schools (5% of those contacted) withdrew from the project, citing discomfort with the use of this survey (likely due to a perceived risk of 'othering' or stigmatising some young people, potentially causing them distress, or getting complaints from parents). In addition, Khulisa staff noted that the SRDS questions emerged as a barrier to engagement in conversations with some schools that they were trying to recruit. While this contributed to challenges with school recruitment, overall, we do not think the majority of schools at the earliest stages of the recruitment process were likely to be aware of the nature of the outcome surveys at the point they dropped out or failed to respond. We note that some schools that were successfully recruited into the project also went on to have concerns about the surveys used in the evaluation (as did some young people). This is discussed in greater detail in the following section on data collection.
  - Overall, we make **no strong recommendation against this challenge**. This is predominantly because a trial funded to examine these outcomes would need to measure them in some way. An alternative would be to rely on other sources of data to gauge impact on offending (i.e. linking with local police data). However, we note that while this may reduce schools' perceived risk of causing distress to young people by asking them to self-report negative behaviours, it would still carry the risk of stigmatising/othering young people (they would still be exposed to the fact that the project aims to reduce offending) and being perceived as intrusive by young people and their parents.

- *Another identified challenge was around schools'* ***concerns about the control group.*** Although no schools formally gave this as their reason for withdrawal, Khulisa staff reported that some schools they tried to recruit were uncomfortable with randomisation. The reasons for this are likely a lack of comfort with being unable to offer the programme to all young people who might benefit from it and the perceived risk of distress caused by having young people who are receiving the programme and those who are not co-existing in the same school environment in proximity to one another.
    - Overall, we make **no strong recommendation against this challenge**. This is partly because few of the 60 approached schools raised it as a problem and partly because a control group is simply necessary to conduct an RCT. An alternative would be to randomise at the school level and conduct a cluster RCT. However, we note that i) this would require even more schools to participate in the project, and the experience of this pilot trial is that this is challenging, ii) this would plausibly introduce additional challenges for school recruitment (as not all schools would be guaranteed to receive the programme, removing one of the main incentives to participate in the project) and iii) this would plausibly introduce challenges around differential attrition at the school level. Control group schools may be reluctant to cooperate with an evaluation throughout the trial period if they are not receiving anything (i.e. the programme) to justify these efforts.

8.2.2. Young people – consenting to the study and baseline data collection

What happened?

Eligible young people who had been referred to the programme were invited to one group data collection session held in their school. At these sessions, Lab researchers talked young people through the project and the information sheet. Young people were given an opportunity to provide their consent to be involved in the project, and they were then asked to complete the baseline data collection survey on a school computer.

Overall, we retained just under three-quarters of the sample between the referral and recruitment stages.

Table 10: Young people recruited into the trial

Cohort	Number of referrals	Number who consented to the evaluation and completed baseline data	Conversion rate
School 1: Cohort 1	25	17	68%
School 1: Cohort 2	28	18	64%

School 1: Cohort 3	27	22	81%
School 2: Cohort 1	18	14	78%
Overall	98	71	72%

Table 11: Characteristics of young people recruited into the trial

Pupil characteristics (categorical)	**Treatment group**		**Control group**	
	n/N (missing)	**Count (%)**	**n/N (missing)**	**Count (%)**
Free school meals	17/34 (0)	17 (50%)	13/31 (0)	13 (42%)
Female	15/34 (0)	15 (44%)	13/31 (0)	13 (42%)
Male	19/34 (0)	19 (56%)	18/31 (0)	18 (58%)
Black	20/34 (0)	20 (59%)	19/29 (2)	19 (66%)
Asian	Under 5	-	Under 5	-
White	5/34 (0)	5 (15%)	Under 5	Under 5
Mixed/Multiple ethnicity	6/34 (0)	6 (18%)	5/29 (2)	5 (17%)
Other minority ethnicities	Under 5	-	Under 5	-
Pupil characteristics (continuous)	**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**
Age	33/34 (1)	15.2 (0.65)	27/31 (4)	15.2 (0.60)
Baseline Strengths and Difficulties Questionnaire total difficulties	34/34 (0)	14.6 (7.1)	31/31 (0)	15.2 (6.0)

Please note that:

- *The total N does not match that presented in the participant flow table. This is because we do not have the baseline characteristics of pupils who withdrew from the trial.*
- *We do not report absolute values in cells with a participant count lower than 5 due to anonymity considerations.*

Key challenges and recommendations

Although our data on reasons for non-attendance at the onboarding/baseline data collection session are limited, we suspect that the following factors reduced our referral-to-recruitment conversion rate:

- *Inevitably,* **typical reasons for missing school in general** (e.g. illness, medical appointments and unauthorised absence) and **competing priorities within school** (e.g. school-approved activities such as trips, important lessons that young people did not want to miss or teachers who did not want them to miss class) will have reduced the number of referred young people attending the baseline data collection session and being formally recruited into the trial.
    - *Recommendation 12 – Provide a longer window and more opportunities for baseline data collection:* For this trial, we only conducted one baseline data collection session per cohort at baseline, which meant that young people missing that first session could not participate in the project. We think it would be wise for future evaluators to either run multiple initial recruitment/baseline data collection sessions or to provide a survey self-complete option for young people who were not able to attend the first session. Either option would provide multiple opportunities for young people to participate and invariably drive up the referral-to-recruitment conversion rate. We note that the former option would carry risks around adding to the school burden, whereas the latter option may have negative consequences for acquiring truly informed consent from young people and for outcome data quality.

- *We also suspect that* ***mismatched expectations – or a lack of clarity – about what the project involved and its objectives*** may have reduced the number of referred young people attending the baseline data collection session and being formally recruited into the trial. One parent withdrew their child from the project at this stage, having looked into the project more and understanding the project's association with violence/crime. Some young people attended the beginning of the onboarding/data collection session but exited partway through. This could be a result of hearing more about the project and/or beginning to complete the survey. For these young people, we do not know the precise reason for their withdrawal. Our qualitative work suggests that some young people had the perception that the programme was for 'bad people'. This perception seemed to be driven mainly by the SRDS survey questions. While there is no direct evidence that this deterred these particular young people from joining the programme, it is a plausible contributing factor. Note that we report concerns relating to the SRDS from young people (who were successfully recruited into the study) in the following section focused on post-programme data collection.
    - *Recommendation 13 – Strengthen pre-session communication to reduce mismatched expectations:* While both Khulisa and the Lab made deliberate efforts to ensure that messaging at every stage was transparent about the project's aims – and simultaneously made clear that the programme is *not* exclusively for young people involved in antisocial and offending behaviour – it is evident that this message was not always received or understood as intended. We recommend that any future evaluation in this context include more deliberate and structured communication prior to onboarding and data collection. This must make the relationship between the programme and

evaluation to youth violence clear, and it could include i) a short explainer video or slide deck aimed at young people and parents, using plain language to describe the project's purpose, ii) a brief summary of the types of survey questions asked and why they are included and iii) clear, consistent framing around how young people were selected to be involved. However, we note that clearer and more accessible communication might still lead to low engagement.

8.3. Randomisation

8.3.1. Feasibility of and adherence to randomisation

What happened?

Randomisation occurred in batches for each cohort after students were referred to the programme and assessed for eligibility and after consent had been obtained and baseline assessment had taken place. The vast majority of young people onboarded into the study were successfully randomised.

Table 12: Young people randomised

Cohort	Number who consented to the evaluation and completed baseline data	Number randomised	Conversion rate
School 1: Cohort 1	17	17	100%
School 1: Cohort 2	18	18	100%
School 1: Cohort 3	22	20	91%
School 2: Cohort 1	14	14	100%
Overall	71	69	97%

Key challenges and recommendations

Some relatively low-level challenges were observed in the conduct of randomisation:

- *Two young people completed baseline data collection but were not put forward for randomisation.* For one young person, this was because they were found to be ineligible for the project, as they did not present with either externalising or internalising behaviour issues (according to school staff). For the other young person, this was because in instances where we recruited more young people than required, we randomly selected a subset to go forward to involvement in the evaluation and

randomisation (to keep Face It group sizes manageable whilst maintaining balance in sample sizes between the treatment and control groups).
    - Overall, we make **no strong recommendation against this challenge**. This is partly because the issue only presented itself on a very small scale, and it was either a conscious decision by the evaluator or an otherwise inevitable human error (rather than anything systematically problematic with the evaluation approach).

- *One young person received the programme despite not being randomised.* Randomisation was adhered to despite a small number of requests to move participants between control and treatment groups (which were refused). We observed one instance of the programme being delivered to a young person who was not randomised (as they did not consent to participate in the evaluation and did not complete baseline data collection).
    - Overall, we make **no strong recommendation against this challenge**. This is partly because the issue only presented itself on a very small scale and partly because, ultimately, in this case, it was not problematic for the trial. Due to our monitoring system, we were able to identify this and exclude the person from the trial, meaning this didn't introduce internal validity issues. In addition, this young person was effectively occupying a vacant slot on the programme, and so they were not preventing other young people from participating in the evaluation or the programme. We recommend that future evaluators cross-reference attendance lists against randomisation output, as we did, to monitor for potential issues.

8.3.2. Acceptability of randomisation

Overall, we find that randomisation is acceptable, in part, to practitioners, teachers and young people, despite some initial reservations.

Generally, school staff members were willing to reluctantly accept randomisation as a feature of the project, given that it was a condition for funding and proceeding with programme delivery. We note that randomisation was cited by one staff member as their only reservation in terms of participating in a project similar to this one in the future. Hesitancy around randomisation was caused by:

- Frustration with the fact that they could not guarantee a place to particular students who they felt would really benefit from the programme
- A concern about disappointing young people in the control group and getting their hopes up about participating in the programme. They said it felt wrong to identify eligible young people, tell them about this great programme and then disappoint them when they were not selected for it.

The facilitators echoed these views and reported that randomisation was particularly problematic in the school where the programme was run multiple times, as young people taking part in the first programme reported how good it was. They also believed that the baseline survey gave the impression that young people had a place in the programme itself, and more needs to be done to improve young people's understanding of the randomisation process.

Young people's views on randomisation varied depending on whether they were in the treatment or control group, but they did not always straightforwardly echo the views of the schools or facilitators. Those in the treatment group thought that their peers who had not been selected were annoyed, disappointed or angry. In contrast, those in the control group who were interviewed did not seem to be *bothered*. Some reported an initial disappointment that they got over relatively quickly. Some reported that those who were given a place probably needed to do the programme more than they did, as they were able to manage their emotions already. This lends support to the facilitator's view that some young people did not understand the randomisation process.

Overall, we make **no strong recommendations against these challenges**. This is partly because randomisation was successful in the schools it was attempted in and did not result in serious relationship management challenges between Khulisa, the Lab and schools. It is also partly because randomisation is necessary to conduct a high-quality impact evaluation in this context. As we noted in the previous section, an alternative approach would be to randomise at the school level and conduct a cluster RCT. However, this carries additional risks and is not a straightforwardly superior approach. Alternatively, using a wait-list design would likely be substantively more acceptable to school staff and young people. However, we note that this option is not available as part of YEF-funded trials due to the YEF's commitment to assessing the impact of programmes on offending in the longer term.

8.3.3. Nature of control group support

As part of this trial, we did not seek to restrict what alternative services were received by the control group (and so we compared Face It to business-as-usual within the participating schools). We identified what services the control group were receiving via the post-programme outcome survey. The alternative services appeared to be a mix of mentoring programmes and socio-emotional learning programmes with a similar focus to Face It. The most reported alternative service was Peer Mentoring (11), followed by Greenhouse Basketball and Table Tennis Mentoring (8), which aim to provide inspirational sports coaching and mentoring for children to develop key life skills. Six young people reported receiving Blueprint, which focuses on employability skills and learning about different career options. While many of these programmes focus on similar outcomes to Face It and seek to support similar groups of young people, they focus on different activities (e.g. mentoring and sport) and are less intensive.

8.4. Data collection

8.4.1. Feasibility of post-test data collection, attrition rates and retention in evaluation

What happened?

After the treatment group completed the Face It programme, an equivalent group data collection session was held in the young people's schools. They were asked to complete the same survey again using school computers. We attempted to collect outcome data from all young people, regardless of whether they were randomly assigned to the treatment or

control group and regardless of how much of the programme they had completed (an intention-to-treat design).

Overall, the data collection approach for young people was found to be feasible, and we successfully collected data for all randomised cohorts at the post-programme point. However, attrition rates were substantial and higher than would be acceptable for a full-scale RCT aiming to minimise internal validity threats and bias.

Table 13: Young people's retention in the evaluation

Study stage	Overall numbers	Treatment group numbers	Control group numbers
Referred	99	-	-
Consented to the evaluation and completed baseline data	69	-	-
Randomised	67	36	31
Completed post-test data collection	43 (36% attrition, i.e. the proportion of randomised young people who did not submit responses to the post-programme survey)	21 (42% attrition)	20 (35% attrition)

- *Note that these figures may differ from those presented elsewhere in this report due to the fact that some pupils withdrew from the study, and we do not have access to their data.*

Attrition rates (the proportion of randomised young people who successfully completed the survey) and completion rates (the proportion of young people beginning the survey who successfully completed it) did not differ substantially across specific measures used in our survey. In particular, given the findings previously reported, it is worth noting that there was no evidence that the SRDS questions were less likely to be completed by young people than the questions from other measures.

At the three-month follow-up point, attrition had increased to 46% overall (50% in the treatment group and 42% in the control group).

In terms of outcome data collection with parents, survey completion was very low. Only two parents completed the post-programme survey, with five completing the three-month follow-up survey. For this reason, we did not conduct any analysis on these responses. Overall, this was not surprising to us, as past school-based trials suggest that acquiring parental engagement with evaluation can be challenging.

Table 14: Characteristics of young people retained in the trial

Pupil	Treatment group	Control group

characteristics (categorical)	n/N (missing)	Count (%)	n/N (missing)	Count (%)
Free school meals	10/20 (0)	10 (50%)	9/20 (0)	9 (45%)
Female	12/20 (0)	12 (60%)	9/20 (0)	9 (45%)
Male	8/20 (0)	8 (40%)	11/20 (0)	11 (55%)
Black	10/20 (0)	10 (50%)	12/19 (1)	12 (63%)
Asian	Under 5	-	Under 5	-
White	5/20 (0)	5 (25%)	Under 5	Under 5
Mixed/Multiple ethnicity	Under 5	-	Under 5	-
Other minority ethnicities	Under 5	-	Under 5	-
Pupil characteristics (continuous)	**n/N (missing)**	**Mean (SD)**	**n/N (missing)**	**Mean (SD)**
Age	20/20 (0)	15 (0.65)	17/20 (3)	15.1 (0.66)
Baseline Strengths and Difficulties Questionnaire total difficulties	20/20 (0)	17.4 (6.0)	20/20 (0)	14.5 (6.1)

Please note that:

- *The total N does not match that presented in the participant flow table. This is because we do not have the baseline characteristics of pupils who withdrew from the trial.*
- *We do not report absolute values in cells with a participant count lower than 5 due to anonymity considerations.*

In terms of qualitative data collection, this was relatively straightforward with respect to young people, and we managed to interview the number of young people we set out to.

Key challenges and recommendations

- Although our data on reasons for non-attendance at the post-programme data session are limited, as with baseline data collection, we believe that the main reasons are **typical reasons for missing school in general** (e.g. illness, medical appointments and unauthorised absence) and **competing priorities within school** (e.g. school-approved activities such as trips, important lessons that young people did not want to miss or teachers who did not want them to miss class).
    - Our recommendation here is the same as in the previous section (Recommendation 12); we strongly advise future evaluators to lengthen the

window for data collection and provide additional opportunities for young people to attend data collection sessions.

8.4.2. Acceptability of data collection

Acceptability of the Self-Report Delinquency Scale

As expected, the majority of young people's feedback on the survey focused on the SRDS questions. Overall, our qualitative work suggested a broadly negative response to these questions, with evidence of reactions varying from confusion to potential distress.

Some young people were puzzled by SRDS questions, as this young person explained:

'What the hell are these questions … people like us, we're not going to rob a house … It was just mad because, like, we're all young, innit, and then they're all asking us about did you steal a car or something?'

One young person said that some of the questions reminded her of things that she used to do that she didn't want to remember, but it is not clear whether she was referring to the SRDS questions.

Some young people also believed their peers would be unlikely to answer these types of questions truthfully for fear of getting into trouble with teachers, as they were not persuaded that the survey was anonymous. As this young person explained:

'*I didn't know if I wanted to personally answer those questions because I felt like I was exposing myself, if that makes sense, to the things that I've done'.*

A few young people also thought that some may try to answer 'yes' to the SRDS questions to get on the programme; this was also reported by one of the school staff members. A few young people and a facilitator also thought that young people in the intervention group may be more likely to answer truthfully at the follow-up surveys, after they got to know and trust the programme facilitators.

Feedback from the young person participatory group on the SRDS questions reflected the feedback of young people interviewed, with some saying that being asked these questions would make them '*feel like a criminal*' and that people were insinuating that they had been involved in criminal activities, while others thought the questions were too removed from reality to be taken seriously.

Generally speaking, the reactions of facilitators and school staff to the SRDS were more negative than those from young people. Facilitators said that young people could get upset by being asked these questions, and it felt wrong to start a programme that is meant to make young people feel better about themselves by asking questions that seem to criminalise them. The fact that young people can decide not to complete the survey was not seen as sufficient to deal with these ethical issues because, in a school environment, pupils may assume they have to do it if they want to be included in the programme. The SRDS questions were discussed with only one school staff member, who reported that some young people were surprised by the questions and became agitated during the survey. However, they did not think the questions created additional challenges and could see why they were needed.

Acceptability of the remainder of the survey

Data here are limited, as the majority of the feedback focused on the SRDS. There were a few positive comments from young people about the rest of the survey. For example, some of the questions made young people reflect, and the survey was seen as a way of giving young people a voice.

Acceptability of survey administration

In terms of the survey administration, some young people liked completing the survey online using a computer. Young people typically said that doing the survey in a group was fine or even fun, but they were also conscious of the fact that in a confined space, they could (and did) look at each other's answers. A few young people said that it could be distracting to do the survey in a group, as some of them started to misbehave. Facilitators and one member of school staff believed this was because answering the SRDS questions had unsettled some students.

Not all young people were persuaded that the survey was truly anonymous, although it was not clear why. For example, a couple said that they did not believe it was anonymous because they had to enter their name. Feedback from the young person participatory group suggests that if students accessed the survey with a login instead of entering their name, they may be more reassured about confidentiality. They also thought that the consent form helped to reinforce the message about confidentiality. However, they wondered whether confidentiality can always be guaranteed when criminal behaviour is reported.

Overall, our approach to data collection seemed broadly acceptable, although we were limited by the spaces schools had available, and some pupils had concerns about anonymity.

Acceptability of qualitative data collection

Young people who took part in the qualitative interview said this was a nice or okay experience, and some noted that the questions discussed were more *'normal'* than the survey questions. The typical reason young people took part was to get out of lessons, but a few also mentioned wanting to share their experiences and wanting to be helpful.

Only a few commented on what may put young people off from agreeing to do an interview. They mentioned finding the experience too intrusive, being afraid to say something wrong or that would get them into trouble and being put off by the fact that the interview is recorded.

One young person suggested that the word 'interview' may be off-putting for some, and a description that makes the interview sound less formal, such as '*friendly question time*', may be more appealing. Another wondered whether some may be more willing to take part if they could do the interview with a computer.

8.4.3. Mechanisms

Overall, there were no problems collecting data (relating to young people's demographics, circumstances and needs) that would be required to explore subgroup effects. These data came from referral information, which schools had easy access to.

8.5. Conclusions

Overall, we conclude that it is possible to recruit young people into an RCT in this context and to conduct randomisation. Collecting outcome data is feasible, although retaining young people throughout the trial period is challenging. The main evaluability barrier observed was in the retention of young people and study attrition. We recommend that future evaluators lengthen the window for baseline data collection and provide additional opportunities for young people to attend data collection sessions.

Several elements of the evaluation were not acceptable to all school staff, facilitators and young people – particularly randomisation and the SRDS. Overall, we think these are necessary elements of a robust impact evaluation assessing the impact on offending. We do not think there are easy alternatives for an evaluator in this context, and we think the best mitigation is to strengthen pre-evaluation communication to reduce mismatched expectations.

9. Findings: Outcomes

9.1. Introduction

As part of the pilot study, we set out to understand whether Face It shows sufficient promise in terms of improvements in key outcomes (resilience, emotional regulation, social and emotional well-being, school attendance and school exclusions) to justify a subsequent efficacy trial.

9.2. Effect sizes and direction of effect

This pilot trial was primarily designed to investigate deliverability, acceptability and evaluability rather than impact. Given that the pilot trial was not designed to be sufficiently well powered to detect impact (and that, in addition to this, we substantially undershot our target of collecting outcome data for 160 young people, primarily due to issues with school recruitment), we only briefly address the quantitative outcome findings from this evaluation.

9.2.1. Post-programme

Overall, all point estimates are in a positive direction (i.e. all represent a 'good' outcome). However, as expected, none are statistically significant. It is not possible to draw strong conclusions from this analysis, given the lack of statistical power and the breadth of the confidence intervals.

Table 15: Post-programme outcome analysis

Outcome	Point estimate	Cohen's d	Confidence interval (around point estimate)
Behavioural difficulties (Strengths and Difficulties Questionnaire)	−1.458	0.008	−4.610; 1.695
Offending (Self-Report Delinquency Scale [SRDS] – variety score)	−1.217	−0.441	−3.063; 0.629
Offending (SRDS – volume score)	−6.136	−0.301	−15.93; 3.659
Victimisation (Problem Behaviour Frequency Scale)	−0.976	−0.075	−5.821; 3.868

Resilience (Children's Hope Scale)	1.480	0.091	−1.617; 4.577
Emotional regulation (The Emotional Regulation Questionnaire)	2.614	0.417	−1.271; 6.498
Social and emotional well-being (Short Warwick–Edinburgh Mental Well-being Scale)	1.868	0.387	-0.957; 4.693

9.2.2. Three-month follow-up

At the three-month follow-up period, all comparisons remain not statistically significant. The direction of effect for behavioural difficulties and victimisation has changed, suggesting worse outcomes for those in the treatment group. However, as noted above, it is not possible to draw strong conclusions from this analysis, given the lack of statistical power and the breadth of the confidence intervals.

Table 16: Three-month follow-up outcome analysis

Outcome	Point estimate	Cohen's d	Confidence interval (around point estimate)
Behavioural difficulties (Strengths and Difficulties Questionnaire)	1.067	−0.165	−2.219; 4.353
Offending (Self-Report Delinquency Scale [SRDS] – variety score)	-0.267	−0.306	−1.836; 1.301
Offending (SRDS – volume score)	−1.639	−0.286	−8.379; 5.102
Victimisation (Problem Behaviour Frequency Scale)	3.343	0.165	−3.745; 10.430
Resilience (Children's Hope Scale)	2.201	−0.18	−0.869; 5.272
Emotional regulation	0.736	0.009	−2.875; 4.347

(The Emotional Regulation Questionnaire)			
Social and emotional well-being (Short Warwick–Edinburgh Mental Well-being Scale)	1.126	0.098	−1.902; 4.155

9.3. Perceptions of impact

9.3.1. Young people

In general, young people were very positive about the impacts of the programme, as they perceived it to support them in managing their emotions and their behaviour. Some reported that this had a positive impact on their attitudes and experiences at school, relationships, social skills and well-being. The perceived intensity of the impact varied. One young person said it had been a '*once-in-a-lifetime'* experience, which affected her more than she expected, and she hoped that the changes she experienced would last for a long time. Another talked about the programme having made a *'seven out of ten'* difference to her life. It was evident that, to a greater or lesser extent, the programme had helped young people feel better about themselves and feel they had more control over their lives.

A typical perceived outcome of the programme (reported largely spontaneously by young people) was learning how to **manage their emotions**, that is, learning what triggers feelings of anger and being given tools and strategies to manage these situations. Young people talked about knowing themselves better, being calmer and being more mindful of the impact they have on others. The extent to which young people perceived this to help them varied, but all talked about the programme having taught them something about self-regulation. As these young people explained:

'I understand how to manage myself more, I'm doing better. It's only like a slight increase, though, not like a drastic boost'.

'If somebody's speaking to me in a certain way, before I would get a bit annoyed, but now I'm a bit more calm … on the inside'.

'Before, I could lash out a bit at my family and stuff, but now, after that, I seem calmer'.

Some linked this outcome to **better behaviour at school**. They thought that they were less reactive and were getting fewer negative points and detentions. They talked about being more respectful of teachers, seeing them as people and not getting angry at them. As this young person explained:

'I stopped giving attitude and being rude and talking back to teachers'.

To illustrate the positive impact on her behaviour, during the interview, a young person showed her report card to the researcher, with a considerable reduction in negative behaviour points since taking part in the programme. She also commented on how her form tutor and head of year noticed her marked improvement in behaviour and are very proud of her for the changes

she has made, and she commented that it is nice that somebody has realised she is actually doing '*good stuff*'.

Young people were asked whether better behaviour at school resulted in improved school attendance. A couple said that because they were less likely to get into trouble and more likely to remain focused on their work, they were more likely to want to go to school.

When probed, however, a few young people said their behaviour at school had not changed, and they did not think there was anything the programme could have done to change how they felt about school. Others did not think the programme affected their attendance.

A few young people commented on how the programme helped to improve their **social skills and relationships**. Some young people met new friends through the programme, but possibly more importantly, they believed their social skills and relationships had improved because they were better able to control their emotions and were more confident. One young person talked about being shy at the beginning of the programme but feeling more confident as it went on. Another talked about having learnt to project her voice more to express her opinions.

Young people also typically reported the programme having a positive impact on their **well-being**, talking about feeling more confident and happier.

9.3.2. School staff

Some school staff members reported that it was difficult to attribute any changes specifically to the programme because of everything else that affects young people's lives. They were nevertheless very positive about the possible impact of the programme and thought it may have helped to create the foundations for young people to establish a more positive relationship with school staff and become more positive about school. They noted that after the programme, young people were happier, better at regulating their emotions, less likely to get into fights and more generally less trouble, reflected in young people receiving fewer sanctions.

9.3.3. Perceived mechanisms of change

As discussed earlier, the evidence shows a high level of acceptability of the programme among **young people** in terms of content, delivery and the principles underpinning the programme, which supported a high level of engagement with and learning from the programme. From young people's accounts, it seems that what helped them engage and supported them in learning included:

- **Trusted bonds with facilitators:** Facilitators' skills and attitudes helped young people quickly established trusted relationships that greatly supported engagement and learning.
- **Face It provided a 'safe space'**: They could express their feelings without being judged; the check-ins with the facilitators were cited as an example of this. Part of this seems to be:
    - A clear message that young people were valued – for example, they made the rules together and were supported to learn from each other

    - A calm environment where young people were not rushed and shouted at, and they could work in ways that made them feel more at ease, for example, while listening to music, chatting with others and choosing the activities they preferred
    - Young people valued the opportunity to learn while having fun, in ways that were very different and, from their perspective, more effective than traditional school learning methods.

- **Useful tools/strategies:** Young people found the content very relevant, as it covered issues that were important to them (e.g. controlling emotions, valuing yourself). Some young people became increasingly aware during the programme of how these issues affected them, their relationships with others and their school experiences. The practical tips for putting the learning into practice were reported as being helpful. As discussed earlier, the impact of these became evident after the programme, when young people started to apply the learning and could see how it helped them, for example, to anticipate and deal with the triggers of anger and to understand how what they say and do may affect others.
- **Duration and intensity:** The sessions being concentrated in a single week supported engagement and the learning consolidation process.

The potential drivers of change reported by **school staff** partly reflect what young people reported as reasons why Face It was a positive experience for them, and include:

- **Trusted bond with facilitators:** Good relationships with the programme facilitators
- **Face It provided a 'safe space':** Being given a safe space to express themselves, reflect on their behaviour and share experiences and views with others while feeling accepted
- **Useful tools/strategies:** Being given learning tools and strategies to cope
- **Duration and intensity:** Having a week-long block provided the opportunity to address issues more openly and honestly
- **Mixing young people with different needs** – A mix of extroverted and introverted young people was seen as supporting peer learning.

9.4. Conclusion

Consistent with the aims of the pilot trial, we are unable to make robust conclusions about the impact of the programme. However, all point estimates of effects at the post-programme stage are positive in direction. This is echoed by the qualitative feedback, where young people and school staff reported that Face It can support improved emotional regulation, improved behaviour and improved social skills and well-being. Key factors seen to lead to these perceived impacts were identified as being the trusted bond with Khulisa facilitators, the programme providing a 'safe space', the programme providing useful and practical tools and strategies (to address coping and emotion management), the duration and intensity of the programme, and the way the programme mixes young people with different needs.

10. Conclusions and recommendations

10.1. Key findings

We primarily set out to establish whether Face It was **deliverable**, **acceptable** and **evaluable**. We also aimed to investigate whether Face It shows sufficient promise in terms of improvements in key outcomes (**outcomes**). A summary of our key findings against each of these is:

Deliverability

- In the recruited schools, it was possible to generate a sufficient number of referrals and to achieve high initial take-up of the programme.
- Ongoing engagement and attendance were relatively high.
- Overall, Face It was delivered with a high level of fidelity. Fidelity was assessed using facilitator self-reports of key Face It activities conducted across the five-day programme. Generally speaking, fidelity was high, with over 80% of key activities being conducted.
- The biggest challenge was in the recruitment of schools. Khulisa experienced significant challenges in recruiting schools, and unfortunately, only **two schools** were successfully recruited (**40%** of the target number of schools and **3%** of those approached). Of the two schools that were successfully *recruited*, one dropped out of the project before delivery could begin with the second cohort. Therefore, we only **retained one** school throughout the project.

Acceptability

- Overall, we conclude that the Face It programme is acceptable to and valued by young people and school staff, having received very positive feedback through the satisfaction survey (for young people) and qualitative interviews.
- There are mixed findings about the intensity of the programme, though these tend to be overcome once a school commits to the programme, it is delivered, and young people and teachers perceive the benefits.

Evaluability

- Overall, we conclude that it is possible to recruit young people into an RCT in this context and to conduct randomisation.
- While collecting outcome data is feasible, retaining young people throughout the trial period is challenging, and attrition was high at the post-programme time point (36% overall).
- Several elements of the evaluation were not acceptable to all school staff, facilitators and young people – particularly randomisation and the SRDS. Overall, we think these are necessary elements of a robust impact evaluation assessing impact on offending. We do not think there are easy alternatives for an evaluator in this context.

Outcomes

- Consistent with the aims of the pilot trial, we are unable to make robust conclusions about the impact of the programme. Given that the pilot trial was not designed to be sufficiently well powered to detect impact (and that, in addition to this, we substantially undershot our target of collecting outcome data for 160 young people, primarily due to issues with school recruitment), no quantitative findings are statistically significant.
- All point estimates of effect at the post-programme time point are positive in direction (i.e. if they were statistically significant, they would suggest the programme had reduced behavioural difficulties, offending and victimisation and improved resilience, emotional regulation, and social and emotional well-being). By the three-month follow-up period, the direction of effect had changed for behavioural difficulties and victimisation (suggesting these were worse in the treatment group).
- Initial improvements for the key programme outcomes are suggested by qualitative feedback, where young people and school staff reported that Face It can support improved emotional regulation, improved behaviour, and improved social skills and well-being. Key factors seen to lead to these perceived impacts were identified as the trusted bond with Khulisa facilitators, the programme providing a 'safe space', the programme providing useful and practical tools and strategies (to address coping and emotion management), the duration and intensity of the programme, and the way the programme mixes young people with different needs.

10.2. Performance against progression criteria

Table 16 shows the performance of Face It against the pre-specified progression criteria, reflecting the findings discussed above.

Table 17: Performance against monitoring and progression criteria

Criterion	Description	Red, amber, green (RAG)	Status (RAG)
Recruitment I	Number of schools successfully recruited to the evaluation in the first 3 months of the pilot	Red: 3 or fewer Amber: 4 Green: 5	2
Recruitment II	Proportion of the target number of eligible young people (160) recruited within	Red: <60% of target Amber: 60–79% of target Green: 80–100% of target	44% (largely a function of issues with school recruitment)

	schools in the first 3 months of the pilot		
Randomisation	Proportion of recruited young people randomised to control or treatment groups	Red: <50% Amber: 50–79% Green: 80–100%	97%
Retention	Proportion of young people in the intervention arm completing the intervention	Red: <50% Amber: 50–79% Green: 80–100%	88%
Fidelity	Assessed by the Lab through comparing the facilitators' self-report of activities conducted during the 5-day programme with Face It's list of key activities.	Red: <50% of sessions meet at least 80% of criteria Amber: 50–79% of sessions meet 80% of criteria Green: 80–100% of sessions meet 80% of criteria	75%
Outcomes I	Data completeness for young people's self-reported outcome variables at post-intervention for treatment and control groups	Red: <35% Amber: 35–89% Green: 90–100%	63% retained (36% attrition)
Outcomes II	Data completeness for young people's self-reported outcome variables at 3 months post-intervention for	Red: <35% Amber: 35–89% Green: 90–100%	54% retained (46% attrition)

	treatment and control groups		
Outcomes III	Directional change in outcome variables for treatment and control groups	Red: at least one outcome measure indicates negative results and no positive results[13] Amber: null or mixed results. Green: at least one outcome measure indicates positive results and no negative results	Null or mixed results

10.3. Evaluator's judgement on readiness for trial

While this pilot trial makes a number of positive conclusions about the programme's deliverability (it can be delivered with fidelity) and acceptability (many young people really enjoy the programme and perceive it to be beneficial to them), we conclude that the five-day intensive Face It programme is **not ready for a full-scale trial**.

The central reason for this is that recruiting a sufficiently large number of schools to enable a well-powered efficacy trial in this context would not currently be possible within any reasonable recruitment window. While we think that a different evaluation approach focusing on different outcomes would have been more acceptable to schools and had greater success, this would still not overcome the programme-based reasons for schools' non-participation.

However, an RCT **may be feasible** if some fundamental changes are made to the recruitment approach, the programme itself and the approach to the evaluation (see recommendations below).

We conclude that the programme is **largely evaluable** (once schools are recruited, many aspects of the evaluation and evaluation procedures were straightforward and worked well), with the caveat that there was relatively high study attrition (which we see as the second main challenge indicated by this project). We believe this could be overcome in any future

[13] Null results: 95% confidence intervals that cross 0; negative results: 95% confidence intervals that are entirely below 0; positive results: 95% confidence intervals that are entirely above 0.

studies by making some straightforward changes to the data collection approach (see recommendations below).

10.4. Recommendations for programme delivery

Our findings and suggestions from the key stakeholders suggest a number of recommendations to improve the recruitment approach for the programme and delivery of the programme itself:

- *Recommendation 1 – Cast the school recruitment net wider from the outset:* Given the low conversion rate (3% of contacted schools participated in the programme), it would be necessary to approach many more schools when recruiting for any future trial. The figures from this trial suggest that we would have had to approach approximately 170 schools (rather than 60) to achieve our target of recruiting five schools. This would, however, introduce other difficulties in terms of the time and resources required to reach out to many more schools and manage communications arising from that. **Casting the net wider would be necessary to recruit a sufficient number of schools and produce a trial which is sufficiently well-powered to detect programme impact.**
- *Recommendation 2 – Have longer lead-in times for school recruitment and more opportunity to lay the groundwork:* A longer recruitment window would have been more straightforward to manage and would also have given more opportunity for Khulisa to raise the profile of its organisation and programme among a broader network of schools. One of the Khulisa staff members believed that, while Khulisa is effective in engaging schools with which it has established a trusted relationship, to reach new schools, Khulisa's visibility would need to be increased (e.g. by participating in community events and school fairs and improving its marketing and social media presence), which requires time. **Allowing more lead-in time would support stronger relationships with schools, increase programme visibility and improve recruitment among schools that may not be familiar with Khulisa. This will support producing a trial sufficiently well powered to detect programme impact.**
- *Recommendation 3 – Align the timing of programme delivery to when it is most convenient for schools and students:* Given the difficulties reported around expecting strong engagement from schools in the autumn and summer terms, it would be better for a future trial to aim to deliver in the spring term (avoiding the onboarding/settling in of the autumn term and the exams of the summer term). **Delivering the project in the spring term may maximise schools' and students' capacity to engage, improving recruitment and supporting a sufficiently well-powered trial.**
- *Recommendation 4 – Shorten the programme to ease schools' concerns:* We understand that there is a less intensive three-day version of Face It offered by Khulisa. This may be more appealing to schools that are concerned about taking pupils who are struggling out of class for long periods of time. **Offering a shorter version of the programme may make participation more appealing to schools, improving recruitment and supporting a sufficiently well-powered trial.**
- *Recommendation 5 – Offer the programme to a younger cohort:* Schools may be more amenable to the intensity and duration of the Face It programme if it is being

offered to younger groups, where missing lessons may be perceived to have less of a negative impact. **Offering the programme to a younger cohort may make it more acceptable to schools, improving recruitment and supporting a sufficiently well-powered trial.**

- *Recommendation 6 – Offer flexibility in venues:* Khulisa staff recommended offering schools an alternative external venue if they struggle to find a classroom for the duration of the programme. **Offering more flexibility in venue arrangements may reduce logistical barriers for schools, increasing participation and supporting a sufficiently well-powered trial.**
- *Recommendation 7 – Provide clearer guidance to schools:* It would be useful to further emphasise the requirement to have a split of pupils exhibiting externalising/internalising problems in communications with schools during the onboarding process and the reasons it is important. More advice and guidance around how to identify pupils with these characteristics using existing school data and teacher judgement are likely to be helpful. **Providing clearer guidance may improve the consistency and appropriateness of pupil selection, supporting programme delivery and delivering the programme in the way in which it is hypothesised to be most effective.**
- *Recommendation 8 – Share more information about the programme in advance:* Both facilitators and school staff said young people who are offered a place should have a better understanding of what the programme involves. A participant suggested that it would be helpful for Khulisa to run an assembly to explain what Face It involves and motivate students to take part. We agree that these measures could be helpful, although we would caution that in the context of an RCT, raising awareness of the programme should only be conducted with those assigned to the treatment group prior to formally beginning programme sessions (to mitigate risks around causing disappointment, resentful demoralisation and differential attrition in the control group). **Sharing more information with pupils in the treatment group prior to the programme may improve motivation and attendance, increasing participant recruitment, reducing data loss and potential bias, and supporting a sufficiently well-powered trial.**
- *Recommendation 9 – Carefully communicate about the aims of the project to parents:* In our communication with young people, we took great care to simultaneously i) be transparent about the overall aims of the project (which included exploring the impact of a school-based programme on offending behaviours), ii) make it clear that Face It is also intended to impact a range of other outcomes beneficial for young people in general and iii) emphasise that young people were not selected because they were suspected to have gotten into trouble. Future work should ensure that this messaging gets through to parents who, understandably, may have concerns without this context. **Strengthening communication with parents may reduce the risk of opt-outs, increasing participant recruitment, reducing data loss and potential bias, and supporting a sufficiently well-powered trial.**
- *Recommendation 10 – Make the time commitment clear up front:* It is clear that for some young people, the fact that attending the programme would imply missing other activities and catching up with academic work only became clear partway through their involvement with the programme. To reduce non-attendance or programme drop-out, the implications of programme attendance should be made clear to pupils prior to being offered a place on Face It. **Being clear about the time commitment**

up front may reduce drop-outs and support consistent attendance, improving programme fidelity and supporting a sufficiently well-powered trial.

- *Recommendation 11 – Reduce the time commitment to ease young people's concerns:* As noted above in the section on school recruitment, there is evidence that schools are concerned about the amount of time Face It demands out of the school timetable. There is some evidence that this commitment can be concerning for young people, too. Shortening the programme may alleviate these concerns. We understand that there is a less intensive three-day version of Face It offered by Khulisa, which may be more acceptable to young people. **Reducing the time commitment may improve both school and pupil engagement, contributing to stronger recruitment and a sufficiently well-powered trial.**

The majority of our recommendations here focus on mitigating challenges with school recruitment. While we believe that implementing this full list of recommendations would improve school recruitment in any future trial, we think that Recommendation 4 (*Shorten the programme to ease schools' concerns)* requires particular attention. We are aware that Khulisa runs a **three-day version of the Face It programme**, and this may be a more suitable candidate for any future trial. However, there are inevitable trade-offs here, and a shorter version of the programme may be less impactful (and we note the feedback from young people that they ended up appreciating the length of the five-day programme).

10.5. Recommendations for future evaluation

Our findings and suggestions from the key stakeholders suggest a number of recommendations to improve the approach to evaluating the programme:

- *Recommendation 12 – Provide a longer window and more opportunities for baseline data collection:* For this trial, we only conducted one baseline data collection session per cohort at baseline, which meant that young people missing that first session could not participate in the project. We think it would be wise for future evaluators to either run multiple initial recruitment/baseline data collection sessions or to provide a survey self-complete option for young people who were not able to attend the first session. Either option would provide multiple opportunities for young people to participate and invariably drive up the referral-to-recruitment conversion rate. We note that the former option would carry risks around adding to school burden, whereas the latter option may have negative consequences for acquiring truly informed consent from young people and for outcome data quality. **Providing multiple opportunities for baseline data collection may increase participant recruitment, reduce data loss and potential bias, and support a sufficiently well-powered trial.**
- *Recommendation 13 – Strengthen pre-session communication to reduce mismatched expectations:* While both Khulisa and the Lab made deliberate efforts to ensure that messaging at every stage was transparent about the project's aims – and simultaneously made clear that the programme is *not* exclusively for young people involved in antisocial and offending behaviour – it is evident that this message was not always received or understood as intended. We recommend that any future evaluation in this context include more deliberate and structured communication prior to onboarding and data collection. This must make the relationship between the

programme and evaluation to youth violence clear, and it could include i) a short explainer video or slide deck aimed at young people and parents, using plain language to describe the project's purpose, ii) a brief summary of the types of survey questions asked and why they are included and iii) clear, consistent framing around how young people were selected to be involved. However, we note that clearer and more accessible communication might still lead to low engagement. **Strengthening pre-session communication may improve young people's and parents' understanding and trust in the project, supporting better engagement, data quality and a sufficiently well-powered trial.**

There are a number of evaluation-based challenges where we made no strong recommendations, largely because we concluded those challenges were baked into the context in which we were conducting our trial (e.g. a pilot trial with offending as a key outcome), and the available mitigations were limited. However, we note that future trials with a different funding context may benefit from:

- *Avoiding self-report measures on sensitive topics to improve school recruitment and acceptability of data collection:* In our trial, the SRDS was off-putting to some schools at the recruitment stage and continued to be a point of contention with programme managers and facilitators, school staff and some young people. A future trial could focus on the more positive or 'strengths-based' intended outcomes of the Face It programme.
- *Using a wait-list design to improve school recruitment and acceptability of randomisation:* Using a wait-list design would be likely to be substantively more acceptable to school staff and young people. However, this would prohibit any investigation of long-term impact.

10.6. Limitations and lessons learnt

There are a few limitations of the pilot evaluation which should be considered in interpreting the results we present:

- *Missing ImpactEd data:* We had initially planned to use the ImpactEd platform to collect data on school attendance and exclusions (as outcome measures). Unfortunately, we were not able to access these data. Facilitating access required a series of administrative steps to be undertaken within schools, which our two recruited schools were not able to complete. We do not know why this was the case, but we suspect it was because school staff were busy and because the project and evaluation were largely concluded by the time we began to step up our efforts to chase for data access (and so schools perceived limited benefit to investing further resources into the work). Overall, we would suggest that future evaluators secure this access earlier rather than later in the project life cycle. We would not recommend conducting this while schools are being formally recruited, as this may add to schools' perceived burden of being involved in the project and exacerbate existing school recruitment challenges.
- *Missing data on study attrition:* In retrospect, the evaluation team did not do enough to investigate the reasons for young people withdrawing from the project and the reasons for absence from data collection sessions. We suggest that future evaluators

follow up with school staff immediately after a formal withdrawal is received and immediately after data collection sessions to clarify these reasons.

- *Missing data on parents' perspectives:* As noted above, acquiring parental engagement with the project was challenging. While this wasn't a substantive weakness of our project, which was always primarily designed to make use of young people's self-reported outcome data, triangulation of data across multiple sources, including from parents, would strengthen any future impact evaluation. For the qualitative work in particular, we attempted to make use of existing opportunities when parents would be interacting with the school (e.g. parents' evenings) to support improved engagement from parents, but this was unsuccessful. For quantitative surveys, teachers were asked to email survey links to parents, but again, response rates were very low. Overall, we suspect that acquiring this sort of engagement is just challenging. Teacher reports ultimately may be a better source of data (in addition to young people's self-reports), though we note that this would add to the school burden.
- *Sources of fidelity data:* Our fidelity or adherence data are based on practitioner self-reports on a checklist (rather than independent observation) and therefore may be subject to potential bias.

Annexes

Annex A: Summary of Face It programme using the TIDieR framework

Name: Provide a name or phrase that describes the intervention.	Face It
Why: Describe any rationale, theory, or goal of the elements essential to the intervention.	Due to the high levels of trauma typically experienced by Khulisa's participants, Face It is scaffolded around trauma-informed methodologies like Bruce Perry's neurosequential model (Perry 2006, 2009). According to this model, in order to manage the impact of trauma in the brain, participants first need help to 'regulate' their emotions, to calm and ground them sufficiently to reduce their level of arousal (amygdala hijack). This then enables them to 'relate' in a way that enables them to trust others to keep them safe to try new things. Only then is it possible to 'reason' with them, to work with the cognitive part of the brain. The process of how change happens set out by Bruce Perry's model also ties into our Theory of Change. Khulisa's theory of change supports the importance of six clusters of social and emotional skills outlined in the EEF's Spectrum Framework to support young people's social and emotional wellbeing and positive longer-term life outcomes (Education Endowment Foundation, 2017). These competencies are inter-connected and at times overlapping. The aim of Khulisa's programme is to develop this broad spectrum of social and emotional skills, but improved

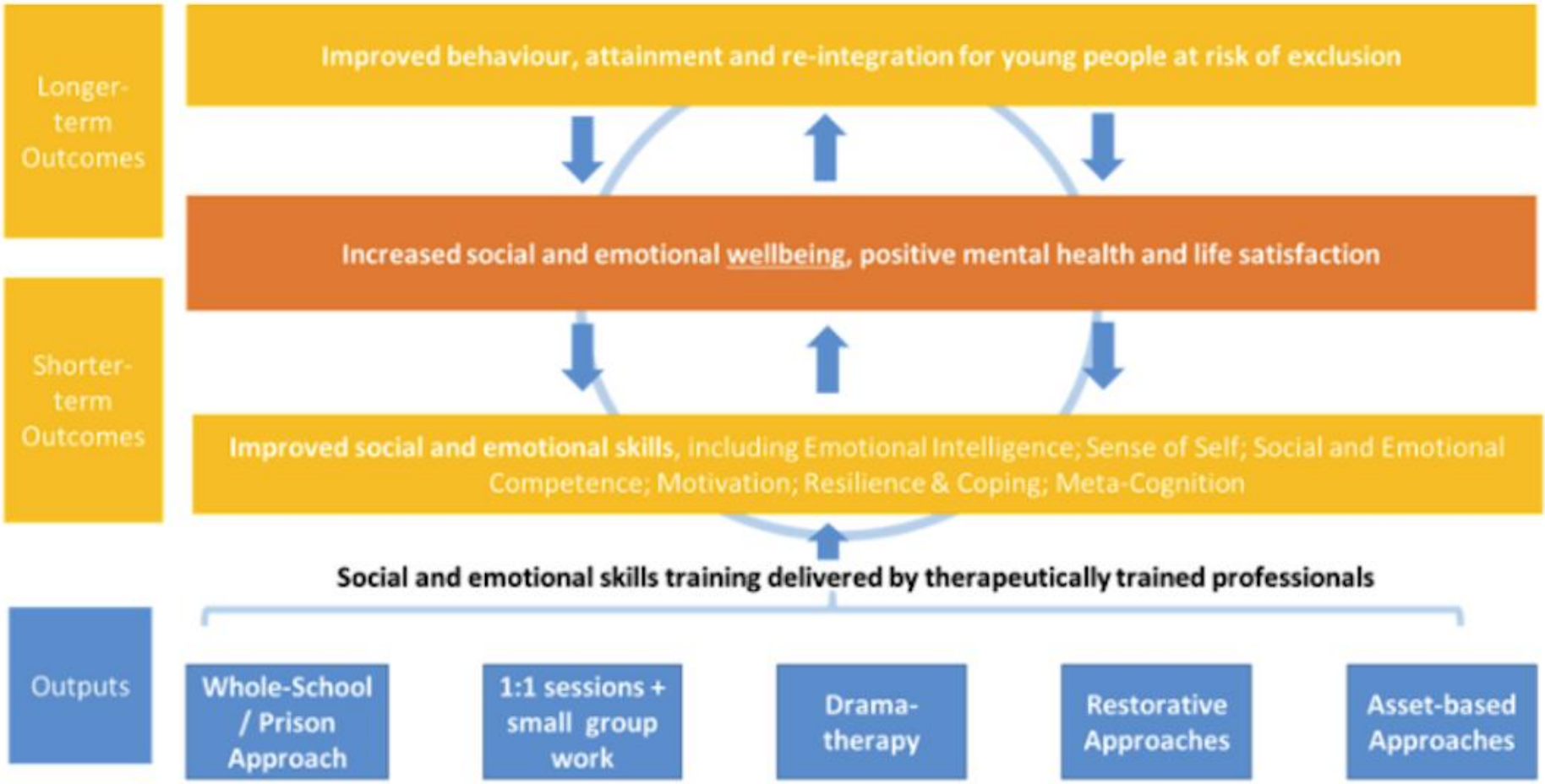
	resilience, coping skills, and self regulation are prioritised as key short-term outcomes as in order to improve social and emotional wellbeing, we must first improve young people's coping skills, emotional regulation and resilience. Heading straight for the 'Reasoning' part of the brain is unlikely to be successful if the young person is dysregulated, stressed and disconnected. First, the person needs help to regulate and calm their flight / fight / freeze responses and build resilience and coping skills to lay the foundation for them to engage with more challenging parts of the programme and wider life, in order to improve levels of wellbeing.
What - Materials: Describe any physical or informational materials used in the intervention, including those provided to participants or used in intervention delivery or in training of intervention providers. Provide information on where the materials can be accessed.	Before they ever deliver a programme, facilitators are given access to: 1. The full programme methodology 2. The scaffolding process 3. An example high level lesson plan that gives a skeleton overview of the process and content of a programme, to enable them to gradually immerse themselves in the detail that sits behind it. 4. A recommended reading list on each of the theoretical concepts that underpin the programme. In the set up phase of the programme, they have access to the following documents: ● Our participant profile- shared with schools to support the referral process. ● A draft letter to parents - sent out by schools. ● A safer recruitment doc that explores our initial safeguarding processes and procedures with Partners ● A Programme Venue Health and Safety Risk Assessment Supporting materials for the delivery of the programme: ● Lesson plan templates - these are edited for each programme. ● A programme fidelity checklist highlighting the core elements every Face It programme should have ● A list of icebreakers and games. ● The poems and games we use on the programme. ● 1 page guides to key activities. ● Handouts to share with participants as part of the programme activities.
What - Procedures: Describe each of the procedures, activities, and/or processes used in the intervention, including any enabling or support activities.	**Pre-programme one-to-one session:** Discussion of pre-programme survey **Pre-programme group session**: Facilitators build a sense of safety, and empower young people to decide to go ahead or not go ahead with the programme. **5-day intensive programme:** ● Day 1 - Participants are introduced to and contribute to the group code, they work on identifying their habits

	and triggers. ● Day 2 - Participants continue to work on understanding the power of triggers, and their physiological and emotional responses when triggered. ● Day 3 - Participants work on understanding different types of violence, power and control, and work on understanding how unmet needs can contribute to reactions. ● Day 4 - Participants practice and embed effective strategies to manage challenging situations and emotions. ● Day 5 - Participants build their self-esteem and are encouraged to connect with their future aspirations. **Post-programme group session:** The focus of this session is reflection on the programme and participants' personal strengths. **Post-programme one-to-one session:** Discussion of personal progress and post-programme survey.
Who: For each category of intervention provider (such as psychologist, nursing assistant), describe their expertise, background, and any specific training given.	Khulisa ensures that all facilitators are therapeutically trained in art or dramatherapy so that they have the skills and training required to create 'safe containment' for the group and to continually assess the pace and content of programmes based on participants' needs. There is no minimum level of therapeutic training, but each programme has someone who is therapeutically trained. All programme managers and associate facilitators receive a stringent 3-day training programme to ensure that they are well-versed in programme content and methodology before they ever deliver a programme. They will also complete levels 2 and 3 safeguarding training. Each programme manager observes and supports 8 programmes as a co-facilitator before they lead a programme of their own. All facilitators are given access to our written methodology and scaffolding process along with a programme manual setting out key activities corresponding to specific parts of the neurosequential model. Facilitators receive ongoing support via monthly clinical supervision, fortnightly/monthly supervision with their line manager, monthly group supervision and peer to peer training support every 6 weeks (during half-terms as there will be no delivery). In addition during half-terms/summer provision, Face It offers a weekly optional online session for staff to practise activities together. Facilitator backgrounds include counselling and youth work, drama and movement therapy, education, education psychology, and dance movement therapy.
How: Describe the modes of delivery (such as face to	**Pre-programme one-to-one session:** One 60 minute in-person session

face or by some other mechanism such as internet or telephone) of the intervention and whether it was provided individually or in a group.	**Pre-programme group session:** One 2.5 hour in-person session **5-day intensive programme**: One intensive 5-day programme delivered in person each day from 9:30-15:30, Monday-Friday **Post-programme group session:** One 2.5 hour in-person session **Post-programme one-to-one session:** One 60 minute in-person session
Where: Describe the type(s) of location(s) where the intervention occurred, including any necessary infrastructure or relevant features.	School-based Khulisa shares a list of programme venue room requirements with schools as part of our service level agreements.
When and how much: Describe the number of times the intervention was delivered and over what period of time including the number of sessions, their schedule, and their duration, intensity, or dose.	The following intervention components are spread over a period of approximately 6 weeks. This estimates one session a week, with a break for half term. **Pre-programme one-to-one session:** One 60 minute session **Pre-programme group session:** 2.5 hours **5-day intensive programme**: One intensive 5-day programme delivered each day from 9:30-15:30 Monday-Friday **Post-programme group session:** 2.5 hours **Post-programme one-to-one session:** One 60 minute session
Tailoring: If the intervention was planned to be personalised, titrated or adapted, then describe what, why, when and how.	Modification within programme description based on ongoing facilitator assessment of group dynamics. For example, facilitators choose an initial icebreaker from a predefined list of options, depending on the dynamic of the group and whether they aim to calm or energise participants.
Modification: If the intervention was modified during the course of the study, describe the changes (what, why, when, and how).	See report.
How well (planned): If adherence or fidelity was assessed, describe how and by whom, and if any strategies were used to maintain or improve fidelity, describe them.	Our programme delivery is often emergent, and dependent on the emotional regulation and capacity of Khulisa's participants. Throughout each day of the programme various icebreakers, games and energisers are used to assist in the regulation of emotions. This ensures Khulisa's Facilitators can safely contain participants, whilst still providing an opportunity for learning and education. Our programme balances the flexibility afforded by this emergent process with a very robust methodology and scaffolding process based on Bruce Perry's **neurosequential model** (Perry 2006, 2009). To further ensure programme fidelity and to quality assure our programmes we have identified and

	closely monitor core milestones and activities every Face It programme should contain. This means that as different as the configuration of programme activities in response to specific needs of the group might be, we will have visibility of and an ability to ensure every Face It programme touches on key milestones that we have identified are central to effecting change. These are reviewed every six weeks as a measure of accountability and quality assurance for the team.
How well (actual): If actual adherence or fidelity was assessed, describe the extent to which the intervention was delivered as planned.	See report.

Annex B: Khulisa Theory of Change



Annex C: Outcome measures in detail

Type of outcome	Outcome measured	Instrument	Completed by	Number of items	Age suitability (young person)	Subscales used	Scoring[14]	References
Primary	Behavioural difficulties	Strengths and Difficulties Questionnaire (SDQ) - post-programme self-assessment	Young person	25	11-17	All subscales including: ● Emotional symptoms ● Conduct problems ● Hyperactivity/inattention ● Peer relationships problems ● Prosocial behaviour	**Total difficulties score:** A score from 0-40 is generated by summing scores from all the subscales, except the prosocial subscale. While the total difficulties score is the primary outcome, we will also examine the total difficulties score when broken down into the externalising score (the sum of the conduct and hyperactivity scales), and the internalising score (the sum of the emotional and peer problems scales).	Goodman,1997 [15]
Secondary	Behavioural difficulties	SDQ - 3-month follow-up self-assessment	Young person	25	11-17	All subscales including: ● Emotional symptoms ● Conduct problems ● Hyperactivity/inattention ● Peer	**Total difficulties score:** A score from 0-40 is generated by summing scores from all the subscales, except the prosocial subscale. Summed score of subscales give a score ranging from 0-40s	Goodman,1997 [16]

[14] Data will be collected and outcome measures scored by the Ending Youth Violence Lab

[15] Goodman R (1997) The Strengths and Difficulties Questionnaire: A Research Note. Journal of Child Psychology and Psychiatry, 38, 581-586.

[16] Goodman R (1997) The Strengths and Difficulties Questionnaire: A Research Note. Journal of Child Psychology and Psychiatry, 38, 581-586.

						relationships problems ● Prosocial behaviour		
	Behavioural difficulties	SDQ - parent assessment	Parent/care giver	25	4-17	All subscales including: ● Emotional symptoms ● Conduct problems ● Hyperactivity/inattention ● Peer relationships problems ● Prosocial behaviour	**Total difficulties score:** A score from 0-40 is generated by summing scores from all the subscales, except the prosocial subscale.	Goodman,1997 [17]
	Offending	The Self-Report Delinquency Scale	Young person	19	10-17	Does not have subscales	**Variety of delinquency score:** Sum the number of items the respondent answers 'yes' to: • Yes = 1 • No = 0 Produces a score that ranges from 0-19. **Volume of delinquency score:** Summing the point values when respondents report a number of times. Point values are assigned as follows:	Thornberry and Krohn, 2000 [18]

[17] Goodman R (1997) The Strengths and Difficulties Questionnaire: A Research Note. Journal of Child Psychology and Psychiatry, 38, 581-586.

[18] Thornberry, T.P., & Krohn, M.D. (2000). The self-report method for measuring delinquency and crime. Measurement and Analysis of Crime and Justice, 4, 33-83.

							• Once = 1 • Twice = 2 • 3 times = 3 • 4 times = 4 • 5 times = 5 • Between 6 and 10 times = 6 • More than 10 times = 11	
	Victimisation	Problem Behaviour Frequency Scale	Young person	10	Adolescents	● Overt victimisation ● Relational victimisation	A score ranging from 6-36 achieved by summing scores of the two subscales	Farrell et al., 2016[19]
	Resilience	Children's Hope Scale	Young person	6	8-16	All subscales ● Agency ● Pathways	A score ranging from 6-36 achieved by summing scores of the two subscales	Snyder et al., 1997[20]
	Emotional regulation	The Emotional Regulation Questionnaire	Young person	10	10-18	● Cognitive reappraisal ● Expressive suppression	Summed score of subscales. The higher the score, the greater the use of emotion regulation strategies, conversely lower scores represent less frequent use of such strategies	Ioannidis & Siegling, 2015[21] Gross & John, 2003[22]

[19]Farrell, A. D., Sullivan, T. N., Goncy, E. A., & Le, A. T. H. (2016). Assessment of adolescents' victimization, aggression, and problem behaviors: Evaluation of the Problem Behavior Frequency Scale. *Psychological assessment*, *28*(6), 702.

[20] Snyder, C. R., Hoza, B., Pelham, W. E., Rapoff, M., Ware, L., Danovsky, M., ... & Stahl, K. J. , "The development and validation of the Children's Hope Scale," *Journal of Pediatric Psychology*, Vol. 22, No. 3, 1997, pp. 399–421.

[21] Ioannidis, C. A., & Siegling, A. B. (2015). Criterion and incremental validity of the emotion regulation questionnaire. Frontiers in Psychology, 6, Article 247. https://doi.org/10.3389/fpsyg.2015.00247

[22] Gross, J.J., & John, O.P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. Journal of Personality and Social Psychology, 85, 348-362.

								Gullone & Taffe, 2012[23]
	Social and emotional wellbeing	The Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS)	Young person	7	11+	Does not have subscales	The SWEMWBS is scored by first summing the scores for each of the seven items, which are scored from 1 to 5. The total raw scores are then transformed into metric scores using the SWEMWBS conversion table resulting in a score ranging from 7-35.	Melendez-Torres et al., 2019[24]
	School attendance	ImpactEd data	Admin data	N/A	N/A	N/A	Number of absences during the pilot period	N/A
	School exclusions	ImpactEd data	Admin data	N/A	N/A	N/A	Number of exclusions during the pilot period	N/A

[23] Gullone, Eleonora; Taffe, John (2012). The Emotion Regulation Questionnaire for Children and Adolescents (ERQ–CA): A psychometric evaluation.. Psychological Assessment, 24(2), 409–417. doi:10.1037/a0025777

[24] Melendez-Torres, G., Hewitt, G., Hallingberg, B. et al. Measurement invariance properties and external construct validity of the short Warwick-Edinburgh mental wellbeing scale in a large national sample of secondary school students in Wales. Health Qual Life Outcomes 17, 139 (2019). https://doi.org/10.1186/s12955-019-1204-z

Annex D: Ethics and data protection

Ethics

Overview

This trial was self-assessed as being high risk due to the inclusion of high risk participants in the form of vulnerable young people. As a result we sought ethical approval from an independent panel of external experts with experience of working with vulnerable children and experience with safeguarding and child protection.

The independent ethics review committee (ERC) reviewed the following information:

- Ethical review form
- Consent forms and information sheets for young people and parents/caregivers of young people
- Student interview topic guide
- Safeguarding and distress protocol

The ERC discussed any issues raised by the research with The Lab with the aim of finding solutions that meet ethical requirements. The reviewers and the project manager agreed solutions to any outstanding issues. The ERC was happy to approve the project with the inclusion of these amendments.

We conducted a separate internal ethical review for the materials for the young person participatory group, as we judged this to be a separate package of lower risk work. This is because we intended to ask young people to comment on our proposed materials and approach rather than on the intervention in detail. This means that any personal reflections and experiences will be limited

The participatory group materials were approved through an internal ethics review process.

Informed Consent

The Head Teacher or other suitable member of each participating school's Senior Leadership Team provided written consent for the school to participate in this study. This took place prior to the Lab/Khulisa receiving any referral data.

We provided parents/caregivers of young people with an information sheet and withdrawal form, providing the opportunity to withdraw their child from the evaluation prior to seeking consent from young people. Any parents/caregivers who didn't want their child's data to be used in the evaluation were able to withdraw their child from the trial prior to any data collection.

We obtained written consent from young people to take part in the evaluation prior to baseline data collection. We also obtained written consent at every data collection activity (embedded within surveys, at the beginning of YP interviews, at the beginning of all other stakeholder interviews - parents/caregivers, facilitators, key school contact)..

We took steps to ensure consent is fully informed, including comprehensive, but accessible, information sheets, going over the information sheet content verbally, and giving plenty of opportunity for participants to ask any clarifying questions. The information sheets included information about the nature and purpose of the study, the organisations involved, what data will be collected, what will happen to it and why, and offer participants the opportunity to opt out of the research. The information sheets were sent via email and included a link to the privacy notice and explained that we will ensure confidentiality and anonymity in reports, providing all information in clear, accessible language.

We reviewed the information sheets and consent forms with the young person participatory group to ensure that the information is clear and accessible.

At any point until 1st May 2024, participants or their caregivers could withdraw their or their child's participation from the trial. If participants withdraw from the trial before any of their personal data is shared with the Lab, then the young person would not participate in the evaluation and no personal data will be shared.

We made it clear to participants that we will use their information to inform the findings of our evaluation, which would be incorporated into a report, or other publicly publishable materials. However, no identifying information will be disclosed in any such materials.

Safeguarding

Given the potential vulnerability of the young people involved in this evaluation, we have ensured that we have a stringent safeguarding plan in place throughout.

All researchers with any contact with children will have an enhanced DBS check, will have completed NSPCC Safeguarding training, and will familiarise themselves with the BIT Group Safeguarding Policy and the project safeguarding and distress protocol.

Researchers conducting in-person data collection in schools will comply with all school requirements for visitors.

Data protection

We followed appropriate data protection processes in accordance with BIT processes, including completing a Data Protection and Security Checklist and Data Protection Impact Assessment which have both been reviewed and approved by BIT's legal team.

The legal basis for processing personal data was legitimate interest and the legal basis for processing special category data was scientific interest. Article 6(1)(f) of UK GDPR states that "*processing is necessary for the purposes of the legitimate interests pursued by the controller or by a third party except where such interests are overridden by the interests or fundamental rights and freedoms of the data subject which require protection of personal data, in particular where the data subject is a child*."

The Lab has determined that there is a genuine purpose to process this data. The purpose of processing personal data is to assess whether it will be possible to conduct a full-scale randomised controlled trial evaluation of the Face It intervention, through assessing feasibility, acceptance, and evaluability of the intervention, and estimating initial outcomes of the intervention. Answering this question will help us, YEF, and the wider education and youth violence prevention sector to better understand effective support to vulnerable young people at risk of exclusion and becoming involved in violence. This will inform the evidence around what works to improve the health, behaviours and life outcomes of young people, particularly those at risk of, or who have engaged in, violent behaviours. Data processing is necessary to complete a robust evaluation. The Lab does not consider that collecting and gathering data for this trial will interfere with individuals' interests, rights or freedoms. The data subjects will include: at-risk youth, caregivers of at-risk youth, the delivery team at Khulisa, and school staff responsible for the relationship with Khulisa.

All data shared with BIT was processed in line with its data protection policy. A summary of this policy can be found below. In the analysis, BIT promotes data quality and security through the following measures.

- All variables clearly named, coded and labelled before analysis.
- Checks on the data received carried out for valid values, range, and consistency against already held data.
- Any modifications to datasets recorded in the analysis code.
- Original raw datasets will never be amended.
- Access to the project data was restricted to project personnel.
- All data stored by BIT will be backed up.

Data management

All quantitative and qualitative data was stored in a secure Google Folder where access is restricted to only researchers conducting the analysis. Data will be deleted upon completion of the project.

After the project has been completed, data will be shared with YEF, pseudonymised, and stored in the YEF Evaluation Data Archive. All projects funded by YEF store data in this way. Data in the archives is separated from any identifiable information.

Approved researchers may apply to access YEF data via the Office of National Statistics (ONS) secure research service. This will be accessed via their own project space created in the ONS secure research environment by the ONS. They may apply to the Department of Education and Ministry of Justice to access the linked National Pupil Database-National Police Computer data, and if successful it will be made available here to combine with the YEF evaluation data, using pupil matching reference numbers. Researchers will only be able to access pseudonymised data. All results will be published in a de-identified form.

This enables YEF to assess the impact of their projects long-term, as well as to conduct quality assurance, reanalysis and methodological exploration across the outputs and results published in the YEF funded evaluation reports; and pooled analysis of data from multiple interventions, to explore what works in different geographic areas and for particular subgroups (e.g. gender, ethnicity, child who have experienced periods of care etc) that may be identified from the data collected from the projects themselves or from the linked administrative datasets.

Further information about this process can be found on the YEF website.[25]

Quantitative data

Survey data
We used SmartSurvey to collect the survey data. SmartSurvey produces a spreadsheet where one row is a survey response. This was used to code the survey outcomes using the methods outlined in the outcome measures table.

Surveys asked participants to record their name. This enables us to link survey responses with demographic data and other outcome measures. Once survey responses have been linked, participants' names will be removed.

ImpactEd and programme administrative data
Khulisa is responsible for providing us with both the ImpactEd and programme administrative data. All data shared with the lab by Khulisa will be received via a secure transfer link (Virtru or Quatrix).

[25]https://youthendowmentfund.org.uk/wp-content/uploads/2021/07/YEF-Data-Guidance-Projects-and-Evaluators.pdf

Programme administrative data includes the *referral data* and the *programme delivery data* (e.g. attendance sheets, fidelity checklists). *Referral data* was collected via an online form (located on FormAssembly) completed by school staff. Khulisa downloaded the data in a spreadsheet and shared the relevant data with the lab. *Programme delivery data* was collected via Khulisa, and shared with the lab.

Qualitative data

Interview transcripts
Interview recordings were uploaded to McGowan for transcription. All interview recordings were transcribed verbatim to ensure accuracy and facilitate subsequent analysis. Transcripts were anonymised by assigning unique identifiers to each participant, replacing their names or any identifying information. Transcripts, observation notes, and any additional relevant documents were securely stored in a password-protected file area. Access to the data was restricted to only project team members involved in the analysis. Recordings will be deleted after completion of the project.

BIT data protection policy summary

The General Data Protection Regulation (GDPR) imposes certain obligations upon Behavioural Insights Limited (BIT), and other companies within the group, as Controllers and / or Processors in relation to processing Personal Data.

BIT takes these obligations seriously. BIT is committed to respecting the rights of all individuals whose personal data it processes:

1. **In relation to data security,** BIT has implemented appropriate measures to ensure the secure storage and handling of Personal Data, including obtaining a Cyber Essentials Plus certification and developing a comprehensive Data Handling Protocol.
2. **In relation to data protection and privacy rights,** our data processing activities are conducted according to the principles relating to the processing of Personal Data set out in the GDPR, including that Personal Data shall be processed lawfully, fairly and in a transparent manner, and in a manner that ensures the security of the Personal Data. BIT has policies and procedures in place to ensure compliance with these principles.

More information on how we handle Personal Data in relation to projects we are working on is detailed below.

BIT is registered with the UK ICO under the terms of the Data Protection Act 2018. Our registration number is ZA038649.

Privacy by design

BIT conducts all trials and research projects with a privacy by design approach to protect and maintain the privacy and security of research participants' and research subjects' data. We work closely with clients, government departments and research partners when designing interventions to ensure that a privacy by design approach is implemented and respected.

Our data protection and data security policies and procedures reflect necessary legislative requirements and set out the standard to which BIT staff should work when dealing with Personal Data, including:

- Attendance at mandatory data protection training for all employees;
- Identifying data requirements from the outset of each project;
- Minimising use of Personal Data where possible and ensuring we have the right to handle any Personal Data where successful project delivery is reliant on using it;
- Putting in place data processing agreements with all clients and suppliers to clarify data handling arrangements ahead of any data being transferred;
- Complying with all relevant data residency requirements and implementing appropriate technical and organisational measures, to protect data and avoid unauthorised access, internally and externally;
- A clear internal reporting process in the event of a data breach, to consider the nature of the breach and identify any necessary action, including whether the breach should be reported to the relevant authorities, i.e. the Information Commissioner's Office in the UK or the Office of the Australian Information Commissioner;
- Clear procedures on retention and destruction of Personal Data to avoid keeping hold of Personal Data longer than necessary for the purposes of each project; and
- Implementing robust investigation and reporting procedures in relation to any data breach or security issues that arise both within our own systems and those of our clients, partners and suppliers.

Data Protection Officer

The BIT group of companies has appointed a Data Protection Officer (DPO) who is the first point of contact for any issue regarding data protection and data security. The DPO can be contacted via email at legal@bi.team or by writing to us at:

Legal Team, Behavioural Insights Limited, 58 Victoria Embankment, London, EC4Y 0DS, United Kingdom.

Annex E: Power calculations

Power calculations using a within-schools design can be found below.

Within-schools trial		
n	***n* (per arm)**	**MDES**
160	80	0.45

Power calculation was determined on the basis the following assumptions:

- Power: 0.8
- Significance level: 0.05
- Level of randomisation: individual-level (i.e. individual students will be randomised)
- Number of trial arms: 2 (i.e. intervention and control)
- Outcome measure: SDQ
- Standard deviation: 6.5 in both groups

Stata code:
power twomeans 10, n(160) power(0.5) sd(6.5)

Annex F: Evaluation team experience

- **Tom McBride** is the Director of the Ending Youth Violence Lab and has over 15 years of experience in research and evaluation roles. He is the former Director of Evidence at the Early Intervention Foundation and Head of Strategic Analysis at the Department for Education. Tom will have overall responsibility for the delivery and quality of this work.

- **Jack Martin** is an Assistant Director within the Ending Youth Violence Lab and has over 8 years of experience working at the Early Intervention Foundation and sits on the Government's Trials Advice Panel. Jack will oversee the delivery of the work and support, supervise and quality assure the work of the project team.

- **Alice Worsley** is an Associate Policy Advisor based in the Health and Wellbeing team at BIT. She has worked on projects across a range of policy areas and methodologies. Alice will project manage and coordinate the project.

- **Patrick Taylor** is a Principal Research Advisor and leads BIT's education and youth evaluation work, supporting the design, improvement and evaluation of complex interventions in these fields. Patrick will provide support and quality assurance for the pilot evaluation.

- **Naomi Jones i**s a highly experienced social research consultant who specialises in helping organisations to design, commission, deliver and use research better, with over 18 years applied research experience. Naomi was formerly head of social attitudes at NatCen, where she led a mixed-method research team and oversaw the British Social Attitudes Survey. Naomi will lead the qualitative evaluation.

- **Lilli Wagstaff** is a quantitative research advisor in the Home Affairs and Security team at BIT and leads the evaluation and day-to-day delivery of a number of projects focusing on policy areas including reducing violence and recidivism. Lilli will lead the quantitative evaluation.

- **Niall Daly** is a Research Advisor in the Health and Wellbeing team at BIT, specialising in trial design, implementation, and quantitative data analysis across a range of projects within the health space. He will support the quantitative evaluation.

- **Tess Moseley-Roberts** is an Associate Policy Advisor in the Home Affairs, Security and International Development team at BIT. She has worked across peacebuilding, anti-corruption and education in the international space and a number of UK criminal justice projects. Tess will support the qualitative evaluation.

- **Dr Nick Axford** is an Associate Professor at the University of Plymouth with a specialism in evidence-based prevention and early intervention to improve child well-being, with a particular focus on developing or adapting interventions and evaluating their effectiveness and implementation. Nick has been a Member of the Early Intervention Foundation Evidence Panel since 2014 and will be providing expert advice and challenge on the design and delivery of the evaluation on a consultancy basis.