Khulisa: Face It Programme

Impact Evaluation Report (Final)

February 2024

www.impacted.org.uk

Contents

Executive Summary

Purpose of the Report

This report follows an interim report, authored in July 2022, which summarised the results from nine cohorts of Face It participants from the years 20/21 to 21/22. Some of the key findings of this report included:

- **A substantial increase across pupil wellbeing, resilience and emotional regulation outcomes over the course of the programme**
- A decrease in attendance following the intervention, within a context of national falling school attendance, suggesting a more positive trend for Face It participants with less decrease in attendance (5.2%) than the national average (10.1%).
- Broadly lower outcomes for SEND and Year 10 pupils compared to other sub-groups

However, the interim report was limited by the disruption and constraints imposed by the Covid-19 pandemic and the relatively small sample size. This report will analyse data from 25 cohorts to determine the impact of the programme on participant wellbeing, resilience and emotional control. As with the interim report, it will test the statistical significance of any changes and benchmark the impact of the programme against national and scale averages for each of these measures. We will also analyse the impact of the programme on 23 cohorts alongside a constructed control group via Propensity score matching to determine the impact of the programme on school attendance and school exclusions.

Key findings

Pupil progress across social, emotional and wellbeing outcomes

- **There was a statistically significant improvement observed across all three social and emotional outcomes: wellbeing, resilience and emotional control.**
- 55.4% of participants reported an increase in their wellbeing over the course of the programme. This change in wellbeing was statistically significant ($p<0.001$, $n=166$), which suggests that a genuine change in wellbeing occurred, rather than this being the result of sampling error. Participants also finished the programme with score above the national average.
- 66.5% of participants reported an increase in their resilience over the course of the programme. This change in resilience was statistically significant ($p<0.001$, $n=167$), which suggests that a genuine change in resilience occurred.
- 63.5% of participants reported an increase in their emotional control over the course of the programme. This change in emotional control was statistically significant

(p<0.001, n=167), which suggests that a genuine change in emotional control occurred.
- 88% of participants reporting using the coping skills post-programme.

Pupil progress across engagement outcomes

- **Average attendance declined over a 6-month period for both the participating group and the control group.** The average attendance of participants fell by -8.1 percentage points (from 92.5% to 84.4%), and the average attendance of individuals in the matched control group fell by -5.0 percentage points (from 91.3% to 86.3%), leaving both participating and control pupils below the national average of 92%.
- While there appears to be slightly more decrease for participating pupils, this finding was not found to be statistically significant(p>0.05, n=345). This suggests that **the discrepancy between participating and control individuals might be better explained by 'sampling error'** (i.e. normal, random fluctuations in the data).
- The **number of exclusions recorded for participating pupils increased by 7%** between the pre- and post-intervention windows (a 6-month period). However, this change was **not statistically significant** (p>0.05 n=174), meaning that the observed change may be due to chance.

Recommendations

- Over half of participants reported an increase in wellbeing, resilience, emotional control and use of the coping skills following the intervention. **This suggests Face It is effective in meeting the shorter-term Theory of Change outcomes around social and emotional wellbeing and these aspects of the programme should be retained.**
- There was only marginal difference in improvement levels of wellbeing, resilience, emotional control and use of the coping skills for different subgroups (apart from emotional control, in which PP students reported substantially higher post-intervention scores). **This suggests Face It is suitable and useful for its target participants and recruitment processes should be retained.**
- Due to the national rise and concerns with attendance and exclusions across the academic years of the evaluation, it is difficult to infer the impact of the programme on attendance and exclusions, though the participating and control pupils were below national average for attendance, participating pupils did not make statistically significant decreases or increases compared to control pupils. **Further investigation would be needed to fully understand any changes observed.**

1. Introduction

Overview of the *Face It* Intervention

'Face It' is an intensive therapeutic group programme which been running in schools and Pupil Referral Units across the UK for the past decade. It is specifically designed to support the wellbeing and social and emotional skills of young people aged 11-18 who have been identified as at risk of offending, exploitation or exclusion. Through a combination of experiential techniques, the programme aims to aid young people to learn new skills to identify individual triggers or root causes behind negative behaviour and to develop alternative, more positive, responses which will ultimately improve the young person's outcomes. As such, the programme not only seeks to help young people improve their wellbeing, it also aims to close the social and emotional skills gap by supporting the most disadvantaged young people in developing the skills required for success and dealing with adversity.

Young people are referred to the *Face It* programme by key school contacts using a participant profile guide and referral form, often when universal offers have proved inadequate. A high proportion of *Face It* participants are registered as vulnerable by the school. The following statistics about the participant population as a whole (outside of this evaluation sample) were provided to ImpactEd by Khulisa:

- 86% of participants in this sample size had at least 1 marker of disadvantage[1]
- 53% of participants are eligible for Pupil Premium
- Half of participants had either a history of exclusion and/or were at risk of exclusion

Previous Khulisa programme evaluations demonstrate behaviours are frequently the result of Adverse Childhood Experiences, such as, bereavement, domestic violence, neglect, sexual exploitation, drug/alcohol abuse, and mental health issues.

ImpactEd and Khulisa: *Face It* evaluation

ImpactEd and Khulisa have been working to evaluate the impact of the Face It programme on participating pupils' social, emotional and wellbeing outcomes as well as their engagement in school (through attendance, attainment, exclusion, and behavioural data). An interim report was authored in July 2022, which assessed the impact of the Face It programme on the social, emotional and wellbeing outcomes and school engagement of nine cohorts. The core analysis for this report was based

[1] The measures of disadvantage we use to calculate this figure are: EAL status, FSM, Pupil premium, previous exclusion, involved in criminal activity, risk of being involved in criminal activity, SEN and whether a child is a looked after child. This also includes things like whether the child has MH needs, experience of ACEs, i s on an education health care plan, is a child in need and/or receives support from social services.

on 52 pupils for social, emotional and wellbeing outcomes and 54 pupils for attendance data. The interim report found a **substantial increase** across pupil social, emotional and wellbeing outcomes; all nine cohorts increased in resilience and all but one cohort reported improvements in wellbeing. Trends for Face It participants saw an average decrease in attendance of 5.2% compared to a national linear average of 6.74%, perhaps suggested a slightly more positive trend for Face It participants, notable given their at-risk profile. However, the report noted significant limitations in the light of continued disruption to both programme delivery and school capacity caused by the Covid-19 pandemic.

This report follows on by assessing the impact of the programme on the same social, emotional, wellbeing and school engagement outcomes, but with an expanded sample and use of more extensive metrics to assess school engagement. Twenty-three cohorts are considered in this report across four schools for the engagement analysis and twenty-five cohorts across five schools for the social and emotional (SEL) analysis: Park View School (PVS), Manchester Communication Academy (MCA), City of London Academy Southwark (COLA), Saracens High School (SHS) and Morley College (for SEL analysis only). The evaluation spans from March 2021 to July 2023, incorporating three months before the first evaluated cohort in February 2021 and six months after the last evaluated cohort in January 2023 (adjusted for school term dates).

2. Evaluation Design & Methodology

2.1 Theory of Change and core outcomes

The measures utilised were developed in line with Khulisa's *Face It* Theory of Change (Fig. 1):

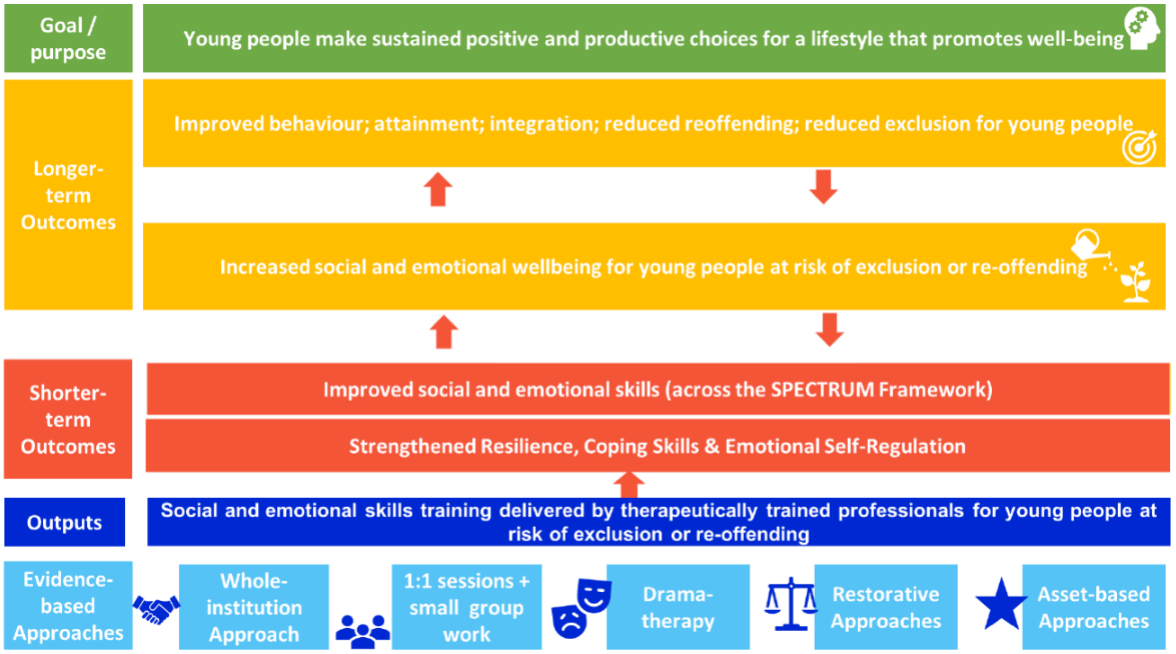


Figure 1: Khulisa's Theory of Change

The premise of Khulisa's programmes, and the above Theory of Change, is to build the basic foundational social and emotional skills critical to mitigating the negative behavioural effects of trauma, specifically improved resilience, coping skills, and self regulation, which will in turn improve general social and emotional wellbeing. Khulisa's Theory of Change, which supports the importance of the six clusters of social and emotional skills (Fig. 2) for social and emotional wellbeing and positive longer-term life outcomes.

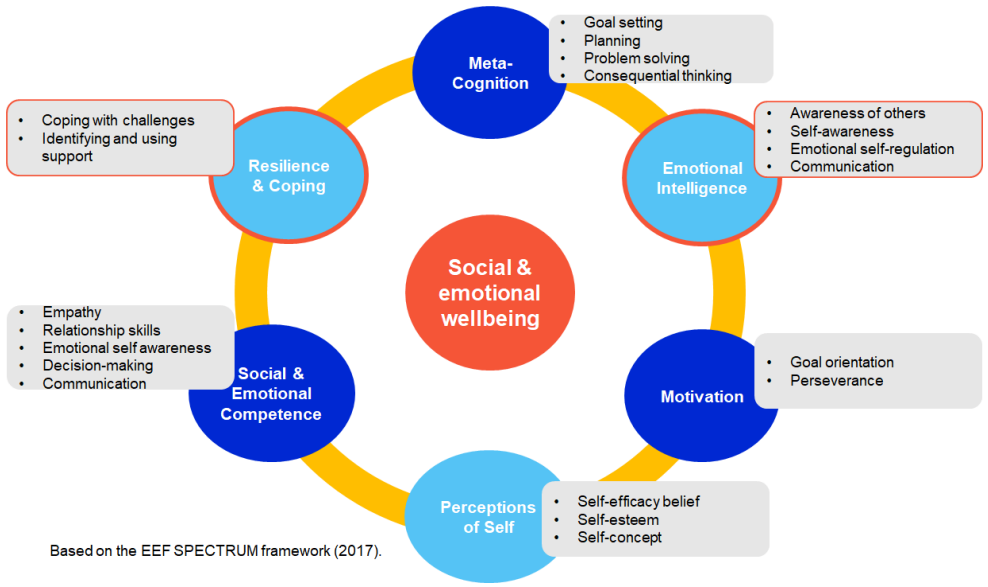


Figure 2: Khulisa's social and emotional skills framework

Khulisa operate under the assumption that improved social and emotional skills form the basis for improved longer-term social and emotional wellbeing, and positive life outcomes. These longer-term outcomes are not limited to the absence of crime, exclusion or negative/violent behaviour. They include meaningful engagement with others and society and positive and productive life choices. Khulisa believe this can look different for every individual and focus more on flourishing and thriving rather than compliance and desistance.

This report will analyse data from twenty-five cohorts total, alongside a constructed control group via Propensity Score Matching for the engagement analysis in particular, to determine:

- The impact of the programme on participant wellbeing
- The impact of the programme on participant resilience
- The impact of the programme on emotional control, including utility of the *Face It* coping skills taught during the intervention
- The impact of the programme on school attendance
- The impact of the programme on school exclusions

2.2 Measures and analysis methods used

The evaluation followed a pre-/post- test design, meaning we used the same measures to collect participant data before and after they completed the evaluation. The core outcome measures for this evaluation were:

	Outcome	Measure	Analysis
Social and emotional outcomes	**Wellbeing**	Shortened Warwick-Edinburgh Mental Wellbeing Scale	Basic descriptive statistics Significance testing
	Resilience	Children's Hope Scale	
	Emotional control	Emotional Regulation Questionnaire	
School engagement	**Attendance**	School-held data on attendance, pulled through ImpactEd platform	Descriptive statistics Significance testing Propensity Score Matching
	Exclusions	School-held data on exclusions, pulled through ImpactEd platform	
	Attainment	Not able to be collected	N/A

Further detail is provided on each of the measures used to assess changes in social and emotional skills, and wellbeing below:

- **Wellbeing (Shortened Warwick-Edinburgh Mental Wellbeing Scale)**

This is a widely used validated scale which measures general mental wellbeing. Wellbeing is measured on a scale from 1 to 5, which is then converted into a standardised score running from 7 to 35 to enable comparison with UK population benchmarks. Benchmarks are available from a number of studies.

- **Resilience (Children's Hope Scale)**

This is another validated scale which measures children's perceptions that their goals can be met. Benchmarks are available from the original study, and we have also contextualised this against more recent data from survey respondents on the ImpactEd platform. Resilience is measured on a scale from 1 to 6.

- **Emotional Regulation (Emotional Regulation Questionnaire)**

This is a validated scale which measures the degree to which people regulate their emotions. Benchmarks are available from the original study, and we have also contextualised this against more recent data from survey respondents on the ImpactEd platform. Emotional regulation is measured on a scale from 1 to 5.

The following provides an overview of the analytical methods used:

- **Basic descriptive statistics**

Descriptive statistics are used to describe or summarize the characteristics of a sample or data set, such as a variable's mean, standard deviation, or frequency. In this report, we have calculated the mean as the average level observed across the dataset.

- **Significance testing**

Statistical significance refers to a determination made by the analyst that the results in the data are not explainable by chance alone. This test provides a p-value, which is the probability of observing results as extreme as those in the data, assuming the results are due to chance alone. Two-tailed paired samples T-tests were conducted for all available outcomes, as this data was matched (pre-/post-) thus change could have occurred in either direction. For this report, a p-value of less than 0.05 was deemed to be statistically significant.

- **Propensity score matching**

Propensity score matching is a quasi-experimental non-equivalent group design that constructs a control group ‘retrospectively’ using individuals that did not receive a treatment. Participants are matched to non-participants on an individual basis. This involves first estimating propensity scores for each individual in the group, then specifying a ‘caliper radius’, which defines the maximum permissible distance between two propensity scores that can be considered a ‘match’. Propensity score-matched pairs are then constructed by matching participants and non-participants in a way that minimises the distance between their propensity scores inside this caliper radius.

2.3 Sample

The data used for the analysis was taken from twenty-three Face It cohorts across four schools for the engagement analysis and twenty-five cohorts across five schools for the social and emotional analysis. For all social and emotional outcomes bar wellbeing, the sample consists of 167 pupils for whom we had matched baseline (i.e. pre-intervention) and endline (i.e. post-intervention) survey data. In the case of wellbeing, the sample is instead 166, with one of the pupils from one school not having completed the wellbeing post-intervention survey. The sample differs for engagement data, relying on availability of school-held data as opposed to matched surveys. Attendance and exclusions data was available for 174 participants and 184 matched control pupils. The details for each sample are described in more detail in the relevant parts of Section 3 and Section 4.

As this was not a randomised experiment and participating and control pupils were not randomly assigned to their groups, when comparing participating and control groups in the analysis of attendance and exclusions data, we had to make sure both groups were properly matched and weighted within our sample to minimise any bias in the results. Therefore, we created two comparable groups using Propensity Score Matching (PSM) to identify and match individual respondents across the two groups. This statistical matching technique helped make our groups more comparable and reduce the potential bias of confounding

variables, mimicking randomisation and reducing any treatment assignment bias in the results. Matching was based on the following variables: Year Group, PP, SEND, EAL, LAC and FSM.

Our sample derived from PSM included 174 participating and 184 control group pupils. The details for each sample are described in more detail in the relevant parts of Section 3 and Section 4.

2.4 Limitations

- **2021/2022 and 2022/2023 have seen particularly low school attendance across the board, thus rendering it difficult to infer the direct impact of Face It on school attendance or exclusions.**
- The evaluation design looked only at selected aspects of the EEF Spectrum Framework (Wellbeing, Resilience and Emotional Control), rendering it difficult to make a judgement call as to the effectiveness of the programme in making improvements across the entire Spectrum. Further research into some of the other aspects (e.g Motivation, Self-efficacy and Metacognition) would be useful in more comprehensively understanding which aspects of social and emotional wellbeing Face It is most impactful for.
- The limited sample sizes within certain sub-groups (such as non-binary students, 11-year-olds, and 18-year-olds) posed a challenge in deriving dependable generalisations that could be consistently applied across sub-groups.
- The evaluation windows were short (3 months pre and 6 months post) thus only enabling us to examine the short-term impact of the programme. Further research could be conducted with a similar methodology and evaluation design to better understand the long-term impact of the programme. Khulisa is planning a future study evaluating the impact of the programme 3,6 and 12 months after the programme to understand this further.
- The evaluation design focussed on quantitative data and validated surveys. Given the highly individualised nature of each participant's profile and needs (under an umbrella of at-risk students), it may be useful in future to incorporate qualitative elements to better understand the challenges faced by participants and their experience of the programme.
- Schools measure behaviour differently and, in this case, it was not possible to reliably standardise the behaviour points collected for the four participating school in a manner that analysis or conclusions could be drawn.

- It proved particularly challenging collecting attainment data from the four participating schools. It was therefore not possible to measure the impact of the intervention on participant attainment.

3. Social, Emotional and Wellbeing Outcomes

3.1 Sample

Schools

The data used for the analysis in this section of the report was taken from twenty-five Face It cohorts across five schools. This sample is approximately the same across the four different outcomes of interest in this section, consisting of 167 pupils for whom we had matched baseline (i.e. pre-intervention) and endline (i.e. post-intervention) survey data. In the case of wellbeing, the sample is instead 166, with one of the pupils from one school not having completed the wellbeing post-intervention survey. The breakdown of respondents by school is shown in Fig. 3.

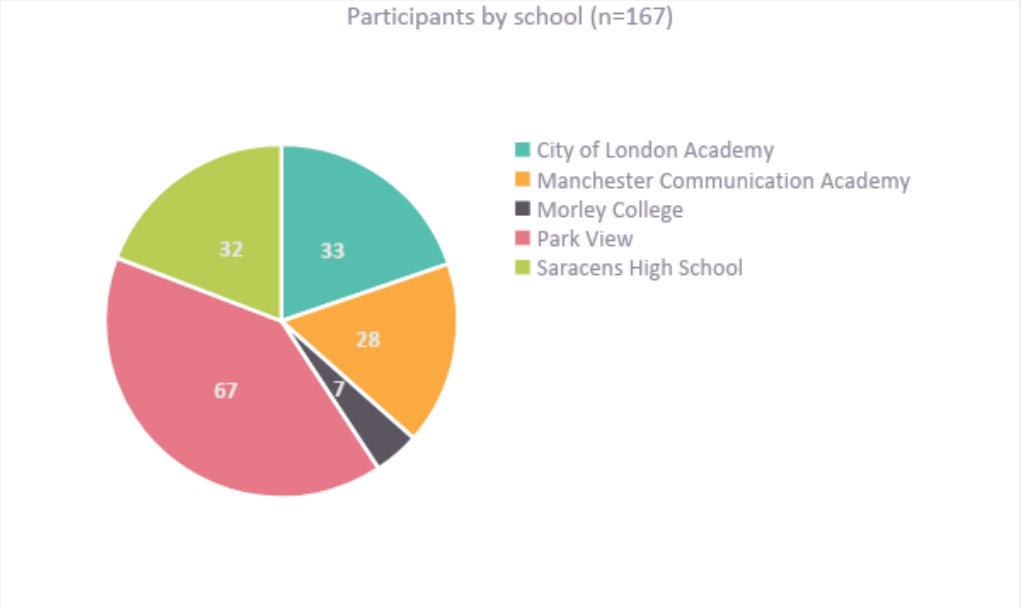


Figure 3

Gender

The sample includes male (50.9%), female (48.5%) and non-binary (0.6%) participants, with the overwhelming majority being made up of the first two categories, which are approximately equal.

SEND

The sample also comprises both pupils that have been identified as requiring SEND support, and those that have not. Pupils requiring SEND support make up 19.2% of the overall sample, with pupils not requiring SEND support making up 80.2%. One pupil was listed as 'unknown', and we have excluded this pupil from SEND subgroup analysis below.

Pupil Premium

The sample includes both participants that are eligible for Pupil Premium, and those that are not. The former group (PP eligible) constitutes 53.3% of the overall sample, and the latter group (non-PP eligible) constitutes the remaining 46.7%.

Age

The sample includes participants whose age ranges from 11-18 years old. The vast majority of the sample falls between the ages of 12 and 14, as indicated in Fig. 4. The small sample sizes at the extreme ends of this age range (particularly 11-, 17- and 18-year olds) should be borne in mind when interpreting some of the findings presented below.

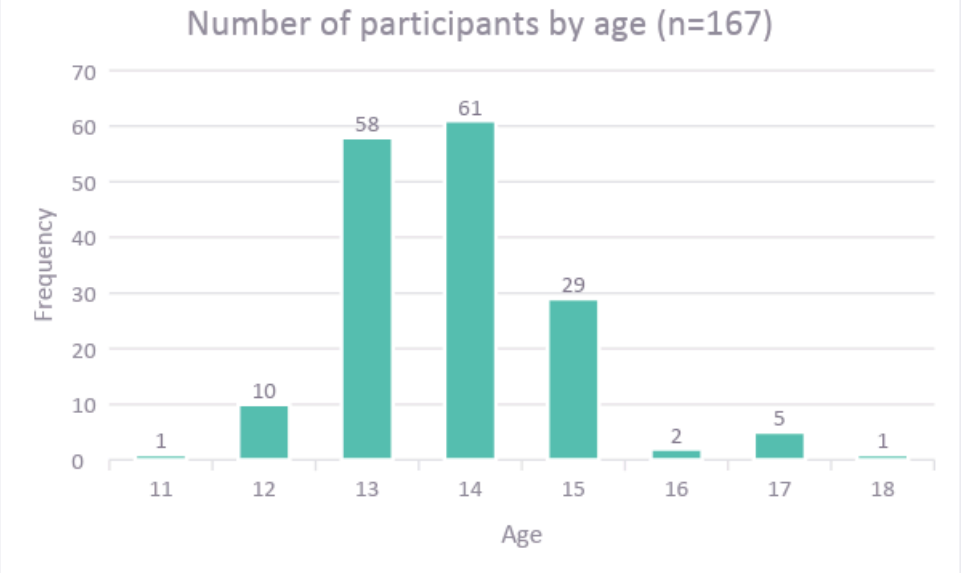


Figure 4

3.2 Wellbeing

Overview

Participant wellbeing was assessed using the Short Warwick-Edinburgh Mental Health and Wellbeing Scale (SWEMWBS). Participants completed this survey before the Face It programme, and then completed it again afterwards. This enabled us to understand the change in wellbeing over the course of the programme. Our key findings are presented below.

- Participants' wellbeing increased 8% on average over the course of the programme, with the average per-question SWEMWBS score rising from 3.24 to 3.50.
- **Notably, the post score is above the ImpactEd Platform national average for 22/23 (based on nearly 42,000 participant responses) of 3.14.**

- 55.4% of participants reported an increase in their wellbeing over the course of the programme.

- This change in wellbeing was statistically significant ($p<0.001$, $n=166$), which suggests that a genuine change in wellbeing occurred, rather than this being the result of sampling error.

Subgroup analysis

To understand the impact of the Face It programme on different subgroups, we split the data by gender, Pupil Premium status, SEND and age. We examined the changes in wellbeing scores pre- and post-intervention for each of these different groups.

Male and female students saw similar changes in wellbeing over the course of the programme (+0.18 and +0.33 respectively). Non-binary students saw a substantially larger increase in wellbeing over the course of the intervention (+1.86) but since there was only one individual in this category, we should be extremely tentative about extrapolating from this data to a broader conclusion about Face It's impact on individuals that identify as non-binary. Overall, 50% of male pupils reported an increase in their wellbeing compared to 59% of female pupils. Fig. 5 shows the pre- and post-intervention wellbeing scores for each reported gender.

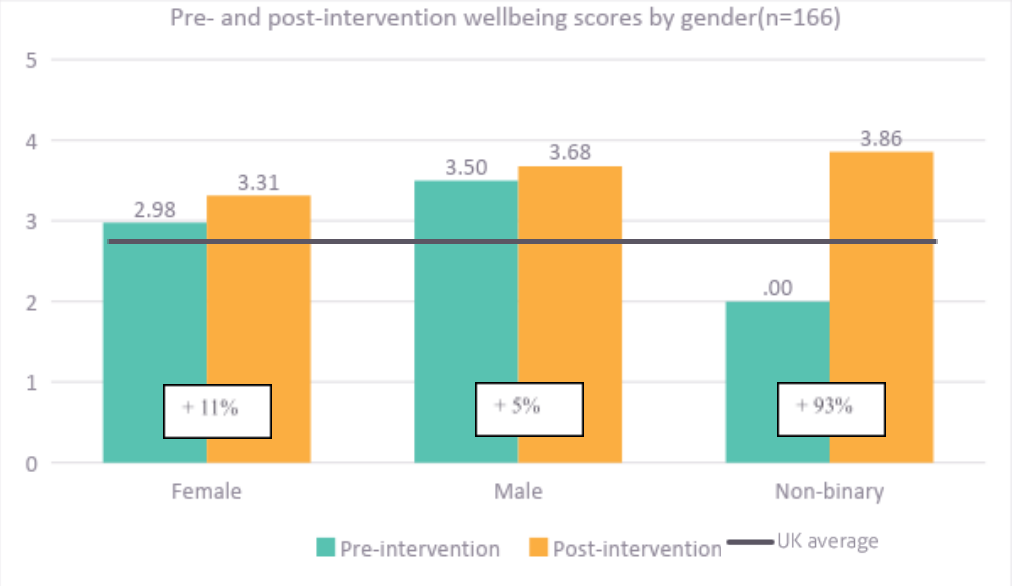
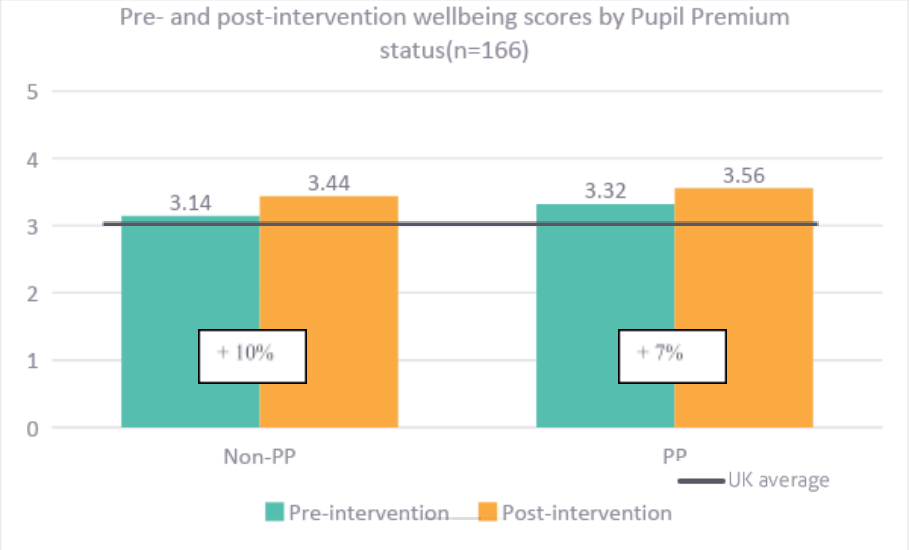


Figure 5

Boys appear to be doing better than girls at the start of the programme. Girls typically started the programme reporting lower scores than boys. This observation aligns with nationwide data indicating that girls tend to have comparatively lower wellbeing scores than boys, as

reported by ImpactEd (2021) and Education Policy Institute (2021). However, while boys typically finished the programme reporting higher scores on average, girls demonstrated a greater propensity to report an improvement, both in terms of likelihood and magnitude. These results suggest that girls may benefit from extra support for their wellbeing during this time.

Pupil Premium students reported increases in wellbeing over the course of the intervention that were marginally larger than those reported by non-Pupil Premium students (+0.30 vs. +0.24). This is a minor difference, and so the intervention seems to affect Pupil Premium and non-Pupil Premium students similarly (Fig. 6). Overall, 59% of PP pupils reported an improvement in their wellbeing compared to 51% of non-Pupil Premium pupils.



A similar trend was observed when analysing the data by SEND. Both SEND and non-SEND pupils reported small, positive changes in wellbeing over the course of the programme, but the impact on non-SEND pupils was nearly twice the size of the impact on SEND pupils (+0.16 vs. +0.30). This is illustrated in Fig. 7. Overall, 44% of SEND pupils reported an improvement in wellbeing compared to 59% of non-SEND pupils.

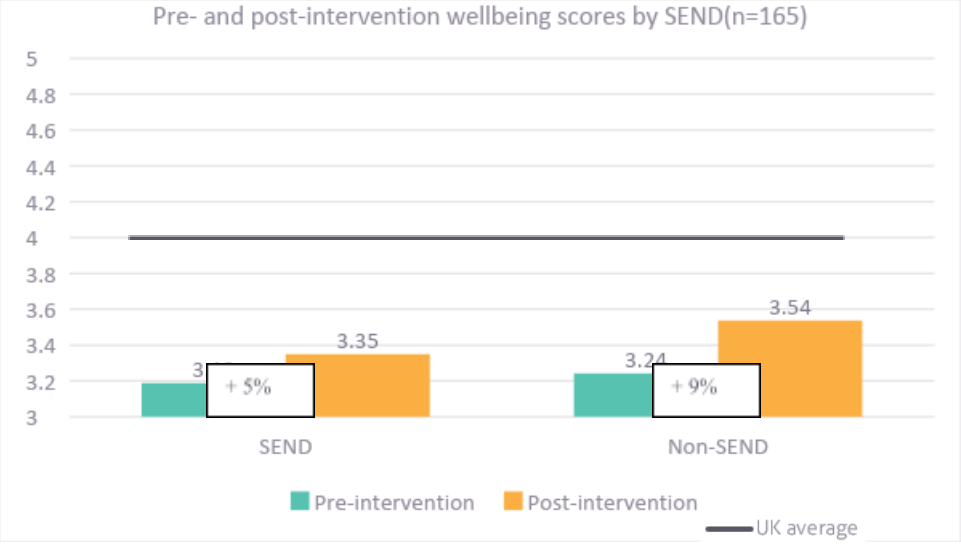


Figure 7

Segmenting the data into the different ages of participants allowed us to understand the impact that Face It was having on different year groups. The impact for each age is small but positive, with the notable exception of 11-year-old participants, who saw a decrease in wellbeing over the course of the intervention (Fig. 8). There is a pronounced increase in the impact of the programme on wellbeing for 16- and 17-year-olds, however this sample size is so small we cannot draw conclusions from this.

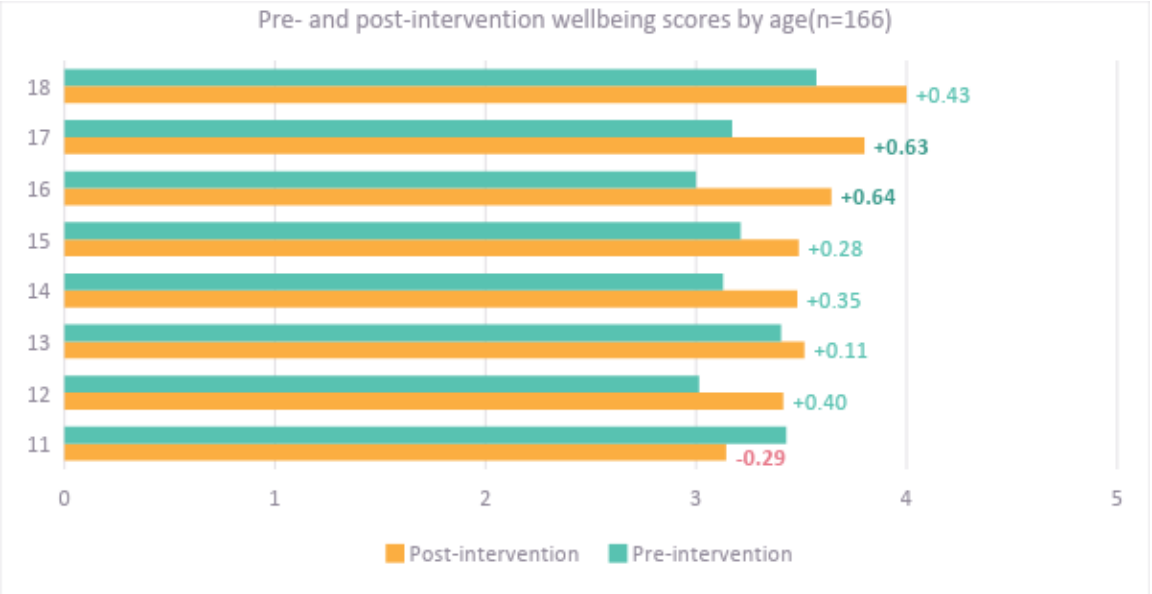


Figure 8

3.3 Resilience

Overview

Participant resilience was measured using the Children's Hope scale. Participants took this survey before the intervention, and again afterwards. Comparing these pre-intervention and

post-intervention scores allows us to understand changes in resilience that occurred over the course of the programme. Our key findings related to resilience are presented below:

- Participants' resilience increased 11% on average over the course of the programme, with the average per-question Children's Hope score rising from 3.57 to 3.96.
- This post score was still below the 22/23 ImpactEd Platform national average of 4.20. However, it was closer to the national average for PP pupils of 4.06 and above the national average for SEND pupils of 3.86, which are likely more representative of Khulisa's cohort.
- 66.5% of participants reported an increase in their resilience over the course of the programme.
- **This change in resilience was statistically significant** ($p<0.001$, $n=167$), which suggests that a genuine change in resilience occurred, rather than this being the result of sampling error.

Subgroup analysis

The impact of the Face It programme on resilience varied by gender, with female participants reporting larger changes in resilience than their male peers. The average increase in resilience was +0.28 for male participants, but 75% higher for female participants, who reported an average increase in resilience of +0.49 (Fig. 10). Again, non-binary participants saw the largest increase of resilience over the course of the programme (+1.16), but the same caveat from section 4.2 applies: there was only one individual in the dataset that identified as non-binary, and so we should be tentative about inferences made from this data. Overall, 60% of male pupils reported an increase in their resilience compared to 73% of female pupils.

Although this suggests that the impact the programme has on a participant will vary with the gender of the participant, it should also be noted that female participants reported much lower baseline resilience scores than male participants, though subsequently went on to make more progress. Again, this trend might be better explained by the fact that Face It simply has a greater impact on those that need it most (i.e. those with lower baseline scores), rather than having a greater impact on females per se.

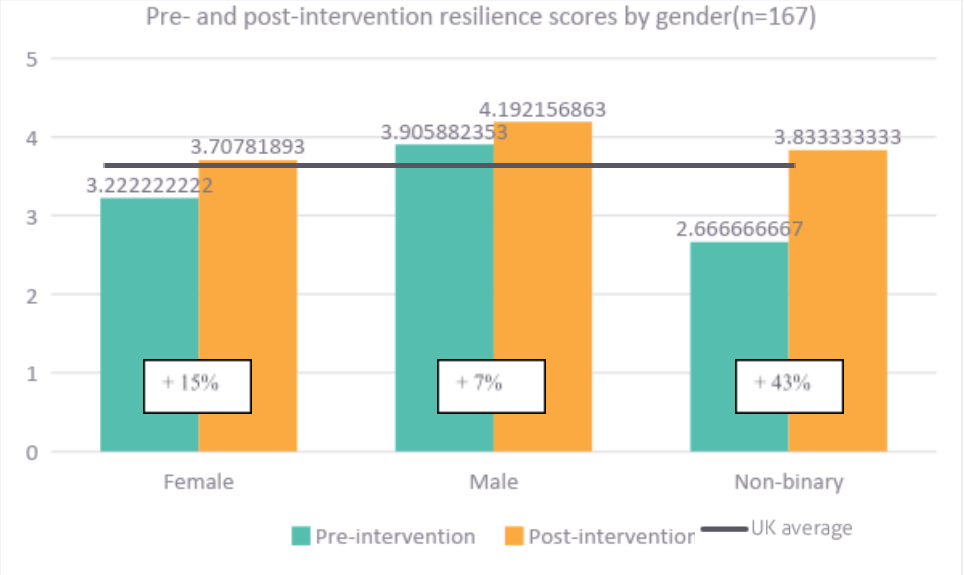


Figure 9

The impact of the programme on resilience varied only slightly with Pupil Premium status, with both groups seeing a reasonable increase. Non-PP participants saw an average increase in resilience of +0.42, and PP participants saw an average increase in resilience of +0.35 (Fig. 11). Overall, 64% of PP pupils reported an improvement in their resilience compared to 69% of non-PP pupils.

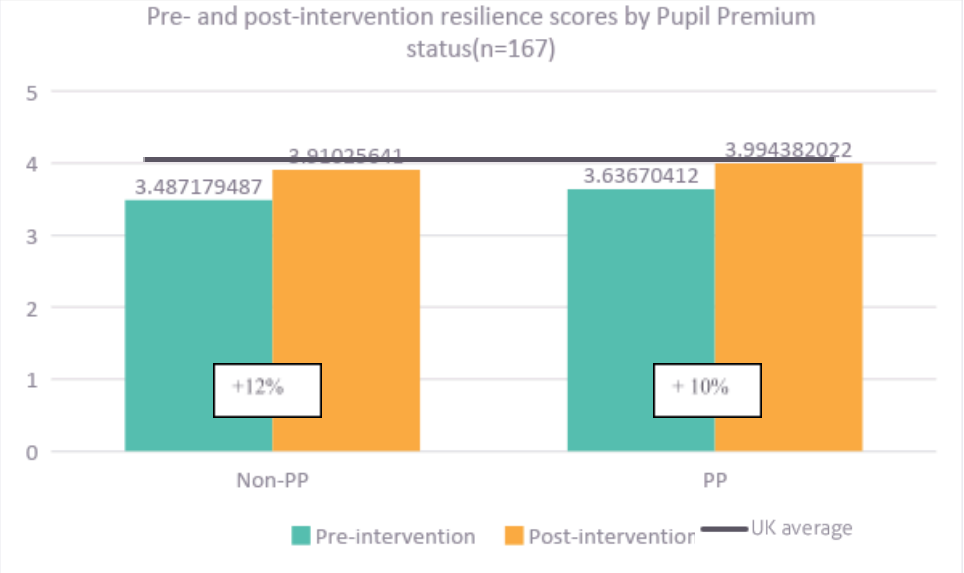


Figure 10

A similar trend is observed with SEND vs. non-SEND participants: the impact of the programme on resilience varied little with the SEND status of a participant. Non-SEND participants reported an average increase in resilience of +0.41, while SEND participants reported an average increase in resilience of +0.32 (Fig. 12). Overall, 56% of SEND participants reported an improvement in their resilience compared to 69% of non-SEND participants.

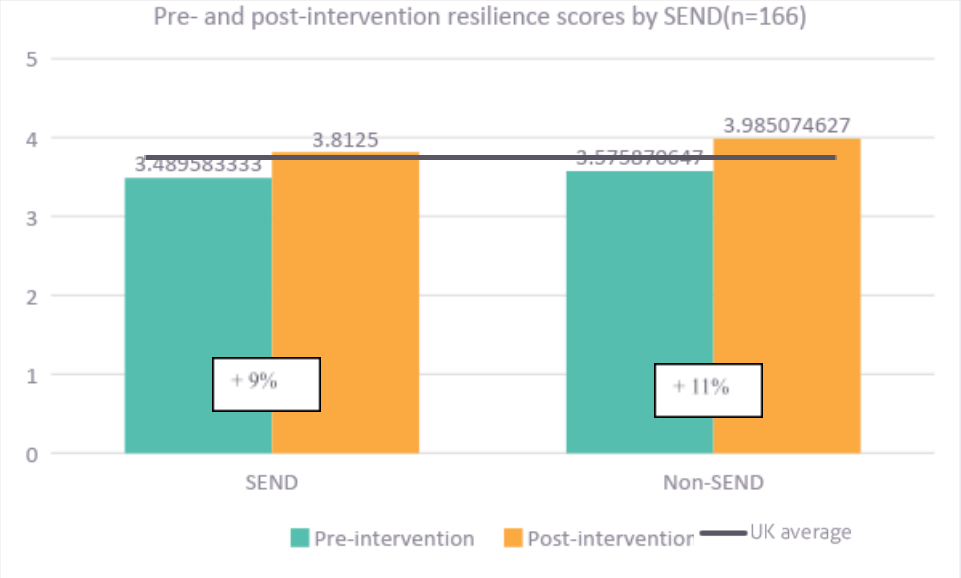


Figure 11

The way that Face It's impact on resilience varied with age was very similar to what was observed in the case of wellbeing: all age categories observed a positive increase in resilience, with the exception of 11 year olds, whose resilience decreased on average by -0.67. There is a pronounced jump in Face It's impact on resilience for 17- and 18-year-olds, who reported large increases of +0.77 and +1.17 respectively (Fig. 13), though those sample sizes are very small so no conclusions should be drawn from them.

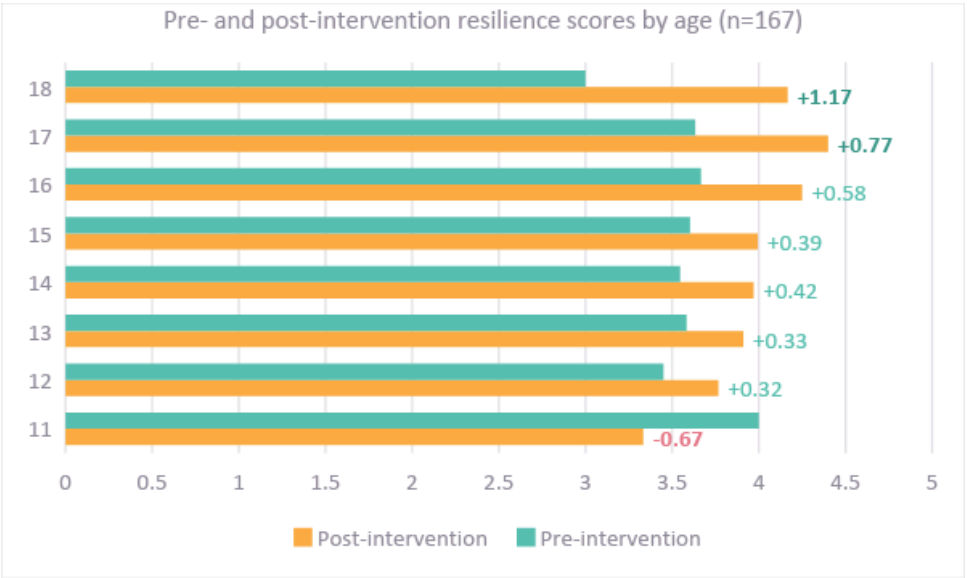


Figure 12

3.4 Emotional Control

Overview

Participants' emotional control was measured using the Cognitive Reappraisal subscale of the Emotional Regulation Questionnaire. Participants took this survey both before and after the intervention, and comparing these pre-intervention and post-intervention scores allowed us to understand the changes in emotional control that occurred over the course of the programme. Our key findings related to emotional control are presented below:

- Participants' emotional control increased 8.8% on average over the course of the programme, with the average per-question Cognitive Reappraisal score rising from 2.96 to 3.22. There is no relevant national average or benchmark for this measure.

- 63.5% of participants reported an increase in their emotional control over the course of the programme.

- This change in emotional control was statistically significant ($p<0.001$, $n=167$), which suggests that a genuine change in emotional control occurred, rather than this being the result of sampling error.

Subgroup analysis

The impact of the Face It programme on emotional control did vary by gender, but in a slightly less pronounced way than it did in the cases of wellbeing and resilience. Female participants reported larger average changes in emotional control than their male peers. The average increase in emotional control was +0.21 for male participants, but 48% higher for female participants, who reported an average increase in emotional control of +0.31 (Fig. 15). Overall, 64% of male pupils reported an increase in their emotional control compared to 63% of female pupils.

Again, non-binary participants reported the largest increase of the three groups, with their emotional control scores increasing from 3.00 to 3.50 over the course of the programme. However, this is based on a sample size of one so more data would be required to draw any conclusions from this.

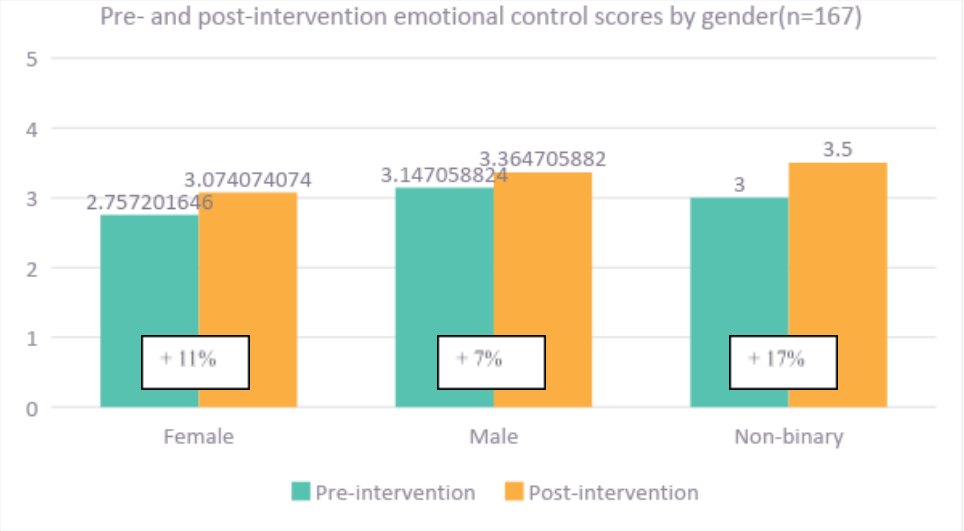


Figure 13

Although Face It's impact on wellbeing and resilience did not vary substantially by Pupil Premium status, the opposite seems to be true in the case of emotional control. Pupil Premium participants reported an average increase in emotional control of +0.33, some 74% higher than the average increases reported by their non-Pupil Premium peers (+0.19). This is particularly interesting given that the baseline scores for each group are identical (Fig. 16). Overall, 70% of PP pupils reported an increase in their emotional control compared to 56% of non-PP pupils.

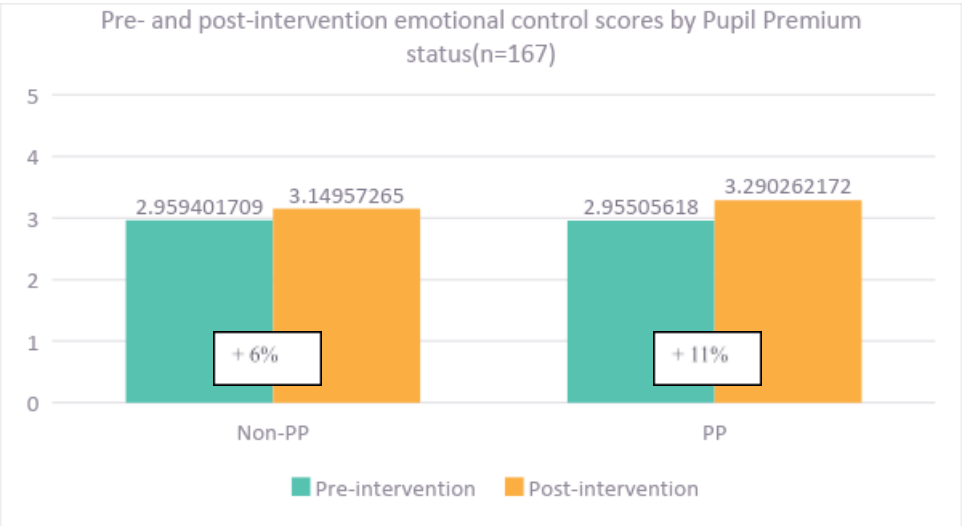


Figure 14

The impact of Face It on emotional control did not vary by SEND status. SEND and non-SEND pupils reported near-identical average increases in emotional control over the course of the programme (+0.27 and +0.28 respectively), as show in Figure 17. Similarly, 63% of SEND pupils reported an improvement in their emotional control compared to 64% of non-SEND pupils.

This suggests that Face It is equally effective at increasing emotional control for both SEND and non-SEND pupils. Moreover, as stated above, this change was statistically significant ($p<0.001$, $n=167$) which suggests a genuine change in emotional control occurred, though it should be remembered that without a control group we cannot be certain that this increase was caused by the programme and not other background factors.

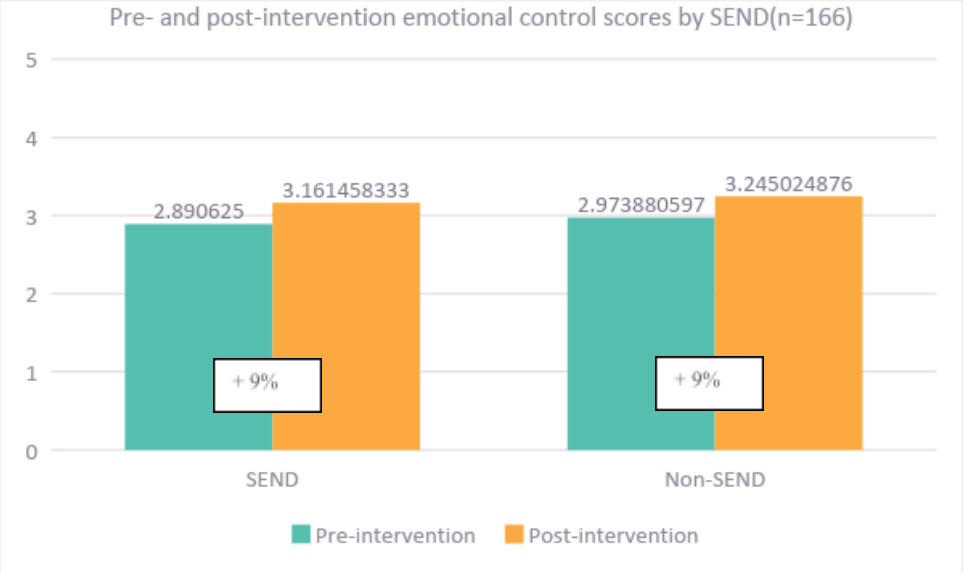


Figure 15

Segmenting the data by age allows us to understand the impact of the programme on emotion control for different year groups. Fig. 18 shows that the impact of Face It on emotional control was uniformly positive, with all year groups reporting an average increase in their emotional control scores. Unlike in the cases of wellbeing and resilience, even 11-year-old participants saw substantial increases in emotional control. There is a notable jump in the impact of the programme on emotional control for 16-year-olds, who reported an average increase in their emotional control of +0.83, some 66% higher than the next highest scores (11 year olds and 17 year olds – though we should keep in mind that the sample of 17- and 18-year olds was minimal).

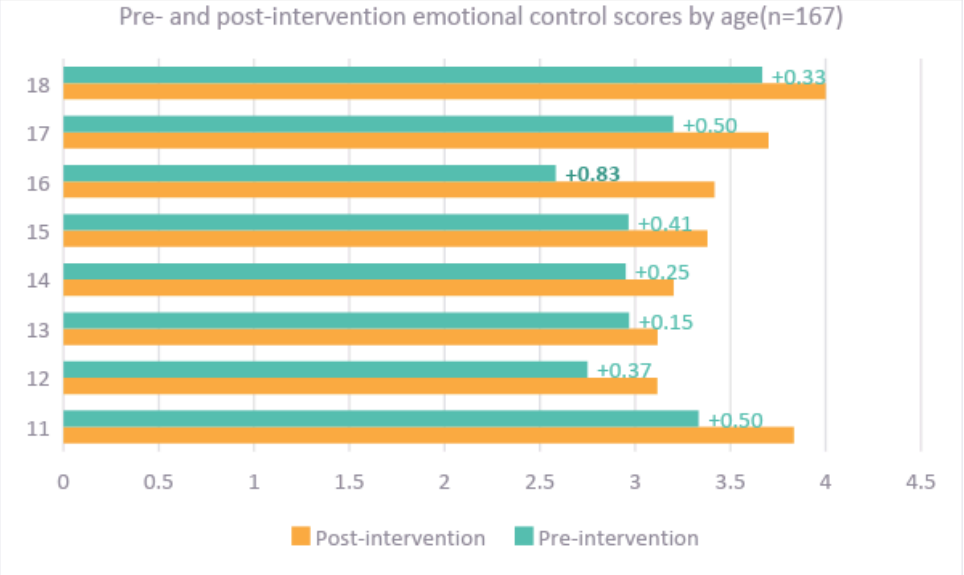


Figure 16

3.5 Coping Skills

Overview

Context

This evaluation did not use a validated scale to understand the impact that Face It had on participants' coping skills. However, here we include some top-level analysis of the extent to which participants responded in the post-programme survey that they had used the coping skills.

From a sample of 167 responses to the post-programme survey over the 21/22 and 22/23 cohorts, 88% of participants said that they had used the coping skills over the course of the programme, with just 12% saying that they had not.

Subgroup analysis

To understand whether specific demographics were more likely to have used the coping skills, we segmented the data according to gender, Pupil Premium eligibility, SEND and age.

- Male participants were slightly more likely to have used the coping skills than female participants (89.4% males used the coping skills vs. 86.4% females).

- Non-Pupil Premium participants were slightly more likely to have used the coping skills than Pupil Premium participants (89.7% of Non-PP participants used the coping skills vs. 86.5% of PP participants).

- Participants who had been identified as needing SEND support were just as likely to use the coping skills as those who had not been identified as needing SEND support (87.5% vs. 88.0% respectively).

- Self-reported use of the coping skills varied considerably by age, but there is a clear trend of use decreasing with age. 12 year olds were mostly likely to report having used the coping skills (90.0% of respondents) followed by 13 year olds (89.5%). 14 and 15 year olds were slightly less likely, with 87.1% and 82.8% of respondents reporting having used the coping skills respectively. This trend is illustrated in Fig. 20. (Note that all age groups with sample sizes smaller than 5 were excluded from this analysis.)

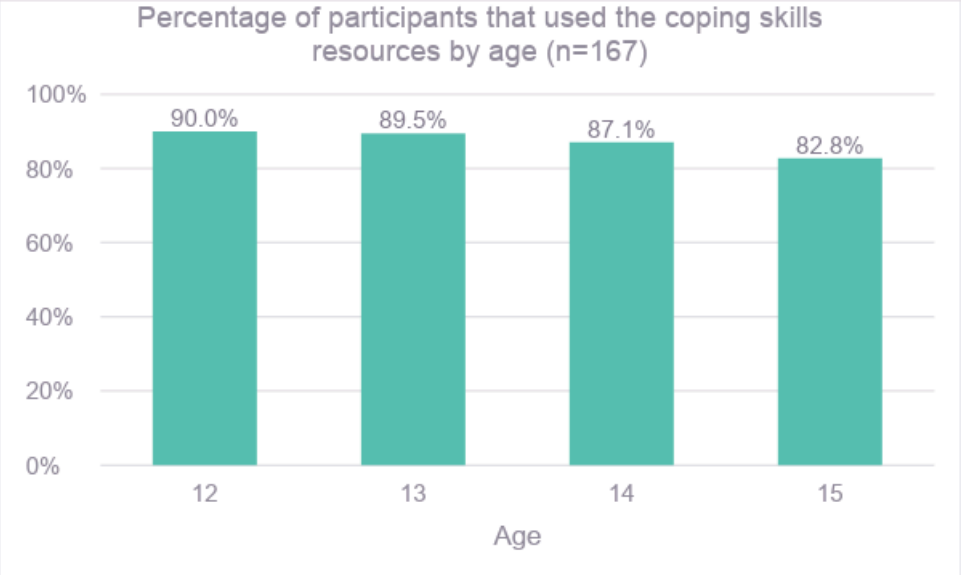


Figure 17

4. Engagement Outcomes

4.1 Sample

Our analysis of the engagement outcomes differs somewhat from our analysis of the social, emotional and wellbeing outcomes in that here we have been able to construct a control group of similar, non-participating pupils against which we can compare the changes observed in the participant group. We used a propensity score matching approach to create a matched sample (n=358), made up of both participating (n=174) and control group pupils (n=184) across four schools. Morley College is not included in the sample as we were not able to collect this data from their school Management Information System. Details on the method used to construct this control group can be found in section 2.

The participating sample is made up of pupils who fully or partially completed the programme (more than 50%) between September 2021 and July 2023. Whilst the total number of pupils who partially or fully completed the programme is 195, we were only able to include 174 in the sample. This is because issues related to the technical setup of the evaluations on the ImpactEd platform meant that data from 21 pupils was not able to be collected (e.g. because a pupil may have moved on or data was not stored as expected). We will be working with Khulisa to resolve these issues moving forward.

Following the analysis of the propensity-score matched sample, Z-scores were calculated to identify outliers, using the formula 'Z-score = (Data Point - Mean) / Standard Deviation'. If a data point had a z-score greater than 3 or less than -3, it was considered an outlier. 8 outliers were identified in the participating group and 12 outliers were identified in the control group. All outliers were removed, and the analysis was carried out on this new sample (n=338). However, removing the outliers made very little difference to the overall results, so we have reported on the original analysis that includes the whole PSM sample (n=358).

The matched sample is described below. Please note that the sample size in the attendance analysis is slightly less, since we used participants with matched baseline and endline data only.

Gender

The sample of participating pupils includes 87 (50%) female and 87 (50%) male participants. The sample of control group pupils is made up of 95 (52%) female and 89 (48%) male participants.

SEND

The sample of participating pupils includes 48 (28%) SEND participants and 126 (72%) non-SEND participants. The sample of control group pupils is made up of 50 (27%) SEND participants and 134 (73%) non-SEND participants.

Pupil Premium

The sample of participating pupils includes 106 (61%) PP participants and 68 (39%) non-PP participants. The sample of control group pupils is made up of 116 (63%) PP participants and 68 (37%) non-PP participants.

Age

The sample of participating pupils includes 11 (6%) Year 8 pupils, 80 (46%) Year 9 pupils, 57 (33%) Year 10 pupils, and 26 (15%) Year 11 pupils. The sample of control group pupils includes 17 (9%) Year 7 pupils, 27 (15%) Year 8 pupils, 43 (23%) Year 9 pupils, 54 (29%) Year 10 pupils, and 43 (23%) Year 11 pupils.

4.2 Attendance

Overview

To understand the impact that the Face It programme was having on participants' attendance at school, we used the ImpactEd platform to integrate with participating schools' management information systems (MIS) and analyse attendance records over the course of the intervention. We compared pre- and post-intervention attendance for the three months pre- and six months post- each cohort's intervention.

Our key findings are as follows:

- On average, **attendance fell -8.1 percentage points**, from 92.5% during the pre-intervention period, to 84.4% during the post-intervention period. This **finding was statistically significant** ($p<0.05$, $n=170$), suggesting that a genuine decline in attendance occurred, rather than this being the result of random fluctuations in the data.

This finding is visualised below (Fig. 21).

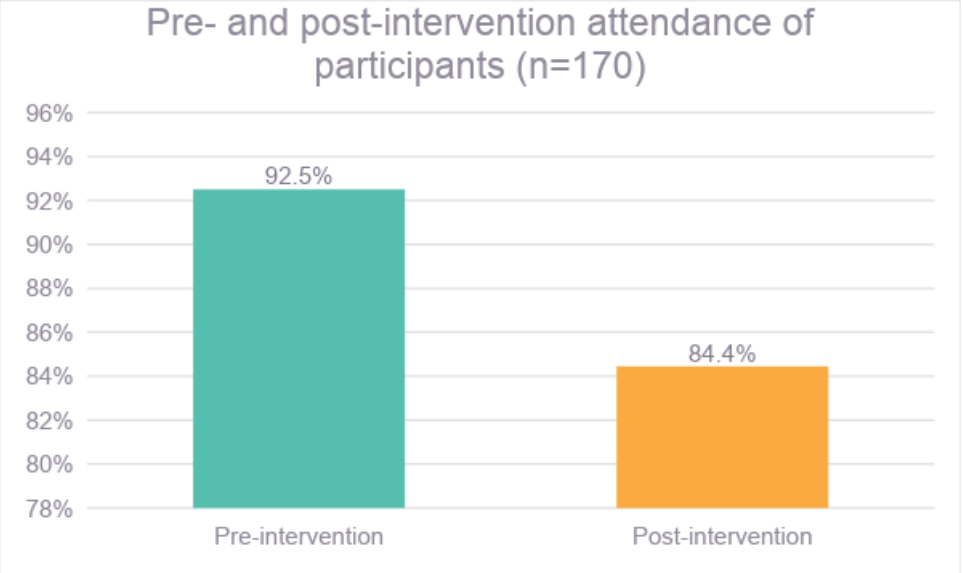


Figure 18

Context (Control Group and National Average)

Comparing the attendance data for Face It participants to the national average contextualises the results, and comparing it to a propensity score-matched control group allows us to understand whether the changes observed in attendance over the course of the programme can be attributed to the programme itself, or whether they are something that is happening to both participants and non-participants alike.

Our key findings are as follows:

- In the 22/23 academic year, the average attendance rate across state-funded UK schools was 92%. Although this was in line with participating pupils' attendance rates pre-intervention (92.5%), **post-intervention participating pupils' attendance rates fell -7.6 percentage points below the national average.**

- **Average attendance declined over the course of the programme for both the participating group and the control group.** The average attendance of participants fell by -8.1 percentage points (from 92.5% to 84.4%), and the average attendance of individuals in the matched control group fell by -5.0 percentage points (from 91.3% to 86.3%).

- Although this suggests that the attendance of the participant group declined more severely than that of the control group over the course of the intervention, a difference in difference analysis shows **this finding to be not statistically significant** (p>0.05, n=345). This suggests that **the discrepancy between participating and control individuals might be better explained by 'sampling error'** (i.e. normal, random fluctuations in the data).

This finding is visualised below (Fig. 22).

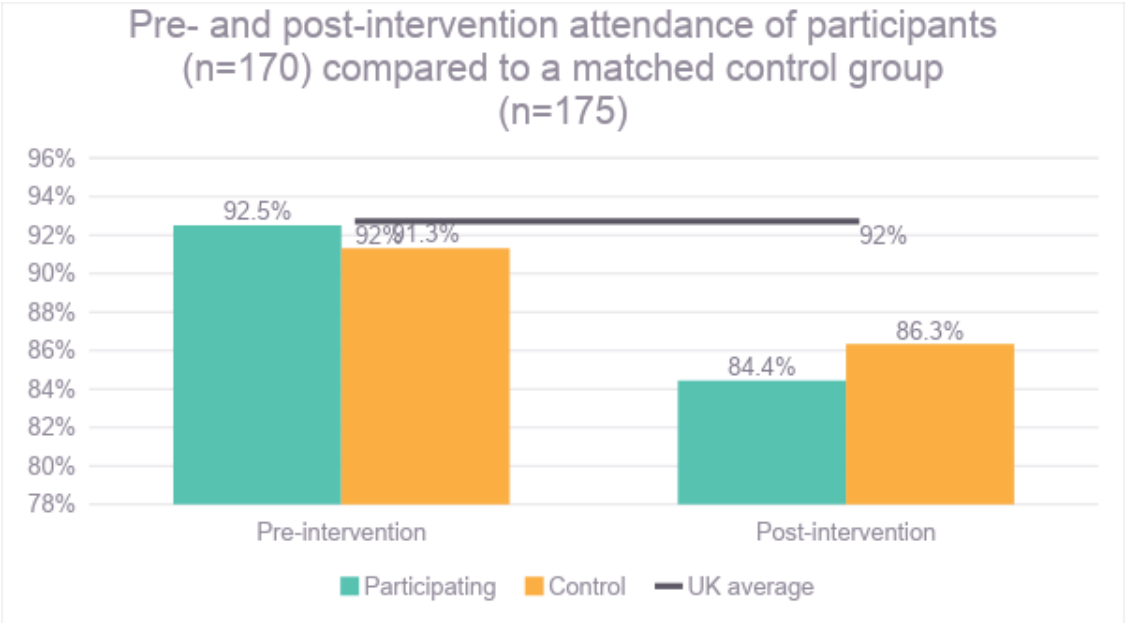


Figure 19

Subgroup analysis

To understand the impact of the Face It programme on different subgroups, we split the data by gender, Pupil Premium status, SEND, and age. We examined the changes in attendance metrics pre- and post-intervention for each of these different groups.

Our key findings are as follows:

Gender

- Both **male and female participating pupils demonstrated almost the same decline average attendance rates** following the intervention, with males' average score falling by -5.85 percentage points, and females' average score falling by -5.80 percentage points. The change for both males ($p<0.05$, $n=85$) and females ($p<0.05$, $n=85$) was **found to be statistically** significant meaning that suggesting that a genuine decline in attendance occurred, rather than this being the result of random fluctuations in the data.

- This decline in average attendance was broadly in line with the matched control group; **matched males' average attendance rates fell by -5.81 percentage points**, whilst **matched females' average attendance rates fell by -4.19 percentage points**. Difference in difference analysis shows the difference between the participating and control group males ($p>0.05$, $n=168$), and females ($p>0.05$, $n=177$) **not to be statistically significant**, meaning that the discrepancy between participating and control individuals might be better explained by 'sampling error' (i.e. normal, random fluctuations in the data).

These results are shown below (Fig 23).

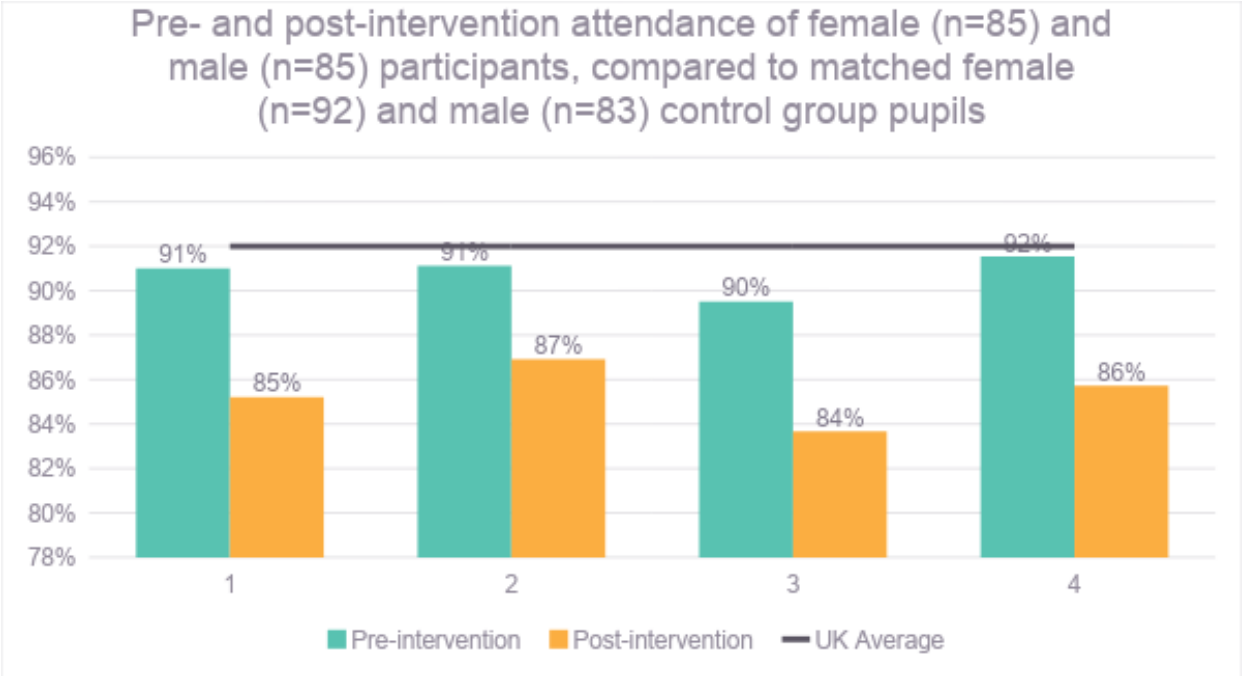


Figure 20

Pupil premium

- **Participating PP pupils reported the biggest decrease in attendance scores**, falling by -**6.17 percentage points** from 89.88% to 83.71%. **Non-PP pupils' average attendance scores fell by -5.18 percentage points**, from 90.91% to 85.73%. The change for both PP ($p<0.05$, n= 106) and non-PP ($p<0.05$, n= 64) pupils was found to be **statistically significant** suggesting that a genuine decline in attendance occurred.

- This decline in average attendance was broadly in line with the matched control group; **matched PP pupils' average attendance rates fell by -5.35 percentage points**, whilst **matched non-PP pupils' average attendance rates fell by -4.45 percentage points**. Difference in difference analysis shows the difference between the participating and control group PP pupils ($p>0.05$, n=217), and non-PP ($p>0.05$, n=128) **not to be statistically significant**, meaning that the discrepancy between participating and control individuals might be better explained by 'sampling error' (i.e. normal, random fluctuations in the data).

These results are shown below (Fig 24).

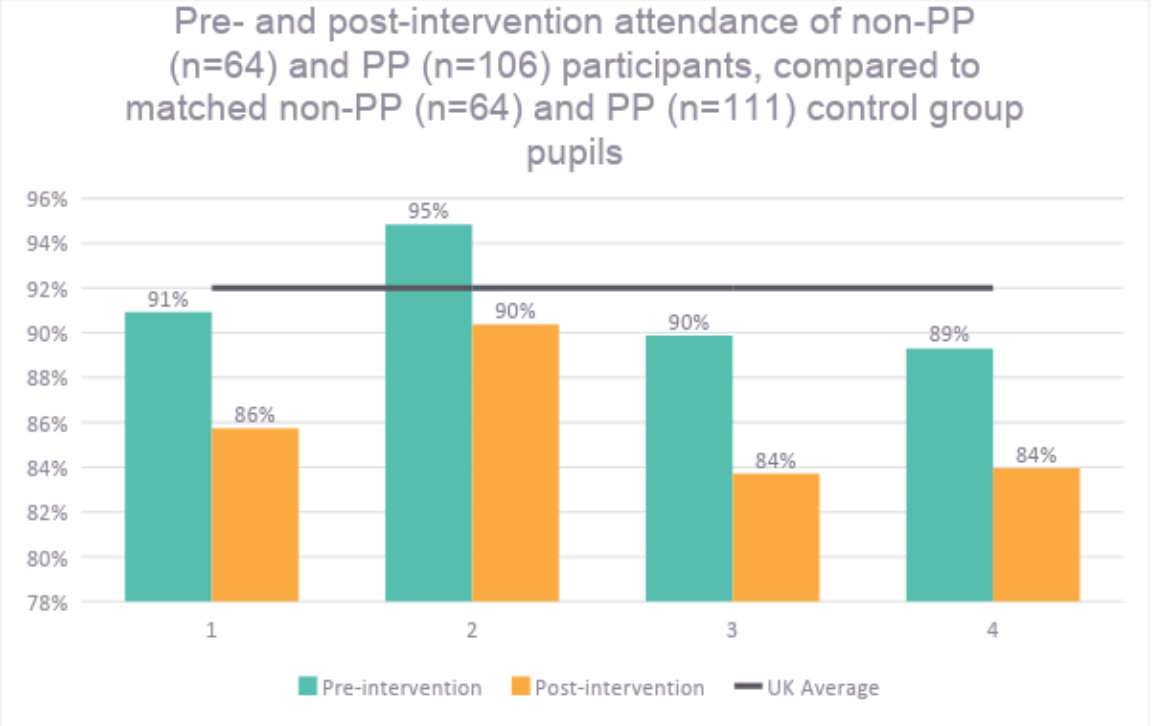


Figure 21

SEND

- **SEND pupils reported a bigger drop in attendance following the intervention than non-SEND pupils**; -7.85 percentage points compared to -5.08 percentage points for non-SEND participants. SEND pupils also started and ended the programme with lower attendance rates than non-SEND pupils; SEND participants' average attendance rates were 83.87% before the programme, and 76.02% after the programme, whilst non-SEND participants' average attendance rates were 92.53% before the programme, and 87.47% after the programme. The **fall in attendance rates was statistically significant for both groups** (SEND participants, $p<0.05$, $n=47$, non-SEND participants, $p<0.05$, $n=123$).

- The decline of SEND and non-SEND participants' attendance rates was **broadly in line with the decline of matched SEND and non-SEND control group pupils' attendance rates**. SEND control group pupils' rates fell by -7.20 percentage points, and non-SEND control group pupils' rates fell by -4.23 percentage points. Difference in difference analysis showed the difference between the participating and control groups was **not statistically significant** (SEND participants, $p>0.05$, $n=95$, non-SEND participants, $p>0.05$, $n=250$), meaning that the discrepancy between participating and control individuals might be better explained by 'sampling error' (i.e. normal, random fluctuations in the data).

These results are illustrated below (Fig 25).

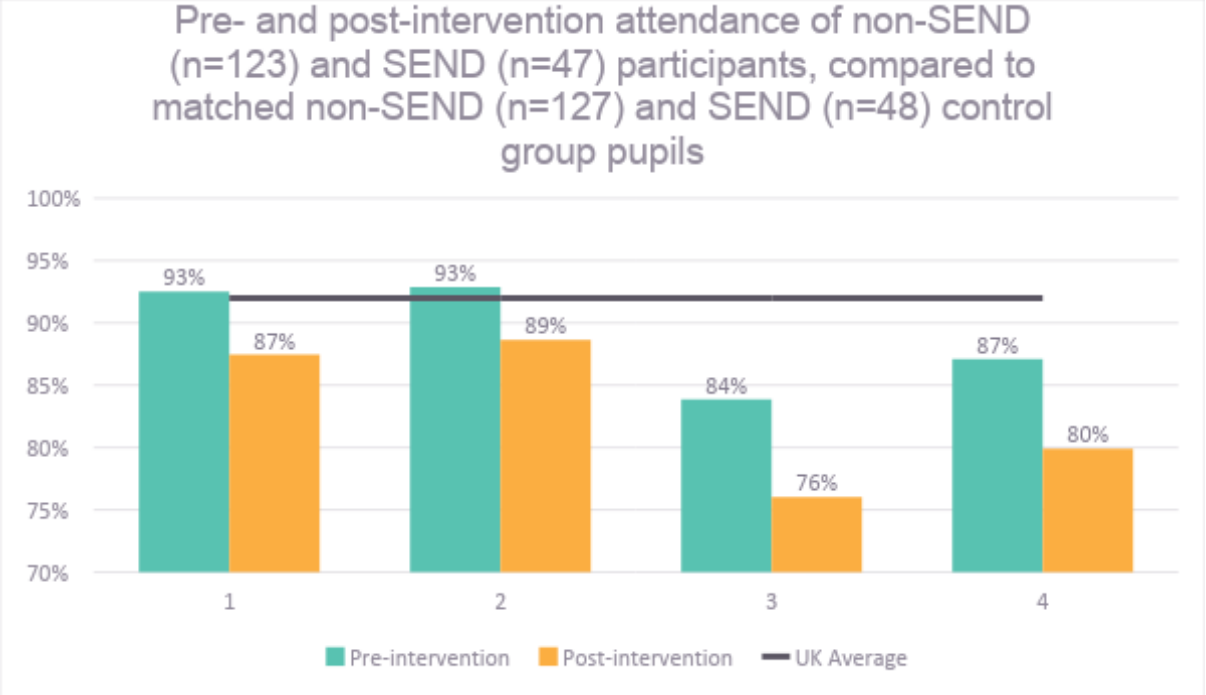


Figure 22

Year Groups

- Average attendance rates declined over the course of the programme for all Year Groups; however, there were quite substantial differences between them. The **decrease in attendance rates gets larger as the pupils increase in age**. Year 8 participating pupils had the least decrease (-2.26 percentage points), followed by Year 9 (-4.79 percentage points), Year 10 (-5.67 percentage points) and Year 11 (-14.70 percentage points).

- Year 11 participating pupils' attendance rates were measured a long way below the national average before and after the programme; pre-intervention they were measured at 76.28% and post-intervention at 61.68%. This means that **Year 11 pupils ended the intervention with average attendance 30 percentage points below national levels**.

- This **trend is broadly concurrent with national trends for Y11 attendance**; FFT Datalab analysis revealed a +5.3 percentage point increase in the percentage of Year 11 school sessions missed between 18/19 and 21/22.[2] The same analysis indicated a 0.9 percentage point reduction in missed Y11 sessions for 22/23, a result attributed to a reduction in illness-related absence likely due to the removal of COVID isolation requirements.

[2] https://ffteducationdatalab.org.uk/2023/06/year-11-attendance-this-year/

- The change in attendance scores **was statistically significant for Year 9 (p<0.05, n=78), Year 10 (p<0.05, n=55), and Year 11 (p<0.05, n=26) pupils**, but not for Year 8 (p>0.05, n=11) pupils.

- We see a **similar pattern in attendance rates in the matched control group data**. Year 11 pupils experienced the worst decline (-9.59 percentage points), followed by Year 10 pupils (-5.47 percentage points). Year 8 pupils were next with a decline of -0.42 percentage points, whilst Year 9 control group pupils' attendance scores remained relatively stable, seeing a drop of just -0.66 percentage points. Difference in difference analysis of each year group showed that the **differences between the participating and control group pupils was not statistically significant** (Year 8, p>0.05, n=38, Year 9, p>0.05, n=120, Year 10, p>0.05, n=103, Year 11, p>0.05, n=67), meaning that the discrepancy between participating and control individuals might be better explained by 'sampling error'.

These findings are shown in the graph below (Fig. 26).

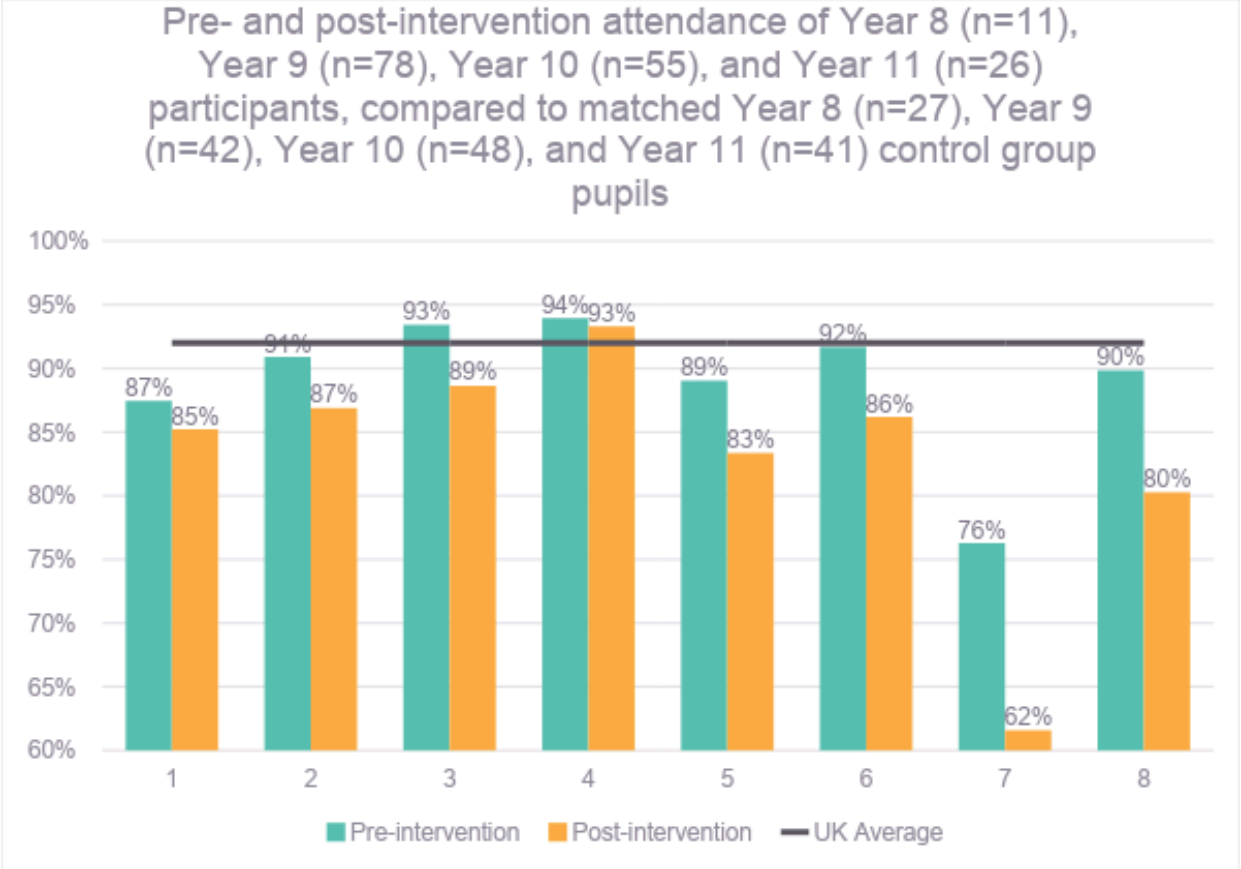


Figure 23

4.3 Exclusions

Overview

Similar to attendance, we analysed the impact of Face It on the number of exclusions participants received by examining data from schools' MIS using the ImpactEd platform. We compared pre- and post-intervention exclusions for the period three months pre- and six months post- the intervention, examining each cohort separately. We have reported on the total number of exclusions pre- and post-intervention, and the percentage change in the total number of exclusions.

- The **number of exclusions recorded for participating pupils increased by 7%** between the pre- and post-intervention windows. However, this change was **not statistically significant** (p>0.05 n=174), meaning that the observed change may be due to chance. These results are shown in the graph below (Fig. 27).

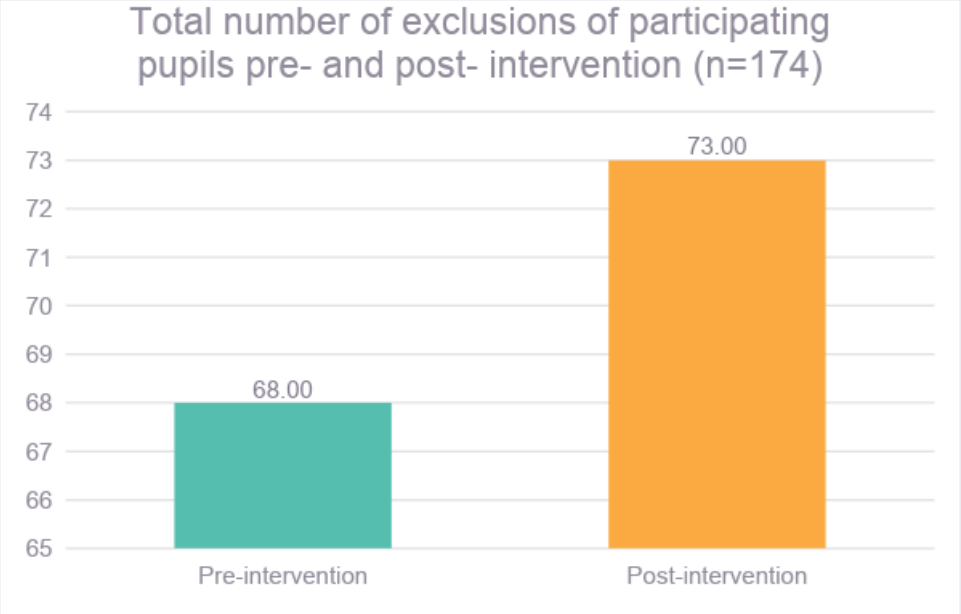


Figure 24

It should be noted that the number of exclusions in a fixed period is typically very small, which makes trends in the data more susceptible to random fluctuations. To be more confident of the impact of the Face It programme on exclusions, we would need to collect this data over a much longer time.

Context (Control Group)

Comparing the exclusions data for Face It participants to a propensity score-matched control group allows us to understand whether the changes occurring can be attributed to the intervention, or whether they are something that is happening to both participants and non-participants alike.

- Whilst **the percentage increase in exclusions was the same for participating (+7%) and control group pupils (+7%)**, control group pupils received a lot less exclusions points during the period measured. Whilst participating pupils ended the programme with a total of 73 exclusions between them, control group pupils had only

15. The difference between the two groups was **not statistically significant** (p>0.05, n=358). Results are shown in the graph below (Fig. 28).

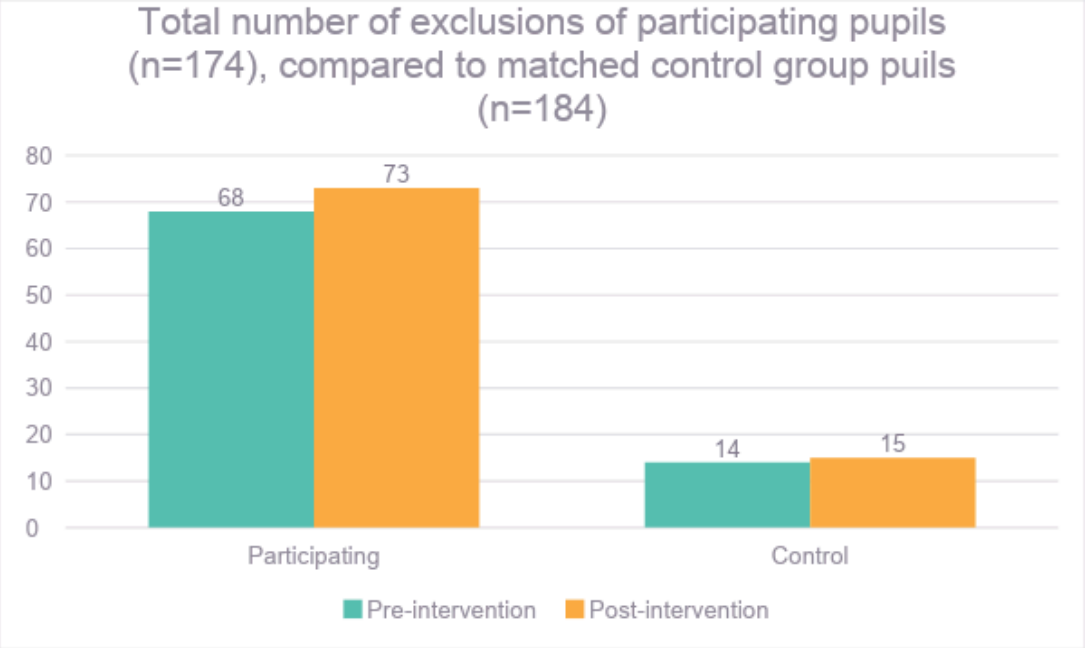


Figure 25

Subgroup analysis

To understand the impact of the Face It programme on different subgroups, we split the data by gender, Pupil Premium status, SEND, and age. We examined the percentage change in the total number of exclusions pre- and post-intervention for participating and control group pupils in each of these different groups.

Our key findings are as follows:

Gender

- The difference in the change in exclusions between male and female pupils is stark. Whilst **male pupils' saw a -16% decrease** in the number of exclusions received, **female pupils saw a +68% increase.** Neither change was statistically significant (male, p>0.05, n=87, female, p>0.05, n=87).

- The **same trend was seen in the matched control male and female groups**. Whilst the number of exclusions received by male control group pupils during the two time points fell from 14 to 9, the number of exclusions received by female control group pupils increased from 0 to 6. However, difference in difference analysis showed that **the difference between the male and female control and participating groups was not statistically significant** (male, p>0.05, n=176, female, p>0.05, n=182).

Findings are seen in the graph below (Fig. 29).

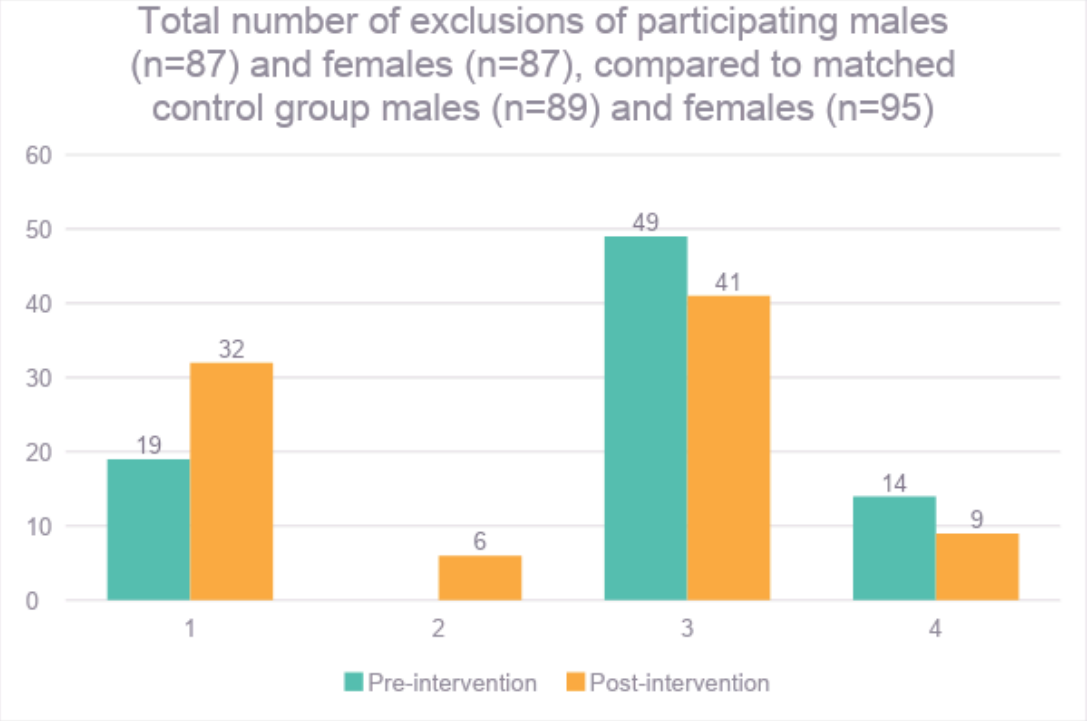


Figure 26

Pupil premium

- The total number of exclusions received by participating **PP pupils increased by 2%**, compared to participating **non-PP pupils who saw an increase of 16%.** Changes were **not statistically significant** for either group (PP, p>0.05, n=106, non-PP, p>0.05, n=68).

- The difference observed between the participating and control PP pupils was stark. While **participating pupils' exclusions scores increased by +2%, matched control group PP pupils' exclusions scores decreased by -26%.** The change in non-PP participating and control groups was relatively similar, with both increasing by a total of 4 and 5 respectively. The differences between the participating groups were **not statistically significant** (PP, p>005, n=222, non-PP, p>0.05, n=136), meaning that we cannot rule out the possibility that observed were due to chance.

These findings are shown below (Fig. 30).

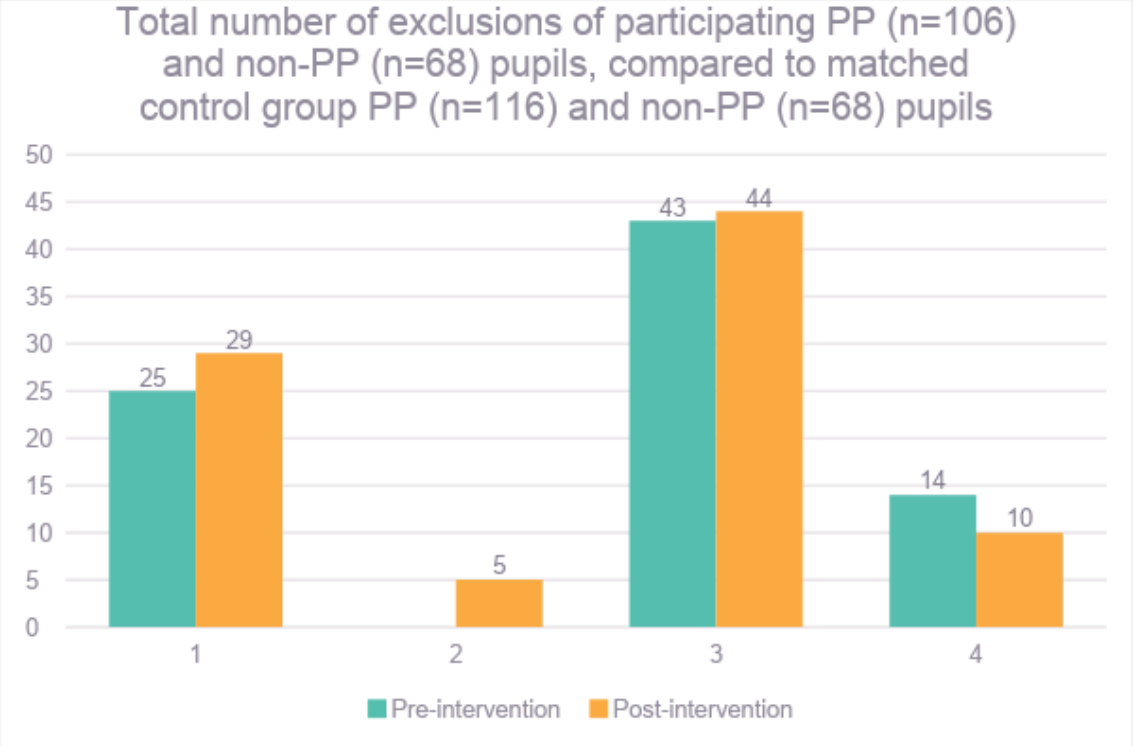


Figure 27

SEND

- Whilst the total number of exclusions collected by **non-SEND participating pupils** before and after the intervention **increased by +13% from 47 to 53**, the number of exclusions collected by **SEND pupils decreased by -5% from 21 to 20**. The changes observed by the two groups were **not statistically significant** (non-SEND, p>0.05, n=126, SEND, p>0.05, n=48).

- The number of exclusions also decreased for SEND control group pupils, but by a larger proportion, decreasing by -64% from 14 to 5. Conversely the number of exclusions increased for non-SEND control group pupils from 0 to 10. This means that **the same pattern was observed across participating and control SEND and non-SEND groups**. However, the difference between the participating and control groups was **not statistically significant** (non-SEND, p>0.05, n=260, SEND, p>005, n=98) meaning that the changes observed may be due to natural fluctuations in the data.

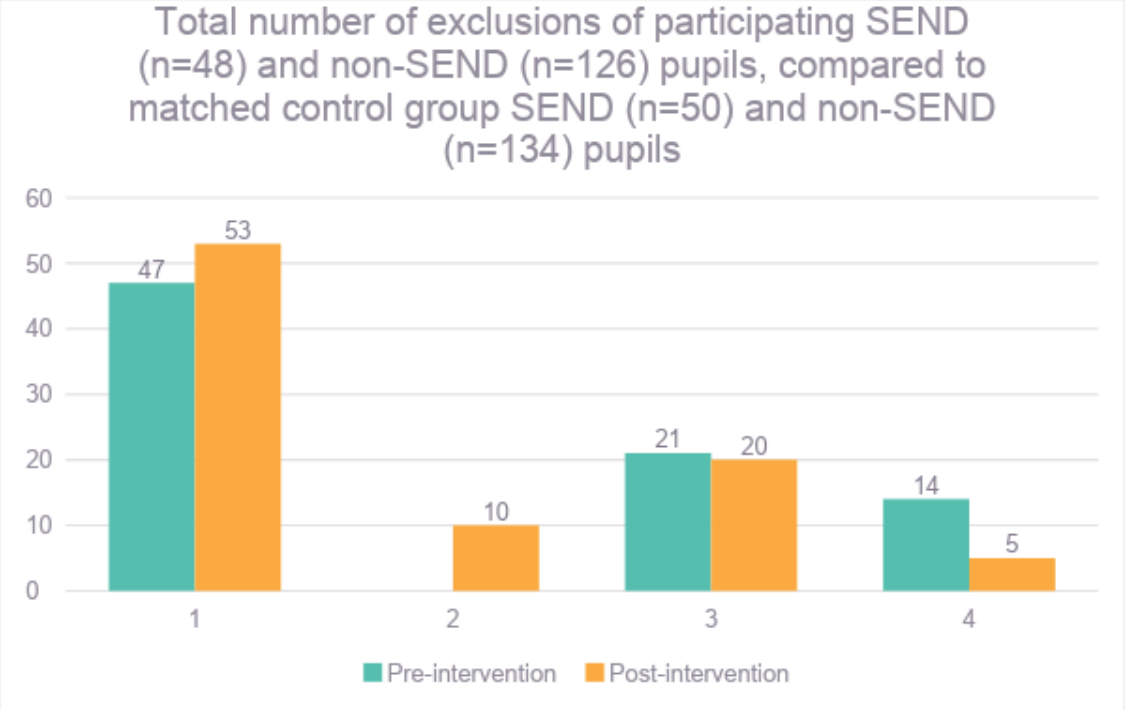


Figure 28

Year Groups

- **Year 11 and Year 8 pupils saw a decrease in the number of exclusions received**, of -33% and -14% respectively. Conversely, **Year 10 and Year 9 pupils saw an increase in the number of exclusions they received**, of +31% and +7% respectively. **None of the changes observed were statistically significant.**

- There was **no pattern across Year Groups between the participating and control group pupils**. While Year 8 participating pupils' exclusions decreased from 7 to 6 (-14%), Year 8 control group pupils' exclusions remined stable at 5 (+0%). Year 9 participating pupils' exclusions increased from 45 to 48 (+7%), Year 9 control group pupils' exclusions increased from 2 to 3 (+50%). Year 10 participating pupils' exclusions increased from 13 to 17 (+31%), Year 10 control group pupils' exclusions decreased from 7 to 6 (-14%). Finally, Year 11 participating pupils' exclusions decreased from 3 to 2 (-33%), Year 11 control group pupils' exclusions increased from 0 to 1. **None of the differences between the control and participating year group sub-groups were statistically significant** (Year 8, $p>0.05$, $n=38$, Year 9, $p>0.05$, $n=123$, Year 10, $p>0.05$, $n=11$, Year 11, $p>0.05$, $n=69$).

Results are shown in the graph below (Fig. 32).

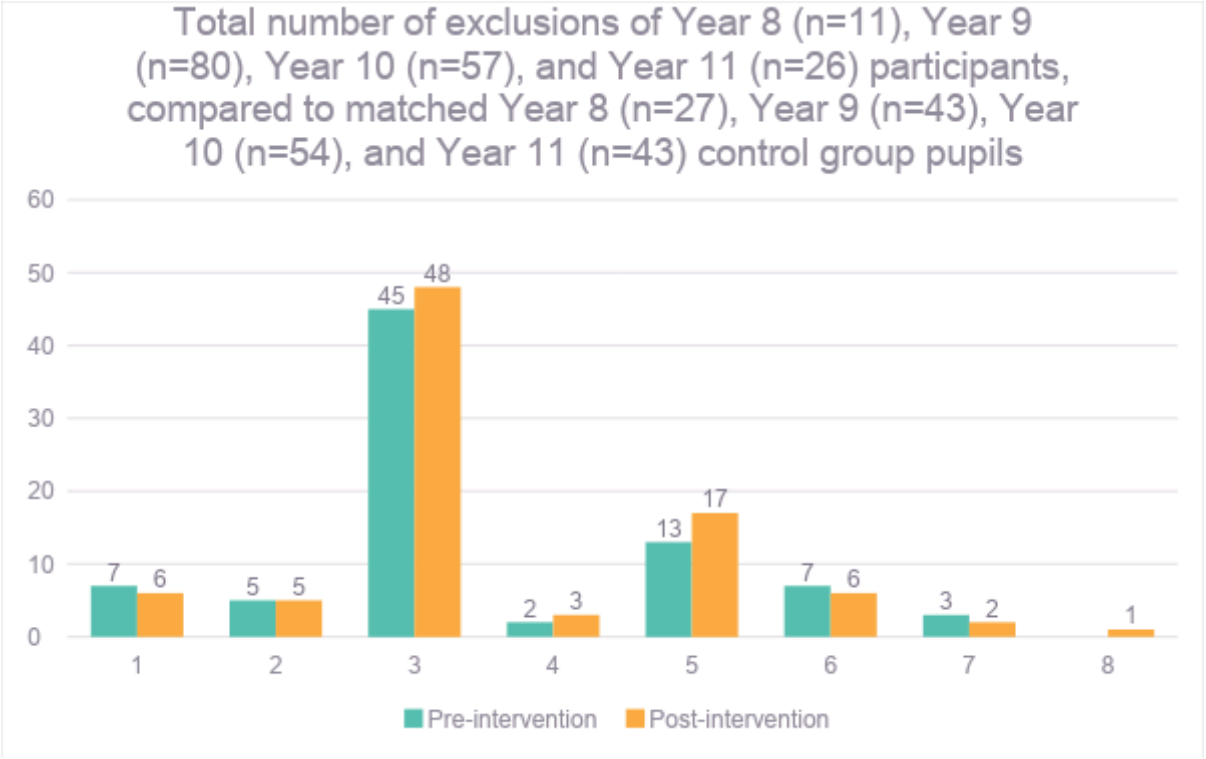

Total number of exclusions of Year 8 (n=11), Year 9 (n=80), Year 10 (n=57), and Year 11 (n=26) participants, compared to matched Year 8 (n=27), Year 9 (n=43), Year 10 (n=54), and Year 11 (n=43) control group pupils
60
50
40
30
20
10
0
7
6
5
5
45
48
2
3
13
17
7
6
3
2
1
1
2
3
4
5
6
7
8
Pre-intervention
Post-intervention

Figure 29

5. Conclusion

This report aimed to analyse data from twenty-five cohorts total, alongside a constructed control group via Propensity Score Matching for engagement data, to determine:

- **The impact of the programme on participant wellbeing**

55.4% of participants reported an increase in their wellbeing over the course of the programme. This change in wellbeing was statistically significant ($p<0.001$, $n=166$), which suggests that a genuine change in wellbeing occurred, rather than this being the result of sampling error. All subgroups reported an improvement in wellbeing.

- **The impact of the programme on participant resilience**

66.5% of participants reported an increase in their resilience over the course of the programme. This change in resilience was statistically significant ($p<0.001$, $n=167$), which suggests that a genuine change in resilience occurred. All subgroups reported an improvement in resilience.

- **The impact of the programme on emotional control, including utility of the *Face It* coping skills taught during the intervention**

63.5% of participants reported an increase in their emotional control over the course of the programme. This change in emotional control was statistically significant ($p<0.001$, $n=167$), which suggests that a genuine change in emotional control occurred. 88% of participants reporting using the coping skills post-programme. All subgroups reported an increase in emotional control.

Overall, this suggests Face It is effective in meeting the shorter-term Theory of Change outcomes around social and emotional wellbeing and these aspects of the programme should be retained. Furthermore, there were only small variances in improvement levels of wellbeing, resilience, emotional control and use of the coping skills for different subgroups (apart from emotional control, in which PP students reported substantially higher post-intervention scores). This suggests Face It is suitable and useful for its target participants and recruitment processes should be retained. It also indicates that the coping skills taught in the programme provide extended benefit for all participant subgroups.

Looking at the impact of the programme on school engagement measures, we see that:

- **Average attendance declined over the course of the programme for both the participating group and the control group.** The average attendance of participants fell by -8.1 percentage points (from 92.5% to 84.4%), and the average attendance of individuals in the matched control group fell by -5.0 percentage points (from 91.3% to 86.3%), leaving both participating and control pupils below the national average of 92%. While there appears to be slightly more decrease for participating pupils, this finding was not found to be statistically significant($p>0.05$, $n=345$). This suggests that **the discrepancy between participating and control individuals might be better explained by 'sampling error'** (i.e. normal, random fluctuations in the data).

- The **number of exclusions recorded for participating and control pupils increased by 7%** between the pre- and post-intervention windows (a 6-month period). However, this change was **not statistically significant** (p>0.05 n=174), meaning that the observed change may be due to chance.

Recommendations

- Over half of participants reported an increase in wellbeing, resilience, emotional control and use of the coping skills following the intervention. **This suggests Face It is effective in meeting the shorter-term Theory of Change outcomes around social and emotional wellbeing and these aspects of the programme should be retained.**
- There was only marginal difference in improvement levels of wellbeing, resilience, emotional control and use of the coping skills for different subgroups (apart from emotional control, in which PP students reported substantially higher post-intervention scores). **This suggests Face It is suitable and useful for its target participants and recruitment processes should be retained.**
- Due to the national rise and concerns with attendance and exclusions across the academic years of the evaluation, it is difficult to infer the impact of the programme on attendance and exclusions, though the participating and control pupils were below national average for attendance, participating pupils did not make statistically significant decreases or increases compared to control pupils. **Further investigation would be needed to fully understand any changes observed.**

6. Appendix

Evaluation terminology

Academically validated measures

These are scales to measure social and emotional skills linked to academic achievement and long-term life outcomes that have been developed and peer reviewed by academic researchers within the fields of education and psychology. These have been developed to ensure:

- Predictive validity. These skills have been shown to be closely related to desirable life outcomes such as educational achievement, employability and earnings potential, or long-term health and life satisfaction. (In psychometrics, predictive validity is the extent to which a score on a scale or test predicts scores on some criterion measure. For example, the validity of a cognitive test for job performance is the correlation between test scores and, say, supervisor performance ratings.)
- Construct validity. The measure tests for the skill that it says it does, as defined in the literature.
- Test-retest validity. The results stay the same when tests are repeated.

Baseline

The initial assessment of pupils' attainment or social and emotional skills, at the start of an evaluation.

Change over time

The difference between a pupil's baseline result and their final result, either for attainment or social and emotional skills. This indicates progress made during participation in the programme. This will begin to indicate whether the programme has had an impact on pupils, though we must also account for other factors that could lead to this change, which is why we recommend the use of control groups and qualitative analysis.

Control Group

A control group is composed of students who do not participate in the programme and who closely resemble the pupils who take part in the programme in attainment and demographic traits. It is used to get an indication of whether a change in results over the course of the programme can likely be attributable to the programme itself, or whether results were likely to change over time in any case. Also known as a comparison group.

Evaluation

An evaluation is set up to measure the impact of a particular programme. This will involve monitoring the programme over a specified period, for one or more groups, in order to

evaluate the progress participating pupils make. One programme can involve multiple evaluations, and we recommend gathering data across multiple time points to ensure valid and reliable results are generated.

Final

The final assessment of pupils' attainment or social and emotional skills at the end of an evaluation.

Matched Pupils

Matched Pupils are pupils who carried out both a baseline and a final assessment at the start and end of the evaluation. It can be useful to consider results from Matched Pupils only because this means only including those pupils who participated in the full duration of the programme.

Outcomes

We use outcomes to refer collectively to any social and emotional skills and academic attainment scores that are being measured over the course of an evaluation.

Participating pupils

The group of pupils participating in the evaluation, and not forming part of a control group.

Social and emotional skills

The term ‘social and emotional skills’ refers to a set of attitudes, behaviours, and strategies that are thought to underpin success in school and at work, such as motivation, perseverance, and self-control. They are usually contrasted with the ‘hard skills’ of cognitive ability in areas such as literacy and numeracy, which are measured by academic tests. There are various ways of referring to this set of skills, such as: non-cognitive skills, twentieth century skills and soft skills. Each term has pros and cons; we use social and emotional skills for consistency but we recognise that it does not perfectly encapsulate each of the skills that come under this umbrella.

Statistical analysis terminology

Statistically significant

A result has statistical significance when it is very unlikely to have occurred given the null hypothesis. In other words, if a result is statistically significant, it is unlikely to have occurred due purely to chance.

P Value

A p-value is a measure of the probability that an observed result could have occurred by chance alone. The lower the p-value, the greater the statistical significance of the observed difference. Typically a p-value of ≤ 0.05 indicates that the change was statistically significant. A p-value higher than 0.05 (> 0.05) is not statistically significant and indicates strong

evidence for the null hypothesis; i.e. that we cannot be confident that this change did not occur due purely to chance.

Education terminology

Pupil Premium (PP)

The pupil premium grant is designed to allow schools to help disadvantaged pupils by improving their progress and the exam results they achieve. Whether a child is eligible for Pupil Premium funding is often used by schools as an indicator of disadvantage.

SEND

A child or young person has special educational needs and disabilities if they have a learning difficulty and/or a disability that means they need special health and education support; this is usually shortened to SEND.

www.impacted.org.uk

Partners and supporters

United Learning The best in everyone™

THE YOUNG ACADEMY

CHALLENGE PARTNERS

SUPER BEING

nesta

TeachFirst INNOVATION PARTNER 2018/19

LEEDS BECKETT UNIVERSITY

MANCHESTERTSA Teaching School Alliance

Improving pupil outcomes by working with schools to address the evaluation deficit.

Get in touch
hello@impacted.org.uk

ImpactEd