

UK Resilience Programme Evaluation: Final Report

Amy Challen, Philip Noden, Anne West
and Stephen Machin

This research report was commissioned before the new UK Government took office on 11 May 2010. As a result the content may not reflect current Government policy and may make reference to the Department for Children, Schools and Families (DCSF) which has now been replaced by the Department for Education (DFE).

The views expressed in this report are the authors' and do not necessarily reflect those of the Department for Education.

ACKNOWLEDGEMENTS

We gratefully acknowledge the contribution of the Local Authority staff and of all the school staff and pupils involved in this evaluation.

We also owe thanks to Dr. Karen Reivich and Dr. Jane Gillham of the Penn Resiliency Project, who provided extremely valuable advice on the evaluation design.

EVALUATION OF THE UK RESILIENCE PROGRAMME

There are increasing concerns about children's well-being in the UK, their behaviour, and the low academic attainment of a large fraction of the population.¹ The Every Child Matters agenda stressed schools' potential and duty to promote pupils' well-being. In September 2007, three local authorities (South Tyneside, Manchester and Hertfordshire) piloted a programme with Year 7 pupils in 22 of their schools, with the aim of building pupils' resilience and promoting their well-being: the UK Resilience Programme. More schools have since started teaching the programme.

This evaluation aims to investigate whether the programme (previously trialled in small samples) can be delivered at scale; whether it has an impact on children's well-being, behaviour, attendance and academic attainment.

The first interim report was published in April 2009 and gives an overview of the UK Resilience Programme and its implementation, describes the evaluation, and offers preliminary findings about programme impact, as well as detailed case studies on the first year of programme implementation. The report also contains a bibliography and descriptions of previous research on the Penn Resiliency Program (the curriculum on which UKRP is based), and describes the curriculum in detail.

The Second Interim Report was published in June 2010 and contains detailed case studies from the third year of implementation, as well as updated information about programme impact.

In this final report we will refer back to the two interim reports², which provide more detail and background information.

The First Interim Report can be found online at:

<http://www.education.gov.uk/publications//eOrderingDownload/DCSF-RR094.pdf>

The Second Interim Report can be found online at:

<http://www.education.gov.uk/publications/eOrderingDownload/DFE-RR006.pdf>

The qualitative work for this report was carried out by Dr. Philip Noden and Prof. Anne West.³

The quantitative work is by Amy Challen⁴ and Prof. Stephen Machin⁵.

¹ See, for example, the recent UNICEF report "An overview of child well-being in rich countries" which puts the UK at the bottom of a list of 21 advanced countries:

http://www.unicef-irc.org/publications/pdf/rc7_eng.pdf

² Referred to as Challen et al. (2009) and Challen et al. (2010). All three evaluation reports are by the same authors.

³ Both Education Research Group, Department of Social Policy, London School of Economics and Political Science.

⁴ Centre for the Economics of Education and Centre for Economic Performance, London School of Economics and Political Science.

⁵ Centre for the Economics of Education and Centre for Economic Performance, London School of Economics and Political Science, and Department of Economics, University College London.

CONTENTS

1. Executive Summary	4
2. Policy and Delivery Implications	6
3. The UK Resilience Programme	8
4. The Evaluation Design	13
5. Quantitative Analysis	16
Summary of quantitative findings	16
Data collection	17
Explanation of measures	18
Treatment and control groups	20
Programme impact on symptoms of depression	22
Programme impact on symptoms of anxiety	29
Behaviour	30
Life satisfaction	31
Absence from school	32
Academic attainment	33
Conclusion	35
Implementation and policy issues	37
Tables	39
6. Qualitative Findings	63
Summary of qualitative findings	63
Data and methods	64
Brief description of the UKRP	65
Pupils' reported use of the UKRP skills	66
Participants' reflections on the UKRP	73
Organisation of the UKRP within schools	76
Conclusion	79
7. Appendix A: Questionnaires used	82
8. Appendix B: References for empirical evaluations of PRP and UKRP	83

1. Executive Summary

Evaluation of the UK Resilience Programme, Final Report

The UK Resilience Programme (UKRP) aims to improve children's psychological well-being by building resilience and promoting accurate thinking. Three local authorities launched it in the academic year 2007-08, with workshops delivered to Year 7 pupils in secondary schools. This report presents findings from the UKRP evaluation, commissioned by the Department for Children, Schools and Families (now the Department for Education).

Methodology

Information on pupils' well-being was collected through questionnaires administered to pupils who had participated in the first year of UKRP workshops and to a control group. Pupils were surveyed before and after the programme. The quantitative work examines the impact on the original cohort of pupils (those in workshops in 2007-08) over a three year period. In addition, interviews with pupils, facilitators (workshop leaders) and school managers were carried out in 10 of the 22 secondary schools involved in the programme at the end of the academic year 2007-08, and follow-up interviews were carried out in 9 of these 10 schools in the autumn term of 2009-10. The interviews explore participants' experiences of the programme, and how schools were implementing the programme in the pilot year and then two years later.

Key findings

- The quantitative work found a significant short-term improvement in pupils' depression symptom scores, school attendance rates, and academic attainment in English. There was some impact on anxiety scores and maths attainment, but this was inconsistent and concentrated in a few groups of pupils.
- The size of the impact varied by how workshops were organised. Weekly workshops showed a larger impact than those timetabled fortnightly.
- The impact also varied by pupil characteristics. Pupils who were entitled to free school meals; who had not attained the national targets in English or maths at Key Stage 2; or who had worse initial symptoms of depression or anxiety; were all more likely to experience a larger measured impact of the workshops on their depression and anxiety scores. There was little difference by pupil characteristics on the absence rate.
- On average the effect of the workshops lasted only as long as the academic year, and had faded by the one-year follow-up questionnaire in June 2009. However, there was still an impact for certain groups at follow-up, particularly for pupils who had not attained the national target levels in English or maths at Key Stage 2. There was no impact on any of the outcome measures by the two-year follow-up in June 2010.
- There was no measured impact of workshops on behaviour scores or life satisfaction scores.
- Return visits to nine of the case study schools in autumn 2009 revealed that seven of the nine schools were continuing to deliver the UKRP to all Year 7 pupils.
- Facilitators were extremely positive about the ideas underlying the programme and about the training they had received. Most reported that they used the skills themselves.
- Facilitators found the curriculum materials too didactic and thought they could be improved.

- Pupils were generally positive about the programme. Interviews for the First Interim Report suggested that pupils had applied UKRP skills in real life situations, and some interviewees showed a good understanding of elements of the programme.

2. Policy and Delivery Implications

Here we list potential policy implications of the results presented in this report and the findings of the interim reports. These are aimed at schools or local authorities which use the programme or are considering doing so. Many of these points should be seen as considerations rather than recommendations, but they do highlight issues around the implementation of the programme.

- 1) The UK Resilience Programme did have a small average impact on pupils' depression scores, school attendance, and English and maths grades, but only in the short run (up to one-year follow-up). There was no average impact on any measure at two-year follow-up. This means that any improvements in pupils' psychological well-being, attendance and attainment were short-lived, and by the time of the two-year follow-up (June 2010) pupils who had participated in UKRP workshops were doing no better on these outcomes than pupils who had not. This suggests that a single set of UKRP lessons is not enough to permanently change pupils' outcomes on average.
- 2) The impact of the programme varied by pupil characteristics, and was much stronger for more deprived and lower-attaining pupils and those who started the year with worse psychological health, particularly girls with these characteristics. Thus even if there is no average impact of the programme beyond the short run (i.e. an impact when measured over all pupils), it appears that some pupils benefitted substantially more, and for longer. These findings suggest that the improvements experienced by these pupils were more likely to be meaningful in terms of the impact on their lives, perhaps longer term as well as in the short run.
- 3) While our quantitative findings suggest there was initially a statistically significant gain in the mental health and well-being of pupils, and many interviewees believed the programme was having a positive impact on their pupils, schools and facilitators should keep in mind the possibility that the programme could have a negative effect for individual pupils.
- 4) A preferred model of delivery for the UKRP, based on the recommendations of the course developers and the findings of this study, might involve 18 weekly sessions delivered to groups of no more than 15 pupils.
- 5) For the UKRP to thrive within schools it is extremely important that the programme has backing from the school's senior management.
- 6) This backing is all the more important when schools face competing pressures such as the need to improve standards of attainment. This may also prove to be the case in relation to financial pressures on schools, for example, if there is a deterioration in pupil: adult ratios.
- 7) There was evidence of a drift, in some schools, towards the programme being delivered by members of the school auxiliary staff. This will clearly reduce the size of the pool from which facilitators may be drawn which would have an impact on the quality of staff who may train as facilitators.
- 8) The role of facilitator can be emotionally demanding due to the distressing nature of some real life problems raised by pupils. Staff need to be adequately prepared for and supported throughout the programme in order to deal with these issues.
- 9) Facilitators were very positive about the quality of training they had received for the UKRP. They had reservations however about the quality of teaching materials provided

for the programme. If the materials are not regarded as being of a sufficiently high quality, facilitators may seek alternative resources and clearly this may constitute a threat to programme fidelity.

10) The UKRP was intended to be a universal programme, but some schools have chosen to target pupils for inclusion in workshops. It is not clear which model is preferable, and this will probably depend on the situation of each school. However, the following points are worth bearing in mind:

- Based on the quantitative analysis, certain groups of pupils appeared to benefit more from the workshops, particularly those who did not achieve the national target level in English and maths at Key Stage 2, pupils with SEN, and pupils who started the school year with higher levels of depression or anxiety symptoms.
- However, the measured impact on these pupils is the impact of the programme delivered to 'universal' or mixed workshop groups, not of groups consisting entirely of targeted pupils. One cannot therefore assume that the same impact would be obtained if workshop groups were targeted.
- Although facilitators and other school staff often appeared to assume that higher ability pupils were naturally more resilient, or had fewer problems, almost all facilitators claimed to use the UKRP skills themselves. It therefore seems unlikely that higher ability pupils or those with better initial psychological well-being would be unable to benefit from the skills.
- Even if pupils were to be targeted for inclusion in workshops, it is important that they should be targeted appropriately. Previous research suggests that school staff tend to identify pupils with behaviour problems rather than those with emotional difficulties, yet the programme is primarily designed to address the latter. The process of targeting would also need to be carefully considered.
- Participation in programmes perceived to be targeted and remedial can attract stigma for those who participate. Universal programmes avoid this.
- The measures used in the quantitative evaluation are sensitive to differences in the severity of symptoms of depression and anxiety, but are not good at distinguishing between children who have few or no symptoms. For instance, they would not be able to detect any improvements in well-being for pupils who showed no initial symptoms of depression, although this would not necessarily mean that these children did not benefit.
- The skills pupils used most (as reported by both pupils and facilitators) were the interpersonal skills around negotiation and assertiveness, and techniques for self-control (see Chapters 5 & 6 of the First Interim Report). Since all pupils are likely to experience conflict and problems around everyday social interactions it is likely that all pupils could benefit from the workshops, at least in these areas.

3. The UK Resilience Programme

The UK Resilience Programme is the UK implementation of the Penn Resiliency Program, a well-being programme that has been trialled more than 13 times in different settings. The UKRP was taught from September 2007 in three participating local authorities, and those workshops that took place in mainstream schools form the subject of this evaluation.

This section outlines the PRP curriculum used and its implementation in the UK, with particular reference to the evaluation.

The Penn Resiliency Program

The Penn Resiliency Program (PRP) is a curriculum developed by a team of psychologists at the University of Pennsylvania. Its original aim was to prevent adolescent depression, but it now has a broader remit of building resilience and promoting realistic thinking, adaptive coping skills and social problem-solving in children. The primary aim of the programme is to improve psychological well-being, but it is possible that any such improvement could also have an impact on behaviour, attendance and academic outcomes. Thirteen controlled trials have found PRP to be effective in helping protect children against symptoms of anxiety and depression, and some studies have found an impact on behaviour. The skills taught in PRP could be applied in many contexts, including relationships with peers and family members, and achievement in academic or other activities.⁶

PRP is a manualised intervention comprising 18 hours of workshops. (“Manualised” means that no additional materials or resources are required to lead the workshops.) The curriculum teaches cognitive-behavioural and social problem-solving skills. Central to PRP is Ellis’s *Activating-Belief-Consequences* model that beliefs about events mediate their impact on emotions and behaviour.⁷ PRP participants are encouraged to identify and challenge (unrealistic) negative beliefs, to employ evidence to make more accurate appraisals of situations and others’ behaviour, and to use effective coping mechanisms when faced with adversity. Participants also learn techniques for positive social behaviour, assertiveness, negotiation, decision-making, and relaxation.

The manualised nature of the curriculum and the intensive training required before using it allows facilitators to be drawn from a wide range of professions and agencies including teachers, learning mentors, teaching assistants, psychologists and health professionals. The training for the original cohort of teachers lasted around 8-10 days, with the first half of the course focusing on teaching trainees the adult-level Cognitive Behavioural Therapy (CBT) skills, and the second week on familiarising them with the students’ curriculum and practising how to communicate it to pupils.⁸

⁶ Some PRP studies have found some effects on behaviour (e.g. Jaycox et al. 1994; Roberts et al. 2003), and academic attainment has not yet been evaluated in a (published) PRP study, although we will examine both of these outcomes. The grounds for believing that the programme could have an impact on behaviour, peer relationships and academic attainment are the suggested links between these outcomes and psychological well-being (see, for example, Kaslow, Rehm and Siegel, “Social-Cognitive and Cognitive Correlates of Depression in Children”, *Journal of Abnormal Child Psychology*, 1984). Moreover, social skills feature prominently in the PRP curriculum, and these are the skills most commonly said to be used by pupils, according to pupils and facilitators (see Sections 5 and 6 of the First Interim Report, Challen et al. 2009).

⁷ See, for example, Ellis’s “Emotional Disturbance in a Nutshell”:

http://www.rebt.org/emo_disturbance.pdf (accessed 17/02/11)

⁸ More recent training events for the PRP have been cut down to 5-7 days.

Additional information on the content of each PRP lesson can be found in Annex C of the First Interim Evaluation Report, Challen et al. (2009).

Additional information on PRP can be found online at:

<http://www.ppc.sas.upenn.edu/prpsum.htm>

Summary of PRP Research

Overall, the controlled trials that have been conducted of PRP suggest that it could prevent symptoms of depression and anxiety in universal, targeted and clinic samples, and some studies have found some evidence of a reduction in disruptive behaviour (see e.g. Jaycox et al. 1994; Roberts et al. 2003; most studies did not measure behaviour). However, there are some inconsistent findings, and a meta-analysis of the PRP research, which includes both published and unpublished research, finds very mixed results across studies (Brunwasser, Gillham & Kim, 2009). Some studies found no effect on depressive symptoms, while others found an effect on some groups but not others. In an earlier review of the PRP studies (Gillham, Brunwasser & Freres, 2008), the PRP team further find a link between measured impact and the level of training and supervision of the workshop facilitators, suggesting that despite the manualised curriculum, facilitator quality is important and treatment heterogeneity (differences in the *quality* of the programme delivered) is likely. In addition, the sample sizes used in prior PRP studies are relatively small, and scaling-up is a common evaluation problem, with the efficacy of an intervention frequently decreasing as the number of subjects involved increases.⁹

For a bibliography of empirical evaluations of PRP and UKRP please see Appendix B below.

For a more detailed summary of previous research on PRP, please see Annex B at the end of the First Interim Report (Challen et al., 2009), an executive summary of the programme and research by members of the Penn team who developed the programme.

UK implementation

The UK Resilience Programme is the first larger-scale use of the PRP curriculum. It was first implemented in 2007-08 as 18 hours of workshops for Year 7 children in 22 UK secondary schools, along with additional workshops with children of similar ages in other settings such as Pupil Referral Units and special schools. The first year of workshops involved around 2000 pupils.

Subsequently, some of the original schools have gone on to offer workshops to subsequent cohorts of pupils, and some have had additional staff members trained in teaching the programme. Additional schools, which were not involved in the first year of workshops, have started teaching the programme, while some of the original UKRP schools no longer do so.

However, only the workshops which took place in 2007-08 in the 22 mainstream secondary schools are the subject of this evaluation.

⁹ See, for example, Weisz, Donenberg, Han & Weiss (1995), "Bridging the gap between laboratory and clinic in child and adolescent psychotherapy", *Journal of Consulting and Clinical Psychology*, 63, 688-701.

Curriculum materials

The American curriculum materials required 'translation' into British English. LA staff read the materials and suggested changes, then these were looked over again by a British children's author. As many cultural references as possible were changed, but in some cases this would have involved changing artwork so it was not possible. The result was a set of materials that was largely anglicised but still had an American feel, and these were the materials used in the first year of workshops.

Further changes have been made since the first year of workshops, and more changes have been proposed to address issues such as cultural fit and teaching style.

Selection

Schools

Three local authorities opted to become involved in the UK Resilience Programme, and in December 2006 they made presentations to potentially interested schools to promote the programme. In some cases this was to a selected group of schools the LA thought would be most interested and most appropriate for the intervention; in others all local secondary schools were invited to get involved. Not all eligible schools chose to take up the programme. In addition, schools had to abide by certain restrictions when timetabling workshops. These included ensuring that only trained staff were timetabled to teach the subject, and that classes did not contain more than 15 pupils, resulting in a doubling of staffing and rooms for these classes for most schools. These could be difficult conditions to meet, and so might increase the self-selection of schools into the programme.

Facilitators

The first cohort of 90 workshop facilitators was trained in Philadelphia in the United States from 23rd July to 3rd August 2007. The majority were teachers, but other staff included learning mentors, teaching assistants, local authority staff and one school nurse. About 65 facilitators were school-based, while almost all of the others were employed by the local authorities. Interested schools were allocated a number of places by their local authority and the selection procedure for facilitators varied by school. In some schools particular individuals were offered places by the senior manager responsible, while in others all staff were invited to apply and then a selection procedure took place. A number of schools did not fill their places, and there were other places offered to staff outside of schools such as local authority officers. Facilitators were thus largely self-selected, although some also had to go through a selection procedure at their school and others were strongly encouraged to participate despite their reluctance to volunteer. Once selected, future facilitators registered and completed an online positive psychology program called *Resilience Online*.¹⁰ This introduced them to the principles of CBT, and encouraged reflection on their own emotional responses and behaviour. A few people were unable to attend the training in Philadelphia for health or other reasons, and some of these places were filled at the last minute by others from within the local authority.

Thus in practice, schools and local authorities were self-selected. Facilitators were also self-selected, although some may also have had some selection imposed by their school. Because of the absence of centralised selection, and of the involvement of the PRP team, one might expect that facilitator quality would be more variable than if an open application

¹⁰ See for details: <http://www.reflectivelearning.com/>

system had been used. However, the self-selection might result in facilitators being particularly highly motivated (perhaps increasing the success of the workshops), making it harder to extrapolate results outside of the sample. Facilitators were asked to give up the first two weeks of their summer holidays in order to train in Philadelphia, and had to prepare a large amount of new material in order to teach the workshops. One would therefore expect that these individuals and schools were highly motivated and enthusiastic, and had a strong belief in the importance of the subject being taught. This could limit the validity of extrapolating our results outside of the sample, as other participants might not be so committed.

In subsequent years the training has been scheduled to require less holiday time and more term time, requiring head teachers to agree to provide cover for their staff while they are on the training course. The number of days of training has also been reduced each year, and now stands at about 5-7 days. However, it still requires participants to give up some weekends or holidays in order to participate.

Training

As mentioned above, the training of the first cohort of facilitators took 10 days (five days of training, a weekend off, then another five days' training). In the first week trainees became familiar with the adult-level CBT skills, and in the second week they studied the PRP curriculum and practised teaching it to others.

Because of cancelled flights, one LA group arrived late and missed the first three days of training. The trainers worked with them to catch up on material missed, but it was felt that they did not receive the same social experience as those who had arrived four days earlier.

Workshops

Most schools had already planned how to deliver the workshops before the end of the summer term 2007, but some had not, and many had to form or revise their plans in September. They were asked to form UKRP groups of not more than 15 pupils, and to schedule the classes during the normal school day. The majority timetabled the programme by splitting an ordinary teaching class in two to get two UKRP groups taught by two facilitators simultaneously, but there were variations on how this was achieved. Many schools did not include as many pupils in workshops as they had originally intended, and there were only seven schools that managed to include all Year 7 pupils. This was achieved by having some workshops facilitated by trained LA staff, with the exception of one school with an unusually small Year 7 intake which used only their own staff and yet managed to cover the full year group. Another school used class sizes larger than 15 in order to include all Year 7 pupils.

Two schools were unable to start workshops until January 2008 due to problems involving timetabling and/or the support of senior management. Almost all of the others started their first workshops in September, with the rest starting in October. This varied within school as well as by school. Two schools were obliged to restart their workshops at October half term, because the lessons that the groups were timetabled against were setted (e.g. science set 3), and teachers had decided that the initial settings based on Key Stage 2 results and primary school reports were not appropriate. One school came across this problem but did not change the sets in order to preserve the UKRP groups.

Most schools taught lessons of one hour, and since the UK Resilience Programme was meant to be timetabled for a minimum of 18 hours this would normally take up about half a year of lessons. Some schools taught this fortnightly, meaning that their workshops lasted all

year, and some weekly, with the first set of workshops finishing around February or March. Many of the latter schools then went on to do a second set of workshops which lasted until July.

In most cases, UKRP lessons were fitted into an already full Year 7 timetable, and schools chose different lessons to replace. In the large majority of cases this was PSHE/citizenship/Learning to Learn, but some schools replaced other lessons such as English or science.

There were some changes of workshop facilitators during the year, due to departure of the facilitator; illness; and maternity leave. These workshop groups were taken over by other facilitators. There were also pupils who changed workshop group, because of class or set changes or behaviour issues, although these were relatively few.

Supervision and support

Facilitators teaching workshops were asked to participate in a series of nine one-hour conference calls with a PRP trainer and approximately 10 other facilitators, to provide continued support and training once they had started the workshops. Calls were weekly or fortnightly at the start of the year, becoming less frequent as time went on. Attendance on these calls was generally good, but after the first few calls most facilitators did not find them to be particularly helpful in offering support.

Some schools chose to use team teaching in the first year of workshops, in order to provide greater support to staff teaching the unfamiliar curriculum for the first time. This was particularly popular with facilitators who were learning mentors, as they were less likely to have experience of teaching classes or large groups. In most schools facilitators met regularly (formally or informally) to discuss their workshops, and in some cases facilitators would plan lessons together. Again, this was particularly popular where one facilitator was not a teacher. Each LA also held termly meetings, but these were more for organisation than for support.

Further cohorts

The second cohort of 67 workshop facilitators was trained in Cambridge (UK), from 16th to 25th July 2008. The training period was shortened to eight days, and several new schools sent staff to be trained, in addition to staff from the original schools and LA staff.

A third year of training took place separately in each LA in July 2009, with about 160 staff trained over eight days. Subsequent training has taken place from July 2010, involving 5-7 days of training for at least 160 more staff, and training places for a further 140 staff are planned for 2011.¹¹

The staff trained subsequent to 2007 have run further workshops in schools and in other contexts such as children's homes and Pupil Referral Units, but these workshops will not be evaluated by LSE.

¹¹ The reduction in training time is not as great as it appears: the original 10-day training course consisted of short days of 9am to 3pm, partly to accommodate the jetlag of the majority of participants. Subsequent training sessions have used longer days.

4. The Evaluation Design

As mentioned above, the PRP curriculum has been evaluated a number of times, but in small samples and with a high degree of control from the developers of the curriculum. Since a major problem in policy evaluation can be a decline in programmes' efficacy after scaling up (i.e. what happens when a small, select programme is rolled out), it is hard to draw any conclusions about whether these programmes would function well if used in schools more widely. The total sample of students involved in previous evaluated interventions was roughly 2000, which is about the same size as the workshop group in this evaluation. This evaluation therefore adds considerably to the evidence on the efficacy of the PRP curriculum.

Design

The evaluation has been designed as a controlled trial, with 'treatment' (i.e. workshop) and control pupils in each of the 22 participating schools.¹² Pupils could not be randomised into treatment or control groups because of timetable constraints, but the schools agreed that the method of selecting which pupils received workshops should be arbitrary, e.g. choosing the form group that fitted the timetable slot available, rather than choosing pupils they thought would benefit most. It was hoped that this would result in 'as-if' randomisation, with workshop and control pupils being similar on observable and unobservable characteristics. However, this is not true statistical randomisation and we will use statistical testing to determine whether it has in fact worked. When splitting a class in two to make a workshop group, schools also agreed to do this in an arbitrary way e.g. alphabetically.

Control groups

Those schools which wished to include all of their Year 7 pupils in workshops (seven schools) used the year-ahead group as the control group. In order to obtain baseline (start of Year 7) measures from these pupils they would have had to be surveyed in September 2006, before the project had begun. Because of this, only measures taken when the pupils were at the end of Year 7 are available for this group: we have no baseline for them, only a follow-up measure. Those schools with within-year control groups will have baseline measures for both workshop and control pupils. Six schools have both within-year and year-ahead control groups, and the remaining nine schools have a within-year control group only (see Table 1 for details by local authority, and Table 2 for details by workshop timing).

There is a possibility of externalities or spillovers, where workshop participants within a school influence the outcomes of those not involved in workshops through social interactions or other channels. Positive spillovers would bias downwards the estimate of the effect of the programme. Depending on the mechanism through which externalities operate, it is possible that the two control groups will produce different results if, for instance, the main channel is through pupils' peer interactions and these are more likely to occur within a year group than between year groups.

There are no control pupils outside of the workshop schools. This is not necessarily a problem, as the most appropriate control group is arguably made up of pupils in the same school. However, if programme schools are different from other schools (e.g. more

¹² We will use the language of 'treatment' and 'control' throughout, as this is the standard terminology in our field of research for any policy intervention. It is not meant to suggest that the UKRP was a true psychological treatment administered to those deemed to be in need of it, and indeed this was not the case.

concerned about pupil well-being) it is possible that this would understate or overstate the programme effect, as (for instance) they might have a positive effect on pupil well-being in the control group independently of the programme.

There are roughly 2000 children in the workshop group for the UKRP evaluation and up to 4000 in the pooled controls.

Table 1: Control and Treatment group details by Local Authority

	LA 1	LA 2	LA 3	Total
UKRP schools	6	9	7	22
Programme Cohort	Year 7 (2007-8)	Year 7 (2007-8)	Year 7 (2007-8)	Year 7 (2007-8)
Year 7 Coverage	33-100%	15-70%	11-100%	49%
Workshop pupils (#)	755	516	681	1952
Control group: Year 7	162	1137	714	2013
Control group: Year 8	960	130	1063	2153
Control group: pooled	1122	1267	1777	4166
Facilitators	30	24	32	86
Facilitators who taught evaluation workshops	24	23	24	71

Table 2: Control and Treatment group details by workshop timing and design

	Design 1	Design 2	Design 3
Workshop timing	Sept 07-Feb 08	Feb 08-July08	Sept 07-July 08
Start months	Sept - Nov 07	Feb - April 08	Sept 07 - Jan 08
End months	Jan - April 08	June - July 08	May - July 08
Workshop frequency	1 lesson/week	1 lesson/week	1 lesson/fortnight
UKRP schools	12	8	11
Programme Cohort	Year 7 (2007-8)	Year 7 (2007-8)	Year 7 (2007-8)
Workshop pupils (#)	480	395	1077
Control group: Year 7	1296	626	880
Control group: Year 8	654	459	1499
Control group: pooled	1950	1085	2379
Facilitators	29	24	41

Note: schools, facilitators and control group pupils will sum to more than the totals reported in Table 2 because schools ran multiple workshops and many of these had different timings. For instance, schools that ran workshops from September – February (design 1) often went on to run another set from February – July (design 2).

Alternative treatments

Since schools have had to make room for UKRP workshops within an already full curriculum, control group pupils will have received some lessons that treated pupils will not. In most cases this will be 18 hours of the Year 7 PSHE curriculum, but some schools displaced other lessons such as English, science or maths (see Table 3 for details). One school was reorganising its timetable and as part of this created a new UKRP slot, so there is no direct comparison in the control group. Moreover, class sizes for UKRP groups were not meant to be larger than 15 (and in 80% of classes this was the case), whereas in most cases the alternative treatment had class sizes of around 30.

Table 3: Alternative treatments by Local Authority (# of schools)

	LA 1	LA 2	LA 3	Total
UKRP schools	6	9	7	22
alternative treatments				
PSHE, citizenship, Learning 2 Learn, thinking skills, pastoral, or Access lessons	5	7	4	16
Science	0	0	1	1
RE	0	1	1	2
English & modern languages	0	1	0	1
English, science or maths	1	0	0	1
UKRP designated slot	0	0	1	1

Measurement

Pupil well-being is measured using depression and anxiety inventories and other validated psychological questionnaires. Behaviour is assessed using a behaviour questionnaire filled out by both pupils and teachers (please see the Appendix for further information on the psychological and behavioural questionnaires, and see the explanation of measures in the quantitative section below for detail on scoring). We also use data from the National Pupil Database/Pupil Level Annual Schools Census (NPD/PLASC) on pupils' demographic characteristics and their prior attainment, in addition to academic attainment data obtained from 14 schools. Table 4 shows the data elements used in the evaluation and the source for each.

Table 4: Quantitative data available, and source

Measure	Data	Source
Fidelity	Workshop group size Hours available for workshops Use of untrained facilitators Workshop attendance Conference call attendance	Schools Schools Schools Schools PRP team
Participant Satisfaction	Pupil satisfaction survey Facilitator satisfaction survey	Q Q
Psychological outcomes	Children's Depression Inventory (CDI) Revised Children's Manifest Anxiety (RCMAS) (Huebner) Brief Multidimensional Students' Life Satisfaction Scale (SLSS)	Q Q Q
Behavioural outcomes	Self-report Goodman SDQ (pupil) Teacher Goodman SDQ (teacher) Attendance Rates	Q Q NPD
Academic attainment	Prior attainment (Key Stage 2) Attainment at secondary school	NPD Schools
Other Relevant	Age, gender, ethnicity, FSM, SEN, Gifted and Talented, in-care, census characteristics of home area, other pupil-reported characteristics from pupil questionnaires.	NPD & Q
Q=questionnaire; NPD=National Pupil Database		

5. Quantitative analysis

Summary of Quantitative Findings

The first UKRP workshops were delivered to Year 7 pupils in 22 participating schools in 2007-8. In the First Interim Report (Challen et al., 2009) and the Second Interim Report (Challen et al., 2010) we provided an assessment of the impact of the workshops up to July 2008 and July 2009, finding that they had a small but significant average impact on pupils' depression scores. We also found differences in the size of the effect of the programme based on the timing and frequency of the workshops (weekly workshops had more of an impact), and by pupil characteristics (lower attaining and more disadvantaged pupils gained more, as well as pupils who started the year with worse symptoms).

In this final quantitative analysis we look at the same cohort of pupils and examine the impact of the programme at the two-year follow-up point in June 2010, comparing this to the impact seen at earlier follow-up periods.

We find an average short-run improvement in pupils' depression symptom scores, school attendance, and attainment in English as a result of the workshops. However, this improvement had faded by one-year follow-up for the depression score and for absence from school. There was still an average impact on English grades by one-year follow-up. There was some impact on anxiety scores, but this was inconsistent and concentrated in a few groups of pupils. There was also a measured impact on maths scores at one-year follow-up but not in the short run (immediately post workshops).

With the exception of the impact on academic attainment, the impact of the workshops lasted only as long as the academic year in which the workshops took place, and had faded by the one-year follow-up questionnaire in June 2009. However, there was still an impact for certain groups at one-year follow-up, particularly for pupils who had not attained the national target levels in English or maths at Key Stage 2. There was still a measured impact on English and maths scores at one-year follow-up. There was no impact on any of the outcome measures by the two-year follow-up survey in June 2010.

There was no measured impact of workshops on behaviour scores or life satisfaction scores.

The size of the impact varied by how workshops were organised: weekly workshops showed a larger impact than those timetabled fortnightly.

The impact also varied by pupil characteristics. Pupils who were entitled to free school meals; who had not attained the national targets in English or maths at Key Stage 2; or who had worse initial symptoms of depression or anxiety; were all more likely to experience a larger measured impact of the workshops on their depression and anxiety scores, particularly girls with these characteristics. There was little difference by pupil characteristics on the absence rate. The impact for these groups was also less likely to have faded by the one-year follow-up.

The short-run average improvement in absence rates appears to be similar across different ways of organising workshops, and is also similar for most groups of pupils. It is equivalent to an improvement of about 1.5 more school days attended over the course of the year.

The chapter concludes with implementation issues to consider, based on the quantitative research.

Introduction

For the purposes of comparison, we are providing similar analyses as for the 2009 and 2010 interim reports. However, we now include the most recent follow-up data, and evaluate the effect on additional outcomes (behaviour, life satisfaction and some measures of academic attainment).¹³ All pupils who were in Year 7 in 2007-08 are included in the analysis below if they completed enough questions from the questionnaires or have sufficient records available from the National Pupil Database (for absence data) or from their schools (for academic data).

However, not all implementations of the UKRP intended for inclusion in the evaluation appear to have produced comparable intervention and control groups. Having similar workshop and control groups is important: unless we are sure that the pupils included in the workshops and those in the control group were similar to begin with, we cannot be sure whether any differences between the two groups at the end of the workshops were due to the effects of the programme, or were simply due to pre-existing differences between them. In the analysis we therefore present results for both the full sample, and for the group (based on workshop timing) that has well-balanced intervention and control groups.

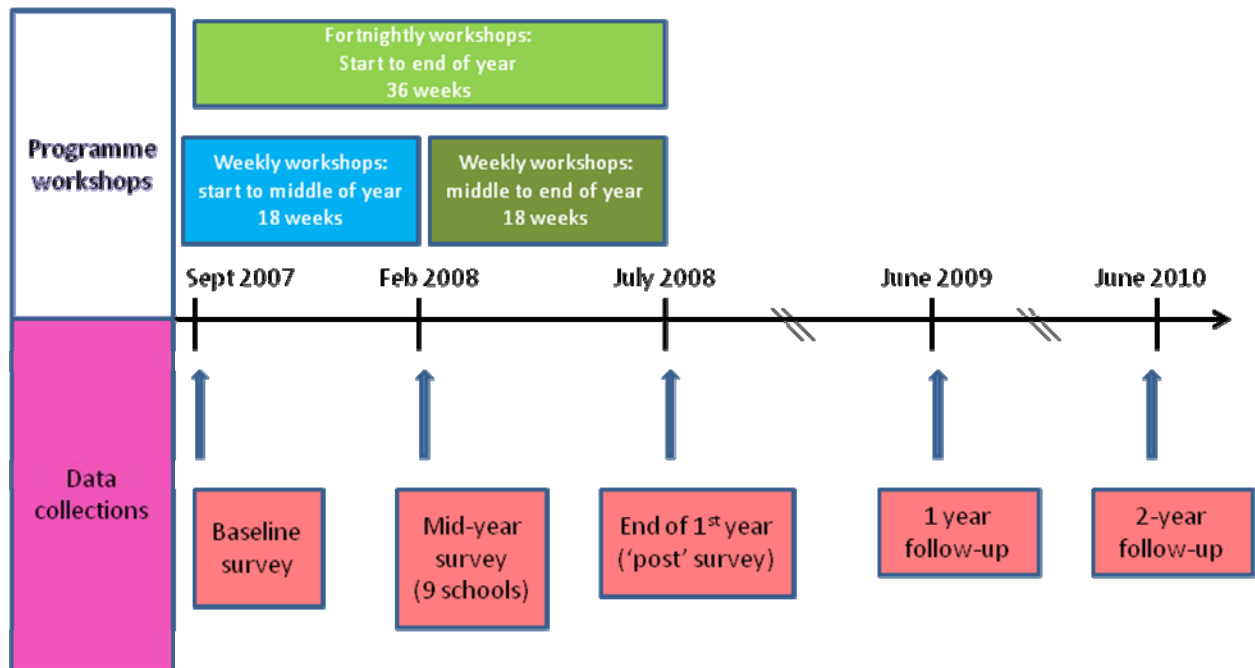
Data collection points

Year 7 pupils were surveyed at the beginning of the school year before the start of the intervention, at the end of the intervention and at the end of the academic year (two or three times in one year, depending on workshop timing) – see Figure 1: Timeline of workshops and data collections.¹⁴ There were further follow-up surveys in July 2009 and July 2010, when the pupils in the workshop cohort were in Years 8 and 9 respectively.

¹³ Another change is that we are modifying one aspect of the specification we are using. In previous years we have clustered the standard errors in the regressions at the pupil level. For this to give a true idea of programme impact we have to assume that there is no correlation between the outcomes of pupils who are in the same schools or classes. This is a strong assumption, and is probably unrealistic. In this report we cluster the standard errors at the level of treatment assignment, which allows for the possibility of correlation between pupils in the same classes, year groups or schools. This does not effect the size of the coefficients we estimate, but it does tend to increase the size of the standard errors and so the significance level of the coefficients, that is, we are now less likely to find statistically significant effects. However, this specification is more likely to be the correct one.

¹⁴ Only 9 of 22 schools were surveyed in the mid-year data collection (Wave C, around February 2008), as this only involved schools that were starting or finishing a set of workshops around this time.

Figure 1: Timeline of workshops and data collections



Explanation of measures

Symptoms of depression are measured using the Children's Depression Inventory (CDI). This scale has 27 items in the original version, but this study omits the item about suicidal ideation and so it contains only 26.¹⁵ Each of these is scored as 0, 1 or 2 depending on the severity of the symptom: 0 indicating no symptoms of depression on that item according to the child's response; and 2 indicating strong symptoms. Item scores are then summed to create a total score between 0 and 52, where higher scores indicate worse symptoms.¹⁶ However, since the scale primarily measures deviations from well-being, rather than degrees of positive well-being, the distribution of the total score is highly skewed. A large number of pupils have very low scores: over 10% score 0 or 1, and over 50% score 7 or lower, and this is true for each of Waves B to F (September 2007 to June 2010). We therefore encounter a 'ceiling effect' on depression symptoms scores, as pupils scoring 0 in the baseline at the beginning of the year cannot improve their scores.

Symptoms of anxiety are measured using the Revised Children's Manifest Anxiety Scale (RCMAS). This scale contains 28 items, plus 9 more to form a 'lie scale' used to detect responses motivated by social desirability (though the latter scale is not used in the present analysis). Each item of the main scale asks about whether a symptom of anxiety is typical of the child or not, and is scored as 1 if the response is 'yes' and 0 if 'no', giving a maximum possible summed score of 28 with higher scores indicating worse symptoms. Again, the

¹⁵ The item on suicidal ideation is often omitted when using the inventory in universal (as opposed to psychiatric) populations. The item was not deemed appropriate or necessary for use in schools, so has been omitted at all stages of this study.

¹⁶ If more than 10% of items are unanswered then the assessment is considered invalid. When up to 10% of items are missing these scores can be replaced by the mean of the non-missing items in order to create a total score.

distribution of scores is highly skewed, with over 10% of pupils scoring 0 or 1 and over 50% scoring 8 or lower.

Behaviour is measured using the self-report and teacher-report versions of the Goodman SDQ. The SDQ total difficulties score is comprised of 20 items, each scored 0, 1 or 2 according to the perceived severity of the symptom. This gives a minimum possible score of 0 and a maximum of 40, with higher scores indicating more (and more severe) symptoms.¹⁷ The distribution of scores is highly skewed, particularly for the teacher version: 50% of all pupils score 5 or lower on the teacher SDQ; and more than 50% score lower than 11 on the pupil version.

Life satisfaction is measured using the Huebner Brief Multidimensional Students' Life Satisfaction Scale, which has five items asking about satisfaction with particular domains of a child's life and one asking about overall life satisfaction.¹⁸ This is scored on a 7-point scale, giving a minimum possible summed score of 6 and a maximum of 42, with higher scores indicating greater life satisfaction. Again the distribution is skewed, with over 50% of pupils scoring 35 or more. Note that on this scale higher scores indicate greater well-being, unlike the other four scales for which the reverse is true.

Annual absence from school is measured as the fraction of school sessions for which pupils were absent during the academic years 2006-7, 2007-8 and 2008-09. This is the sum total of authorised and unauthorised absences during the year.¹⁹

Academic attainment in English, maths and science is measured in sublevels, such as 3b, 5.5 etc. Key Stage 2 attainment in these three subjects is obtained from the National Pupil Database, and attainment throughout the first three years of secondary school is provided by the schools. There may therefore be some variation between schools in the way pupils are graded. However, we have a control group within each school this is not so great a problem, so long as we account for the school a pupil is at, and we do this in the analyses below. Schools record attainment grades differently: some (most of our sample) use alphanumeric sublevels, e.g. 5b, 7c, while some use decimals such as 5.3, 4.2. We include all of these in the analysis by converting the former into decimal sublevels, with 'a' becoming .8, 'b' .5 and 'c' .2 (this follows the practice of one school which had recently modified its marking system, and appears to fit the data well).

Unfortunately, only 14 of the 22 evaluation schools responded to our request for academic data in time for it to be included in this report. As a result the sample we are using to assess pupils' academic outcomes is limited to pupils at these schools. Since these 14 schools are not representative of the sample as a whole we are reluctant to claim that our findings from this sample offer conclusive evidence on academic impact. However, we present some evidence below.

The data on pupils' gender, SEN status, entitlement to free school meals, Key Stage 2 attainment, and absence are obtained from the National Pupil Database/Pupil Level Annual Schools Census (NPD/PLASC).

¹⁷ The assessment is valid if at least 3 items of each of the four difficulties subscales have been completed.

¹⁸ The domains are: family, friends, school, oneself, and where the respondent lives.

¹⁹ We use this measure of absence because the distinction between authorised and unauthorised absence is not reliably recorded in the data, but the total number of absences should be (according to the guidance notes for the NPD tables).

Treatment and control groups

Tables 5-8 present the treatment and control group means for 11 variables for the full sample and for the three different experiments separately, including the p-values from mean-comparison tests. These tests indicate how different the two samples are at the baseline (September 2007), and therefore tell us how well matched the workshop and control groups were before any of the workshops began. They provide an estimate of how likely it is that an equally large difference between the two means could have arisen by chance if the samples were in fact similar. The first six variables in the upper panel of each table are potential outcome variables. The first five are psychological and behavioural variables obtained from the questionnaires we administered in September 2007; the sixth is annual absence, obtained from NPD/PLASC for the academic year 2006-07, i.e. the year before the workshops and the year before these pupils transferred to secondary school. The five variables in the lower panels are socio-demographic and educational attainment variables obtained from the NPD and matched into our dataset (although attainment in English and maths are also outcome variables). The data on Key Stage 2 attainment comes from the KS2 attainment tables in the NPD for summer 2007. The other variables are taken from PLASC for the academic year 2007-08, unless this record is missing, in which case values from the previous year (2006-07) or the following year (2008-09) are used. In this way we minimise the number of pupils for whom we lack demographic information.

Table 5 presents statistics for the full Year 7 cohort when they were surveyed at baseline. This includes all pupils who were in Year 7 in 2007-8, in either workshop or control groups, and who completed enough questions and questionnaires throughout the course of the data collections to be included in the evaluation sample (so this does not include the year-above control groups, who do not have a baseline measure). It is clear that there are significant differences between the workshop and control groups: the p-values of tests of equality of means between the treatment and control groups are less than 0.1 for 7 out of 11 variables, suggesting that there is (at most) a 10% chance that the observed differences between these groups would have arisen by chance if they were in fact similar. (The smaller the p-value, the less likely that such a difference could have arisen by chance.) Importantly, there are significant differences between treatment and control at baseline for both the depression score ($p=0.03$) and the anxiety score ($p=0.08$), which are the first two of the outcome variables we will consider in our analysis below. The two groups do seem similar in terms of the mean absence rate from school in the previous year ($p=0.74$). However, the groups also differ significantly in terms of demographic and attainment characteristics (the five variables in the bottom panel of Table 5). Although we cannot reject that they have a similar composition in terms of gender ($p=0.18$), the full treatment group and the control group are significantly different at the 5% level for the others: on average the treatment group as a whole has lower levels of SEN and FSM eligibility, and higher prior attainment than the control group.

Since we have data on these characteristics, we can control for them in our analysis, and thereby control for any differences between treatment and control groups that are not due to the causal impact of workshop participation. However, given that there are these observable differences between the two groups, it is likely that there are also differences between them that we cannot observe (we do not have data on them), and therefore cannot control for. If these unobserved differences are correlated with assignment to the treatment or control group, then if we use poorly matched groups we cannot be sure that differences in outcomes between treatment and control are due to the programme and not due to other factors. This is why it is important to have well-matched treatment and control groups, and why we will not only use the full sample in our analysis.

The following three tables present the same information as Table 5 but for each configuration of workshops separately. Table 6 presents descriptive statistics for pupils who were in 'Group 1' workshops, relative to the control group (the control group is the same as for the full sample, and for the other two groups of workshops discussed below). These are workshops that took place at least once a week from the beginning of the school year, and finished around the middle of the school year (roughly September 2007 to February 2008).

The treatment and control groups are well matched: it is clear from Table 6 that most of the p-values from the mean comparison tests are well above 0.1. In fact, the only two variables presented for which there are statistically significant differences in means are Key Stage 2 maths attainment and absence rate in the previous year. On average, pupils in the treatment group had significantly higher KS2 maths scores than those in the control group. The difference between the two groups is not large (0.13 of a level), but it is interesting that on average pupils in the control group did not attain the national target of level 4 in Key Stage 2 maths, suggesting that they have somewhat below average attainment. This difference could have resulted from the way schools assigned class groups to UKRP workshops: if the assigned classes were setted, for instance in science or English sets, the treatment group's academic attainment might differ from the school average. This was not necessarily deliberate, as many schools simply had to choose classes to participate in UKRP workshops based on which ones fitted the timetable. However, this could be a problem for evaluating the programme impact on academic outcomes, as if pupils did not have similar attainment before the workshops it is more difficult to attribute subsequent differences to the impact of the programme.

Also worrying is the difference in absence rates, as this is another outcome variable we will be examining. Pupils in these workshops had a higher absence rate in their last year of primary school than pupils in the control group. We suspect that this was largely due to pupils in these two groups coming from different schools in different proportions, meaning that they are likely to have different average levels of attendance and attainment. (Schools are not equally represented in both groups, as the proportion of the year group included in workshops varied greatly by school, ranging from 11% to 100% of the cohort.) That is to say, the differences are due to between-school differences in these variables rather than within-school differences. Nevertheless, in our analysis below our preferred specification uses pupil fixed effects, using multiple observations of the same pupil over time to control for all the characteristics of the pupil which are fixed through time, and this should control for differences in KS2 attainment and prior absence rates. Moreover, we control for the school that each pupil attended in September 2007 and allow for different trends by school, and this should take account of many differences due to schools.²⁰

The second treatment group consists of weekly workshops that started mid year and finished near the end of the school year (roughly February to July 2008), and baseline statistics for

²⁰ There could still be a problem if baseline differences give rise to different trends, and these trends are correlated with workshop/control assignment. For instance, if it were the case that more academically able children become happier at secondary school while less able children become less happy, and a school assigned only its top set maths pupils to workshops, we might attribute the improvement in these pupils' subjective well-being to the workshops when in fact it was due to their academic ability. However, we do not think that the mismatch in maths attainment is a particular problem for us, at least when the outcomes is the depression symptoms score, as we find that lower-attaining pupils actually seem to benefit more from the workshops. Given that the descriptive statistics suggest that there is a smaller proportion of low-attaining pupils in our treatment groups this mismatch would in fact bias downwards, not upwards, our estimate of the average treatment effect. That said, this is likely to be more of a problem when the outcome variable is academic attainment.

this sample and the control group are presented in Table 7. Here it is clear that the pupils in the workshops are on average significantly different from the control group pupils on the psychological measures, on one of the behavioural measures, and in terms of their absence rate (the p-values are less than 0.05 for all but one of the first six outcome variables, meaning that there is less than a 5% chance in each case that such a large difference between the treatment and control groups would have arisen by chance if they were in fact similar). We know that some schools selected pupils for this second batch of workshops based on perceived psychological need, so it is not surprising that they look different to the control group. However, they seem reasonably well matched on the five variables in the lower panel – the socio-demographic and attainment measures.

Table 8 presents measures for the third group of workshops, which took place fortnightly and lasted from the start of the academic year to (near) the end. This treatment group seems well matched to the control group in terms of the psychological scores at baseline, with the exception of life satisfaction ($p=0.09$, with the control group scoring higher than the treatment group, meaning that on average they report being more contented). However, the difference in the absence rate is significant at 0.1%. Moreover, these pupils are very different in terms of all five socio-demographic and attainment variables: the treatment group is significantly more female, higher attaining, less likely to be entitled to free school meals, and less likely to have special educational needs than the control group. Many of the schools that timetabled workshops in this pattern included all of their Year 7 pupils in workshops and so do not have a within-year control group; moreover, they are on average less deprived and higher attaining than the sample as a whole. This could make comparisons more difficult, as treatment and control pupils are in different schools, so are probably more different to begin with, and may be subject to different environments and events throughout the course of the year. It is therefore not surprising that treatment and control pupils appear to be significantly different at baseline. As mentioned above, in our analysis we can control for many differences between schools, but with so many obvious differences between the pupils in this treatment group and the control group this is probably not the best sample to rely on to estimate the programme impact.

Thus, because of the significant baseline differences between the treatment and control groups when using the full sample, we will present some analyses on both the full sample and on the first group of workshops (the September to February weekly group of workshops). We can be reasonably confident of any results we obtain that are robust to using both of these two samples.

Programme impact on symptoms of depression

We first present an analysis of the treatment effect using an econometric specification which adds variables to control for pupil characteristics, which should go some way to correcting any bias produced as a result of mismatched treatment and control groups (Table 9). We show how the results obtained differ for the three different experiments using this same framework (Table 10). Then we will use our preferred specification and examine heterogeneity in treatment effect by pupil characteristics (Table 11). We follow a similar pattern for analyses of the programme impact on the other outcome variables.

The overall treatment effect for all three groups of workshops combined is presented in Table 9.²¹ Columns 1-4 present the short-run effect of the workshops, measured either in the

²¹ This can be compared with Table 9 of the Second Interim Report (Challen et al. 2010), and Table 28 of the First Interim Report (Challen et al. 2009).

middle of the year immediately after the end of the Group 1 workshops; at the end of the academic year (immediately post for Workshop Groups 2 and 3, and 4 months later for Group 1); or both. Columns 5-8 present the results for the one-year follow-up, with data collected in June-July 2009. Columns 9-12 present the results for the two-year follow-up, with data collected in June-July 2010.

The method we are using to estimate programme impact is called difference-in-differences. We obtain this estimate by subtracting the mean depression score after the end of the programme from the mean depression score at baseline for the control and treatment groups separately, then taking the difference between these two to obtain the overall effect of treatment. In essence, this measures how the treatment group has changed relative to how the control group has changed.

The coefficient on 'Treated*PolicyOn' in Table 9 gives the difference-in-differences estimate, using the standardised depression symptoms score as the outcome variable.²² Columns 1, 5 and 9 only control for the month in which the questionnaire was filled in (a control for reporting and seasonal effects), for the three follow-up points respectively. The negative coefficient on 'Treated*PolicyOn' in column 1 suggests an improvement in the treatment group's depression scores relative to the control group at the short-run follow-up point (February or July 2008), and this is significantly different from zero.^{23 24} The match between the treatment and control groups (in terms of the depression score) is shown by the coefficient on the 'Treated' variable. As suggested by the descriptive statistics in Table 5, pupils in the combined treatment group score on average at least 0.1 of a standard deviation higher (worse) than those in the control group on the depression inventory (given this specification). The single asterisk following the coefficient indicate that this difference is significant at 10%, i.e. there is only a 10% chance that it could have arisen by chance if in fact the treatment and control values were equal. The size of this difference (about 0.1 of a standard deviation) is not very large, but given that we would not expect the programme to have an impact which is much larger than this difference, this could be a problem in interpreting the results of the analysis.

Columns 5 and 9 use the same specification as column 1, but here the data used are for the one-year and two-year follow-up points. As in column 1, the coefficient on 'Treated' in both of these later follow-ups suggests a pre-existing difference between the treatment and control groups of about 0.1 of a standard deviation, significant at 10%. However, the coefficient on 'Treated*PolicyOn', which gives us the estimate of programme impact, is positive for both of these periods. This means that the treatment group had a slightly worse depression score at these points than the control group, although this difference is very small and is not significantly different from zero, meaning that we should really treat it as no difference (i.e. an effect of zero).

²² Standardising scores involves subtracting the mean score and dividing by the standard deviation to give a standardised score with a mean of 0 and a standard deviation of 1. This does not change the results we obtain, but makes interpretation and comparison of the coefficients easier.

²³ Remember that for the depression score, a higher score indicates worse symptoms, so a decline in the scores is an improvement.

²⁴ If the difference between two values is statistically significant this means that it is unlikely that an equally large difference would arise by chance if in fact the two values were the same. For instance, if a difference is significant at the 10% level ($p \leq 0.1$), this means that there is only a 10% chance that it could have arisen by chance; if it is significant at the 1% level ($p \leq 0.01$), there is only a 1% chance of its arising by chance, etc. Here and in the tables below, the level of significance of a coefficient is indicated by asterisks after it: one asterisk means that it is significant at the 10% level ($p \leq 0.1$); two asterisks that it is significant at 5% ($p \leq 0.05$); and three that it is significant at 1% ($p \leq 0.01$).

These results mean that, based on this very simple specification, there does appear to be a significant average effect of the programme on the depression score soon after the end of the programme (column 1), but that there is no measured effect later on, at one-year and two-year follow-up (columns 5 and 9).

However, this does not take into account other factors which may differ between the two groups and could affect the measured impact. The advantage of using this econometric specification is that we can control for some other factors, such as gender, free school meal eligibility, special educational needs etc., which could have some impact on the outcomes we are evaluating, independently of any impact of the workshops. This is particularly important given that Table 5 suggests that there are important differences on average between the treatment and control groups in terms of these characteristics. Columns 2-4, and the equivalents at one- and two-year follow-up (columns 6-8 and 10-12), add more controls for pupils' demographic characteristics and school attended. The controls used are listed in the grid beneath the table.

Looking at all three periods, the coefficient on 'Treated' increases slightly when controls for pupils' demographic characteristics are added, but drops again once school fixed effects are included (columns 3, 7 and 11). This is because, as we suggested above, many of the differences between pupils in the treatment and control groups here are due to between-school differences rather than within-school differences, so once we control for a pupil's school at baseline the difference is reduced. However, the coefficient is still significantly different from zero at both short-run and one-year follow-up (although not at two-year follow-up), suggesting that these controls might not fully account for baseline differences between pupils.

For the short-run follow-up, the coefficient on 'Treated*PolicyOn' gets slightly larger as more controls are added, implying that there is an average improvement as a result of the workshops once we account for the differences between pupils in the control and treatment groups. Columns 4, 8 and 12 present specifications using pupil fixed effects. This works by using multiple observations of the same pupil over time to control for all the characteristics of the pupil which are fixed through time. This should control for many of the differences in composition between the treatment and control groups (such as gender, prior attainment etc.), including those for which we do not have data (e.g. family climate, so long as this does not change over time). For this reason it is our preferred specification. Based on this, we obtain a programme impact of 0.09 of a standard deviation in the short run, significant at 10% (column 4), with no detected effect at one-year follow-up (column 8) or at two-year follow-up (column 12). This represents a small short-run programme effect.²⁵

We know that there are differences in composition between the different treatment groups (see Tables 5-8), so it is worth examining treatment effects separately by treatment group using this econometric specification, which allows us to control for pupil characteristics. Table 10 presents difference-in-differences estimates of the treatment effects with the three groups of workshops entered separately.²⁶ Again, the coefficient on 'Treated*PolicyOn' gives the difference-in-differences estimate, i.e. the average impact of having been in UKRP workshops on a pupil's (standardised) depression symptoms score at the end of the treatment. The specifications used are the same as for Table 9, and the columns represent the same time periods: columns 1-4 concern the short-run programme impact in February or

²⁵ In specifications using school attended at baseline or pupil fixed effects we also control for school trends.

²⁶ Compare Table 10 of the Second Interim Report and Table 30 of the First Interim Report.

July 2008; columns 5-8 are for the one-year follow-up in June-July 2009; and columns 9-12 are for the two-year follow-up in June-July 2010.

As suggested by the simple comparison of baseline characteristics, the start-mid year treatment group (Group 1, top panel) seems well-matched to the control group in terms of the depression symptoms score, hence the absence of a significant coefficient on the 'Treated' variable in the first panel in both time periods. Moreover, adding controls to the regression does not change much the size of the coefficient on 'Treated*PolicyOn' or that on 'Treated' for this group of workshops, implying that the workshop and control groups are reasonably well balanced on the control characteristics too. Based on this group of workshops, we obtain a treatment effect of 12% of a standard deviation improvement in pupils' depression scores in the short run when pupil fixed effects are added in column (4), a small treatment effect.^{27 28} This has disappeared for the one- and two-year follow-ups, where the estimated effect is essentially zero.

The second panel gives the same information for the second group of workshops. The problem of the mismatch between the treatment and control groups for this experiment is shown by the positive and significant coefficient on the 'Treated' variable: this means that pupils who were in the treatment group scored higher (worse) on the depression symptoms inventory even before the treatment had started. This suggests that the as-if randomisation did not successfully create comparable treatment and control groups, and so taking comparisons of these groups before and after the workshops will not provide a true estimate of the treatment effect.

Both of the first two treatment groups show an improvement in depression scores, although this only remains significant for Group 1 workshops in the short run (at 10%), once pupil fixed effects have been included. Interestingly, the measured impact of the Group 2 workshops appears to have increased by one-year follow-up, with a point estimate of 0.2 of a standard deviation when pupil fixed effects are included (column 8, panel 2). It is possible that the impact of an intervention could increase over time. Reasons for this include better use of the programme skills over time as pupils gain more practice; a 'sleeper' effect where an impact on well-being is only seen with a lag; or the programme preventing a worsening in the treatment group which does appear in the control group. While these may all be plausible explanations for the difference, given the mismatch between treatment and control at baseline we are not confident of the estimated impact for the Group 2 workshops, as this effect might simply be due to mean reversion.

These effects compare with a positive coefficient on 'Treated*PolicyOn' (i.e., a worsening of scores in the treatment group relative to the control group) in the first two specifications for the short run for the start-end year treatment group (Group 3), although this becomes slightly negative in columns 3 and 4 as more controls are added. The short-run programme impact for this set of workshops is zero, as only the coefficient in the first two (largely uncontrolled) specifications are significantly different from zero. However, at one- and two-year follow-up this has grown to about 0.1 of a standard deviation, which is reasonably robust to the inclusion of controls, though is not significantly different from zero once pupil fixed effects are included. This implies that pupils who were in this treated group had actually got slightly worse relative to the control group one or two years after the end of the workshops, although this cannot be interpreted as a programme effect since the inclusion of control variables reduced this difference to zero. Again, we suspect that this result might be partly due to

²⁷ Pupil fixed effects control for all characteristics of pupils that are fixed over time.

²⁸ Remember that a higher score indicates more and worse symptoms, so a negative coefficient means an improvement in these regressions.

mismatch between the treatment and control groups: the coefficient on 'treated' for these workshops jumps around as the specification changes, implying that they are not matched on a number of characteristics, and that these characteristics have some relationship with the outcome variable. Indeed, Table 8 suggests that the Group 3 treatment and control groups differ in terms of demographic characteristics. This is in contrast to the Group 1 workshops, for which the coefficient on 'Treated' is always small and insignificant and does not change much with different specifications.

Thus we do seem to see some treatment heterogeneity by the three groups of workshops, although the p-value of a chi-squared test of constant treatment effect is greater than 0.1 for all three pupil fixed-effect specifications, meaning that it is possible that such different estimates of the treatment effect could arise by chance. Combined with the evidence from Tables 5-8, this seems to suggest that although the 'as-if' randomisation may have worked for the Group 1 workshops, it was not successful in the other two samples. We know that some schools overtly or covertly decided to select pupils to participate in workshops, rather than randomly assigning them as had been agreed, and that this was particularly evident in the mid-end year treatment group.²⁹

We suggest that the as-if randomisation worked well enough for the first out of these three workshop sets, but apparently failed on the other two to some extent. We will therefore present some analysis which uses both the full sample and the Group 1 sample plus the control group, as we believe that this provides a more reliable picture of the impact of the workshops: if the results are robust to using these two samples, we can be more confident that our estimates are correct.

This is not to say that there might not be differential effects caused by different frequency and timing: fortnightly workshops (as in Group 3) were extremely unpopular with teachers, as it proved difficult to teach with such a long gap between lessons.³⁰ In comparison, Group 1 and 2 workshops were held weekly with no obvious problems relating to timing. For this reason it might be sensible to consider workshops from Groups 1 and 2 as one type of programme, while Group 3 workshops constituted a programme that was slightly different. Moreover, it is possible that since pupils are new to the school in September 2007 when the weekly Group 1 workshops begin, pupils could be easier to influence and so these workshops might have had a greater potential for impact than the weekly Group 2 workshops which began in February. School transitions can be difficult and can have an adverse impact on subjective well-being; starting the programme when pupils first came into the school might help to mitigate these problems more effectively than waiting until half way through the academic year. If this is the case, we would expect a larger programme impact for workshops starting earlier in the year, such as the Group 1 and Group 3 workshops. These two mechanisms combined could result in weekly workshops beginning in September being the most effective, and that appears to be what we find. However, we have no way of disentangling workshop timing and frequency from a host of other variables that could impact the treatment effect, so we can only speculate as to why there appears to be treatment heterogeneity between the three groups of workshops.

²⁹ Since pupils were new to the school in September 2007, even if schools did attempt to select pupils for workshops on the basis of psychological distress they would have been less successful at this point as they did not know the pupils well enough. By February 2008 they might have known pupils well enough to select more effectively.

³⁰ See Challen et al. (2009), p86.

Heterogeneity of impact by pupil characteristics

Now we can go on to examine heterogeneity in the impact of treatment by pupil characteristics, that is, whether different groups of pupils appear to be affected more or less by the workshops. We run the same specifications on our two different samples to check for the robustness of the results.

Table 11 presents evidence on the heterogeneity of treatment effects by pupil characteristics, using the full sample. These results are generated from separate specifications relative to controls, each column within a box representing a separate regression. Here we use the full sample, which includes all three treatment groups, and is very similar to that used in the 2010 Second Interim Report.³¹ The specifications we use are slightly different, but columns 1-6 of Table 11 are comparable with Table 11 of the 2010 report.³² Columns 1-3 present the combined estimated treatment effect immediately after the end of the start-mid year workshops (around February-April 2008) and at the end of the academic year (July 2008). Columns 4-6 of Table 11 present the estimated impact in June 2009; and columns 7-9 present results for June-July 2010.

These results compare to the overall average treatment effects in Table 9 of -0.092 (0.052) for the short-run impact; -0.021 (0.073) for the one-year follow-up; and 0.028 (0.084) for the two-year follow-up.

For this sample we cannot reject that the size of the treatment effect on girls and boys is equal ($p=0.80$).³³ However, pupils with special educational needs seem to benefit significantly more from the workshops than those without SEN, and this result appears to be driven by the differences between boys with and without SEN (there is no difference for girls). Moreover, this difference persists to the one- and two-year follow-up points. However, because of the large standard error, the coefficient for boys with SEN is not significantly different from zero at any point (although the short-run result for all children with SEN is), so we cannot claim that there was a strong programme impact for this group. The other groups for whom there appears to be a significant difference in programme impact are girls with FSM entitlement, and who did not attain the national target of Level 4 in English or maths. We find that these girls have an average short-run decrease in their depression score of about one-third of a standard deviation (which is also significantly different from zero), and that this appears to be significantly different from girls who were in workshops but who were not entitled to FSM, or who had higher Key Stage 2 attainment. However, even here the p-value of the test of equality between the two groups is only below 0.1 for girls who attained below level 4 in maths, so the evidence is not strong. Moreover, although the impact on these three groups of girls appears to persist to one-year follow-up, the standard errors on the estimates are large enough that the coefficients are not significantly different from zero. Thus we can only suggest that they experienced a short-run improvement in their depression scores. Splitting the sample by the baseline quintile of the depression score, it appears that

³¹ There will be some differences in the samples used as we are adding another wave of data from July 2010, so we may exclude some pupils from the previous samples who did not have questionnaires in 2010. By using the same pupils at each follow-up point in this report we can more easily compare the programme impact through time.

³² One difference in the specification is the use of standard errors clustered at the level of treatment assignment, rather than at the pupil level as we used in previous reports.

³³ The p-values reported are from a test of equality of the coefficients immediately above: if the p-value is greater than about 0.1 this implies that the coefficients are not significantly different from each other and therefore that pupils with (e.g.) FSM entitlement are not more or less likely to benefit from the workshops than those without.

much of the average short-run effect comes from an improvement in those who were in the worst scoring quintile at baseline (5th quintile), notably girls. However, even for this group, there is no significant impact at one- or two-year follow-up.

As we showed above, the full sample is not balanced by outcome or demographic variables at baseline (see Table 5). Since there appear to be differential treatment effects by pupil characteristics, any differences in the composition of the treatment and control groups could result in an unrepresentative estimate for the average treatment effect. For this reason, in Table 12 we also present these same regressions on treatment heterogeneity for the Group 1 workshops, for which treatment and control groups are well balanced at baseline.

Table 12 presents the programme impact using only the Group 1 workshops and the control group. These results compare to the overall average treatment effects in Table 10 of -0.123 (0.071) for the short-run impact; -0.052 (0.114) for the one-year follow-up; and -0.033 (0.125) for the two-year follow-up.

Girls appear to have a larger and more significant programme effect than boys, particularly in the short-run, but the p-value from the test of equality of the two coefficients ($p=0.22$ for the first two periods) suggests that they are not significantly different from each other. Only at two-year follow-up does the p-value fall below 0.1, and here neither coefficient is significantly different from zero.

Further disaggregating by pupil characteristics, we find that there is apparently no difference in impact by free school meals eligibility or SEN. However, it is interesting that differences between some subgroups appear to be larger at the one-year follow-up than immediately post workshops. It appears that girls who did not achieve level 4 English or maths at Key Stage 2 (KS2) and who participated in workshops improved their depression scores by about half of a standard deviation compared to the control group (a moderate-to-large programme effect), and that (unusually) this effect persists at one-year follow-up, although it is only significantly different from zero for English attainment.³⁴ Girls in workshops who did achieve the national targets at KS2, obtained an improvement of 0.2 of a standard deviation in depression scores relative to the control group in the short-run, and this had faded out by one-year follow-up (and was again zero at two-year follow-up). The p-value of a test of equality between these groups of girls comes close to 0.1 in the one-year follow-up period, but is never smaller than this, so the difference is not strong. There was no such difference for boys.

Once we split the sample by the quintile of the baseline (September 2007) depression score, we find important differences in measured impact by baseline score. Pupils who scored in the worst (highest) 20% of scores in September 2007 (quintile 5) improved significantly relative to the control group, and the effect was particularly large (70% of a standard deviation) for girls in the 5th quintile. This improvement is not maintained at one-year follow-up (the coefficients are moderate and negative, but are not significantly different from zero), but, curiously, reappears at two-year follow-up.

Overall, the picture seems somewhat mixed, with significantly negative coefficients appearing in different quintiles across time, although the most consistent coefficients are in the 5th quintile – they are always negative and are at least 0.2 of a standard deviation in size, even if their large standard errors mean that they are rarely significantly different from zero. This could partly be because of small sample sizes meaning that there is low power to detect

³⁴ These two groups – girls who did not attain the national targets at KS2 – also have quite large and negative coefficients at two-year follow-up, but these are not significantly different from zero.

small effects, but it also appears likely that the measured gains from the programme are concentrated among those who had the worst subjective well-being at the start of the programme. Note that this is unlikely to be due to mean reversion, as we have controlled for this in the specification. The Penn Resiliency Program was originally conceived as a depression prevention programme for targeted (at-risk) adolescents, so it is perhaps not surprising that pupils closer to this category might respond the most. However, given the strong ceiling effect imposed on pupils who obtain low (good) scores at baseline due to the skewed distribution of the depression scale we are using, it is also possible that pupils with higher (worse) scores simply have further to go in terms of possible improvement. That is to say, the scale we are using to measure symptoms of depression is only sensitive around the part of the distribution where there are a number of symptoms of depression, and not sensitive to changes for children who are already reporting few or no symptoms.

Taking the results from these two samples together, we suggest that there is likely to be some treatment heterogeneity by SEN status, prior attainment, and baseline depression score, with lower attaining and initially more depressed children benefitting more. Given that the largest and most consistent treatment effect is found when the sample is split by baseline depression score it is possible that the large effects found for lower attaining pupils could be due to the fact that they are more likely to have poor psychological scores, rather than any direct effect of SEN or low attainment. We tested this hypothesis in the following way: we further split the treated sample into four groups (pupils with low initial depression scores and low maths attainment; pupils with low depression and high maths; pupils with high initial depression and low maths; and with high depression and high maths), and run the same regressions. We find that there is a significant treatment effect for each of these groups except pupils with low depression and high maths (regressions not shown; this result applies whether using the full sample or the Group 1 treatment only). This suggests that controlling for baseline depression score there is a differential impact on depression scores for low-attaining pupils over and above that which might be due to their starting with a worst depression score.

Overall, the analysis of the depression scores suggests some treatment heterogeneity by organisation or timing of treatment, part of which may have been caused by selection of pupils into the workshop groups, although it is also possible that the timing and frequency of the workshops could have an impact. There is also significant heterogeneity of treatment impact by pupil characteristics, with lower-attaining pupils and those who began the year with worse depression scores apparently gaining more from workshops in the short run. At one- and two-year follow-up the only impacts that appear to persist are those for low-attaining girls and those with the worst baseline for subjective well-being.

Programme impact on symptoms of anxiety

Table 13 is the equivalent of Table 9, but uses the anxiety (RCMAS) score as the outcome variable. The general pattern of results is similar to that seen for the depression symptoms score, which is not surprising given that these two scores are strongly correlated and the symptoms they measure often occur together. However, Table 13 suggests that there is no impact of treatment on anxiety score on average.³⁵ The significant and positive coefficient on 'Treated' suggests that the treatment group has a worse anxiety score at baseline than the control group, implying that they are not well matched along this dimension. This is what we saw in Table 5, where it is clear that the treatment and control groups are not well matched

³⁵ The equivalent in the 2010 evaluation report is Table 14.

on anxiety symptoms at baseline. Table 14 unpacks this further for the three treatment groups separately.³⁶

As for the depression score, the difference between the treatment and control groups for the start-mid year treatment is small and insignificant, while those treated from the middle of the year to the end score significantly higher than the control group at the baseline and the mid-year measurement dates. Once the pupil fixed effects specification is used there is no significant treatment effect for any of the three groups for any of the three follow-up periods.

Table 15 shows the equivalent specifications with the anxiety score as the outcome as are shown in Table 11 for the depression score, using the full sample with all three treatment groups. Here there are no coefficients that are significantly different from zero, and only some suggestion that treatment effects might be different for different groups of pupils: in the short run, pupils eligible for free school meals appear to experience a greater reduction in their anxiety scores than those who are not ($p=0.12$), while girls who did not achieve level 4 in Key Stage 2 maths appear to experience a greater reduction in anxiety scores than those who did achieve level 4 ($p=0.09$). Looking at the results split by the baseline quintile of the anxiety score (second part of Table 15), the coefficients are more negative (implying a greater reduction in anxiety) for pupils who started with a higher anxiety score, but no coefficients are significantly different from zero and based on the p-values of tests of equality of the coefficients we cannot say that they are statistically different from one another.

Since the results obtained from the Group 1 sample – those workshops for which treatment and control groups are well matched – are the same as for the full sample, we do not present these tables here. Thus overall we find no statistically significant impact on anxiety scores for any group of pupils, once we have taken into account pupil characteristics.

Behaviour

We have two behaviour measures, both based on scores from the Goodman Strengths and Difficulties Questionnaire. The first is pupil-reported behaviour, and the second is teacher-rated behaviour. The teacher who filled out the form for each pupil was usually their form tutor, but in some cases class teachers, heads of year, or other staff completed them. The samples we have for each questionnaire will be different: if a pupil was absent on the day they were to fill in the questionnaire then we will not have a self-reported SDQ score for them, but their teacher could still have filled in a teacher SDQ questionnaire. However, response rates for teacher questionnaires, though generally high, varied significantly: most schools got at least 80% of teacher questionnaires returned, but some were not willing to insist that teachers complete them and so response rates could be as low as 30%. Also, when a form tutor refused to fill in questionnaires we will generally be missing surveys for the entire form group. Two schools failed to complete any baseline teacher surveys before it was too late for the responses to be valid. The sample for the teacher-reported SDQ score is therefore different (and smaller) than that for the other, pupil-reported measures we have presented so far.

We might also expect different results from teacher-reported and pupil-reported questionnaires. Teachers, even form tutors, will only see their pupils in specific contexts, and so will not have as much information about a pupil's behaviour as the pupil themselves. However, teachers might be more objective about a pupil's behaviour, not least because they are likely to be more aware of what is normal behaviour for a child of a given age.

³⁶ The equivalent in the 2010 report is Table 15.

Pupil's responses might have less objectivity, and their validity will depend importantly on a pupil's level of self-awareness, but they do at least observe their own behaviour in different contexts and so can comment on their behaviour outside of school. The two measures thus provide us with different, and complementary, information about pupils' behaviour.

Table 16 presents the same econometric specifications as Tables 9 and 13, but uses the self-reported behaviour score as the outcome variable. Table 17 does the same for the teacher-reported behaviour score, using a slightly smaller sample as a result of difficulties in getting teachers to complete questionnaires. The coefficient on 'Treated' in both tables shows that there is not a significant difference in behaviour scores between treatment and control groups at baseline – as was also shown in the descriptive statistics in Table 5. However, there is also no measured impact of treatment on either score. Part of the reason for this might be measurement problems: even more pupils score 0 (meaning no obvious problems) on the teacher-reported behaviour score than do so on the depression and anxiety inventories, meaning that there is simply no room for improvement for these pupils (the self-reported behaviour scores have fewer zeros, but most pupils still score low on this measure). Moreover, it is unlikely that the measure is very sensitive at the low-scoring (few problems) end of the scale, so even if a pupil does not score zero then it would be difficult to ascertain whether they have improved or not. However, it is also plausible that the programme simply has no effect on the externalising symptoms that the Goodman SDQ aims to measure. The Penn Resiliency Program was designed primarily to prevent emotional difficulties rather than behavioural ones, and although the two may occur together and the programme could have an impact on behaviour too, it is perhaps less likely than an impact on symptoms of depression.³⁷ We do not find any impact of the programme on any of the subgroups we have looked at, so we do not include tables presenting these analyses.

Life satisfaction

Table 18 presents the same specifications for the full sample as Tables 9, 13, 16 and 17, but here the outcome is the life satisfaction score. As suggested in Table 5, the treatment and control groups are not equally matched at baseline: the treatment group reports significantly lower life satisfaction on average than the control group, hence the negative and significant coefficient on 'Treated' in the first two columns of each panel.³⁸ Only once the school at baseline is controlled for does this difference between treatment and control fall to zero, suggesting that these differences are in large part due to between-school differences (differences between pupils at different schools) rather than within-school differences (differences between pupils at the same school). Although the coefficients on 'Treated*PolicyOn' are all negative, suggesting a decrease in life satisfaction as a result of going through the workshops, they are not significantly different from zero once we use the pupil fixed effects specification (and indeed are only marginally significant in columns 7 and 11). Thus there does not appear to be any programme impact on life satisfaction scores.³⁹

³⁷ It is worth noting that at least one of the previous PRP studies that found some impact on behaviour did so in a sample that was at risk of some form of mental illness (Jaycox et al., 1994). It may be easier to find an effect when using at-risk or targeted samples both because of the greater need for intervention (a real programme effect), and because of fewer measurement problems: there will be less risk of a ceiling effect (because fewer children will be scoring zero), and the participants are more likely to have scores in the range in which the instrument is sensitive.

³⁸ For the life satisfaction score, a higher score indicates greater life satisfaction.

³⁹ Since the results in other samples and for all subgroups are very similar we have omitted these tables here.

Absence from school

Absence from school is measured as the fraction of school sessions for which pupils were absent during the academic years 2006-7, 2007-8 and 2008-9. This is the sum total of authorised and unauthorised absences during the year.⁴⁰ Since workshop pupils participated in workshops in the academic year 2007-8, if there is an effect of the workshops on attendance we would expect to see fewer absences for these pupils relative to the control group in this academic year, and possibly also in subsequent academic years. The data on absence was obtained from the National Pupil Database, and is likely to be less subject to reporting biases than the psychological or questionnaire measures we have used in the analyses above. In addition, since we can obtain data on a child's attendance at school for as long as they attend a state school in England, there is a much lower level of attrition than for the other measures. There is also much less danger of selective attrition from not filling in the questionnaires due to absence or refusals, and so we can be less concerned about the possibility of sample selection bias. However, one disadvantage of using these data is that data for the academic year 2009-10 was not available at the time of writing and so we are only able to evaluate the impact to July 2009.

Table 19 is the equivalent of Tables 9, 13 etc., presenting estimated treatment effects for the pooled workshop groups with absence as the outcome variable. Here the pooled sample appears balanced between treatment and control groups on this outcome, with the coefficient on 'Treated' small and insignificant (this tallies with the evidence presented on this in Table 5, where the full sample appears balanced on absence rates at baseline). The coefficient on 'Treated*PolicyOn' for the short run is not significant until we control for school at baseline and school trends in columns 3 and 7, and remains so when we use a pupil fixed effects specification in column 4, but here we obtain an estimate of programme impact on attainment that is 16% of a standard deviation and significant at 10%. There is no average effect of treatment at one-year follow-up.

Table 20 presents the same specifications with the treatment split into the three groups of workshops. Here, the treatment and control groups appear to be well-balanced at baseline, with the exception of the Group 3 workshops (fortnightly workshops) for which the treatment group appears to have a significantly lower absence rate than the control group. If the programme reduces absence we should see a negative coefficient on the Treated*PolicyOn variable, as indeed we do for all three groups, although this is only significant for Group 1 workshops in the short run. However, using the column 4 pupil fixed effects specification we cannot reject the hypothesis that all three treatment effects are equal (the chi-squared test of constant treatment effect gives us a p-value of 0.47 for the short-run and 0.17 for the one-year follow-up).

Examining the treatment effect by pupil characteristics for the full sample shows no overall pattern of differential treatment effect by the characteristics we use to split the sample (Table 21), although there may be more of an impact for boys than for girls. Splitting the sample by quartiles of the baseline absence rate suggests that the impact does not depend on the initial level of absence, but is fairly constant across the distribution. If there is a difference here, it appears to be that workshop pupils in the top quartile of the absent rate (i.e. those with little or no absence in the baseline year) did not change relative to the control group, while pupils in the next two or three quartiles did see some improvement, at least in the short run. This could be another ceiling effect in measurement: about 27% of pupils have no

⁴⁰ We are using total absences rather than trying to distinguish between authorised and unauthorised absences because this measure is more reliable (according to NPD guidelines).

absence at all in the baseline year, meaning that the first quartile consists of pupils who cannot improve their absence rates, so unless the control group is gets significantly worse in the following two years we will not be able to find an effect. (The high proportion of pupils with baseline absence rates of zero was the reason that we have split this sample into quartiles rather than quintiles here: the latter did not fit well the distribution of absence rates.) So overall, we find an average treatment effect of 0.16 of a standard deviation improvement in absence rates in the short run, with no effect at one-year follow-up, and find that this effect is fairly homogeneous across different groups of pupils.

To think about what kind of impact this is in practice, we can use the data from the evaluation sample. The mean absence rate for these pupils in 2006-7, when they were in the final year of primary school, was about 0.04 or a 4% absence rate. This had risen to 0.07 in 2007-8, and to 0.08 in 2008-9. The standard deviation in which the effect is measured is that of the absence rate in 2006-7, which has a standard deviation of about 0.06. The median number of possible school sessions for these pupils in 2006-7 (for which attendance is recorded) was 308 (306 in 2007-8; 310 in 2008-9). If UKRP workshops reduced absence by 0.16 of a standard deviation on average in the first year, this is equivalent to $0.06 \times 0.16 \times 308 = 3.0$ more sessions attended over the course of the year, or about 1.5 more school days. This might not seem like a large improvement with respect to the total number of sessions possible (it is only 1% of the 308 possible sessions for the year), but it is moderate relative to the overall absence rate. Relative to the 7% absence rate in 2007-8 it represents an improvement of 0.14 or 14%.

It is worth noting that whereas the psychological measures are assessed at a point in time, and can therefore be deemed to be 'pre' and 'post' measures for the workshops, the absence measure is accumulated over the course of the year and so includes the time period when pupils in the treatment group were attending workshops. Pupils generally enjoyed the workshops, and some teachers commented that attendance was higher on UKRP days because pupils did not want to miss the lessons. If this is the case, it is possible that the effect observed could be entirely due to increased attendance on days with UKRP lessons. This is important both because one would expect the effect to wear off as soon as the workshops finished, and moreover we might not see so much value in increased attendance at school if pupils only attend this additional 'fun' lesson rather than other subjects as well. As shown in the tables above, it appears that this impact does not persist to the second year, which means that the impact could be entirely due to increased attendance on UKRP days (although we have no evidence whether or not this is the case). However, even if pupils only improved their attendance on UKRP lesson days, they will also have attended other lessons on these days.

Academic attainment

Tables 22, 23 and 24 report the impact of the UKRP workshops on pupils' academic attainment in English, maths and science (respectively). As mentioned above, only 14 of the 22 UKRP evaluation schools responded to the request for academic data in time for it to be included in this report, and this sample of schools is not representative of all schools in the sample. Moreover, because of the reduced sample of schools we have fewer observations on which to run our analysis – about 2300 pupils in total. Having said this, one advantage of the academic data, like the absence data, is that we have data on pupils even if they had poor school attendance or chose not to complete the questionnaires, so we would expect attrition to be lower than for the questionnaire measures.⁴¹ This means that any measured

⁴¹ Clearly, if a pupil's absent rate is too high academic data will not be available for them, as teachers

impact we obtain is less likely to have been biased by the composition of the sample. However, we do not have data on the academic attainment of pupils who have left the school, so there will still be some attrition.

Table 22 reports the impact of the UKRP workshops on attainment in English. The coefficient on 'Treated' is positive and significant at 10% in the first columns of each panel, but falls to zero once pupil characteristics and then school at baseline are controlled for. Interestingly, the coefficient on 'Treated*PolicyOn' does not change much, regardless of the specification used. In the short run, the estimated impact of the workshops is 0.37 of a standard deviation – a moderate to large policy effect. At one-year follow-up it is 0.22 of a standard deviation, and at two-year follow-up it is positive but not significant (i.e. it is essentially zero). This suggests a positive impact in the short run, which fades over time. In Table 23 we see that the impact on maths scores is essentially zero in the short run (about 0.10 but not significant), rising to 0.24 at one-year follow-up and possibly around 0.19 at two-year follow-up, although the coefficient on this in our preferred specification in column 12 is not significantly different from zero. This is an odd pattern of results, as we usually find that any programme impact is strongest in the short term, fading as time goes on. One problem with this outcome is that treatment and control pupils are clearly not balanced at baseline: in the first columns of each panel (columns 1, 5 and 9), the coefficient on 'Treated' is about 0.2 of a standard deviation, and although this is reduced when pupil characteristics are included, once baseline school is included this returns close to the original value (except at one-year follow-up, where it remains zero). So it would seem that maths attainment is not equal between treatment and control groups here, with pupils in workshops having attained significantly higher grades at Key Stage 2 than pupils in the control group. Table 24 gives results for attainment in science, and here again we see from the coefficient on 'Treated' that pupils in workshops scored about 0.2 of a standard deviation higher in Key Stage 2 science than pupils in the control group. However, once pupil characteristics and the school they attend are taken into consideration this coefficient drops to zero. There is no measured impact of the workshops on attainment in science in any period (the coefficient on PolicyOn is always small and non-significant).

These scores are standardised, that is, we have transformed them by subtracting the baseline mean and dividing by the standard deviation in order to obtain results in terms of standard deviations of academic grades measured in sublevels. To get an idea of how large these effect sizes are, the standard deviation of KS2 English attainment in this sample is about 0.7, while that for maths is 0.8. So an effect of the order of 0.4 of a standard deviation, which is what we see for the English as an outcome in the short run, is equivalent to about $0.4 \times 0.7 = 0.28$ of a level, which is equal to about one sublevel of Key Stage 2 attainment. Assuming that one sublevel at KS2 can be applied to KS3 attainment, this suggests that a pupil who would have attained an English grade of 5b at the end of Year 7 actually attains 5a as a result of participation in UKRP workshops. The gain at the end of Year 8 is smaller, however: it is about half of a sublevel. For maths, there is no significant impact at the end of Year 7, but at the end of Year 8 the impact is roughly equivalent to about two-thirds of a sublevel ($0.24 \times 0.8 = 0.20$). There is no average gain in scores for attainment in science.

It is curious that we find different results by subject: there is clearly a strong impact on English, but this is not so clear for maths and science (especially once we consider the possible mismatch between the treatment and control groups). This is particularly surprising given that these academic scores are highly correlated with each other: the correlation coefficient between maths and science scores in this sample in years 7-9 is about 0.75-0.81;

will not be able to assess their progress. So missing data can still present a problem, but is likely to be less of a problem than for the questionnaire measures.

between English and maths about 0.65-0.70; and between English and science about 0.65-0.70. It is possible that English attainment is more sensitive to pupils' effort, so any improvement in pupils' psychological well-being or persistence is more likely to result in improved performance in English than in maths and science. To provide very crude evidence in support of this, we have data from tests of cognitive ability that pupils took in the first few weeks of Year 7 for 6 of the 14 schools for which we have academic data.⁴² Based on these, we find that the correlation between mean cognitive ability test score at the start of Year 7 and attainment during Key Stage 3 is significantly higher in maths (correlation coefficient 0.83-0.87), than in English (0.65-0.69), with the correlation with science attainment somewhere in the middle (0.68-0.81). If it is the case that attainment in English depends slightly less on ability and slightly more on effort than attainment in maths, then one would expect to find that interventions that improved pupils' attitude to learning would more easily impact their English scores than their maths scores. However, this would not explain why we found no impact on science scores and an inconsistent impact on maths scores.

In addition, in the pupil survey we conducted at the end of workshops, pupils were asked whether they thought the skills they had learnt would help them in English, science and maths. Of about 1500 pupils who responded to these questions, 39% said that they agreed or strongly agreed that the workshops would help them with their work in English, while only 31% and 28% said that they would help in maths and science respectively. Moreover, at one of the case study schools that can be said to have 'embedded' the programme into their curriculum and in 2010 was still teaching it to Year 7 pupils, the school has chosen to teach the workshops as part of the English curriculum, believing that it displays a good fit with elements of the curriculum (see case study section below for further details). So it is possible that UKRP lessons, and the skills taught in the lessons, are more obviously applicable to work in English than to work in maths or science, though clearly this is speculative.

This perhaps provides some crude evidence as to why we might expect to see more of an impact on English scores than on the other attainment scores. However, we are still only using the attainment data from 14 of the original 22 schools, and these are not representative of all the schools. We will therefore not provide further analysis here as to do so would be misleading on the basis of such a selected sample. There does appear to be some impact on academic attainment based on the sample we currently have available to us, and the size of this is small to moderate. The most consistent (and strongest) result appears to be on the English score, and we hypothesise that this could be due to attainment in English depending more on attitude to learning than does attainment in other subjects, and on the UKRP lessons themselves promoting skills relevant to the English curriculum.

Conclusion

We find that there is a short run impact of the UK Resilience Programme on depression scores, school attendance, and English scores. However, the average impact has faded by one-year follow-up for the depression score and for absence from school. There is still an average impact on English grades by one-year follow-up, and for the maths score there is no impact in the short-run but a significant impact at one-year follow-up. There is no impact on

⁴² These 6 schools used Cognitive Ability Tests to assess pupils at the beginning of Year 7. Three more schools in this sample used MidYIS tests, which give similar results to the correlations reported here: they have a much stronger correlation with attainment in maths than in English, with science somewhere in between. For more information on CAT see:

<http://shop.gj-assessment.co.uk/home.php?cat=310>

For more information on MidYIS see: <http://www.cemcentre.org/midyis/introduction>

any of the measures we are using by the time of the two-year follow-up (although we do not yet have data to assess this for the absence rate).

We find some heterogeneity in the measured effects of the UK Resilience Programme, in terms of how the workshops were organised, the outcomes assessed, and the impact on different groups of pupils. Workshops that were timetabled weekly and which started at the beginning of the academic year appeared to have a larger impact on depression scores and absence than those which either started later or were timetabled fortnightly, but these differences were usually only significantly different from the one-year follow-up period, i.e. in the short run the impact looks similar.⁴³

Furthermore, we found variation in the impact of treatment by pupil characteristics: in general, lower attaining and more disadvantaged pupils appear to gain more from the workshops, and in some cases the programme impact has not faded by one-year follow-up for these groups. Specifically, the impact of the workshops on anxiety and depression scores was larger for pupils entitled to free school meals; for pupils who had not attained the national target levels in Key Stage 2 exams; and for pupils with worse initial scores for symptoms of depression or anxiety. The impact on these outcomes also seemed to be greater for (some) girls, e.g. girls with low prior attainment. Interestingly, there did not seem to be so much heterogeneity in the impact of workshops when the outcome was the absence rate. To the extent that there was some heterogeneity this was concentrated in other groups: the less disadvantaged (not FSM; higher attaining) and those with moderate absence rates at baseline. However, very few of these differences were strongly statistically significant, and overall the average impact of the workshops on absence was fairly evenly distributed across groups of pupils.

Some of the observed heterogeneity might be at least partly due to the lack of sensitivity of the measures we are using. For instance, the psychological and behavioural measures we are using may be good at detecting change above a certain level of symptoms, but are unable to detect improvements in those who already have good psychological well-being or more ordinary behaviour. The same applies to the absence rate: since 27% of the sample has no absence at the baseline, unless the control group pupils' absence rate increases markedly there is no room to measure an impact. This is probably not the case for the academic data: we use Key Stage 2 results in English, maths and science as the baseline for our analysis, and there is more variation in grades (in sublevels) here than for the other measures. However, we do not have this data for all schools involved in the evaluation, so we cannot report full analyses here.

In summary, we find some impact of participation in UK Resilience Programme workshops on depression scores, absence rates, and academic attainment. The impact is small, and relatively short-lived: for no outcome measure does it persist until two years after the end of workshops. We also find no impact on behaviour scores, whether measured by pupil self-reports or by teacher reports, or on life satisfaction scores. We find some heterogeneity in impact by the organisation or timing of workshops, and by pupil characteristics.

⁴³ We should bear in mind that the association of these workshops with a greater programme impact is not necessarily causal.

Implementation and policy issues

Here we present some possible policy implications based on the results reported above and from the findings of the interim reports. These points highlight issues to examine if considering implementing the programme, based on the evaluation results and the results of prior PRP trials.

- 1) The UK Resilience Programme did have a small average impact on pupils' depression scores and school attendance, but only in the short run. It also had some impact on English grades in the short run and at the one-year follow-up point, and on maths scores at the one-year follow-up. There was no impact on any measure at two-year follow-up. This means that any improvements in pupils' psychological well-being, attendance and attainment were short-lived, and by the time of the two-year follow-up (June 2010) pupils who had participated in UKRP workshops were doing no better on these outcomes than pupils who had not. This suggests that a single set of UKRP lessons is not enough to permanently change pupils' outcomes on average.
- 2) The impact of the programme varied by pupil characteristics, and was much stronger for more deprived and lower-attaining pupils and those who started school with worse psychological health, particularly girls with these characteristics. Thus even if there is no average impact of the programme beyond the short run (i.e. an impact when measured over all pupils), it appears that some pupils benefitted substantially more, and for longer. These findings suggest that the improvements experienced by these pupils were more likely to be meaningful in terms of the impact on their lives, both in the short run and perhaps in the longer run too (our evaluation measures stop at two-year follow-up so we cannot know what happens later than this).
- 3) The UKRP was intended to be a universal programme, but some schools have chosen to target pupils for inclusion in workshops. It is not clear which model is preferable, and this will probably depend on the situation of each school. However, the following points are worth bearing in mind:
 - Based on the quantitative analysis, certain groups of pupils appeared to benefit more from the workshops, particularly those who did not achieve the national target level in English and maths at Key Stage 2, pupils with SEN, and pupils who started the school year with higher levels of depression or anxiety symptoms.
 - However, the measured impact on these pupils is the impact of the programme delivered to 'universal' or mixed workshop groups, not of groups consisting entirely of targeted pupils. One cannot therefore assume that the same impact would be obtained if workshop groups were targeted.
 - Some schools that did run workshops entirely with targeted pupils reported these as being very difficult to manage and not very successful compared to more mixed groups.
 - The same applies to levels of academic attainment: many facilitators commented that SEN groups or lower set groups did not go well, or that the presence of more able or more literate pupils aided the success of the lessons.
 - Although facilitators and other school staff often appeared to assume that higher ability pupils were naturally more resilient, or had fewer problems, almost all facilitators claimed to use the UKRP skills themselves. It therefore seems unlikely

that higher ability pupils or those with better initial psychological well-being would be unable to benefit from the skills.

- Even if pupils were to be targeted for inclusion in workshops, it is important that they should be targeted appropriately. Previous research suggests that school staff tend to identify pupils with behaviour problems rather than those with emotional difficulties, yet the programme is primarily designed to address the latter. The process of targeting would also need to be carefully considered.
 - Participation in programmes perceived to be targeted and remedial can attract stigma for those who participate. Universal programmes avoid this.
 - The measures used in the quantitative evaluation are sensitive to differences in the severity of symptoms of depression and anxiety, but are not good at distinguishing between children who have few or no symptoms. For instance, they would not be able to detect any improvements in well-being for pupils who showed no initial symptoms of depression, although this would not necessarily mean that these children did not benefit.
 - The skills pupils used most (as reported by both pupils and facilitators) were the interpersonal skills around negotiation and assertiveness, and techniques for self-control (see Chapters 5 and 6 of the First Interim Report). Since all pupils are likely to experience conflict and problems around everyday social interactions it is likely that all pupils could benefit from the workshops, at least in these areas.
- 4) The impact of the programme appeared to vary by workshop timing. Even if the association of fortnightly workshops with lower impact is not causal, teaching fortnightly workshops was very unpopular with teachers (Challen et al., 2009, p87). Fortnightly workshops were also less popular with pupils, though again it is not clear that this was causal (Challen et al., 2009, p21). All previous PRP trials have scheduled workshops to take place at least once a week. Timetabling workshops once a week rather than once a fortnight therefore seems preferable, if this is possible.
- 5) About 20% of pupils in workshops in the first year of UKRP were in groups of 16 or more, though only 3% were in groups of 18 or more. Facilitators thought that small class sizes were very important to the success of the programme (Challen et al., 2009, p37). It might therefore be important to use small classes in order to maintain the quality of the programme.

Table 5: Comparison of treatment and control group baseline means – Full sample

Full sample: all experiments pooled	Treatment group	Control group	p-value of test of equality of means
All treatments			
Depression score at baseline	8.55	8.03	0.0269
Standard deviation	6.97	6.14	
number of observations	1607	1546	
Anxiety score at baseline	9.38	8.97	0.0838
Standard deviation	6.81	6.30	
number of observations	1585	1525	
Pupil-reported behaviour score at baseline	11.00	10.81	0.4012
Standard deviation	6.32	6.20	
number of observations	1586	1522	
Teacher-reported behaviour score at baseline	5.37	5.52	0.4864
Standard deviation	5.45	5.42	
number of observations	1429	1197	
Life satisfaction score at baseline	33.71	34.19	0.0320
Standard deviation	6.19	5.99	
number of observations	1539	1454	
Fraction of sessions absent (2006-7)	0.04	0.04	0.7408
Standard deviation	0.06	0.06	
number of observations	1901	1909	
Gender (male=0; female=1)	0.49	0.47	0.1810
Standard deviation	0.50	0.50	
number of observations	1607	1546	
Special Educational Needs	0.24	0.27	0.0317
Standard deviation	0.43	0.44	
number of observations	1607	1546	
Free School Meals	0.22	0.30	0.0000
Standard deviation	0.41	0.46	
number of observations	1607	1546	
KS2 English score	4.07	3.91	0.0000
Standard deviation	0.85	0.93	
number of observations	1561	1502	
KS2 maths score	4.05	3.90	0.0000
Standard deviation	0.86	0.91	
number of observations	1561	1502	

Notes: this table presents baseline means for a range of variables to enable us to gauge the similarity of the treatment and control groups. Here we present means for the full sample: all three treatment groups pooled plus the control group. The last column gives the p-value of a test of equality of means between the two groups: if this is less than 0.1 for a variable the treatment and control groups are said to be significantly different from one another at the 10% level. This means that there is less than a 10% chance that an equally large difference between the two would have arisen by chance if they were in fact the same, i.e. that the two groups are not well matched on this variable at baseline.

Table 6: Comparison of treatment and control group baseline means by workshop timing – Group 1 workshops (weekly start-mid year workshops)

Group 1 workshops plus control group	Treatment group	Control group	p-value of test of equality of means
Treatment: start year - mid year			
Depression score at baseline	8.27	8.03	0.5217
Standard deviation	7.16	6.14	
number of observations	373	1546	
Anxiety score at baseline	9.37	8.97	0.2811
Standard deviation	6.86	6.30	
number of observations	365	1525	
Pupil-reported behaviour score at baseline	10.70	10.81	0.7503
Standard deviation	6.13	6.20	
number of observations	367	1522	
Teacher-reported behaviour score at baseline	5.21	5.52	0.4050
Standard deviation	6.12	5.42	
number of observations	289	1197	
Life satisfaction score at baseline	33.80	34.19	0.2747
Standard deviation	6.24	5.99	
number of observations	351	1454	
Fraction of sessions absent (2006-7)	0.05	0.04	0.0004
Standard deviation	0.07	0.06	
number of observations	446	1909	
Gender (male=0; female=1)			
Gender (male=0; female=1)	0.47	0.47	0.9735
Standard deviation	0.50	0.50	
number of observations	373	1546	
Special Educational Needs	0.23	0.27	0.1113
Standard deviation	0.42	0.44	
number of observations	373	1546	
Free School Meals	0.27	0.30	0.2445
Standard deviation	0.44	0.46	
number of observations	373	1546	
KS2 English score	3.99	3.91	0.1471
Standard deviation	0.95	0.93	
number of observations	366	1502	
KS2 maths score	4.03	3.90	0.0139
Standard deviation	0.93	0.91	
number of observations	364	1502	

Notes: this table presents baseline means for a range of variables to enable us to gauge the similarity of the treatment and control groups. Here we present means for the Group 1 workshops plus the control group. Group 1 workshops were those held weekly that started at the beginning of the academic year. The last column gives the p-value of a test of equality of means between the two groups: if this is less than 0.1 for a variable the treatment and control groups are said to be significantly different from one another at the 10% level. This means that there is less than a 10% chance that an equally large difference between the two would have arisen by chance if they were in fact the same, i.e. that the two groups are not well matched on this variable at baseline.

Table 7: Comparison of treatment and control group baseline means by workshop timing – Group 2 workshops (weekly mid-end year workshops)

Group 2 workshops plus control group	Treatment group	Control group	p-value of test of equality of means
Treatment: mid year - end year			
Depression score at baseline	10.64	8.03	0.0000
Standard deviation	8.14	6.14	
number of observations	175	1546	
Anxiety score at baseline	11.47	8.97	0.0000
Standard deviation	7.34	6.30	
number of observations	172	1525	
Pupil-reported behaviour score at baseline	12.64	10.81	0.0004
Standard deviation	6.92	6.20	
number of observations	166	1522	
Teacher-reported behaviour score at baseline	6.13	5.52	0.2451
Standard deviation	5.64	5.42	
number of observations	116	1197	
Life satisfaction score at baseline	33.10	34.19	0.0323
Standard deviation	6.91	5.99	
number of observations	159	1454	
Fraction of sessions absent (2006-7)	0.06	0.04	0.0000
Standard deviation	0.07	0.06	
number of observations	229	1909	
Gender (male=0; female=1)	0.43	0.47	0.2878
Standard deviation	0.50	0.50	
number of observations	175	1546	
Special Educational Needs	0.28	0.27	0.8004
Standard deviation	0.45	0.44	
number of observations	175	1546	
Free School Meals	0.31	0.30	0.8454
Standard deviation	0.46	0.46	
number of observations	175	1546	
KS2 English score	3.98	3.91	0.3414
Standard deviation	0.93	0.93	
number of observations	168	1502	
KS2 maths score	4.00	3.90	0.1723
Standard deviation	0.92	0.91	
number of observations	168	1502	

Notes: this table presents baseline means for a range of variables to enable us to gauge the similarity of the treatment and control groups. Here we present means for the Group 2 workshops plus the control group. Group 2 workshops were those held weekly that started in the middle of the academic year. The last column gives the p-value of a test of equality of means between the two groups: if this is less than 0.1 for a variable the treatment and control groups are said to be significantly different from one another at the 10% level. This means that there is less than a 10% chance that an equally large difference between the two would have arisen by chance if they were in fact the same, i.e. that the two groups are not well matched on this variable at baseline.

Table 8: Comparison of treatment and control group baseline means by workshop timing – Group 3 workshops (fortnightly start-end year workshops)

Group 3 workshops plus control group	Treatment group	Control group	p-value of test of equality of means
Treatment: start year - end year			
Depression score at baseline	8.31	8.03	0.2790
Standard deviation	6.63	6.14	
number of observations	1059	1546	
Anxiety score at baseline	9.04	8.97	0.8008
Standard deviation	6.65	6.30	
number of observations	1048	1525	
Pupil-reported behaviour score at baseline	10.85	10.81	0.8853
Standard deviation	6.26	6.20	
number of observations	1053	1522	
Teacher-reported behaviour score at baseline	5.32	5.52	0.4002
Standard deviation	5.22	5.42	
number of observations	1024	1197	
Life satisfaction score at baseline	33.78	34.19	0.0920
Standard deviation	6.05	5.99	
number of observations	1029	1454	
Fraction of sessions absent (2006-7)	0.03	0.04	0.0000
Standard deviation	0.05	0.06	
number of observations	1226	1909	
Gender (male=0; female=1)	0.51	0.47	0.0318
Standard deviation	0.50	0.50	
number of observations	1059	1546	
Special Educational Needs	0.23	0.27	0.0299
Standard deviation	0.42	0.44	
number of observations	1059	1546	
Free School Meals	0.18	0.30	0.0000
Standard deviation	0.39	0.46	
number of observations	1059	1546	
KS2 English score	4.12	3.91	0.0000
Standard deviation	0.79	0.93	
number of observations	1027	1502	
KS2 maths score	4.07	3.90	0.0000
Standard deviation	0.83	0.91	
number of observations	1029	1502	

Notes: this table presents baseline means for a range of variables to enable us to gauge the similarity of the treatment and control groups. Here we present means for the Group 3 workshops plus the control group. Group 3 workshops were those held fortnightly that started at the beginning of the academic year. The last column gives the p-value of a test of equality of means between the two groups: if this is less than 0.1 for a variable the treatment and control groups are said to be significantly different from one another at the 10% level. This means that there is less than a 10% chance that an equally large difference between the two would have arisen by chance if they were in fact the same, i.e. that the two groups are not well matched on this variable at baseline.

Table 9: Treatment effects for the three experiments pooled
Outcome: depression score
(From specification pooled across all three treatments relative to controls)

	All treatment and control groups											
	Outcome post workshops/end of year (Feb and July 2008)				Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated*PolicyOn	-0.074* (0.043)	-0.085** (0.041)	-0.169*** (0.046)	-0.092* (0.052)	0.048 (0.045)	0.039 (0.044)	-0.008 (0.055)	-0.021 (0.073)	0.038 (0.061)	0.032 (0.061)	0.044 (0.061)	0.028 (0.084)
Treated	0.111** (0.052)	0.168*** (0.044)	0.127*** (0.046)		0.097* (0.057)	0.147*** (0.048)	0.095* (0.051)		0.095* (0.057)	0.140*** (0.050)	0.065 (0.050)	
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Pupil Fixed Effects	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Number of Pupils	2909	2909	2909	2909	2841	2841	2841	2841	2750	2750	2750	2750
Sample size	7154	7154	7154	7154	5682	5682	5682	5682	5500	5500	5500	5500

Notes: Standard errors (clustered by level of treatment assignment) in brackets; the control variables are dummies for gender (1), special educational needs status (1), free school meal status (1), gifted and talented status (1) and Key Stage 2 maths and English performance (10 dummies). The outcome measure here is the depression (CDI) score standardised to have a mean of 0 and a standard deviation of 1. Columns 3, 4, 7, 8, 11 and 12 contain controls for school trends (school at baseline*post).

Here, and in subsequent tables, the level of significance of a coefficient is indicated by asterisks after it: one asterisk means that it is significant at the 10% level ($p \leq 0.1$ – there is a 10% chance that this could have arisen by chance); two asterisks that it is significant at 5% ($p \leq 0.05$); and three that it is significant at 1% ($p \leq 0.01$).

Columns 1-8 present results comparable with those of Table 28 in Challen et al. (2009), and Table 9 of Challen et al. (2010), although here a different sample of pupils is used and the specification modified slightly.

Table 10: Treatment effects for the three experiments: depression score
(From specification pooled across all three treatments relative to controls)

	Outcome post workshops/end of year (Feb and July 2008)				Treatment: Start year-Mid year Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated*PolicyOn	-0.130** (0.052)	-0.130** (0.052)	-0.123** (0.058)	-0.123* (0.071)	-0.015 (0.072)	-0.028 (0.071)	-0.049 (0.081)	-0.052 (0.114)	-0.021 (0.084)	-0.028 (0.083)	-0.027 (0.088)	-0.033 (0.125)
Treated	0.033 (0.102)	0.070 (0.082)	0.054 (0.082)		0.056 (0.104)	0.085 (0.083)	0.080 (0.085)		0.023 (0.102)	0.062 (0.086)	0.038 (0.083)	
	Outcome post workshops/end of year (Feb and July 2008)				Treatment: Mid year-End year Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated*PolicyOn	-0.153** (0.071)	-0.133* (0.070)	-0.158** (0.073)	-0.124 (0.108)	-0.281** (0.134)	-0.285** (0.130)	-0.204 (0.142)	-0.277* (0.168)	-0.201 (0.164)	-0.205 (0.161)	-0.083 (0.158)	-0.167 (0.208)
Treated	0.254* (0.131)	0.280** (0.109)	0.262** (0.104)		0.438** (0.188)	0.455*** (0.160)	0.376** (0.152)		0.387** (0.187)	0.406** (0.164)	0.297* (0.154)	
	Outcome post workshops/end of year (Feb and July 2008)				Treatment: Start year-End year Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated*PolicyOn	0.103** (0.045)	0.075* (0.044)	-0.042 (0.074)	-0.012 (0.088)	0.132** (0.053)	0.122** (0.052)	0.113* (0.064)	0.113 (0.095)	0.110 (0.077)	0.105 (0.075)	0.173** (0.080)	0.172 (0.116)
Treated	0.061 (0.060)	0.141*** (0.052)	0.070 (0.069)		0.051 (0.065)	0.115** (0.058)	0.011 (0.064)		0.071 (0.065)	0.122** (0.058)	0.014 (0.065)	
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Pupil Fixed Effects	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Sample size	7154	7154	7154	7154	5682	5682	5682	5682	5500	5500	5500	5500
Number of Pupils	2909	2909	2909	2909	2841	2841	2841	2841	2750	2750	2750	2750
p-value of $\chi^2(2)$ test of hypothesis of constant treatment effect	0.000	0.000	0.526	0.588	0.005	0.004	0.092	0.124	0.141	0.126	0.179	0.295

Notes: Standard errors (clustered by level of treatment assignment) in brackets; the control variables are dummies for gender (1), special educational needs status (1), free school meal status (1), gifted and talented status (1) and Key Stage 2 maths and English performance (10 dummies). The outcome measure here is the depression (CDI) score standardised to have a mean of 0 and a standard deviation of 1. Columns 3,4,7,8,11 & 12 contain controls for school trends (school at baseline*post).

Table 11: Heterogeneity in treatment effects for all three treatments pooled
Outcome: depression score
(From separate specifications relative to controls)

	Estimated Treatment*Policy Effect								
	Outcome post workshops/end of year (Feb and July 2008)			Outcome at one year follow-up (June 2009)			Outcome at two year follow-up (June 2010)		
	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls
Boys	-0.082 (0.062)			-0.030 (0.089)			-0.007 (0.111)		
Girls	-0.104 (0.071)			-0.020 (0.093)			0.056 (0.102)		
p-value from test of equality	0.796			0.919			0.631		
SEN	-0.191*	-0.194	-0.173	-0.141	-0.161	-0.064	-0.195	-0.213	-0.068
	(0.098)	(0.126)	(0.143)	(0.141)	(0.179)	(0.222)	(0.178)	(0.212)	(0.270)
Not SEN	-0.056	-0.002	-0.123	0.014	0.094	-0.083	0.084	0.133	0.015
	(0.059)	(0.077)	(0.084)	(0.077)	(0.105)	(0.107)	(0.087)	(0.139)	(0.113)
p-value from test of equality	0.210	0.171	0.736	0.291	0.177	0.936	0.136	0.127	0.766
FSM	-0.136	-0.006	-0.287**	-0.031	0.089	-0.197	0.010	0.051	-0.042
	(0.091)	(0.145)	(0.123)	(0.127)	(0.181)	(0.177)	(0.154)	(0.205)	(0.224)
Not FSM	-0.071	-0.077	-0.074	-0.013	-0.007	-0.027	0.037	0.049	0.017
	(0.060)	(0.073)	(0.096)	(0.078)	(0.109)	(0.107)	(0.092)	(0.136)	(0.121)
p-value from test of equality	0.531	0.650	0.165	0.892	0.625	0.362	0.865	0.994	0.813
KS2 English < Level 4	-0.157	-0.044	-0.333*	-0.136	0.019	-0.318	0.000	0.085	0.003
	(0.106)	(0.133)	(0.178)	(0.144)	(0.171)	(0.253)	(0.181)	(0.224)	(0.307)
KS2 English >= Level 4	-0.078	-0.069	-0.100	-0.012	0.008	-0.056	0.015	0.013	-0.003
	(0.057)	(0.074)	(0.081)	(0.077)	(0.107)	(0.103)	(0.088)	(0.139)	(0.111)
p-value from test of equality	0.487	0.857	0.201	0.398	0.952	0.305	0.934	0.748	0.984
KS2 maths < Level 4	-0.187**	-0.051	-0.327**	-0.105	0.046	-0.260	-0.053	-0.053	-0.047
	(0.093)	(0.139)	(0.138)	(0.139)	(0.185)	(0.198)	(0.163)	(0.218)	(0.239)
KS2 maths >= Level 4	-0.072	-0.067	-0.091	-0.014	0.006	-0.050	0.035	0.049	0.004
	(0.055)	(0.069)	(0.083)	(0.078)	(0.107)	(0.104)	(0.091)	(0.140)	(0.117)
p-value from test of equality	0.215	0.909	0.094	0.533	0.843	0.302	0.607	0.660	0.839
Number of pupils	2909	1513	1396	2909	1513	1396	2909	1513	1396
Sample size	7154	3779	3375	5750	2986	2764	5659	2933	2726

(continued)

Table 11 (continued): Heterogeneity in treatment effects for all three treatments pooled

Outcome: depression score

(From separate specifications relative to controls)

	Estimated Treatment*Policy Effect								
	Outcome post workshops/end of year (Feb and July 2008)			Outcome at one year follow-up (June 2009)			Outcome at two year follow-up (June 2010)		
	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls
1st quintile baseline CDI score	-0.001 (0.060)	0.036 (0.093)	-0.058 (0.084)	-0.027 (0.074)	0.138 (0.103)	-0.192* (0.116)	0.001 (0.097)	0.116 (0.150)	-0.125 (0.125)
2nd quintile baseline CDI score	-0.088 (0.075)	-0.147 (0.113)	-0.027 (0.116)	0.049 (0.127)	-0.005 (0.182)	0.098 (0.174)	-0.012 (0.161)	-0.029 (0.249)	0.019 (0.190)
3rd quintile baseline CDI score	-0.085 (0.086)	-0.156 (0.113)	-0.002 (0.116)	-0.052 (0.110)	-0.016 (0.155)	-0.070 (0.153)	0.055 (0.135)	-0.030 (0.204)	0.175 (0.181)
4th quintile baseline CDI score	-0.115 (0.080)	-0.066 (0.105)	-0.193 (0.134)	0.062 (0.123)	0.044 (0.157)	0.072 (0.190)	0.202 (0.147)	0.074 (0.186)	0.362* (0.199)
5th quintile baseline CDI score	-0.237** (0.117)	-0.126 (0.171)	-0.380** (0.164)	-0.083 (0.146)	-0.136 (0.204)	-0.057 (0.220)	-0.157 (0.160)	-0.132 (0.208)	-0.245 (0.251)
p-value of test of equality 1st-2nd quintile coefficients	0.258	0.149	0.814	0.565	0.417	0.132	0.941	0.559	0.518
p-value of test of equality 2nd-3rd quintile coefficients	0.980	0.953	0.869	0.523	0.957	0.443	0.711	0.998	0.507
p-value of test of equality 3rd-4th quintile coefficients	0.771	0.550	0.213	0.447	0.767	0.524	0.418	0.677	0.471
p-value of test of equality 4th-5th quintile coefficients	0.379	0.754	0.321	0.401	0.461	0.623	0.114	0.493	0.041
p-value of test of equality all quintile coefficients	0.282	0.493	0.207	0.895	0.639	0.586	0.599	0.888	0.093
Number of pupils	2909	1513	1396	2909	1513	1396	2909	1513	1396
Sample size	7154	3779	3375	5750	2986	2764	5659	2933	2726

Notes: Each column-box represents a separate regression. The outcome measure here is the depression (CDI) score, standardised to have a mean of 0 and a standard deviation of 1. Regressions include controls for mean reversion (5 dummies – initial quintile of depression score*post), and for school-specific trends (school at baseline*post). Compare Table 11 of the Second Interim Report.

Table 12: Heterogeneity in treatment effects for Group 1 (start-mid year) treatment
Outcome: depression score
(From separate specifications relative to controls)

	Estimated Treatment*Policy Effect								
	Outcome post workshops/end of year (Feb and July 2008)			Outcome at one year follow-up (June 2009)			Outcome at two year follow-up (June 2010)		
	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls
Boys	-0.049 (0.098)			0.087 (0.151)			0.138 (0.163)		
Girls	-0.223** (0.102)			-0.152 (0.150)			-0.166 (0.151)		
p-value from test of equality	0.220			0.219			0.099		
SEN	-0.201 (0.162)	-0.196 (0.205)	-0.210 (0.249)	-0.039 (0.220)	-0.112 (0.262)	0.061 (0.334)	-0.188 (0.278)	-0.213 (0.370)	-0.114 (0.347)
Not SEN	-0.091 (0.088)	0.022 (0.131)	-0.237* (0.122)	-0.007 (0.134)	0.096 (0.177)	-0.132 (0.171)	0.061 (0.129)	0.173 (0.179)	-0.055 (0.160)
p-value from test of equality	0.565	0.386	0.927	0.898	0.464	0.626	0.387	0.339	0.858
FSM	-0.088 (0.138)	0.027 (0.217)	-0.262 (0.203)	0.014 (0.193)	0.004 (0.265)	-0.026 (0.260)	0.163 (0.269)	0.255 (0.315)	0.065 (0.420)
Not FSM	-0.133 (0.085)	-0.083 (0.118)	-0.216 (0.143)	-0.020 (0.128)	0.028 (0.170)	-0.117 (0.168)	-0.045 (0.135)	0.009 (0.179)	-0.110 (0.207)
p-value from test of equality	0.778	0.661	0.865	0.864	0.933	0.764	0.438	0.382	0.735
KS2 English < Level 4	-0.159 (0.205)	0.021 (0.272)	-0.534** (0.270)	-0.050 (0.216)	0.208 (0.221)	-0.616* (0.364)	-0.155 (0.269)	0.019 (0.394)	-0.518 (0.506)
KS2 English >= Level 4	-0.114 (0.077)	-0.074 (0.105)	-0.189* (0.111)	-0.021 (0.121)	-0.029 (0.175)	-0.044 (0.138)	0.010 (0.126)	0.042 (0.172)	-0.032 (0.159)
p-value from test of equality	0.837	0.733	0.236	0.895	0.318	0.124	0.528	0.953	0.313
KS2 maths < Level 4	-0.194 (0.151)	-0.029 (0.253)	-0.393** (0.196)	-0.216 (0.185)	0.037 (0.226)	-0.489 (0.296)	-0.280 (0.231)	-0.259 (0.304)	-0.233 (0.414)
KS2 maths >= Level 4	-0.114 (0.078)	-0.064 (0.105)	-0.210* (0.110)	0.004 (0.123)	0.010 (0.180)	-0.052 (0.138)	0.044 (0.130)	0.106 (0.171)	-0.053 (0.168)
p-value from test of equality	0.624	0.892	0.371	0.278	0.923	0.153	0.155	0.229	0.676
Number of pupils	1804	960	844	1804	960	844	1804	960	844
Sample size	4748	2567	2181	3548	1884	1664	3469	1836	1633

(continued)

Table 12 (continued): Heterogeneity in treatment effects for Group 1 (start-mid year) treatment
Outcome: depression score
(From separate specifications relative to controls)

	Estimated Treatment*Policy Effect								
	Outcome post workshops/end of year (Feb and July 2008)			Outcome at one year follow-up (June 2009)			Outcome at two year follow-up (June 2010)		
	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls
1st quintile baseline CDI score	-0.062 (0.077)	0.001 (0.112)	-0.162 (0.107)	-0.017 (0.097)	0.071 (0.133)	-0.105 (0.145)	0.069 (0.195)	0.250 (0.322)	-0.104 (0.201)
2nd quintile baseline CDI score	-0.171 (0.111)	-0.285* (0.154)	-0.035 (0.165)	-0.067 (0.188)	-0.269 (0.244)	0.202 (0.243)	-0.225 (0.197)	-0.263 (0.287)	-0.068 (0.255)
3rd quintile baseline CDI score	-0.053 (0.145)	-0.073 (0.177)	-0.038 (0.178)	-0.103 (0.169)	0.054 (0.230)	-0.317* (0.175)	0.014 (0.206)	-0.052 (0.306)	0.095 (0.279)
4th quintile baseline CDI score	-0.095 (0.122)	0.105 (0.184)	-0.350* (0.184)	0.106 (0.209)	0.209 (0.243)	-0.042 (0.283)	0.178 (0.267)	0.252 (0.400)	0.080 (0.306)
5th quintile baseline CDI score	-0.453* (0.233)	-0.281 (0.342)	-0.685** (0.285)	-0.200 (0.261)	-0.237 (0.389)	-0.217 (0.310)	-0.567** (0.274)	-0.543* (0.322)	-0.691 (0.441)
p-value of test of equality 1st-2nd quintile coefficients	0.296	0.068	0.415	0.795	0.136	0.266	0.267	0.203	0.907
p-value of test of equality 2nd-3rd quintile coefficients	0.525	0.341	0.990	0.878	0.318	0.049	0.400	0.620	0.624
p-value of test of equality 3rd-4th quintile coefficients	0.827	0.486	0.241	0.379	0.618	0.298	0.606	0.559	0.971
p-value of test of equality 4th-5th quintile coefficients	0.207	0.378	0.255	0.355	0.316	0.683	0.096	0.181	0.207
p-value of test of equality all quintile coefficients	0.448	0.211	0.290	0.869	0.472	0.360	0.434	0.530	0.533
Number of pupils	1804	960	844	1804	960	844	1804	960	844
Sample size	4748	2567	2181	3548	1884	1664	3469	1836	1633

Notes: Each column-box represents a separate regression. The outcome measure here is the depression (CDI) score, standardised to have a mean of 0 and a standard deviation of 1. Regressions include controls for mean reversion (5 dummies – initial quintile of depression score*post), and for school-specific trends (school at baseline*post). Compare Table 31 of the First Interim Report, and Table 12 of the Second Interim Report.

Table 13: Treatment effects for the three experiments pooled

Outcome: anxiety score

(From specification pooled across all three treatments relative to controls)

	All treatment and control groups											
	Outcome post workshops/end of year (Feb and July 2008)				Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated*PolicyOn	-0.051 (0.042)	-0.057 (0.041)	-0.125*** (0.043)	-0.040 (0.048)	0.062 (0.044)	0.055 (0.044)	-0.004 (0.049)	-0.007 (0.070)	0.026 (0.051)	0.025 (0.051)	-0.044 (0.054)	-0.044 (0.076)
Treated	0.108** (0.049)	0.146*** (0.043)	0.107** (0.045)		0.081 (0.053)	0.116** (0.048)	0.070 (0.049)		0.087 (0.053)	0.115** (0.049)	0.061 (0.051)	
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Pupil Fixed Effects	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Number of Pupils	2879	2879	2879	2879	2804	2804	2804	2804	2701	2701	2701	2701
Sample size	7057	7057	7057	7057	5608	5608	5608	5608	5402	5402	5402	5402

Notes: Standard errors (clustered by level of treatment assignment) in brackets; the control variables are dummies for gender (1), special educational needs status (1), free school meal status (1), gifted and talented status (1) and Key Stage 2 maths and English performance (10 dummies). The outcome measure here is the anxiety (RCMAS) score standardised to have a mean of 0 and a standard deviation of 1. Columns 3, 4, 7, 8, 11 and 12 contain controls for school trends (school at baseline*post).

Here, and in subsequent tables, the level of significance of a coefficient is indicated by asterisks after it: one asterisk means that it is significant at the 10% level ($p \leq 0.1$ – there is a 10% chance that this could have arisen by chance); two asterisks that it is significant at 5% ($p \leq 0.05$); and three that it is significant at 1% ($p \leq 0.01$).

Columns 1-8 present results comparable with those of Table 14 of Challen et al. (2010), although here a different sample of pupils is used and the specification modified slightly.

Table 14: Treatment effects for the three experiments pooled: anxiety score
(From specification pooled across all three treatments relative to controls)

	Outcome post workshops/end of year (Feb and July 2008)				Treatment: Start year-Mid year Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated*PolicyOn	-0.123** (0.052)	-0.120** (0.053)	-0.105 (0.064)	-0.102 (0.080)	-0.058 (0.073)	-0.066 (0.072)	-0.057 (0.075)	-0.054 (0.107)	-0.100 (0.081)	-0.102 (0.082)	-0.095 (0.089)	-0.092 (0.125)
Treated	0.020 (0.091)	0.049 (0.078)	0.019 (0.079)		0.034 (0.093)	0.060 (0.080)	0.038 (0.079)		0.005 (0.090)	0.031 (0.082)	0.025 (0.083)	
Treatment: Mid year-End year												
	Outcome post workshops/end of year (Feb and July 2008)				Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
Treated*PolicyOn	-0.057 (0.074)	-0.042 (0.076)	-0.042 (0.073)	-0.050 (0.099)	-0.158 (0.131)	-0.170 (0.129)	-0.079 (0.123)	-0.103 (0.175)	-0.197 (0.187)	-0.205 (0.189)	-0.170 (0.160)	-0.176 (0.228)
Treated	0.263** (0.113)	0.285*** (0.098)	0.226** (0.094)		0.366** (0.160)	0.391*** (0.143)	0.297** (0.141)		0.377** (0.165)	0.394*** (0.149)	0.311** (0.142)	
Treatment: Start year-End year												
	Outcome post workshops/end of year (Feb and July 2008)				Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
Treated*PolicyOn	0.135*** (0.050)	0.114** (0.051)	0.067 (0.060)	0.080 (0.074)	0.147*** (0.052)	0.140*** (0.052)	0.082 (0.073)	0.086 (0.107)	0.122* (0.062)	0.123* (0.063)	0.056 (0.072)	0.057 (0.103)
Treated	0.055 (0.058)	0.103* (0.054)	0.072 (0.070)		0.047 (0.062)	0.088 (0.059)	0.030 (0.070)		0.065 (0.062)	0.096 (0.060)	0.015 (0.073)	
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Pupil Fixed Effects	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Sample size	7057	7057	7057	7057	5608	5608	5608	5608	5402	5402	5402	5402
Number of Pupils	2879	2879	2879	2879	2804	2804	2804	2804	2701	2701	2701	2701
p-value of $\chi^2(2)$ test of hypothesis of constant treatment effect	0.000	0.002	0.129	0.210	0.006	0.004	0.329	0.538	0.020	0.020	0.237	0.486

Notes: Standard errors (clustered by level of treatment assignment) in brackets; the control variables are dummies for gender (1), special educational needs status (1), free school meal status (1), gifted and talented status (1) and Key Stage 2 maths and English performance (10 dummies). The outcome measure here is the anxiety (RCMAS) score standardised to have a mean of 0 and a standard deviation of 1. Columns 3,4,7,8,11 & 12 contain controls for school trends (school at baseline*post).

Table 15: Heterogeneity in treatment effects for all treatments pooled
Outcome: anxiety score
(From separate specifications relative to controls)

	Estimated Treatment*Policy Effect								
	Outcome post workshops/end of year (Feb and July 2008)			Outcome at one year follow-up (June 2009)			Outcome at two year follow-up (June 2010)		
	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls
Boys	-0.061 (0.060)			-0.031 (0.085)			-0.090 (0.099)		
Girls	-0.017 (0.061)			0.010 (0.089)			-0.010 (0.090)		
p-value from test of equality	0.547			0.687			0.489		
SEN	-0.069 (0.098)	-0.099 (0.134)	-0.009 (0.148)	-0.035 (0.138)	-0.066 (0.176)	0.000 (0.196)	-0.103 (0.158)	-0.204 (0.199)	0.114 (0.232)
Not SEN	-0.032 (0.047)	-0.045 (0.065)	-0.028 (0.075)	-0.008 (0.068)	-0.062 (0.086)	0.035 (0.095)	-0.041 (0.075)	-0.087 (0.109)	-0.011 (0.100)
p-value from test of equality	0.696	0.671	0.903	0.836	0.984	0.860	0.693	0.532	0.611
FSM	-0.143 (0.091)	-0.168 (0.125)	-0.142 (0.125)	-0.090 (0.133)	-0.114 (0.175)	-0.118 (0.171)	-0.092 (0.144)	-0.158 (0.197)	-0.032 (0.192)
Not FSM	0.001 (0.049)	-0.023 (0.067)	0.024 (0.080)	0.025 (0.071)	-0.044 (0.098)	0.098 (0.097)	-0.030 (0.076)	-0.097 (0.111)	0.032 (0.104)
p-value from test of equality	0.120	0.205	0.252	0.387	0.711	0.237	0.657	0.744	0.767
KS2 English < Level 4	-0.085 (0.116)	-0.066 (0.143)	-0.072 (0.161)	-0.080 (0.150)	-0.042 (0.182)	-0.126 (0.236)	-0.104 (0.180)	-0.111 (0.224)	0.014 (0.271)
KS2 English >= Level 4	-0.035 (0.048)	-0.068 (0.068)	-0.013 (0.075)	-0.013 (0.068)	-0.079 (0.089)	0.040 (0.093)	-0.061 (0.077)	-0.133 (0.111)	-0.002 (0.099)
p-value from test of equality	0.667	0.992	0.736	0.634	0.827	0.482	0.812	0.919	0.955
KS2 maths < Level 4	-0.136 (0.099)	-0.033 (0.145)	-0.217 (0.135)	-0.111 (0.139)	-0.015 (0.185)	-0.194 (0.183)	-0.124 (0.153)	-0.111 (0.207)	-0.094 (0.200)
KS2 maths >= Level 4	-0.025 (0.047)	-0.075 (0.068)	0.018 (0.073)	-0.002 (0.068)	-0.087 (0.089)	0.073 (0.096)	-0.045 (0.077)	-0.127 (0.109)	0.021 (0.102)
p-value from test of equality	0.254	0.761	0.093	0.425	0.687	0.166	0.614	0.931	0.591
Number of pupils	2879	1497	1382	2804	1450	1354	2701	1394	1307
Sample size	7057	3736	3321	5608	2900	2708	5402	2788	2614

(continued)

Table 15 (continued): Heterogeneity in treatment effects for all treatments pooled

Outcome: Anxiety score

(From separate specifications relative to controls)

	Estimated Treatment*Policy Effect								
	Outcome post workshops/end of year (Feb and July 2008)			Outcome at one year follow-up (June 2009)			Outcome at two year follow-up (June 2010)		
	All	Boys	Girls	All	Boys	Girls	All	Boys	Girls
1st quintile baseline RCMAS score	-0.013 (0.055)	-0.003 (0.085)	-0.037 (0.066)	-0.024 (0.070)	-0.068 (0.096)	0.010 (0.111)	-0.056 (0.086)	-0.056 (0.101)	-0.071 (0.123)
2nd quintile baseline RCMAS score	-0.002 (0.067)	-0.107 (0.088)	0.088 (0.110)	0.022 (0.097)	-0.118 (0.099)	0.132 (0.163)	-0.058 (0.105)	-0.197 (0.120)	0.063 (0.167)
3rd quintile baseline RCMAS score	0.039 (0.075)	0.028 (0.106)	0.040 (0.117)	0.070 (0.109)	0.055 (0.139)	0.092 (0.152)	0.070 (0.118)	0.026 (0.162)	0.137 (0.166)
4th quintile baseline RCMAS score	-0.109 (0.084)	-0.163 (0.120)	-0.040 (0.138)	0.040 (0.111)	-0.106 (0.152)	0.175 (0.169)	-0.047 (0.124)	-0.195 (0.155)	0.102 (0.192)
5th quintile baseline RCMAS score	-0.117 (0.084)	-0.104 (0.127)	-0.146 (0.131)	-0.078 (0.133)	-0.015 (0.169)	-0.168 (0.202)	-0.039 (0.154)	-0.081 (0.195)	-0.032 (0.229)
p-value of test of equality 1st-2nd quintile coefficients	0.892	0.350	0.305	0.670	0.686	0.496	0.986	0.332	0.502
p-value of test of equality 2nd-3rd quintile coefficients	0.668	0.279	0.757	0.721	0.297	0.840	0.404	0.255	0.748
p-value of test of equality 3rd-4th quintile coefficients	0.198	0.197	0.646	0.833	0.403	0.699	0.470	0.275	0.881
p-value of test of equality 4th-5th quintile coefficients	0.941	0.736	0.515	0.474	0.685	0.186	0.968	0.628	0.655
p-value of test of equality all quintile coefficients	0.499	0.548	0.653	0.892	0.869	0.666	0.886	0.727	0.818
Number of pupils	2879	1497	1382	2804	1450	1354	2701	1394	1307
Sample size	7057	3736	3321	5608	2900	2708	5402	2788	2614

Notes: Each column-box represents a separate regression. The outcome measure here is the anxiety (RCMAS) score, standardised to have a mean of 0 and a standard deviation of 1. Regressions include controls for mean reversion (5 dummies – initial quintile of anxiety score*post), and for school-specific trends (school at baseline*post). Compare Table 16 of the Second Interim Report.

Table 16: Treatment effects for the three experiments pooled

Outcome: Self-reported behaviour score

(From specification pooled across all three treatments relative to controls)

	All treatment and control groups											
	Outcome post workshops/end of year (Feb and July 2008)				Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated*PolicyOn	-0.010 (0.044)	-0.015 (0.043)	-0.057 (0.047)	0.023 (0.056)	0.055 (0.043)	0.045 (0.042)	0.024 (0.054)	0.010 (0.073)	0.037 (0.056)	0.032 (0.055)	0.054 (0.049)	0.041 (0.067)
Treated	0.064 (0.050)	0.114*** (0.044)	0.042 (0.044)		0.057 (0.057)	0.107** (0.051)	0.001 (0.049)		0.064 (0.057)	0.110** (0.051)	-0.008 (0.051)	
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Pupil Fixed Effects	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Number of Pupils	2880	2880	2880	2880	2815	2815	2815	2815	2703	2703	2703	2703
Sample size	7069	7069	7069	7069	5630	5630	5630	5630	5406	5406	5406	5406

Notes: Standard errors (clustered by level of treatment assignment) in brackets; the control variables are dummies for gender (1), special educational needs status (1), free school meal status (1), gifted and talented status (1) and Key Stage 2 maths and English performance (10 dummies). The outcome measure here is the self-reported behaviour (Goodman SDQ) score standardised to have a mean of 0 and a standard deviation of 1. Columns 3, 4, 7, 8, 11 and 12 contain controls for school trends (school at baseline*post).

Here, and in subsequent tables, the level of significance of a coefficient is indicated by asterisks after it: one asterisk means that it is significant at the 10% level ($p \leq 0.1$ – there is a 10% chance that this could have arisen by chance); two asterisks that it is significant at 5% ($p \leq 0.05$); and three that it is significant at 1% ($p \leq 0.01$).

Table 17: Treatment effects for the three experiments pooled

Outcome: Teacher-reported behaviour score

(From specification pooled across all three treatments relative to controls)

	All treatment and control groups											
	Outcome post workshops/end of year (Feb and July 2008)				Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated*PolicyOn	0.014 (0.081)	-0.039 (0.073)	-0.047 (0.067)	-0.012 (0.083)	0.069 (0.071)	0.052 (0.071)	0.144* (0.086)	0.141 (0.124)	0.011 (0.084)	-0.026 (0.084)	0.054 (0.118)	-0.015 (0.164)
Treated	-0.087 (0.090)	0.017 (0.073)	-0.061 (0.087)		-0.072 (0.092)	0.024 (0.076)	-0.111 (0.085)		-0.014 (0.088)	0.081 (0.075)	-0.070 (0.090)	
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Pupil Fixed Effects	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Number of Pupils	2455	2455	2455	2455	2433	2433	2433	2433	2248	2248	2248	2248
Sample size	5597	5597	5597	5597	4657	4657	4657	4657	4256	4256	4256	4256

Notes: Standard errors (clustered by level of treatment assignment) in brackets; the control variables are dummies for gender (1), special educational needs status (1), free school meal status (1), gifted and talented status (1) and Key Stage 2 maths and English performance (10 dummies). The outcome measure here is the teacher-reported behaviour (Goodman SDQ) score standardised to have a mean of 0 and a standard deviation of 1. Columns 3, 4, 7, 8, 11 and 12 contain controls for school trends (school at baseline*post).

Here, and in subsequent tables, the level of significance of a coefficient is indicated by asterisks after it: one asterisk means that it is significant at the 10% level ($p \leq 0.1$ – there is a 10% chance that this could have arisen by chance); two asterisks that it is significant at 5% ($p \leq 0.05$); and three that it is significant at 1% ($p \leq 0.01$).

Table 18: Treatment effects for the three experiments pooled

Outcome: life satisfaction score

(From specification pooled across all three treatments relative to controls)

	All treatment and control groups											
	Outcome post workshops/end of year (Feb and July 2008)				Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated*PolicyOn	-0.022 (0.038)	-0.019 (0.038)	-0.038 (0.045)	-0.077 (0.057)	-0.056 (0.046)	-0.051 (0.046)	-0.091* (0.050)	-0.081 (0.071)	-0.078 (0.056)	-0.077 (0.055)	-0.098* (0.056)	-0.093 (0.079)
Treated	-0.086* (0.044)	-0.119*** (0.041)	0.002 (0.048)		-0.097* (0.051)	-0.125** (0.049)	0.012 (0.055)		-0.100* (0.051)	-0.127** (0.049)	0.017 (0.052)	
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Pupil Fixed Effects	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Number of Pupils	2811	2811	2811	2811	2744	2744	2744	2744	2656	2656	2656	2656
Sample size	6861	6861	6861	6861	5488	5488	5488	5488	5312	5312	5312	5312

Notes: Standard errors (clustered by level of treatment assignment) in brackets; the control variables are dummies for gender (1), special educational needs status (1), free school meal status (1), gifted and talented status (1) and Key Stage 2 maths and English performance (10 dummies). The outcome measure here is the life satisfaction score standardised to have a mean of 0 and a standard deviation of 1. Columns 3, 4, 7, 8, 11 and 12 contain controls for school trends (school at baseline*post).

Here, and in subsequent tables, the level of significance of a coefficient is indicated by asterisks after it: one asterisk means that it is significant at the 10% level ($p \leq 0.1$ – there is a 10% chance that this could have arisen by chance); two asterisks that it is significant at 5% ($p \leq 0.05$); and three that it is significant at 1% ($p \leq 0.01$).

Table 19: Treatment effects for the three experiments pooled

Outcome: absence from school

(From specification pooled across all three treatments relative to controls)

All treatment and control groups								
	Outcome in academic year 2007-08				Outcome in academic year 2008-09			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated*PolicyOn	-0.060 (0.050)	-0.060 (0.050)	-0.159** (0.067)	-0.159* (0.094)	-0.059 (0.057)	-0.059 (0.057)	-0.055 (0.066)	-0.055 (0.094)
Treated	-0.015 (0.024)	0.032 (0.022)	0.068*** (0.026)		-0.013 (0.023)	0.031 (0.022)	0.063** (0.026)	
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	No	Yes	No
Pupil Fixed Effects	No	No	No	Yes	No	No	No	Yes
Number of Pupils	3755	3755	3755	3755	3728	3728	3728	3728
Sample size	7510	7510	7510	7510	7456	7456	7456	7456

Notes: Standard errors (clustered by level of treatment assignment) in brackets; the control variables are dummies for gender (1), special educational needs status (1), free school meal status (1), gifted and talented status (1) and Key Stage 2 maths and English performance (10 dummies). The outcome measure here is the fraction of school sessions absent standardised to have a mean of 0 and a standard deviation of 1. Columns 3, 4, 7, 8, 11 and 12 contain controls for school trends (school at baseline*post).

Here, and in subsequent tables, the level of significance of a coefficient is indicated by asterisks after it: one asterisk means that it is significant at the 10% level ($p \leq 0.1$ – there is a 10% chance that this could have arisen by chance); two asterisks that it is significant at 5% ($p \leq 0.05$); and three that it is significant at 1% ($p \leq 0.01$).

Table 20: Treatment effects for the three experiments

Outcome: absence from school

(From specification pooled across all three treatments relative to controls)

		Treatment: Start year-Mid year							
		Outcome in academic year 2007-08				Outcome in academic year 2008-09			
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treated*PolicyOn		-0.217*** (0.076)	-0.217*** (0.076)	-0.256*** (0.077)	-0.256** (0.108)	-0.145** (0.065)	-0.145** (0.065)	-0.186** (0.085)	-0.186 (0.120)
Treated		0.021 (0.042)	0.062 (0.038)	0.078* (0.041)		0.024 (0.041)	0.058 (0.039)	0.076* (0.041)	
		Treatment: Mid year-End year							
		Outcome in academic year 2007-08				Outcome in academic year 2008-09			
Treated*PolicyOn		0.052 (0.116)	0.052 (0.116)	-0.096 (0.118)	-0.096 (0.167)	0.262 (0.177)	0.262 (0.178)	0.224 (0.161)	0.224 (0.228)
Treated		0.041 (0.062)	0.052 (0.053)	0.104* (0.063)		0.044 (0.062)	0.058 (0.052)	0.109* (0.063)	
		Treatment: Start year-End year							
		Outcome in academic year 2007-08				Outcome in academic year 2008-09			
Treated*PolicyOn		-0.024 (0.057)	-0.024 (0.057)	-0.075 (0.121)	-0.075 (0.171)	-0.088 (0.065)	-0.088 (0.065)	-0.030 (0.100)	-0.030 (0.141)
Treated		-0.039* (0.021)	0.017 (0.021)	0.039 (0.025)		-0.037* (0.021)	0.015 (0.022)	0.027 (0.025)	
Month dummies	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	No	No	Yes	No
Pupil Fixed Effects	No	No	No	Yes	No	No	No	No	Yes
Sample size	7510	7510	7510	7510	7510	7456	7456	7456	7456
Number of Pupils	3755	3755	3755	3755	3755	3728	3728	3728	3728
p-value of $\chi^2(2)$ test of hypothesis of constant treatment effect	0.028	0.028	0.225	0.474	0.070	0.070	0.029	0.169	

Notes: Standard errors (clustered by level of treatment assignment) in brackets; the control variables are dummies for gender (1), special educational needs status (1), free school meal status (1), gifted and talented status (1) and Key Stage 2 maths and English performance (10 dummies). The outcome measure here is the fraction of school sessions absent standardised to have a mean of 0 and a standard deviation of 1. Columns 3,4,7,8,11 & 12 contain controls for school trends (school at baseline*post).

Table 21: Heterogeneity in treatment effects for all treatments pooled

Outcome: Absence

(From separate specifications relative to controls)

	Estimated Treatment*Policy Effect					
	Outcome in academic year 2007-08			Outcome in academic year 2008-09		
	All	Boys	Girls	All	Boys	Girls
Boys	-0.219*			-0.088		
	(0.113)			(0.125)		
Girls	-0.090			-0.018		
	(0.099)			(0.105)		
p-value from test of equality	0.197			0.608		
SEN	-0.044	-0.223	0.253	-0.037	-0.075	0.086
	(0.175)	(0.221)	(0.241)	(0.169)	(0.223)	(0.255)
Not SEN	-0.171*	-0.223*	-0.126	-0.029	-0.016	-0.047
	(0.087)	(0.125)	(0.105)	(0.098)	(0.162)	(0.114)
p-value from test of equality	0.413	0.999	0.106	0.960	0.793	0.601
FSM	-0.160	-0.328	0.022	0.071	0.013	0.137
	(0.160)	(0.225)	(0.209)	(0.174)	(0.258)	(0.223)
Not FSM	-0.145*	-0.197	-0.091	-0.100	-0.088	-0.103
	(0.084)	(0.122)	(0.095)	(0.095)	(0.166)	(0.105)
p-value from test of equality	0.917	0.526	0.583	0.363	0.728	0.301
KS2 English < Level 4	-0.103	-0.304	0.249	0.073	-0.037	0.303
	(0.175)	(0.252)	(0.266)	(0.182)	(0.255)	(0.287)
KS2 English >= Level 4	-0.130	-0.176	-0.092	-0.054	-0.032	-0.072
	(0.095)	(0.127)	(0.112)	(0.096)	(0.162)	(0.115)
p-value from test of equality	0.873	0.579	0.197	0.472	0.984	0.189
KS2 maths < Level 4	-0.121	-0.289	0.044	0.053	0.064	0.047
	(0.175)	(0.233)	(0.226)	(0.189)	(0.285)	(0.269)
KS2 maths >= Level 4	-0.127	-0.200	-0.054	-0.045	-0.055	-0.022
	(0.094)	(0.143)	(0.094)	(0.093)	(0.160)	(0.104)
p-value from test of equality	0.967	0.691	0.635	0.601	0.681	0.795
Number of pupils	3763	1973	1790	3763	1973	1790
Sample size	7518	3943	3575	7491	3923	3568

(continued)

Table 21 (continued): Heterogeneity in treatment effects for all treatments pooled**Outcome: Absence**

(From separate specifications relative to controls)

	Estimated Treatment*Policy Effect					
	Outcome in academic year 2007-08			Outcome in academic year 2008-09		
	All	Boys	Girls	All	Boys	Girls
1st quartile baseline absence rate	-0.022 (0.160)	-0.046 (0.222)	0.024 (0.147)	-0.095 (0.166)	-0.128 (0.238)	-0.057 (0.177)
2nd quartile baseline absence rate	-0.302*** (0.111)	-0.383** (0.163)	-0.235 (0.152)	0.054 (0.223)	0.195 (0.431)	-0.054 (0.204)
3rd quartile baseline absence rate	-0.265** (0.113)	-0.249* (0.129)	-0.319* (0.192)	-0.025 (0.138)	-0.014 (0.199)	-0.037 (0.194)
4th quartile baseline absence rate	-0.158 (0.162)	-0.460* (0.255)	0.187 (0.191)	-0.082 (0.176)	-0.158 (0.254)	0.020 (0.248)
p-value of test of equality 1st-2nd quartile coefficients	0.129	0.194	0.209	0.581	0.500	0.992
p-value of test of equality 2nd-3rd quartile coefficients	0.751	0.370	0.676	0.733	0.602	0.944
p-value of test of equality 3rd-4th quartile coefficients	0.508	0.360	0.035	0.784	0.607	0.856
p-value of test of equality all quartile coefficients	0.442	0.506	0.139	0.952	0.888	0.995
Number of pupils	3763	1973	1790	3763	1973	1790
Sample size	7518	3943	3575	7491	3923	3568

Notes: Each column-box represents a separate regression. The outcome measure here is the fraction of school sessions absent, standardised to have a mean of 0 and a standard deviation of 1. Regressions include controls for mean reversion (5 dummies – initial quartile of absence rate*post), and for school-specific trends (school at baseline*post). Compare Table 20 of the Second Interim Report.

Table 22: Treatment effects for the three experiments pooled

Outcome: Attainment in English

(From specification pooled across all three treatments relative to controls)

	All treatment and control groups											
	Outcome post workshops/end of year (Feb and July 2008)				Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated*PolicyOn	0.346*** (0.100)	0.351*** (0.101)	0.359*** (0.100)	0.366** (0.143)	0.211*** (0.075)	0.210*** (0.076)	0.212*** (0.076)	0.216** (0.107)	0.096 (0.108)	0.093 (0.108)	0.098 (0.108)	0.092 (0.153)
Treated	0.210* (0.122)	0.110 (0.078)	-0.036 (0.074)		0.210* (0.122)	0.102 (0.077)	0.026 (0.066)		0.210* (0.122)	0.098 (0.075)	0.088 (0.074)	
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Pupil Fixed Effects	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Number of Pupils	2299	2299	2299	2299	2299	2299	2299	2299	2299	2299	2299	2299
Sample size	4509	4509	4509	4509	4557	4557	4557	4557	4572	4572	4572	4572

Notes: Standard errors (clustered by level of treatment assignment) in brackets; the control variables are dummies for gender (1), special educational needs status (1), free school meal status (1), and gifted and talented status (1). The outcome measure here is attainment in English standardised to have a mean of 0 and a standard deviation of 1 at baseline. Data reported here come from 14 of the 22 schools in the sample.

Here, and in subsequent tables, the level of significance of a coefficient is indicated by asterisks after it: one asterisk means that it is significant at the 10% level ($p \leq 0.1$ – there is a 10% chance that this could have arisen by chance); two asterisks that it is significant at 5% ($p \leq 0.05$); and three that it is significant at 1% ($p \leq 0.01$).

Table 23: Treatment effects for the three experiments pooled

Outcome: Attainment in maths

(From specification pooled across all three treatments relative to controls)

	All treatment and control groups											
	Outcome post workshops/end of year (Feb and July 2008)				Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated*PolicyOn	0.093 (0.083)	0.094 (0.083)	0.105 (0.083)	0.102 (0.117)	0.236*** (0.082)	0.233*** (0.083)	0.234*** (0.083)	0.239** (0.117)	0.193** (0.092)	0.188** (0.093)	0.189** (0.093)	0.188 (0.131)
Treated	0.196* (0.115)	0.103 (0.071)	0.143** (0.067)		0.196* (0.115)	0.091 (0.069)	0.045 (0.070)		0.196* (0.115)	0.088 (0.067)	0.126* (0.073)	
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Pupil Fixed Effects	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Number of Pupils	2301	2301	2301	2301	2301	2301	2301	2301	2301	2301	2301	2301
Sample size	4509	4509	4509	4509	4565	4565	4565	4565	4579	4579	4579	4579

Notes: Standard errors (clustered by level of treatment assignment) in brackets; the control variables are dummies for gender (1), special educational needs status (1), free school meal status (1), and gifted and talented status (1). The outcome measure here is attainment in maths standardised to have a mean of 0 and a standard deviation of 1 at baseline. Data reported here come from 14 of the 22 schools in the sample.

Table 24: Treatment effects for the three experiments pooled

Outcome: Attainment in science

(From specification pooled across all three treatments relative to controls)

	All treatment and control groups											
	Outcome post workshops/end of year (Feb and July 2008)				Outcome at one year follow-up (June 2009)				Outcome at two year follow-up (June 2010)			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Treated*PolicyOn	0.114 (0.113)	0.120 (0.113)	0.123 (0.113)	0.110 (0.161)	0.083 (0.084)	0.085 (0.084)	0.084 (0.084)	0.081 (0.119)	0.121 (0.121)	0.114 (0.121)	0.115 (0.121)	0.113 (0.171)
Treated	0.211* (0.112)	0.122* (0.073)	0.059 (0.076)		0.211* (0.112)	0.123* (0.073)	0.098 (0.067)		0.211* (0.112)	0.099 (0.067)	0.119 (0.075)	
Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes	No	Yes	Yes	Yes
School Fixed Effects	No	No	Yes	No	No	No	Yes	No	No	No	Yes	No
Pupil Fixed Effects	No	No	No	Yes	No	No	No	Yes	No	No	No	Yes
Number of Pupils	2297	2297	2297	2297	2297	2297	2297	2297	2297	2297	2297	2297
Sample size	4499	4499	4499	4499	4549	4549	4549	4549	4577	4577	4577	4577

Notes: Standard errors (clustered by level of treatment assignment) in brackets; the control variables are dummies for gender (1), special educational needs status (1), free school meal status (1), and gifted and talented status (1). The outcome measure here is attainment in science standardised to have a mean of 0 and a standard deviation of 1 at baseline. Data reported here come from 14 of the 22 schools in the sample.

6. Qualitative Findings

Summary of Qualitative Findings

This chapter reports findings from the qualitative element of the evaluation. Ten of the twenty-two UKRP schools were visited in 2007-08 – the first year in which the UKRP operated. Nine of these schools were revisited in 2009-10. We briefly describe the UKRP programme and then present findings drawn from interviews with pupils, UKRP facilitators and senior managers at the case study schools.

Many of the pupils interviewed were able to describe real life circumstances in which they reported that they had used UKRP skills. Some of these pupils showed a fairly sophisticated degree of understanding when applying those skills to real life events. Most numerous were pupils who described circumstances in which they had ‘not risen’ to provocation and therefore presented the UKRP as reinforcing good behaviour. Others described using the ABC model to modify their feelings, or using skills such as assertiveness and negotiation to address everyday problems.

Facilitators reported that pupils sometimes raised serious issues during UKRP sessions - such as bereavement and family illness - and that the role of facilitator could be emotionally demanding. Facilitators were extremely positive about the training they had received. Facilitators who were not qualified teachers (e.g. learning mentors, school nurses) particularly valued the opportunity to practice delivering UKRP sessions during the training events. Facilitators were however less positive about the UKRP course materials. In particular they frequently suggested that there was too much ‘teacher talk’.

Nine of the case study schools were revisited in 2009-10 - the third year of the programme. Seven of the nine schools were delivering the UKRP to the whole Year 7 cohort in that year. Schools organised delivery of the UKRP in different ways. At some schools UKRP sessions were delivered fortnightly, although this arrangement was not popular among facilitators. Only two schools were delivering the programme weekly while also sticking to the recommendations of the course developers regarding the most desirable group size and course length. In some of the schools the UKRP was increasingly delivered by members of the non-teaching staff and a number of reasons, including competing pressures on teachers’ time, were suggested to explain this drift.

The chapter concludes with some lessons to be drawn from the qualitative element of the project, including lessons for practice.

Introduction

This section reports findings from case study visits to 10 out of the 22 UKRP schools. Initial visits were made to the schools during the spring and summer terms of 2007-08 – that is, the first year in which the UKRP was delivered. At this point, the first cohort of UKRP facilitators had completed their training which took place in the USA during the summer holiday of 2007. They were therefore in their first year of delivering the UKRP programme. Nine of these 10 schools were then revisited during the autumn term of 2009-10 – that is, the third year in which the UKRP was delivered. By this time two more cohorts of facilitators had received UKRP training, during a residential training event held in the summer of 2008 and non-residential training events held locally in 2009. The findings provide qualitative data to deepen the understanding of the UKRP and to provide a context for the quantitative results presented earlier in this report. In particular, the qualitative case study element of the research aimed to provide some insight into how the programme was implemented within schools, programme participants' reflections on their experience of the UKRP and also to provide some examples of pupils' use of the UKRP skills.

All of the qualitative fieldwork was conducted earlier in the evaluation and was reported on in the previous interim reports (Challen et al., 2009; Challen et al., 2010). In this section we present selected findings, relating to pupils' use of UKRP skills, because they illustrate some of the skills being taught and therefore complement some of the quantitative findings. In addition we present some key findings related to the implementation of the UKRP and facilitators' reflections on the programme. Earlier reports provide greater detail on some of these issues.

Data and methods

At least three schools were visited in each of the three local authority areas participating in the UKRP. Case study schools were selected to reflect the variation between schools in the proportion of the Year 7 cohort receiving UKRP sessions in the first year of the programme, and also to reflect variation in levels of pupil attainment, rates of eligibility for free school meals and school Contextual Value Added scores⁴⁴. In the first year of the programme, at some of the UKRP schools all facilitators were teachers while at others members of the non-teaching staff (e.g. learning mentors, teaching assistants) were also trained to deliver the UKRP. At a smaller number of schools the UKRP was delivered in part by facilitators who were not employed by the school. Case study schools were also selected to include these different arrangements.

The case study visits involved semi-structured interviews with facilitators, pupils who had attended UKRP groups and senior managers responsible for oversight of the programme within the school. In addition, UKRP coordinators employed by the local authorities were interviewed. Sixty interviews were carried out with staff involved in delivery of the UKRP, while 45 pupils were interviewed who had received UKRP

⁴⁴ Contextual Value Added (CVA) scores were calculated by the DfE (and its predecessors) to assess the amount of progress made by pupils from Key Stage 2 to Key Stage 4, taking into account factors outside a school's control that are known to have an impact on attainment – such as special educational needs, gender, family circumstances and the pupil peer group. The statistical technique used to calculate CVA scores identifies schools in which pupils make significantly more or significantly less progress than would be expected in the average school.

lessons (all pupil interviews took place during the 2007-08 visits). The previous interim reports also give more details of the interviewees and interview topics.

In the first year of the UKRP, four of the ten case study schools were delivering the UKRP to *all* Year 7 pupils. The remaining schools delivered the UKRP to 16-50% of the Year 7 cohort. When nine of the schools were revisited two years later, seven out of nine schools were delivering the UKRP to all Year 7 pupils, at one school the UKRP had been discontinued and at another school this was likely to be the case.

In what follows we provide a brief account of the content of the UKRP in order to explain some of the terminology of the UKRP that was used by interviewees. We then present data relating to pupils' reported use of the skills they had learned during UKRP sessions. This provides some insight into the types of day to day situation in which pupils reported they had used UKRP skills. Participants' reflections on the UKRP are then reported. In particular we report on pupils' enjoyment of UKRP sessions, discussion of personal material during UKRP sessions and the emotional demands this could make on facilitators, facilitators' thoughts on the UKRP training that they had received and finally their views on the UKRP course materials. We then go on to discuss some of the differences in how the UKRP was organised across the case study schools, noting the variation in levels of demand for UKRP training across the schools and different approaches to recruitment, the tendency within some of the schools for the UKRP to be increasingly delivered by members of the auxiliary staff, and also how schools found a 'home' for the UKRP within their timetable.

Brief description of the UKRP⁴⁵

In this brief description, UKRP terminology, some of which is used in later sections of this report, is shown in bold. The first sessions of the UKRP course focus on understanding and using the **ABC model**. The ABC model illustrates that, when faced by an Adversity or Activating event, Beliefs about that event mediate the behavioural and emotional Consequences. Thus, for example, if you are woken by a loud noise during the night (Activating event), Beliefs (e.g. 'it might be a burglar', 'it is just the cat') mediate the emotional and behavioural consequences (e.g. feeling scared, getting out of bed or going back to sleep).

Pupils are encouraged to identify the beliefs that may affect their own emotional and behavioural responses. In particular, in response to an adversity (e.g. receiving a bad mark in a test) they are encouraged to challenge negative automatic thoughts that arise (e.g. 'I always do badly in tests', 'I will never be able to do this') and **Generate Alternatives** (e.g. 'Everyone gets bad marks sometimes', 'If I worked harder I would get better marks').

Pupils are encouraged to challenge negative automatic thoughts by **Evaluating Evidence** (e.g. addressing the question 'have I done better in other tests?'). This is practiced through the **File Game** activity in which pupils are presented with documentary evidence (e.g. diary entries, school reports) about a fictitious character

⁴⁵ This description of the programme was produced by the authors and is intended to contextualise the terminology used in this section of the report. It is not intended to provide the best possible description of the programme.

with a negative thinking style. They then find evidence that the character could use in order to challenge the negative automatic thoughts.

When this understanding is applied to future adversities (e.g. a forthcoming presentation) they learn how to **Put it into Perspective** by considering the best, worst and most likely outcomes of that stressful event.

Pupils then practice these cognitive skills on the **Hot Seat** by responding to hypothetical or real automatic thoughts using the skills of Evaluating Evidence ('That's not true because...'), Generating Alternatives ('Another way of looking at this is...') and Putting it into Perspective ('The most likely outcome is...'). The cognitive skills learned so far are then tested through a team quiz called **Jeopardy**.

The course also includes class and homework tasks. These sometimes involve completing **Thought Bubbles** in cartoon strips to show that what a character believes mediates between a particular activating event and a given outcome. Pupils are also sometimes asked to write down real life problems on a **Problem Pool Card** so that these problems may subsequently (and anonymously) be discussed in class.

The skills are then applied to social situations. Pupils are invited to distinguish between aggressive, assertive and passive responses to situations. The **DEAL model** provides a means of being assertive by Describing the problem, Explaining how you feel, Asking for a change and Listing the benefits that will follow. They also practice negotiating using the maxim **Be wise, compromise**.

Pupils are then introduced to behavioural coping techniques to deal with uncontrollable situations (such as parents arguing) and strong emotions such as anger and sadness. The coping techniques include **Controlled Breathing, Muscle Relaxation, Positive Visualisation, Leaving the Room** and distraction techniques such as **Mental Games**.

Several new skills introduced in the final sessions. These include **Overcoming Procrastination**, which may be caused by inaccurate thinking regarding a task (e.g. 'My report has to be faultless'), by applying cognitive skills (e.g. putting it into perspective) or behavioural skills (such as rewarding a series of intermediate steps). In addition, pupils are introduced to techniques to assist with decision-making and five step approach to problem solving. The problem solving technique is introduced using a scenario in which the pupil is taking a drink at a water fountain when someone bumps into them. They are encouraged to: 'stop and think, look for clues' (e.g. consider why this happened); 'stand in others' shoes' (e.g. look for body language to indicate the motives of others); 'choose your goal' (e.g. are they concerned that this doesn't happen again or do they want to be friends with the person who did it?); make a list of possible strategies; try the chosen strategy.

Pupils' reported use of the UKRP skills

The quantitative results presented earlier in this report provide our analysis of the impact of the UKRP on pupils. In that analysis, psychometric scales provide particularly important evidence. Qualitative interviews were also carried out with pupils so that they could report examples of how they had used the UKRP skills in real life. This qualitative data thus illustrates some pupils' recall of UKRP skills and provides examples of the type of real life circumstances in which pupils reported that they had applied those skills.

Pupils were asked if they had used any of the things they had learned in UKRP lessons to deal with problems in real life. More than four fifths of the pupils were able to provide an example. Their examples tended to focus on tangible problems, most often instances of avoiding arguments, shouting or having fights. They also included cases of being assertive, negotiating agreements or using relaxation techniques.

For the purposes of presentation, we have grouped the 45 interviewees' examples according to their reason for using the skill. (As pupils used their own words to describe their use of UKRP skills it would be difficult to group them according to, for example, the particular skill used.) Reasons for using the skills and the number of examples provided are shown below:

- making themselves feel better (10)
- 'not rising' to provocation (17)
- using assertiveness and negotiation techniques to address problems (5)
- using techniques to overcome procrastination (2)
- rejecting negative beliefs (4)
- not used any skills (4) or insufficient detail provided (3)

Of the 45 pupils interviewed, 38 were able to provide an example of using a UKRP skill in real life. Some of the more compelling examples appear in the first group of cases listed above. Respondents who described using UKRP skills to make themselves feel better sometimes provided a good deal of detail and some subtlety of understanding. This group comprises about a quarter of interviewees. Their examples included using relaxation exercises, and also the ABC model, to change the way the interviewees felt. More frequently, pupils however described instances of 'not rising' to some form of provocation. These interviewees described circumstances in which they had *not* had an argument or *not* shouted or *not* got into a fight. Almost half of the interviewees' responses were placed in this category. Five interviewees gave examples of using assertiveness and negotiation techniques. Other pupils gave examples of overcoming procrastination or rejecting negative thoughts. Four interviewees reported that they had not used any of the things they had learned during UKRP sessions and a further three were unable to provide sufficient detail for their responses to be placed in the loose categories (e.g. 'I've used it but I can't remember when').

Making themselves feel better

Interviewees who used UKRP skills to 'make themselves feel better' used a range of UKRP skills. For example, some described using the ABC model to challenge beliefs about a situation and thus change the consequent feelings. Relaxation techniques were also reported to have been used to deal with strong emotions. Another interviewee had considered the best, worst and most likely outcomes of a pending event in order to reduce worry and another interviewee described using distraction techniques for the same reason. Six of the ten respondents in this group were girls, and those who gave more expansive answers also tended to be girls.

One interviewee described an episode when her two step-brothers had been allowed to go to a major indoor shopping centre but her mother had not allowed *her* to visit the shops in the city centre. The apparent injustice of this situation made her feel angry:

Well, I was really annoyed with me mam, because I wasn't allowed to go out, but then I just invited my friends round and then we went upstairs and then we went to the [shopping] centre later.

So in what way, how did you use the [UKRP]?

I was like – because I got really annoyed at first – but then I realised the reason why I couldn't, because it was really, really wet and I would get really wet.

And did you consciously, like was there a resilience stage, so go on, what did you think?

So I was like thinking in the first place 'why is she not letting us go' because I had been before but every time when I had been before it was like a nice day. So I thought, 'yes, it is wet' and then I invited my friends round and then we went into the [shopping] centre.

Did you say anything to her or was it just the way you decided to think about it?

The way I decided to think about it

...

Just to get it clear. So if you hadn't have changed it, how would you have felt, do you think?

Really annoyed and like I would have like stayed in my room all day and things.

So instead how did you feel?

I felt good because I was still going off with my friends. (Girl)

While the questioning in this example is somewhat leading, it nevertheless provides an instance of using the ABC schema effectively. Similarly, another girl gave an example of changing the way she felt about an event:

And have you found any of the things that you've learned in [UKRP] lessons helpful in real life?

Yeah, because ABC was quite helpful, when you've got a problem with a friend or want to do something, against your mum or whatever you can use it.

So tell me about the ABC, in what way is it helpful?

You can stress your points without seeming rude or anything, it shows all the different views of the situation, what people can interpret from it all, or people could see everything.

Ahuh, right. And can you give me any examples of how you've used it... in real life?

Well the ABC, my mum was, my birthday is going to be on the day of this workshop I'm doing, which is the Lion King and I didn't know it was going to be on my birthday, so obviously I said, 'oh yes, I'd love to do it' and then now I've just found out and I wasn't sure whether I wanted to do it anymore because it'll be on my birthday. So I said to mum how it's made me feel, but I didn't want to cancel it, because I'm still really eager to do it, so she's seen how I feel so she could sympathise and help me.

Right, ahuh. And how did you present it to her, what did you say?

Yeah, I used the ABC, I think it's adversity, consequence and beliefs and, so then I just stated the problem and then I explained how I felt and how we could make it better.

...

Ahuh, and what would you do instead, if you didn't do the workshop?

Well, I really want to do the workshop now, because I spoke to my mum and she actually made a good point, because I'll get to see loads of friends on my birthday, rather than just having two of my best friends, so actually I'm a lot happier now and I want to do it now, rather than just having my normal birthday.

Okay. That's great. And how, without thinking in terms of the ABC, how do you think you would have thought about it previously?

I could get quite moody! I wouldn't get moody, I don't tend to get angry or anything, but I think I'd be a little bit upset that I couldn't do what I usually do, but now, thinking about it from what my mum's said, I can think I'm going to see loads of my friends, absolutely loads, because there's about 50 of us there, so I get to see 50 of my friends on my birthday rather than seeing about ten at my party. (Girl)

These two examples concern day to day problems but they do show that the interviewees understood how their beliefs could affect their consequent feelings. It is of course important to note that, despite the interviewer asking them to speculate as to how they would have behaved if they had not attended UKRP sessions, it is not possible to know how they would have otherwise thought, felt or acted.

At a school with a very deprived intake, another girl described using relaxation techniques to deal with strong emotions. The strength of those feelings was clear even when they were being described to the interviewer:

Have you found what you've learned in [UKRP] lessons helpful at all?

Yeah.

In what way?

Because when we get angry and everything then I cry loads of times when I'm at home and it sets off my asthma so I do the deep breathing and it helps me calm down and my breathing eases off.

So what kind of relaxation stuff do you do to help you with that?

Well I tense my muscles and relax and do deep breathing.

And is that especially useful when you get angry?

Yeah.

Can you give me an example of that?

When my little brothers and my big brother hit me.

Has that happened recently?

Yep, especially when my brother's yelling at me [*interviewee sounds very upset and sniffs*].

And then he hits you sometimes, what do you do about that?

Well I tell my mum but if she's asleep I can get in quite a big mood and I yell and I end up getting sent to bed and then crying and it sets off my asthma so I do deep breathing most of the time. (Girl)

A boy also gave a real sense of being upset by a teacher shouting at his class and described using distraction techniques to deal with his feelings; and similarly a girl described withdrawing and using relaxation techniques in response to an argument between her mother and brother. Others in this group of respondents described using relaxation techniques in response to worries over homework and using the 'putting it into perspective' skill (thinking of the best, worst and most likely outcomes) to reduce stage fright.

Not rising to provocation

A larger group of interviewees described instances of 'not rising to provocation' to exemplify their use of UKRP skills. In short, they described their socially desirable behaviour in the face of provocation to behave otherwise. Ten of the 17 interviewees providing such examples were boys. For example:

Have you found any of the things that you've learnt in [UKRP] lessons helpful?

Yes.

In what way?

Like when I have arguments with someone and like I don't just like hit them or something, I just say what I think.

So have you used that in real life?

Yes.

Can you give me an example?

Like say I had an argument with my brother or something I'd normally like be aggressive, but now I'm not, I just like say what I think.

...

Right, and so how would you have reacted before?

I would have just like shouted really loud and might have pushed him or something.

Right, and what would you do now?

I just like talk with him, say he says something I'd say "don't say that it's out of order" or something. (Boy)

Another boy gave a more specific example:

Have you used any of the things you've learned in resilience lessons?

Yes.

To help deal with problems in real life?

Yeah.

Can you tell me a bit about that?

I've used, like, being more, say you've got a problem, instead of just shouting it out you calm down yourself and then think of a happy place and then just say it to them.

Can you give me an actual example of when you've done that?

My brother was really annoying me and once he dropped some drink on the stairs and then my mum started shouting at me. And I got really angry because I knew it was him. And I kind of thought, and I went up to my brother and went "Can't you just say it was you?" And I just said it nicely. And he went "Yeah okay". I didn't shout and everything else at my mum.

Or shout at your brother?

No. (Boy)

The detail provided to the interviewer in this example gave a convincing sense that the pupil had applied something he had learned from the UKRP to modify his response to a situation. However this was quite rare among interviewees reporting that they had 'not risen to provocation'. For example, other examples related to incidents that could clearly occur to young people several times each day. In short it is less convincing that these interviewees had actually modified their responses or applied a specific UKRP skill rather than relabelling incidents that would have occurred in any case. For example:

Have you found what you've learned in resilience lessons helpful at all?

Yeah the Hot Seat.

Tell me about that.

It's like you've got a problem and then you just take a couple of minutes out to think about it and what you ought to do, what you want to do about it. So like you're in a mood coz you can't watch the telly and you think about it, what you're going to do. Watch it later or something.

Right and have you used any of the things you learned in real life.

Yeah the hot seat.

So go on, can you say a bit more about how you've used that in real life?

When I was having an argument with my big brother about who was the best football team. He was saying, coz United got beat and I was saying it was United's fault, and he was going "Why is it? It's not just United who get to win anything" and then I just took it out and went "All right" and just walked away.

And what would you normally have done?

Started arguing with him and got in trouble and had a fight with him and - gone mad.

So why did you not?

Because I thought about it and I thought it's not worth it getting in trouble just over a little argument. (Boy)

The minor provocations of siblings and classmates featured frequently in the examples falling in this category. Even though some of these responses had a flavour of social desirability (that is, saying what they thought the interviewer wished to hear), we may at least conclude that interviewees were able to recall some of the things they had learned in resilience sessions, to apply what they had learned to their real life situations and that they chose to present the UKRP as reinforcing socially desirable behaviour:

someone tripped me up and then they tripped [my friend] up and then they spoke to me and then I said 'walk away and ignore them.' (Girl)

like if your parents shout at you... you are still not shouting at them back (Boy)

[my sister] used to take all my sweets and all that and I used to shout at her and say 'what are you taking all my sweets for' and now I just say when she starts taking my sweets, 'because they are mine and not yours', something like that (Girl)

My brother was annoying us... he just wouldn't let us go on the way and after I've done like deep breathing... It just helps you calm down when I get angry (Boy)

If I get into an argument with someone before I used to be arguing and end up fighting but now I walk away or I tell them to leave me alone. I use some of the techniques that [the UKRP facilitator] taught us.

Talk me through it. What would you do?

Well if I was in an argument I would either tell the person to just leave me alone or if they didn't do it and they carried on then I'd just walk away or I'd tell a teacher that they're annoying me.

What would you do previously?

Before, normally I'd probably just carry on arguing and end up in sort of like a big argument and would probably end up fighting. (Girl)

Two of the responses in this group did however suggest that it may be possible to apply UKRP techniques in situations in which they may not necessarily have a positive outcome for the pupil. In one case, a girl described ignoring what she saw as insulting and racist remarks about her hair. Her example may be seen as an effective way to respond to bullying (by ignoring it) although it does illustrate the possibility that there may be circumstances in which not responding to provocation may not be in the best interests of the pupil. Indeed, one interviewee recounted incidents in which his younger sister broke his train set and in which he would think, 'it was bound to happen anyway' and 'I've got lots of money so I can buy another one'. This boy presented these thoughts as a means of avoiding shouting or fighting

with a sibling. However the interviewer was left feeling the responses might be reinforcing the passive responses of a vulnerable boy. That is, it is not enough just to recall and apply what they learned in real life, but also apply it in the right circumstances.

Applying other UKRP skills

Examples of using negotiation and assertiveness techniques, as well as methods to overcome procrastination and reject negative beliefs were reported by other interviewees.

However, as with the example of not rising to provocation described above, another boy illustrated that it may be possible to apply the negotiation skills learned in circumstances that may not result in unequivocally positive outcomes:

Have you found any of the things that you have learned in resilience lessons helpful?

Yes

Yes, in what ways?

Like the negotiation – instead of me always like having a fight with my mam, because I want what I want – I have learned like negotiation, so she gets something and I get something out of it.

Right, so can you give me an example of when you've done that at home?

She was saying to tidy my room and I said 'well can I' – because I had been wanting a new game – so I said, 'can I get this new game and then I'll tidy my room' and she said, 'I will think about it'. So I tidied my room and then she said OK I can get the game. (Boy)

Four other interviewees described using assertiveness and negotiation skills to address problems. For example, two girls described using the DEAL assertiveness model (Describe the problem, Explain how you feel, Ask for a change in behaviour, List the benefits that will follow). They each described using the technique in some detail. In one case the girl assertively asked for her own door key while the other asked her mother to spend some 'quality time' with her. Both girls gave expansive answers indicating accurate recall and suggesting that the skills learned had resulted in a change in their behaviour.

The UKRP teaches a method to overcome procrastination and two girls provided examples of applying this technique. One girl suggested that not all resilience sessions had been useful to her but she had been able to identify the negative role of perfectionism in preventing her from starting homework. The other girl, one of several who struggled to recall the terminology of the programme (for example, confusing pessimism with procrastination), could recall the suggestion of breaking a large task down into small steps with intermediate rewards and had applied this when doing a homework project.

Four pupils reported using different techniques to reject negative beliefs about themselves. For example, one girl described how she had felt stupid after doing badly in a maths test but challenged that belief by considering her achievements in other subjects. Similarly, another girl reported rejecting a negative belief after she had not been selected for the school netball team.

Pupils who had not used the skills or provided insufficient detail

Four pupils did however state that they had not used any of the things they had learned in real life and three more were viewed as providing insufficient detail to be

categorised (of these seven, three were girls and four were boys). One boy openly acknowledged the difficulty he found in applying what he had learned to real life:

OK, have you found any of the things that you have learnt in resilience lessons helpful?

Yes.

Yes, in what way?

Friends and all that, like are more important than everything and that's OK. Like say if I have a problem or anything I have told my mates to get the picture and that's OK... Well that's just most of it and then about self talk and everything.

Right tell me about that, what's self talk?

Self talk is about part of your mind where you are like talking to yourself, blaming yourself.

So can you give me an example of blaming yourself?

Well say like with my Dad with my CD, he starts coming and shouting and leaves. Then I would think 'oh yes, I have done this and all that' you have to think positive as well.

So what could you think in that case more positively?

I honestly don't know. It takes you ages to think what that's like. (Boy)

Participants' reflections on the UKRP

Resilience sessions were among the favourite lessons of most of the pupils interviewed (though it should be noted that pupil interviewees were largely selected by members of the UKRP teaching staff). Reasons given for this popularity included the UKRP's focus on real life, the opportunity it offered for pupils to talk about themselves, and activities such as role play. Perhaps importantly also, sessions did not involve much writing.

Pupils sometimes discussed their own problems during UKRP sessions. We were keen to investigate whether pupils felt this had been a positive or negative experience. Importantly, no pupils who were interviewed suggested there had been negative experiences arising from such discussions – for example, due to bullying or teasing resulting from the issues discussed. Indeed, this was reinforced by the interviews with UKRP facilitators, several of whom commented on how sensitive and supportive pupils had been of each other when serious personal issues were raised.

Pupils could, on occasion, raise very serious issues during UKRP sessions and these included, for example, bereavement, family illness, domestic violence or being taken into care. While not all facilitators reported pupils sharing such personal material, the role of facilitator could clearly be very emotionally demanding if pupils felt able to open up.

On rare occasions facilitators needed to notify school child protection officers of information that had been shared during UKRP sessions. One interviewee referred to such episodes in this way:

I said if they said any more I would have to break confidentiality and why don't you stay behind at the end and talk to me about it if you want to.

And what did they do [on each occasion]?

They all stayed and I would write it all down and took it to the child protection officer or head of year.

So they didn't then go on and share what they were going to say in class?

No.

Was that OK?

I would say there was always a slight feeling of shellshock within the class after they'd said... my parents were fighting... One boy said my parents were fighting and it got really bad and my dad got really drunk and I could hear him so I went downstairs and he hit mum and I said 'right, stop there and we'll talk about it separately.' It's kind of not appropriate that everyone in the class knows...

Did it have any effect on the teaching?

Yes, it was hard after something like that to get back into the skill we were looking at that day. And this sounds really weird [laughing] but I'm a Christian and I always wanted to say a prayer, I always thought it would be nice if we had some kind of acknowledgement of it, but you can't, obviously. That was difficult, it was almost like right, we'll stop now and we'll move on to something else and now we're going to be more cheery.

This quotation perhaps also illustrates the sense of powerlessness that facilitators could experience when faced with genuinely distressing problems. Similarly, another facilitator described the satisfaction felt by being a confidante of pupils but also the sense of guilt when feeling unable to 'really help', as she put it. This sense was aroused particularly strongly when a recently bereaved pupil would cry during sessions each time pupils were requested to think of a problem. This led the facilitator to wonder whether the sessions were having a negative impact for that particular pupil.

Facilitators were drawn from both the teaching staff and from non-teachers – for example, learning mentors, teaching assistants and school nurses. As one interviewee remarked of the skills required of a facilitator:

Non-teachers have to know they're comfortable leading a group of 15 and teachers have to know they're comfortable with the 'touchy feely'.

There was some evidence that the different professional backgrounds of facilitators affected their views of the training that they received. In particular, it was quite common for non-teachers to complete the training feeling daunted by the prospect of leading a group of up to 15 pupils, and several commented on the value of practising leading sessions during the training events.

How well do you think the training prepared you for delivering the Resilience Programme?

It – and I'm only a learning support assistant, so I'm not a teacher – so it was very, very daunting... I didn't know what an hour's teaching... would mean, So that was very, very scary. That was very, very daunting.

While such comments were quite common among UKRP facilitators who were not teachers, it was also quite common for them to remark on how those fears had largely dissipated when they had begun to deliver the programme to pupils.

In contrast, some teachers commented favourably on learning about the psychological theory underlying the resilience programme (and also applying it to their own lives) but found sessions in which they practiced delivering UKRP sessions less valuable.

Nevertheless, overall the training was viewed extremely positively. The most extreme positive endorsement of the training was given by an interviewee who remarked:

I'm nearly 65 and I've had, in my life, one or two life changing moments. The resilience training was one of those for me.

There was a suggestion that part of the power of the training experience, among those attending the first training event, arose from it being an intensive, residential course (unlike that experienced by the children). In the first year, facilitators were trained in the United States and therefore the training experience had been something of an adventure for the participants. There was a fervour and excitement about the programme that was discernible among some of the first cohort of interviewees but this was less obvious among interviewees from subsequent cohorts. Nevertheless, even those who had attended subsequent training events in the UK, either on a residential or non-residential basis, spoke very highly of the quality of the training. Encouragingly, most of the facilitators interviewed reported using the UKRP skills which in itself indicates a belief in the value of the techniques taught.

While facilitators were very positive about the training they had received a number expressed some reservations about the course materials provided for the UKRP. The course materials include a heavily scripted manual. In particular, facilitators felt that lessons were too didactic:

I felt there was a lot of talking... kids can cope with a few minutes but then they need to be doing something...

My biggest concern about the course: it's too talk intensive.

There is too much standing talking to the kids and you've lost them, they switch off.

In addition, facilitators suggested that the materials were too American and this could create a barrier for pupils.

Two senior managers observed that the prescriptive lesson plans and heavy scripting were at odds with the style of lessons expected by Ofsted. By the third year of the programme, some schools and facilitators had made adjustments to the teaching materials, including replacing some of the teaching materials with resources from alternative sources. However, interviewees generally recognised the importance of sticking to the programme materials as the evidence suggesting that the course could have a positive impact was only based on lessons being taught as set out in the manual. As one interviewee put it:

[It's] not something we want to mess around with... because my concern is that none of us are trained psychologists or anything like that... we're not qualified or in a position to start altering the content of something that is clearly long established and obviously works.

However, there is clearly a danger that programme fidelity (and therefore possibly also impact) may be threatened if facilitators believe that the teaching materials are not of a sufficiently high quality or suitable for British pupils.

Organisation of the UKRP within schools

Some information about the nine schools that were visited in 2009-10 is shown in Table 25. In our second interim report (Challen et al., 2010) we suggested that, for six of the nine schools, the UKRP could be described as being embedded in the school. In each of these schools it was delivered to 100% of Year 7 pupils, it had been delivered in each of the three previous years and also had a designated home within the curriculum and timetable. One other school was delivering the UKRP to all Year 7 pupils although the programme's position was a little more precarious; it had not been delivered at all in the second year of the UKRP due to staffing problems. Schools in which the programme was most obviously thriving had larger numbers of staff trained to be facilitators and had integrated the programme into the school curriculum. Schools' commitment to the programme could be reflected in a variety of ways. For example, it was reflected in efforts to improve the quality of the teaching materials, in vetting the suitability of facilitators to deliver the course *after* they had received UKRP training (e.g. managers judging whether they had demonstrated empathy and understanding of the course), or in developing monitoring and evaluation procedures within the school.

In one of the schools the UKRP had been discontinued. In the other school, at the time of the case study visit, delivery in the third year had been postponed and, according to a senior manager, it was likely to be discontinued. The schools in which the UKRP was thriving were not necessarily those that had made the most enthusiastic start to delivering the programme in its first year. However, it does seem clear that, to thrive, the programme required continuing management support and a sufficiently large number of trained facilitators to make it sustainable.

Delivery of the UKRP was organised in different ways in the case study schools. The developers of the UKRP require that the programme is only delivered by trained facilitators. They also recommend that the programme is followed by groups of 15 or fewer pupils over 18 one hour sessions. These stipulations introduce additional constraints into school timetables and, in both the first and third years of operation, schools varied in their capacity to accommodate these requirements.

As was noted earlier, case study schools were selected to include schools delivering the UKRP to different proportions of Year 7 in the first year of the programme. In that year four of the ten schools visited delivered the UKRP to all Year 7 pupils while at the other six schools between 16% and 50% of Year 7 pupils followed the course. Two years later nine of the case study schools were revisited. The number of schools delivering the programme to 100% of pupils had thus risen from four to seven.

Some schools with timetables operating on a fortnightly cycle offered one UKRP lesson per fortnight (thus equating to approximately 18 hours of study over the full academic year). However, in all such schools facilitators believed this mode of delivery left too large a gap between UKRP sessions. That is, facilitators reported that, with infrequent contact, pupils were unable to remember what they had learned from one session to the next, which could be further exacerbated by missed lessons due to holidays, illness or days off the timetable, and that this also made it more difficult to build relationships with pupils.

We might therefore regard the preferred implementation model to be delivery of the UKRP in eighteen weekly sessions, to groups of no more than 15 pupils. However only two of the seven schools offering the UKRP in autumn 2009 were able to

Table 25: Information about the nine case study schools revisited in 2009-10

	Proportion of Year 7 cohort receiving UKRP in 2009-10	Total number of staff trained to deliver UKRP (number of whom not working at the school in 2009-10) ⁴⁶	Of whom		Timetable slot	Lessons delivered fortnightly or weekly	Statistically significant (+ or -) CVA scores during the three years prior to Autumn 2009
			Teachers	Non-teachers			
Embedded, predominantly teacher-led programme	100%	10	6	4	UKRP named slot	Fortnightly	+ +
	100%	17	14	3	Part of English programme	Weekly	+ +
	100%	10 (2)	6	2	Part of PSHE (with SEAL and other elements)	Weekly	+ +
Embedded, support staff-led	100%	8	3	5	Part of PSHE (with Thinking Skills and SEAL)	Weekly	-
	100%	10	3	7	Part of PSHE	Weekly	+
	100%	7	2	5	UKRP	Fortnightly	--
UKRP not delivered in all three years	100%	5 (2)	1	2	UKRP	Fortnightly	+ +
	Two classes though yet to begin	7 (1)	2	4	Part of PSHE	Fortnightly	+ +
	Discontinued	3	1	2	Not applicable	Not applicable	--

⁴⁶ Not all trained facilitators were leading UKRP groups in 2009-10.

accommodate such a model in their timetable in 2009-10. That is, most of the schools continuing to deliver the UKRP had either increased the group size, reduced the number of UKRP sessions or were delivering sessions on a fortnightly basis.

Clearly, key organisational decisions that had to be made by schools related to which members of staff would deliver the UKRP and how the programme would be accommodated in the timetable.

Recruitment of facilitators

Methods for recruiting staff to undertake training to deliver the UKRP varied across the schools. At some schools senior managers approached particular members of staff. As one interviewee put it:

I approached people I thought were suitable and they were pleased to be asked.

Other schools operated an application process open to all staff. At one school applicants were then interviewed while at another training places were allocated through a random ballot. In all schools, the other responsibilities of UKRP facilitators could however impose constraints when attempting to timetable the programme.

In three of the schools the programme was delivered, in its third year, primarily by members of the teaching staff while in the other schools it was delivered predominantly by auxiliary staff members (e.g. teaching assistants, learning mentors). Importantly, each of the schools that had persisted with a teacher-led programme had achieved statistically significant positive CVA scores in two of the previous three years. That is, the schools may have been sufficiently confident in the quality of teaching and learning at the school that teachers were able to be deployed delivering the UKRP.

Indeed, at the school in which the UKRP had been discontinued and where levels of attainment were low, the key facilitator had been taken off the UKRP team in order to devote more time to teaching English. (Importantly, the school was under substantial pressure to improve levels of attainment with the possible threat of closure if it should fail. The key facilitator had been promoted to become head of English.) Other factors were also reported to have contributed to the demise of the programme, including the departure of the headteacher and even the school moving into a new building with open plan classrooms. In this school, as in others, it was clear that management backing was essential to the survival and success of the UKRP.

A teacher from another school commented on the pressures teachers could face arising from the drive to improve attainment and the problems this could pose for the UKRP:

It's very difficult for teachers because, as you can see, we've kind of been pushed out of the UKRP because our timetables are, the commitments are quite big elsewhere in terms of exam results and things like that, especially core subjects like myself. So I would say that... it's all non-teaching staff who are doing it now.

At the other schools in its third year, the UKRP tended to be delivered by non-teachers and new UKRP trainees tended to be drawn from the auxiliary staff. Interviewees suggested several reasons why this was the case in addition to the pressure on attainment noted above. One attraction of using non-teachers to deliver the UKRP arose from the difficulty of timetabling two teachers to teach the same class (that is, with a preferred group size of 15, ordinarily a class would be split in half for UKRP sessions). In contrast, non-teachers who frequently did not have a full timetable offered greater flexibility to managers. Another reason for the increase in participation of non-teachers arose from the different contractual

arrangements. That is, the training events took place, at least in part, during school vacations. Whereas auxiliary staff were frequently employed on term-time contracts (and therefore work during the holidays attracted additional pay), for teachers this was not the case. In addition it was suggested that, for teaching assistants who wished to go on to train to become teachers, the UKRP offered an 'ideal stepping stone'.

Fitting the UKRP into the curriculum

As well as deciding which staff should receive UKRP training, schools also had to find a place for the UKRP in the school timetable. For schools timetabling UKRP as a separate subject this offered the programme a degree of security (as it was not dependent on its fit with another subject area) but usually also implied that pupils would receive lessons once a fortnight. That is, as the UKRP was designed to be delivered in 18 one hour sessions, with one lesson per fortnight the UKRP would be spread across the whole school year. An alternative was however to offer the UKRP on a carousel with another subject although this was of course dependent on fitting it in with other course units of appropriate length. In one school this resulted in truncating the UKRP in order to deliver it in one term while in another the carousel created some conflict with the school's theme based key stage 3 curriculum.

In other schools the UKRP had been incorporated into an existing wider programme of PSHE (sometimes incorporating Personal Learning and Thinking Skills⁴⁷ and or elements of the Social and Emotional Aspects of Learning materials⁴⁸).

One school found a particularly interesting solution by having all members of the English department trained to deliver the UKRP. The programme was then incorporated into English lessons. The head of department felt it offered a good fit with the speaking and listening elements of the curriculum. In addition, this arrangement offered other advantages. For example, by delivering the UKRP to all pupils at the same time, pupils could move between English sets without needing to disrupt their UKRP groups. In addition if offered the UKRP a secure curriculum home for future years and, if required, would enable the programme to be delivered as a more intensive block rather than as weekly one hour sessions. It is also notable that the headteacher of this school had attended the UKRP training and, although the headteacher had not yet called upon to deliver sessions to pupils, this demonstrated the level of management backing for the programme.

Conclusion

In this section we have reported our findings drawing on the qualitative data collected in ten case study schools. Visits were made to these schools in the first year that the UKRP was delivered (2007-08) and again in the third year (2009-10).

Many of the pupils interviewed during the 2007-08 session reported that they had used some of the UKRP skills in real life. Some pupils showed a fairly sophisticated level of understanding when they applied the UKRP skills to their own experiences. Most numerous were pupils who described instances when they had 'not risen' to some form of provocation. Many, although not all, of these responses were somewhat sketchy. Other interviewees described using the ABC model and other cognitive skills, as well as behavioural skills such as assertiveness and negotiation. Examples provided by pupils tended to focus on day to

⁴⁷ Personal Learning and Thinking Skills (PLTS) were introduced into the new Key Stage 3 curriculum in 2008.

⁴⁸ The Social and Emotional Aspects of Learning teaching materials were developed for the DCSF as part of its National Strategies.

day problems such as conflict with siblings. Pupils frequently reported that UKRP sessions were among their favourite lessons.

In both years, facilitators who were interviewed tended to be positive about the objectives of the programme and also about their experiences of delivering it. They spoke particularly positively about the quality of the training they had received. However, in both years they were less positive about the course materials. In particular, the course was thought to involve too much 'teacher talk' and many thought it would benefit from a greater range of activities in which pupils could participate.

Seven of the nine schools that were revisited in 2009-10 were continuing to deliver the UKRP to new cohorts of Year 7 pupils. Perhaps this is the best indicator of schools' overall satisfaction with the operation of the programme within schools. In these seven schools, the UKRP was delivered to all Year 7 pupils and more members of staff had been trained as facilitators. In some schools interviewees remarked that there had been a shift from the programme being delivered by teachers to it being delivered by members of the auxiliary staff such as teaching assistants and learning mentors. This provided greater flexibility in timetabling the UKRP. It was suggested that this drift may have resulted from pressure on teachers to focus on attainment and also pay, workload and career development considerations. Schools that continued to offer the UKRP primarily through sessions facilitated by teachers tended to have good track records in pupil attainment. This may have provided the confidence to allow teachers to devote time and energy to the UKRP. The level of demand from members of staff to undertake UKRP training events varied between schools.

Schools accommodated the UKRP within the curriculum and timetable in different ways. Most often the UKRP was incorporated into an existing subject area such as PSHE or in one school as part of the English curriculum. In some schools UKRP operated as a separate subject, most often timetabled for one lesson a fortnight throughout the school year. This offered the advantage of the programme not being dependent on fitting in with other subjects or competing for space with other course units, although a fortnightly delivery model was generally unpopular among facilitators. No single model for timetabling the UKRP would fit all schools. If we consider the ideal mode of delivery to be eighteen weekly sessions for groups of not more than 15 pupils then only two out of the nine schools met these criteria in the third year of delivery.

It is perhaps useful to conclude by drawing attention to some of the lessons that may be drawn for practice from the qualitative element of this study.

- 1) Pupils were able to recall some of the material they had learned in UKRP sessions and some were able to describe episodes in which they had applied UKRP skills to real life events.
- 2) For the UKRP to thrive within schools it is extremely important that the programme has backing from the school's senior management.
- 3) This backing is all the more important when schools face competing pressures such as the need to improve standards of attainment. This may also prove to be the case in relation to financial pressures on schools, for example, if there is a deterioration in pupil: adult ratios.
- 4) The role of facilitator can be emotionally demanding due to the distressing nature of some real life problems raised by pupils.
- 5) There was evidence of a drift, in some schools, towards the programme being delivered by members of the school auxiliary staff. This will clearly reduce the size of

the pool from which facilitators may be drawn which would have an impact on the quality of staff who may train as facilitators.

- 6) Facilitators were very positive about the quality of training they had received for the UKRP. They had reservations however about the quality of teaching materials provided for the programme. If the materials are not regarded as being of a sufficiently high quality, facilitators may seek alternative resources and clearly this may constitute a threat to programme fidelity.
- 7) While our quantitative findings suggest there was initially a statistically significant gain in the mental health and well-being of pupils, and many interviewees believed the programme was having a positive impact on their pupils, schools and facilitators should keep in mind the possibility that the programme could have a negative effect for individual pupils.

7. Appendix A: Questionnaires used

Questionnaires used to survey treatment and control pupils at all data collection points

Outcome: Symptoms of depression

Children's Depression Inventory

UK supplier: Harcourt Assessment

UK copyright holder: Multi-Health Systems Inc.

The version used excludes item 9, as this concerns suicidal ideation and this was not deemed appropriate or necessary.

Manual: CDI Technical Manual, Maria Kovacs 2003, pub. Multi-Health Systems

Outcome: Symptoms of anxiety

Revised Children's Manifest Anxiety Scale

UK copyright holder: Western Psychological Services

The wording of some items was modified slightly for UK English.

Manual: Revised Children's Manifest Anxiety Scale [RCMAS], Cecil R. Reynolds and Bert O. Richmond 2000, pub. Western Psychological Services

Outcome: Life satisfaction

Brief Multidimensional Students' Life Satisfaction Scale

This is not copyrighted, and is available online at:

<http://www.cas.sc.edu/psyc/facdocs/hueblifesat.html>

Scoring instructions are also available here.

Outcome: Behaviour

Goodman Strengths and Difficulties Questionnaire

(1) Self-report version

(2) Teacher-report version

All versions of the Goodman SDQ are available online at:

<http://www.sdqinfo.com/>

The Goodman SDQ can be used free of charge, although it is not possible to modify the wording. For the purposes of evaluating an intervention there is one version (an 'initial' version) that has a reference period of six months, and a follow-up version with a reference period of one month. All data collections subsequent to the baseline use the one-month reference period. When the follow-up version is used it is acceptable to change the word 'clinic' to 'classes' or 'programme', and this has been done, in order to clearly refer to the UKRP. Modifications to the format are acceptable provided the layout is essentially the same.

Scoring details and references are available at the website listed above.

8. Appendix B: References for Empirical Evaluations of PRP and UKRP

- Brunwasser, S. M., J. E. Gillham and E. S. Kim, "A Meta-Analytic Review of the Penn Resiliency Program's Effect on Depressive Symptoms", *Journal of Consulting and Clinical Psychology*, 77(6), 1042-1054
- Cardemil, E., Reivich, K., Beevers, C.G., Seligman, M.E.P., & James, J. (2007). The prevention of depressive symptoms in low-income, minority children: Two-year follow-up. *Behaviour Research and Therapy*, 45, 313-327.
- Cardemil, E.V., Reivich, K.J., & Seligman, M.E.P. (2002). The prevention of depressive symptoms in low-income minority middle school students. *Prevention & Treatment*, 5, np.
- Challen, A. R., S. J. Machin, et al. (2009). UK Resilience Programme Evaluation: Interim Report, Department for Children, Schools and Families.
<http://www.education.gov.uk/publications/eOrderingDownload/DCSF-RR094.pdf>
- Challen, A. R., S. J. Machin, et al. (2010). UK Resilience Programme Evaluation: Second Interim Report, Department for Education.
<http://www.education.gov.uk/publications/eOrderingDownload/DFE-RR006.pdf>
- Chaplin, T.M., Gillham, J.E., Reivich, K., Elkon, A.G.L., Samuels, B., Freres, D.R., Winder, B., & Seligman, M.E.P. (2006). Depression prevention for early adolescent girls: A pilot study of all-girls verses co-ed groups. *Journal of Early Adolescence*, 26, 110-126.
- Cutuli, J.J. (2004). Preventing externalizing symptoms and related features in adolescence. Unpublished honors thesis, University of Pennsylvania, Philadelphia, PA.
- Cutuli, J.J., Chaplin, T.M., Gillham, J.E., Reivich, K., & Seligman, M.E.P. (2007). Preventing externalizing symptoms and related features in adolescents. Manuscript in preparation.
- Cutuli, J.J., Chaplin, T.M., Gillham, J.E., Reivich, K.J., & Seligman, M.E.P. (2006). Preventing co-occurring depression symptoms in adolescents with conduct problems: The Penn Resiliency Program. *New York Academy of Sciences*, 1094, 282-286.
- Gillham, J.E. (1994). Preventing depressive symptoms in school children. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.
- Gillham, J.E., Hamilton, J., Freres, D.R., Patton, K., & Gallop, R. (2006). Preventing depression among early adolescents in the primary care setting: A randomized controlled study of the Penn Resiliency Program. *Journal of Abnormal Child Psychology*, 34, 203-219.
- Gillham, J.E., & Reivich, K.J. (1999). Prevention of depressive symptoms in school children: A research update. *Psychological Science*, 10, 461-462.
- Gillham, J.E., Reivich, K.J., Freres, D.R., Chaplin, T.M., Shatté, A.J., Samuels, B., Elkon, A.G.L., Litzinger, S., Lascher, M., Gallop, R., & Seligman, M.E.P. (2007). School-based prevention of depressive symptoms: A randomized controlled study of the effectiveness and specificity of the Penn Resiliency Program. *Journal of Consulting and Clinical Psychology*, 75, 9-19.

Gillham, J.E., Reivich, K.J., Freres, D.R., Lascher, M., Litzinger, S., Shatté, A., & Seligman, M.E.P. (2006). School-based prevention of depression and anxiety symptoms in early adolescence: A pilot of a parent intervention component. *School Psychology Quarterly*, 21, 323-348.

Gillham, J.E., Reivich, K.J., Jaycox, L.H., & Seligman, M.E.P. (1995). Prevention of depressive symptoms in schoolchildren: Two-year follow-up. *Psychological Science*, 6, 343-351.

Jaycox, L.H., Reivich, K.J., Gillham, J., & Seligman, M.E.P. (1994). Prevention of depressive symptoms in school children. *Behaviour Research & Therapy*, 32, 801-816.

Miller, J.B. (1999). The effect of a cognitive-behavioral group intervention on depressive symptoms in an incarcerated adolescent delinquent population (juvenile delinquents). Unpublished doctoral dissertation, Wright Institute Graduate School of Psychology, Berkeley.

Pattison, C., & Lynd-Stevenson, R.M. (2001). The prevention of depressive symptoms in children: The immediate and long-term outcomes of a school based program. *Behaviour Change*, 18, 92-102.

Quayle, D., Dziurawiec, S., Roberts, C., Kane, R., & Ebsworthy, G. (2001). The effect of an optimism and lifeskills program on depressive symptoms in preadolescence. *Behaviour Change*, 18, 194-203.

Reivich, K.J. (1996). *The prevention of depressive symptoms in adolescents*. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.

Roberts, C., Kane, R., Bishop, B., Matthews, H. & Thompson, H. (2004). The prevention of depressive symptoms in rural children: A follow-up study. *International Journal of Mental Health Promotion*, 6, 4-16.

Roberts, C., Kane, R., Thomson, H., Bishop, B., & Hart, B. (2003). The prevention of depressive symptoms in rural school children: A randomized controlled trial. *Journal of Consulting and Clinical Psychology*, 71, 622-628.

Shatté, A.J. (1997). Prevention of depressive symptoms in adolescents: Issues of dissemination and mechanisms of change. Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia.

Yu, D.L., & Seligman, M.E.P. (2002). Preventing depressive symptoms in Chinese children. *Prevention & Treatment*, 5, np.

Zubernis, L.S., Cassidy, K.W., Gillham, J.E., Reivich, K.J., & Jaycox, L.H. (1999). Prevention of depressive symptoms in preadolescent children of divorce. *Journal of Divorce and Remarriage*, 30, 11-36.

Ref: DFE-RR097

ISBN: 978-1-84775-883-5

© Enterprise LSE Ltd

April 2011