



AN EVALUATION OF 'PARENTS UNDER PRESSURE'

A PARENTING PROGRAMME FOR
MOTHERS AND FATHERS WHO
MISUSE SUBSTANCES

TECHNICAL REPORT

Vicki Hollis, Richard Cotmore, Helen L. Fisher, Paul Harnett & Sharon Dawe

March 2018

NSPCC

EVERY CHILDHOOD IS WORTH FIGHTING FOR

Impact and Evidence series

This report is part of the NSPCC's Impact and Evidence series, which presents the findings of the Society's research into its services and interventions. Many of the reports are produced by the NSPCC's Evaluation department, but some are written by other organisations commissioned by the Society to carry out research on its behalf. The aim of the series is to contribute to the evidence base of what works in preventing cruelty to children and in reducing the harm it causes when abuse does happen.

Contents

Table 1: Measures used in the PuP programme and for the PuP evaluation	7
Evaluation sample	12
Table 2: Comparison of parents who did and did not consent to the NSPCC evaluation based on their demographic data at T1	13
Table 3: Comparison of parents who did and did not consent to the NSPCC evaluation based on their measures data at T1	14
Evaluation attrition	15
Data analysis	15
Findings	16
Question 1: What is the demographic profile and the needs of the parents and children who access this service?	16
Table 4: Characteristics of the parents being accepted onto the PuP programme at T1 based on the categorisation of their scores on the individual PuP measures	16
Table 5: Parents' reports of their child's social-emotional and behavioural problems or deficits/delays in social-emotional competence at the start of the PuP programme. This is based on the categorisation of scores from the two BITSEA subscales	18
Question 2: Are there any differences between the parents (and their children) that start, but do not complete, PuP compared with those who complete the programme?	19
Table 6: The proportion of parents completing PuP and the reasons for non-completion (based on the 166 parents taking part in the evaluation)	19
Table 7: Demographic profile at T1 of the parents completing PuP compared with those who did not complete PuP	21
Table 8: Characteristics of the parents at T1 who did complete ('completers') and did not complete ('non-completers') PuP based on the categorisation of their scores on the PuP measures	22
Table 9: Statistical comparison of the scores on the PuP measures at T1 for the parents who did complete ('completers') and did not complete ('non-completers') PuP	23
Table 10: Parents' reports of their child's social-emotional and behavioural problems and deficits/delays based on the categorisation of BITSEA scores – comparison of parents who did and did not complete PuP	24

Table 11: Statistical comparison of the scores on the BITSEA for the children of the parents who did complete ('completers') and did not complete ('non-completers') PuP	25
Table 12: Logistic regression analysis exploring predictors of programme completion	25
Question 3: Is there evidence of change for the parents and children who complete the programme? How does change occur over time?	26
Figure 1: Change in parents' median scores for each measure across all three time-points. Median scores are shown only for parents who completed the programme and completed the measure at every time-point (number of parents is reported in each figure)	27
Change in parents' median scores for their child's social-emotional and behavioural problems and competencies. Median scores are shown only for parents who completed the programme and completed the measure at every time-point (n=17).	29
Table 13: Change in parents' scores on each measure across all three time-points. Data is only for parents who completed PuP and completed the measure at every time-point	30
Table 14: Post-hoc analysis of change across all three time-points to identify when in the programme significant change occurred. Tests were carried out only for parents who had completed the measure at every time-point	31
Table 15: Correlation between change in DASS21, IMP-I and BITSEA scores between T1 and T3	32
Figure 2a: Change in depression (n=85)	33
Figure 2b: Change in anxiety (n=85)	34
Figure 2c: Change in stress (n=85)	34
Figure 2d: Change in IMP-I (n=79)	35
Figure 2e: Change in AUDIT-C (n=62)	35
Figure 2f: Change in MSPSS (n=84)	36
Figure 2g: Change in friend support (n=84)	36
Figure 2h: Change in significant other support (n=84)	37
Figure 2i: Change in total support (n=84)	38
Figure 2j: Change in BITSEA (n=17)	38
Table 16: Parents' reliable change on the DASS21	40
Table 17: Parents' reliable change on the IMP-I	40
Table 18: Parents' reliable change on the BITSEA	41
Table 19: Reliable and clinical change in DASS21 and BITSEA scores from the start to the end of PuP	42
Table 20: Reliable change in categorical classification of mindful parenting during the PuP programme	43

Question 4. What factors are associated with the change experienced by parents and children during PuP?	44
Table 21: Factors associated with change in DASS21 subscale scores between T1 and T3 for the parents who showed reliable change for each sub-section	45
Table 22: Factors associated with change in IMP-I scores between T1 and T3 for the parents who showed reliable change for each sub-section	47
Question 5. Does the safeguarding status of children change during PuP and are there any predictors of change?	49
Table 23: Ranking of safeguarding status	49
Table 24: Factors associated with change in safeguarding status during PuP	50
Table 25: Change in safeguarding status for the parents who did (n=96) and did not (n=47) complete PuP and for whom we could assess change	50
Question 6: What do we know about the fathers enrolled in PuP?	51
Table 26: Categorisation of fathers' scores for emotional wellbeing, alcohol consumption and social support at T1 (n=10)	51
Conclusion	52
References	53

AN EVALUATION OF 'PARENTS UNDER PRESSURE'

This technical report provides additional detail and information on the evaluation methods, analysis and findings from the NSPCC's evaluation of the Parents Under Pressure (PuP) programme. It should be read alongside the main evaluation report found here <http://library.nspcc.org.uk/HeritageScripts/Hapi.dll/search2?searchTerm0=C7031>

Table 1: Measures used in the PuP programme and for the PuP evaluation

Measure	Description	Reliability and Validity	Scoring and clinical classification	Number who completed measure ¹		
				T1	T2	T3
Child measures						
Brief Infant-Toddler Social and Emotional Assessment (BITSEA) ² (Briggs-Gowan et al, 2004)	A 42-item parent report of social-emotional and behavioural problems and social-emotional competence in their infants (aged 1–3 years).	The BITSEA has been found to be a valid and reliable measure of children’s social-emotional problems and competencies.	Scored using the normed clinical cut-off scores outlined in the BITSEA manual. Scoring of the measure provides a total problem and total competence score. These scores can be used to determine whether the child falls within the top 25th percentile (problem subscale) or 15th percentile (competence subscale), indicating raw scores that are obtained by very few of the children in the normative nonclinical sample. If their scores fall within this range, the child is showing the presence of an unusual number of problems and/or deficits/delays in competence. Clinical classification is based on the age (adjusted for prematurity) and gender of the child.	51	25	26

1 These are not the final numbers for the analysis of change in parents’ scores as not all parents completed the measure at all time-points.

2 The BITSEA was completed for six additional children but this data could not be included in the final analysis as it was completed for children who did not fall within the valid age range (1–3 years) for the measure.

Measure	Description	Reliability and Validity	Scoring and clinical classification	Number who completed measure ¹		
				T1	T2	T3
Parent measures						
Depression, Anxiety and Stress Scale (DASS21) (Lovibond & Lovibond, 1995)	Self-report questionnaire that measures three related negative emotional states: depression, anxiety and tension/stress. Parents are asked to use a 4-point severity/frequency scale to rate the extent to which they have experienced each state over the past week. Scores can be used to determine whether the parent has normal, low, moderate, severe, or extremely severe difficulties in each of these three areas, along with a total score. This evaluation used the 21-item short-version of the original DASS (the DASS21), which has 7 items per scale.	The DASS21 is well validated as a good measure of depression, anxiety and stress.	Normed cut-off scores outlined in the DASS21 manual are used. In this evaluation, we group the clinical severity of symptoms into three categories: normal/ low, moderate/severe, and extremely severe. This is to assist analysis and interpretation.	164	111	92
Alcohol Use Disorders Identification (AUDIT-C) (Meneses-Gaya et al, 2010)	A brief self-report screening test for heavy drinking and/or active alcohol abuse or dependence. Parents are asked to use a 5-point scale to provide some indication of their recent drinking behaviours during the past six months (T1 assessment only) or since the last assessment (T2 and T3). Scores on the measure indicate the level of risk posed by alcohol consumption and can be categorised as: safe levels of alcohol consumption; alcohol abuse; and alcohol dependence. The AUDIT-C comprises the first three questions from the full 10-item version of the AUDIT.	This 3-item shorter version of the AUDIT has been found to be reliable and sensitive to the presence of alcohol problems.	Cut-off scores taken from Meneses-Gaya, et al (2010).	140	86	77

Measure	Description	Reliability and Validity	Scoring and clinical classification	Number who completed measure ¹		
				T1	T2	T3
Daily Parenting Hassles Scale	Parents report on the presence of 10 typical daily hassles/ challenges for families, rating on a scale of 0–4 whether they are not a hassle at all or are a huge hassle. This measure was adapted from the Crnic & Greenberg (1990) 'Parenting daily hassles scale' by the PuP programme developers for use in this service.	This measure has been found to have good internal consistency ($\alpha = .89$) in a sample of 2,971 parents with substance misuse who have accessed the PuP service (mean = 10.29, SD = 9.02; Unpublished data).	Mean scores are used.	154	103	84
Life Events Scale	Parents report on whether they have experienced eight critical life events within the past six months. Events are rated on a scale of 0–2 to indicate whether the event happened and, if so, if they are coping OK or not coping. This measure was only used during the PuP assessment (T1) to explore the parents' experiences prior to PuP. The measure was designed by the programme developers for use in PuP and is based on the 'Recent life events' questionnaire used in the Department of Health's (2000) <i>Framework for the assessment of children in need and their families</i> .	This measure has approached acceptable internal consistency ($\alpha = .69$) in a sample of 2,971 parents with substance misuse who have accessed the PuP service (mean = 5.26, SD = 3.03; unpublished data). However, it would not necessarily be expected that there would be a high correlation between these life events as they refer to discrete experiences.	The total number of events experienced by parents is explored and described in this evaluation, along with a description as to how parents reported coping with these events.	154		
Multidimensional Scale of Perceived Social Support (MSPSS) (Zimet et al, 1988)	Parent-report 12-item measure of how much support the parent feels they receive. There are three subscales measuring support from family, friends and significant others (as defined by the parent). A total support score is also calculated.	Found to be a reliable and valid measure of perceived social support.	There are no population norms for this measure but average scores, along with their standard deviations (SD), have been published based on a sample of college undergraduates (Zimet et al, 1988). For the purposes of this study, we have therefore classified levels of support into high (<1 SD below mean), low (>2 SD below mean) and mid-range (1–2 SD below mean).	166	111	91

Measure	Description	Reliability and Validity	Scoring and clinical classification	Number who completed measure ¹		
				T1	T2	T3
Parenting Support Scale	A four-item self-report questionnaire that measures actual levels of practical and emotional support for the day-to-day tasks of being a parent against ideal levels of practical and emotional support. A score is given to determine the discrepancy between actual and ideal levels. The measure was designed by the PuP programme developers based on the 'Significant Other Scale' (Power et al, 1988).	This measure has not been tested for reliability and validity.	Mean actual and ideal parenting support scores and discrepancy scores are used in this evaluation.	166	111	91
Interpersonal Mindfulness in Parenting – Infant Version (IMP-I). (Laurent et al, 2017)	A 27-item self-report questionnaire that assesses the parent's ability to reflect on their own emotional state, to manage their emotions and to identify and respond to their baby/child's emotional state. The scale has been designed to produce a total score.	The infant version of the interpersonal mindful parenting questionnaire has not yet been validated on a large population sample. However, the total score has been found to have good internal consistency in the 2017 paper by the author and was also found to be good in the current PuP evaluation (a = .86).	There are no population norms or classification cut-off scores for this measure. For the purposes of quantifying the amount of change made by parents throughout this programme, three cut-off scores have been defined to classify the total mindful parenting score: high in mindful parenting (>1.5 SD above the mean); moderate level of mindful parenting (-1.5 to +1.5 SD of the mean); and low with mindful parenting (>-1.5 SD below the mean).	157	106	90

Measure	Description	Reliability and Validity	Scoring and clinical classification	Number who completed measure ¹		
				T1	T2	T3
HITS scale (Sherin et al, 1998)	A four-item, self-report questionnaire asking questions about the frequency a partner physically hurts, insults, threatens or screams at the respondent. Used to screen for the likely presence of domestic violence in the respondent's relationship. For the PuP programme, the HITS questions were incorporated into the demographic questionnaire completed by parents at assessment only.	Found to have good internal consistency and validity and to effectively differentiate between victims and non-victims of domestic violence.	A cut-off score is used to classify the presence of domestic violence in a respondent's relationship.	84		
Severity of Dependence Scale (SDS) (Gossop et al, 1995)	A five-item, self-report questionnaire used to measure a person's degree of dependency on a variety of different substances. Questions relate to behaviours that are indicators of substance dependency and parents were directed to think about their substance use over the past month. Cut-off scores have been identified to determine whether someone is showing dependency on a substance(s). For the PuP programme, these questions were incorporated into the demographic questionnaire completed by parents at assessment only.	Found to be a reliable and valid measure of the degree of dependency and an accurate diagnostic tool of substance dependency.	Cut-off scores can be used to determine whether the parent is showing dependence on substances. Optimal cut-off scores have been explored for a range of substances. For the purposes of this evaluation, a cut-off score of 3+ was used across all substances (this has been found to be the optimal score in many papers) given that some parents were dependent on more than one substance.	95 (only 88 were valid due to missing data)		

Evaluation sample

This evaluation of PuP was based on analysis of the data from the service measures administered to parents by practitioners as part of the PuP programme. These measures were completed during assessment (baseline or T1), half-way through the programme (T2) and upon programme completion (T3). Only the data from parents who met the eligibility criteria for the evaluation (see page 21 of the main [service evaluation report](#) for details) and gave their consent to the evaluation was included in the evaluation.

There were 223 parents who were eligible for the PuP evaluation and some of these may have been joint carers as, in these instances, measures would have been completed by both parents. For inclusion in the evaluation, we asked practitioners to choose the parent with the main caregiving responsibilities so that two parents involved in the same case were not included in the evaluation. However, cleaning of the dataset revealed the inclusion of data from two joint carers involved in the same case, neither of whom the practitioner could define as the main caregiver. In this case we included the data from the mother instead of the father in the evaluation given that the majority of primary caregivers accessing PuP were mothers.

A total of 180 parents (out of 223) who were eligible for the internal NSPCC evaluation provided evaluation consent (81 per cent). Of these, 11 disengaged from the PuP service before any assessment measures were completed so could not be included in the evaluation. There were three additional parents who gave evaluation consent, but their measures data was not entered onto the PuP database and this therefore prevented their inclusion in the evaluation. We therefore had T1 data for 166 parents (74 per cent of those eligible) who participated in the PuP programme and consented to the NSPCC evaluation.

The Early Intervention Foundation's (EIF) Evidence Standards (2017) state that the evaluation sample for a pre/post evaluation design should represent at least 60 per cent of the target population. Our sample at T1 comprised 74 per cent of our target population of parents and therefore exceeds this requirement. However, the proportion of the target population of parents who completed measures at both T1 and T3 was 41 per cent and is therefore much lower.

We compared the parents who did (n=166) and did not (n=57) take part in the evaluation based on their demographic data and their PuP measures data. As the data was significantly skewed, we carried out non-parametric tests to explore differences between the two groups of parents. There were few significant differences between parents based on whether they did or did not consent to the evaluation, as can be

seen in Table 2 and Table 3. However, significantly fewer fathers took part in the evaluation, as did parents with more than one child. Those taking part in the evaluation also received the service for significantly longer than those who did not take part. It is important to note that there may be unobserved demographic differences between the parents that have not been explored in Table 2. The parents who gave their consent to be part of the evaluation also reported significantly greater total problems and significantly fewer social-emotional competencies in their children (aged 1–3 years) than the parents who withheld their evaluation consent. The effect sizes for these tests were small, which suggests there was a small difference between the two groups. We were unable to compare parents’ primary substance of use due to the way in which the data was recorded on the demographic questionnaire and in parents’ case files.

There were no other significant differences between the parents who did and did not take part in the evaluation based on their measures data at T1 (Table 3).

Table 2: Comparison of parents who did and did not consent to the NSPCC evaluation based on their demographic data at T1

	Took part in evaluation (n=166)		Non-evaluation (n=57)		Statistical test (chi-square)		
	N	%	N	%	χ^2	df	p
Fathers	10	6	14	25	15.18	1	<.001
White British, British or White ‘other’ ¹	156	98	52	98	*	1	1.000
Parent had other children aside from the child referred for PUP ²	131	79	13	24	55.61	1	<.001
	Median	IQR	Median	IQR	U	Z	p
Age ³	31	9	29	8	4323.5	-0.194	.846
Average duration of PuP (days) ⁴	199	88	144	131	3015.5	-3.18	<.001

*Fishers Exact test computed as 50 per cent of cells had expected count less than 5.

IQR = Interquartile Range.

Analysis based on available data for: ¹160 parents taking part in the evaluation and 53 who did not; ²166 parents taking part in the evaluation and 55 who did not; ³163 parents taking part in evaluation and 54 who did not; ⁴161 parents taking part in evaluation and 52 who did not.

Table 3: Comparison of parents who did and did not consent to the NSPCC evaluation based on their measures data at T1

Measure (T1)	Took part in evaluation			Non-evaluation			Statistical test (Mann-Whitney)			Effect size
	N	Median	IQR	N	Median	IQR	U	z	p	(r)
DASS21										
Depression	164	8	16.0	56	10	12.0	4476.0	-.283	.777	.02
Anxiety	164	6	12.0	56	11	15.5	3975.5	-1.51	.131	.10
Stress	164	14	17.5	56	16	18.5	4394.0	-.482	.630	.03
Total score	164	28	43.5	56	38	45.0	4266.0	-.793	.428	.05
IMP-I										
Mindful parenting score	159	104	22.0	42	102	21.8	3313.5	-.076	.939	.01
AUDIT-C										
Total score	140	2	4.0	32	2	3.8	2099.0	-.567	.571	.04
MSPSS										
Family support	166	5	2.8	57	5.5	2.3	4413.5	-.758	.448	.05
Friends support	166	4.25	2.3	57	4	2.1	4448.0	-.675	.499	.05
Significant other	166	6	2.3	57	6	2.3	4715.5	-.038	.970	.03
Total	166	4.92	1.6	57	5	1.5	4506.0	-.536	.592	.04
Parenting support scale										
Actual practical support	166	4	3.0	57	4	4.0	4725.0	-.014	.988	.00
Actual emotional support	166	4.5	3.0	57	5	3.0	4598.0	-.320	.749	.02
Total actual support	166	9	5.0	57	9	6.5	4678.0	-.127	.899	.01
Actual-ideal support discrepancy	166	0	3.3	57	0	3.0	4390.0	-.824	.410	.06
Daily parenting hassles										
Total score	154	4	9.0	49	4	11.0	3638.0	-.378	.705	.03
Life Events Scale										
Total score	154	4	4.0	49	4	4.0	3472.5	-.845	.398	.06
BITSEA										
Total problems	57	10	9.5	10	6	6.3	157.5	-2.25	.024	.27
Total competencies	54	17.5	5.3	10	19.5	4.5	152.5	-2.19	.029	.27

Note that the median is reported, and non-parametric analysis was carried out as parents' scores on these measures were not normally distributed. IQR = Interquartile Range.

r effect size: 0.1 = 'small', 0.3 = 'medium', 0.5+ = 'large' (Cohen, 1988)

DASS21: Depression, anxiety and stress scale. IMP-I: Interpersonal mindfulness in parenting scale- infant version. AUDIT-C: Alcohol Use Disorders Identification. MSPSS: Multidimensional Scale of Perceived Social Support. BITSEA: Brief Infant-Toddler Social and Emotional Assessment.

Evaluation attrition

If a parent dropped out of the PuP programme, then they were no longer eligible to take part in the evaluation. As can be seen in Table 1 of the main [service evaluation report](#), we retained 55 per cent of our sample in the evaluation and a total of 91 parents completed evaluation measures at T3. Some of these parents will have completed the T3 measures at the point in which they stopped participating in PuP (for those who finished early), while others will have completed T3 measures when the programme came to a planned end. The EIF (2017) state that a 40 per cent evaluation attrition rate is acceptable for pre/post evaluation designs, and our attrition rate, at 45 per cent, is therefore slightly higher than this. However, we explored the differences between the parents who did and did not complete PuP and found very few differences between the two groups (see pages 31 and 32 of the main PuP evaluation report). The differences we did identify were that parents with lower levels of family support and those who had other children who were not currently in their care were more likely to complete PuP than those with greater family support and those with other children who had not been removed. This suggests that attrition from the evaluation is unlikely to have significantly skewed our findings or rendered our sample unrepresentative of the PuP population. However, it does suggest that parents who potentially have more to gain from the service (in terms of it providing them with additional support and helping them regain or retain care of their children) are more likely to stay in the programme.

Data analysis

The information collected on the demographic questionnaire and the measures data were first explored to look at the profile of parents assessed for the PuP programme. We then statistically analysed this data to explore similarities and differences between the parents who dropped out of the programme and those who completed it. We also explored whether there was any statistically significant change in scores on the measures and safeguarding status of the children over the duration of PuP. The specific details of the analysis carried out for each evaluation question are provided at the start of each section below.

Findings

For each of the six evaluation questions, we first describe the type of analysis carried out followed by the data tables and findings from any statistical analysis. These data tables support the summary of the findings presented within the main PuP evaluation report.

Question 1: What is the demographic profile and the needs of the parents and children who access this service?

The information collected on the demographic questionnaire and the measures data were first explored to look at the profile of parents assessed for the PuP programme. This data is descriptive and draws on the categorisation of scores on the DASS21, AUDIT-C and MSPSS as these are more meaningful than the mean scores. This is because it translates the scores into clinically meaningful categories to better represent the level of need. This data is presented in Table 4. Table 5 presents the data for the categorisation of BITSEA scores.

Table 4: Characteristics of the parents being accepted onto the PuP programme at T1 based on the categorisation of their scores on the individual PuP measures

4a: Emotional wellbeing

DASS21	N completing measure at T1	Extremely severe				
		Normal % (N)	Mild % (N)	Moderate % (N)	Severe % (N)	% (N)
Stress	164	57 (93)	8 (13)	13 (21)	14 (23)	9 (14)
Depression	164	54 (88)	10 (17)	19 (31)	5 (8)	12 (20)
Anxiety	164	58 (95)	6 (10)	6 (10)	13 (22)	17 (27)

DASS21: Depression, anxiety and stress scale.

4b: Alcohol consumption

AUDIT-C	N completing measure at T1	Safe/low risk		
		alcohol intake % (N)	Alcohol abuse % (N)	Alcohol dependence % (N)
Total AUDIT score	140	81 (113)	11 (15)	9 (12)

AUDIT-C: Alcohol Use Disorders Identification.

4c: Social support

MSPSS	N completing measure at T1	Low level of support % (N)	Mid-range % (N)	High level of support % (N)
Family support	166	25 (41)	19 (32)	56 (93)
Friends support	166	35 (58)	27 (44)	39 (64)
Significant other support	166	7 (12)	11 (18)	82 (136)
Total score	166	24 (40)	27 (45)	50 (81)

MSPSS: Multidimensional Scale of Perceived Social Support.

4d: Daily parenting hassles/challenges

	N completing measure at T1	Range	Mean	SD
Daily Parenting Hassles Scale	154	0–34	6.78	6.89

4e: Significant life events experienced within the past year

Life Events Scale	N completing measure at T1	Did not happen in past year % (n)	Happened but coping OK % (n)	Happened and not coping % (n)
Major problem in relationship with partner, member of family or close friend	154	32 (49)	55 (85)	13 (20)
Major money problems	154	43 (66)	49 (75)	8 (13)
Problems with housing	154	56 (86)	34 (53)	10 (15)
Suffered a serious health problem	154	71 (110)	23 (36)	5 (8)
Member of family or close friend suffered a serious health problem	154	58 (89)	35 (54)	7 (11)
Death in family or of a close friend	154	54 (83)	39 (60)	7 (11)
Legal problems	154	73 (113)	22 (34)	5 (7)
Experienced a trauma, such as an accident, assault, house broken into or other traumatic event	154	68 (105)	25 (39)	7 (10)
Total number of life events experienced in past year	Range = 0–8	Mean = 4	SD = 2.56	

4f: Domestic abuse and substance dependency

	N completing measure at T1	Range	Mean	SD	Positive classification of scores
HITS	84	4–20	6.11	3.13	9.5% ¹
SDS	88	0–14	6.50	3.53	90% ²

HITS = domestic violence screening tool. SDS = Severity of Dependence Scale.

¹Likely presence of domestic violence in the respondent's relationship. ²Dependent on substances used within the past month.

As can be seen in Table 4, around two-thirds of parents were scoring in the 'normal' or 'mild' range for stress, depression and anxiety, the vast majority were drinking in the safe/low risk range, and the majority reported high levels of support from a 'significant other' and from their family at the start of PuP. Parents reported a low level of daily parenting hassles/challenges, but they did report experiencing a number of significant life events within the past year; most commonly major problems in a relationship and money problems. Most parents showed substance dependency within the past month and almost one in ten tested 'positive' for likely domestic violence in their current relationship.

Table 5: Parents' reports of their child's social-emotional and behavioural problems or deficits/delays in social-emotional competence at the start of the PuP programme. This is based on the categorisation of scores from the two BITSEA subscales

BITSEA¹	N completing measure at T1	Yes % (N)	No % (N)
Possible problem	49	33 (16)	67 (33)
Possible deficit/delay	47	23 (11)	77 (36)

¹Note that some children whose scores fall within the 'possible problem' and 'possible deficit/delay' range will not have clinically significant problems.

BITSEA: Brief Infant-Toddler Social and Emotional Assessment.

The data presented in Table 5 shows that around a third of parents reported a possible problem in their child's social-emotional and behavioural functioning at assessment, and a quarter reported possible deficits/delays.

Question 2: Are there any differences between the parents (and their children) that start, but do not complete, PuP compared with those who complete the programme?

As presented in Table 6, we explored the percentage of parents who completed PuP and the reasons for non-completion. We then looked at the average amount of time parents received PuP for depending on whether they completed the programme or not. Finally, we explored any similarities and differences between parents who completed the programme and those who did not (Tables 7–11) and looked to see whether there were any predictors of programme completion (Table 12).

Table 6: The proportion of parents completing PuP and the reasons for non-completion (based on the 166 parents taking part in the evaluation)

Programme completion and reasons for non-completion	Frequency (N)	Percent
Work completed – full programme	74	44.6
Agreed work completed – not full programme	32	19.3
Finished early – parent disengaged from the service	33	19.9
Finished early – parent was accessing other services instead	2	1.2
Finished early – professional decision to close the case	4	2.4
Finished early – child no longer in parent’s care	9	5.4
Finished early – mum went to prison	1	0.6
Finished early – caregiver went to detox/focused on targeting their drug use instead	4	2.4
Finished early – lack of NSPCC worker continuity/NSPCC service centre closed	5	3.0
Case ongoing	2	1.2

Most parents (64 per cent) completed either the full PuP programme or an agreed amount of work. Those who did not complete the programme did so for a number of reasons, most commonly that they disengaged from the service (20 per cent).

Calculating the length of time and the number of sessions a parent received PuP for

The length of time the PuP case was open for was calculated based on the time between the recorded ‘workplan agreed’ date and ‘case closure’ date entered into the NSPCC’s case recording system. This was the only way to gain a standardised assessment of programme duration for all eligible PuP cases. As part of the NSPCC’s case recording requirements, the agreed workplan for a case must be entered onto the NSPCC’s case recording system within 28 days of a case being accepted by the NSPCC. There may then be a 1–2-week period where the practitioner meets with the parents and a further 1–2 weeks for them to gain their consent to the programme (if this has not already taken place). The assessment for PuP will then begin, but as can be seen, this may be up to four weeks from the ‘workplan agreed’ date, which will have an impact on the average duration the cases were open to receiving PuP for. In some cases, the parent completed T1 measures but disengaged before a workplan had been agreed and they were therefore deemed to have received PuP for 0 weeks.

The calculated number of PuP sessions received is based on all face-to-face sessions recorded between the practitioner and parent. This includes 1-to-1 sessions, professionals’ meetings, assessment and goal setting sessions, and ad-hoc visits and engagement sessions. Across all parents, the mean number of sessions received by parents was 16, ranging from 2–36.

Comparing parents who completed PuP to those who did not

We first carried out chi-square analysis to compare the demographic characteristics of the parents who did and did not complete PuP (Table 7), along with the categorisation of parents’ scores at T1 (Table 8).

Table 7: Demographic profile at T1 of the parents completing PuP compared with those who did not complete PuP

	Completers		Non-completers		Statistical analysis (chi-square)		
	N	%	N	%	χ^2	df	p
Female	99	93	55	95	*	1	1.000
Single parent	54	52	26	50	0.051	1	.821
Criminal offence	46	59	24	62	0.071	1	.790
Drug abuse (versus alcohol abuse or alcohol and drug abuse)	72	68	47	81	3.236	1	.072
Diagnosed mental health difficulty	45	55	24	52	0.087	1	.768
Victim/perpetrator of domestic violence	51	57	22	48	0.957	1	.328
More than one child	82	77	47	81	0.302	1	.583
Did not have all of their children in their care during the assessment (<i>of parents who had other children</i>)	23	34	6	14	5.105	1	.024
Gender of child (male)	54	51	33	57	0.533	1	.465
Child living with parent at start of assessment	102	98	50	96	*	1	.601

Due to missing data, the percentages for each variable are based on the number of parents in each group for whom we had data on for that variable, and not the total number of parents.

*Fishers exact test was carried out as 25 per cent or more cells have an expected count less than 5.

	Completers		Non-completers		Statistical analysis ¹		
	Mean	SD	Mean	SD	t	df	p
Parent's age	30.7	5.88	30.5	6.08	-.229	159	.819
	Median	IQR	Median	IQR	U	z	p
Child's age (months)	12.5	20	17.0	27	2831	-0.837	.402

¹Parent's age was normally distributed and therefore t-tests were carried out. Child's age was significantly skewed and therefore the Mann-Whitney test was carried out. IQR = Interquartile Range.

As can be seen in Table 7, the only significant difference in the demographic characteristics of the parents who did and did not complete PuP was that those parents who completed the programme were more likely to have had their other children removed from their care.

The only significant difference identified between the scores on the measures for parents who did and did not complete PuP at baseline, as seen in Table 8, was that more of those who completed PuP had lower levels of family support than those who did not complete the programme.

Table 8: Characteristics of the parents at T1 who did complete ('completers') and did not complete ('non-completers') PuP based on the categorisation of their scores on the PuP measures

	Completers				Non-completers				Statistical analysis (chi-square)		
	N	Normal/ Mild % (N)	Moderate/ severe % (N)	Extremely severe % (N)	N	Normal/ Mild % (N)	Moderate/ severe % (N)	Extremely severe % (N)	χ^2	df	p
DASS21											
Stress	105	65 (68)	26 (27)	10 (10)	57	63 (36)	30 (17)	7 (4)	no analysis ¹		
Depression	105	67 (70)	20 (21)	13 (14)	57	58 (33)	32 (18)	11 (6)	2.74	2	.254
Anxiety	105	65 (68)	17 (18)	18 (19)	57	63 (36)	23 (13)	14 (8)	1.00	2	.607
AUDIT-C											
	N	Safe/low risk % (N)	Alcohol abuse % (N)	Alcohol dependence % (N)	N	Safe/low risk % (N)	Alcohol abuse % (N)	Alcohol dependence % (N)	χ^2	df	p
T total score	93	83 (77)	12 (11)	5 (5)	46	76 (35)	9 (4)	15 (7)	no analysis ¹		
MSPSS											
	N	Low support % (N)	Mid-range % (N)	High support % (N)	N	Low support % (N)	Mid-range % (N)	High support % (N)	χ^2	df	p
Family support	106	31 (33)	18 (19)	51 (54)	58	12 (7)	22 (13)	66 (38)	7.39	2	.025
Friends support	106	38 (40)	27 (29)	35 (37)	58	29 (17)	26 (15)	45 (26)	1.76	2	.415
Significant other support	106	7 (7)	10 (11)	83 (88)	58	9 (5)	12 (7)	79 (46)	no analysis ¹		
T total score	106	26 (28)	27 (29)	46 (49)	58	19 (11)	28 (16)	53 (31)	1.28	2	.528

¹25 per cent or more cells had an expected count less than 5 and therefore no statistical analysis could be carried out.

DASS21: Depression, anxiety and stress scale. AUDIT-C: Alcohol Use Disorders Identification. MSPSS: Multidimensional Scale of Perceived Social Support.

Table 9 displays and compares the median scores on the PuP service measures at T1 for both groups of parents using Mann-Whitney tests. The median was used as the data was not normally distributed, with the exception of the IMP-I score, which was normally distributed and therefore a t-test was carried out.

Table 9: Statistical comparison of the scores on the PuP measures at T1 for the parents who did complete ('completers') and did not complete ('non-completers') PuP

Measure	Completers			Non-completers			Statistical test (Mann-Whitney)			Effect size
	N	Median	IQR	N	Median	IQR	U	z	p	(r)
DASS21										
Depression	105	6	16.0	57	8	16.0	2954.5	-.134	.893	.01
Anxiety	105	6	12.0	57	6	12.0	2966.0	-.094	.925	.07
Stress	105	14	18.0	57	14	17.0	2973.0	-.069	.945	.01
Total score	105	28	45.0	57	32	42.0	2954.5	-.133	.894	.01
AUDIT-C										
Total score	93	2	4.0	46	1.5	4.5	2061.5	-.356	.722	.03
MSPSS										
Family support	106	4.75	3.1	58	5.25	2.5	2568.0	-1.747	.081	.14
Friends support	106	4.25	2.3	58	4.63	2.3	2898.0	-.607	.544	.05
Significant other	106	6.00	2.0	58	5.63	2.5	2850.0	-.784	.433	.06
Total	106	4.83	1.7	58	5.04	1.4	2795.0	-.960	.337	.07
Parenting support scale										
Actual practical support	106	4	3.0	58	4	3.3	3016.0	-.202	.840	.02
Actual emotional support	106	4	3.0	58	5	4	2770.0	-1.058	.290	.08
Total actual support	106	9	5.0	58	8.5	6.0	2916.5	-.544	.586	.04
Actual-ideal support discrepancy	106	0	3.0	58	0	4	2991.5	-.288	.773	.02
Daily parenting hassles										
Total score	97	5	9.0	55	4	9.0	2543.5	-.477	.633	.04
Life Events Scale										
Total score	97	4	4.0	55	4	4.0	2513.5	-.595	.552	.05

r effect size: 0.1 = 'small', 0.3 = 'medium', 0.5+ = 'large' (Cohen, 1988).

DASS21: Depression, anxiety and stress scale. AUDIT-C: Alcohol Use Disorders Identification. MSPSS: Multidimensional Scale of Perceived Social Support.

Measure	Completers			Non-completers			Statistical test (t-test)			Effect size
	N	Mean	SD	N	Mean	SD	t	df	p	(g ¹)
Mindful parenting										
Total IMP-I score	102	103	14.75	55	102	14.14	-.300	155	.765	0.05

Note: the total IMP-I score was normally distributed within both groups and therefore permitted the use of parametric tests (t-test). We have used the total mean score instead of the average mean score (average score across all 27 questions).

¹Hedges' g effect size is used as sample sizes are different across the groups. This effect size is very small.

IMP-I: Interpersonal mindfulness in parenting scale- infant version.

As can be seen in Table 9, there were no significant differences in the baseline scores on the PuP service measures between parents who did and did not complete PuP.

Based on parents' reports of their child's social-emotional and behavioural problems and competencies using the BITSEA, Table 10 displays the categorisation of scores to show possible problems and delays for the children of the parents who completed ('completers') and who did not complete ('non-completers') PuP. Note that some of the children whose scores fell within the 'possible problem' and 'possible deficit/delay' range will not have clinically significant problems.

Table 11 displays the median scores for both groups on the BITSEA subscales (median scores are used as the data was not normally distributed).

Table 10: Parents' reports of their child's social-emotional and behavioural problems and deficits/delays based on the categorisation of BITSEA scores – comparison of parents who did and did not complete PuP

BITSEA	Completers		Non-completers		Statistical analysis (chi-square)			Effect size
	N completing measure at	Yes	N completing measure at	Yes	χ^2	df	p	Odds ratio (95% CI)
	T1	% (N)	T1	% (N)				
Possible problem	31	39 (12)	17	24 (4)	1.139	1	.286	2.05 (0.54–7.79)
Possible deficit/delay	30	27 (8)	16	19 (3)	*	1	.722	1.58 (0.35–7.02)

*Fisher's exact test was carried out as 25 per cent or more cells had an expected count less than 5.

BITSEA: Brief Infant-Toddler Social and Emotional Assessment.

Table 11: Statistical comparison of the scores on the BITSEA for the children of the parents who did complete ('completers') and did not complete ('non-completers') PuP

Measure	Completers			Non-completers			Statistical test (Mann-Whitney)			Effect size
	N	Median	IQR	N	Median	IQR	U	z	p	(r)
BITSEA										
Total problems	31	10	9.0	17	10	9.5	228.5	-0.76	.448	.11
Total competencies	30	17	6.2	16	18	4.8	204.5	-0.83	.409	.12

IQR = Interquartile Range. *BITSEA*: Brief Infant-Toddler Social and Emotional Assessment. There were no significant differences in parents' reports of their child's social-emotional and behavioural problems or competencies depending on whether they did or did not complete PuP.

Predicting programme completion

Based on the few differences we identified between parents who completed PuP and those who did not, we explored whether parents' levels of support (based on MSPSS scores), referral to PuP because of a drug (only) problem (yes/no), and having other children who were not currently in their care (yes/no) could significantly predict PuP completion (combining those who received the full programme or an agreed programme of work into one 'completers' group).

Binary logistic regression analysis was carried out, including all predictors in the same model, and checking of the data confirmed it met the assumptions for this test (there were no issues with multicollinearity, there was linearity of the logit, and no individual cases were shown to be having an undue effect on analysis). The findings are presented in Table 12.

Table 12: Logistic regression analysis exploring predictors of programme completion

	B (SE)	Odds Ratio	95% Confidence interval
Constant	1.66 (1.05)		
Parent has other children who are currently in their care	-0.98 (0.55)	0.38	0.13–1.09
Substance misuse involves alcohol (alone or with drugs)	0.78 (0.49)	2.18	0.83–5.67
Family support	-0.31 (0.15)*	0.73	0.55–0.98
Friend support	-0.06 (0.13)	0.94	0.72–1.22
Significant other support	0.20 (0.17)	1.22	0.89–1.69

Note: $R^2 = .10$ (Hosmer & Lemeshow), $.12$ (Cox & Snell), $.16$ (Nagelkerke).

Model $\chi^2(5) = 14.09$, $p < .05$.

* $p < 0.05$

The only significant predictor of programme completion was family support; parents with higher levels of family support at T1 were around 0.7 times less likely to complete PuP than parents with lower levels of family support.

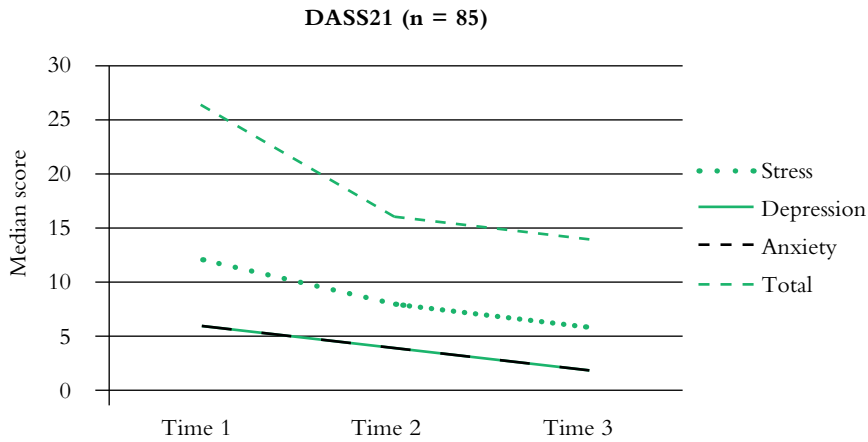
Question 3: Is there evidence of change for the parents and children who complete the programme? How does change occur over time?

To explore the change made by parents on the measures during the course of PuP, we carried out a series of non-parametric analyses as the data was significantly skewed. The analysis of change for each measure only included parents who had completed PuP and had completed that measure at all three time-points. While we have presented the findings in relation to primary and secondary outcomes within the main report, here we group all of the findings together based on the type of analysis carried out (for example, change in scores followed by change in the categorisation of scores, etc).

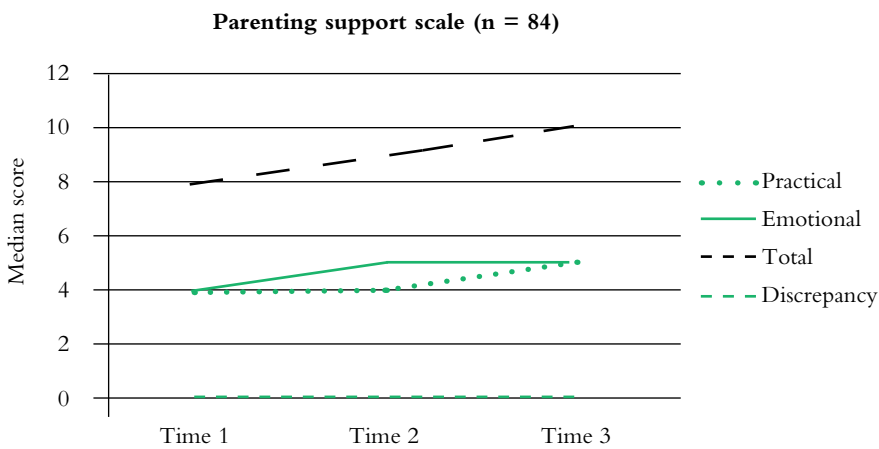
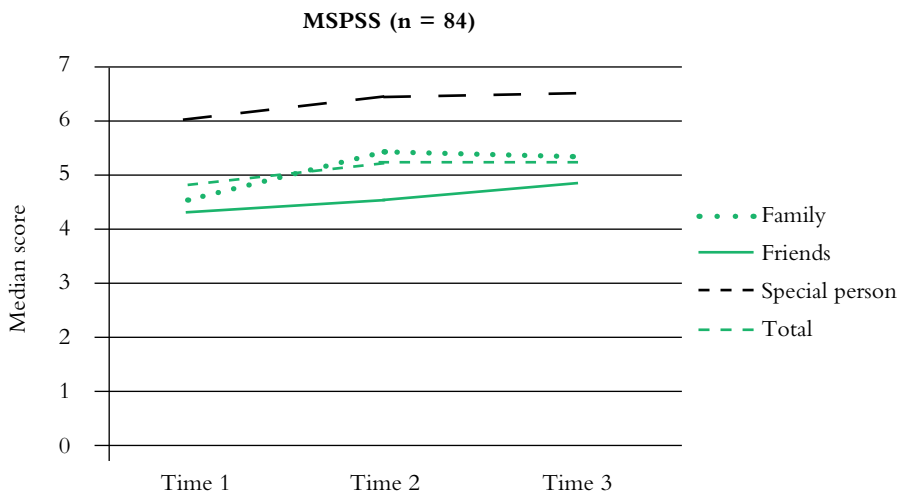
Change in scores over time for the whole sample of parents who completed PuP

Change in median scores on each measure over the three time-points are displayed in Figure 1 below. We first carried out Friedman's tests to explore significant change in scores on the measures over time. Where there was significant change, post-hoc Wilcoxon Signed Ranks tests were carried out to explore whether the change occurred during the first half or latter half of the programme. A Bonferroni correction was applied to adjust for the number of post-hoc tests being carried out on each measure, resulting in a significance level set at $p = .017$. The statistical results are presented in Table 13.

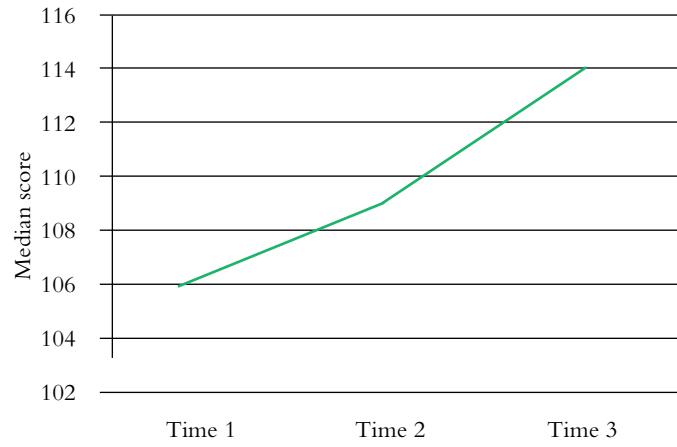
Figure 1: Change in parents' median scores for each measure across all three time-points. Median scores are shown only for parents who completed the programme and completed the measure at every time-point (number of parents is reported in each figure)



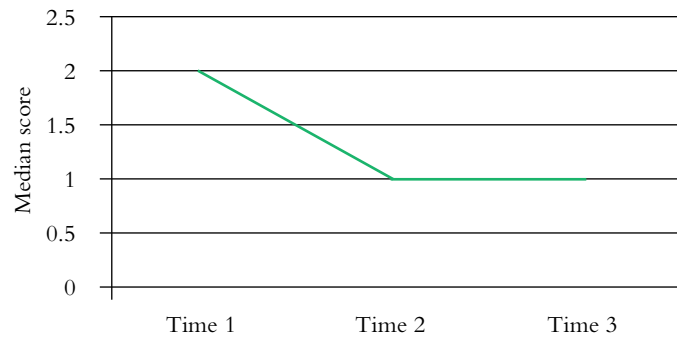
*Note that depression and anxiety scores were the same at each time-point



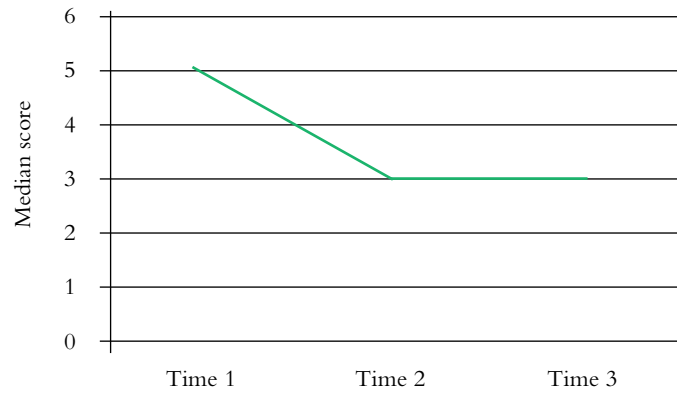
Mindful Parenting (n = 80)



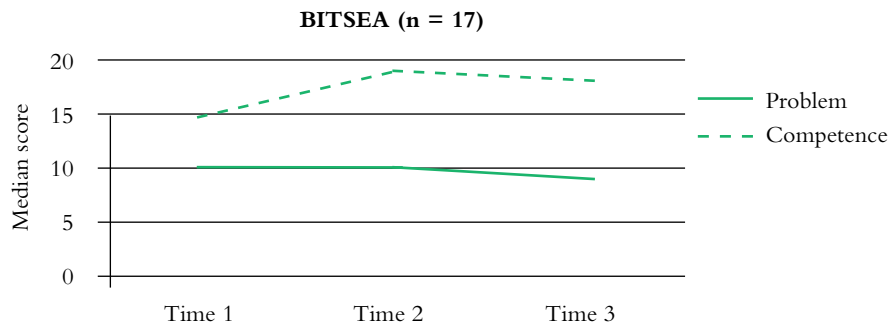
AUDIT-C (n = 62)



Daily Parenting Hassles (n = 75)



Change in parents' median scores for their child's social-emotional and behavioural problems and competencies. Median scores are shown only for parents who completed the programme and completed the measure at every time-point (n=17).



The changes in median scores, as can be seen in the graphs presented in Figure 1, show a pattern of change whereby parents showed gradual improvements over the duration of PuP, but often made the greatest change within the first half of the programme. For the daily parenting hassles scale, alcohol use scale, total social support sub-scale and parents' reports of their child's social-emotional and behavioural problems, the change made during the first half of PuP appears to be sustained during the latter half. Mindful parenting and practical parenting support show a slightly different pattern whereby the greatest increase in scores appears to take place during the latter half of the programme.

Table 13: Change in parents' scores on each measure across all three time-points. Data is only for parents who completed PuP and completed the measure at every time-point

Measure	N	T1		T2		T3		Significant change over time (Friedman)		
		Median	IQR	Median	IQR	Median	IQR	χ^2	df	p
DASS21										
Depression	85	6	15.0	4	8.0	2	8.0	36.55	2	<.001
Anxiety	85	6	13.0	4	9.0	2	6.0	13.93	2	<.001
Stress	85	12	19.0	8	10.0	6	10.0	48.08	2	<.001
Total score	85	26	46.0	16	22.0	14	24.0	31.05	2	<.001
AUDIT-C										
Total score	62	2	5.0	1	4.0	1	3.0	8.80	2	.012
MSPSS										
Family support	84	4.5	2.9	5.4	2.5	5.3	3.0	17.57	2	<.001
Friends support	84	4.3	2.2	4.5	2.8	4.8	2.5	6.75	2	.034
Significant other	84	6	2.0	6.4	2.0	6.5	2.0	6.48	2	.039
Total	84	4.8	1.6	5.2	1.8	5.2	1.8	15.63	2	<.001
Parenting support scale										
Actual practical support	84	4	4.0	4	3.0	5	3.0	3.71	2	.157
Actual emotional support	84	4	3.0	5	3.0	5	2.8	5.02	2	.081
Total actual support	84	8	4.8	9	6.0	10	5.8	5.20	2	.074
Actual-ideal support discrepancy	84	0	2.8	0	2.8	0	1.5	5.00	2	.082
Daily parenting hassles										
Total score	75	5	9.0	3	5.0	3	3.0	15.41	2	<.001
IMP-I										
Total score	80	106	24.3	109	22.5	114	17.8	25.06	2	<.001
BITSEA										
Problem	17	10	11.5	10	7.0	9	8.0	3.12	2	.210
Competence	17	15	6.0	19	3.5	18	4.0	9.39	2	.009

IQR = Interquartile Range

DASS21: Depression, anxiety and stress scale. *AUDIT-C*: Alcohol Use Disorders

Identification. *MSPSS*: Multidimensional Scale of Perceived Social Support.

IMP-I: Interpersonal mindfulness in parenting scale- infant version. *BITSEA*: Brief Infant-Toddler Social and Emotional Assessment.

Table 14: Post-hoc analysis of change across all three time-points to identify when in the programme significant change occurred. Tests were carried out only for parents who had completed the measure at every time-point

Measure	N	T1 - T2		T2 - T3		T1 - T3	
		Z	p	Z	p	Z	p
DASS21							
Depression	85	-4.56	<.001	-2.23	.026	-5.12	<.001
Anxiety	85	-2.90	.004	-2.94	.003	-4.06	<.001
Stress	85	-4.52	<.001	-3.75	.000	-6.17	<.001
Total score	85	-4.50	<.001	-3.37	.001	-5.93	<.001
AUDIT-C							
Total score	62	-2.53	.011	-1.64	.100	-3.53	<.001
MSPSS							
Family support	84	-3.12	.002	-0.98	.329	-3.60	<.001
Friends support	84	-1.80	.072	-0.77	.444	-2.25	.025
Significant other	84	-1.86	.063	-0.29	.774	-1.43	.152
Total	84	-3.09	.002	-0.80	.423	-3.31	<.001
Daily parenting hassles							
Total score	75	-2.47	.014	-3.51	<.001	-3.62	<.001
IMP-I							
Total score	80	-3.54	<.001	-2.66	.008	-4.85	<.001
BITSEA							
Competence	17	-2.74	.006	-.284	.777	-2.017	.044

Note. A Bonferroni correction was applied to adjust for the number of post-hoc tests being carried out on each measure, resulting in a significance level set at $p = .017$

DASS21: Depression, anxiety and stress scale. *AUDIT-C*: Alcohol Use Disorders

Identification. *MSPSS*: Multidimensional Scale of Perceived Social Support.

IMP-I: Interpersonal mindfulness in parenting scale- infant version. *BITSEA*: Brief Infant-Toddler Social and Emotional Assessment.

As can be seen in Tables 13 and 14, parents made significant improvements in all areas assessed during the course of PuP, with the exception of parenting support and child's social-emotional and behavioural problems. For emotional wellbeing (except depression), mindful parenting and daily parenting hassles/challenges, there was significant change during the first half and the latter half of the programme. For parental depression, alcohol use, family support, total support and parents' reports of their child's social-emotional competencies, the significant change occurred in the first half of the programme and no further significant change occurred in the latter half.

Correlation between primary outcome measure scores

To explore the relationship between the areas of change made by parents, we looked at the correlation between change scores on our primary outcome measures over the duration of PuP (T1–T3). Spearman’s rho correlations were carried out as the data was not normally distributed. The results are presented in Table 15. The DASS21 stress, depression and anxiety subscales were all highly and significantly positively correlated with each other, as would be expected (ie, difficulties in one area were associated with difficulties in another). There were also significant negative relationships between DASS21 subscale scores and mindful parenting, showing how levels of mindful parenting increased as DASS21 subscale scores decreased. Change in the BITSEA problem and competencies scores were not significantly correlated with change in any other areas.

Table 15: Correlation between change in DASS21, IMP-I and BITSEA scores between T1 and T3

	Depression	Anxiety	Stress	Mindful parenting	Social-emotional and behavioural problems	Social-emotional deficits/ delays
Depression	1.00	.524**	.752**	-.470**	-.037	.308
Anxiety	.524**	1.00	.615**	-.408**	-.094	.103
Stress	.752**	.615**	1.00	-.453**	0.024	-.104
Mindful parenting	-.470**	-.408**	-.453**	1.00	.241	-.331
Social-emotional and behavioural problems	-.037	-.094	0.024	.241	1.00	-.289
Social-emotional deficits/delays	.308	.103	-.104	-.331	-.289	1.00

Note. Correlation coefficients calculated using Spearman’s rho tests.
 ** $p < 0.01$

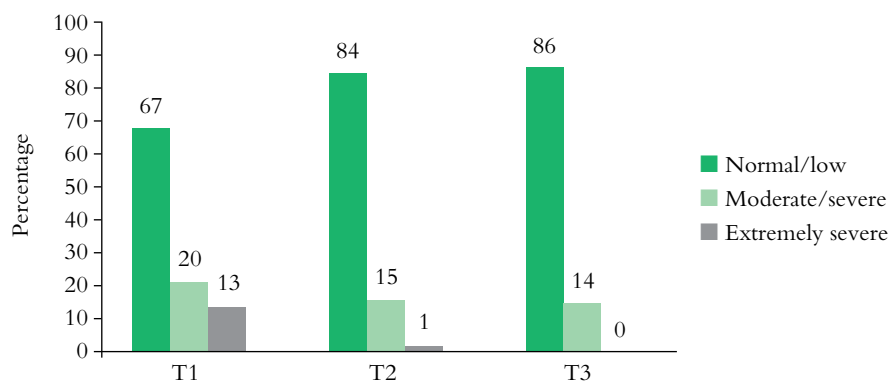
Change in categorisation of need for the whole sample of parents

We next tested for statistically significant change in the categorisation of parents’ scores on the PuP measures over time. This helps to identify clinically meaningful amounts of change. As the data is categorical, skewed, and has been obtained on the same parents at all time-points, it is necessary to use related-samples non-parametric tests, specifically the related-samples Cochran’s Q test (for two-category outcomes), and the related-samples Marginal Homogeneity test (for three-category outcomes). Note that the Marginal Homogeneity test can only be calculated for two samples at a time so here it is calculated for T1 vs T2 and then T2 vs T3. This helps to pinpoint when during the programme there were group changes in clinical

classification of need. A series of graphs showing the change in the proportion of parents in each category at each time-point are displayed first (Figure 2a to Figure 2j), followed by discussion of statistical significance.

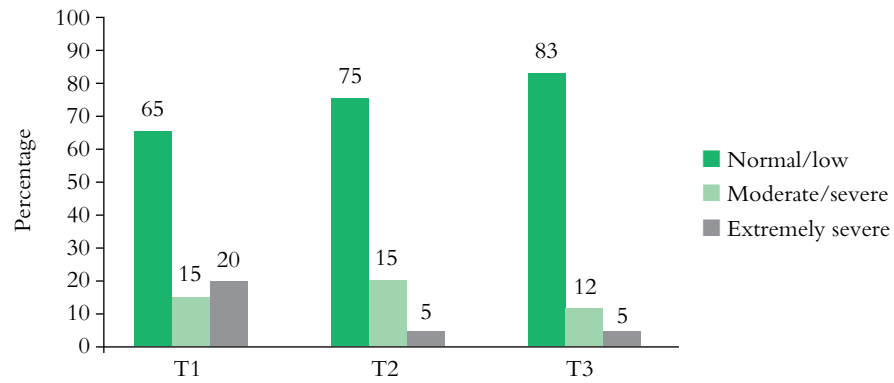
To categorise scores on the IMP-I, parents scoring 1.5 standard deviations below the mean were classed as displaying low levels of mindful parenting, -1.5 to +1.5 standard deviations were classed as displaying moderate levels of mindful parenting, and more than 1.5 standard deviations above the mean were classed as displaying high levels of mindful parenting. Scores ranged from 68–133 (mean = 102.82, SD = 14.43) and most parents were classified within the ‘moderate’ range at T1 (88 per cent).

Figure 2a: Change in depression (n=85)



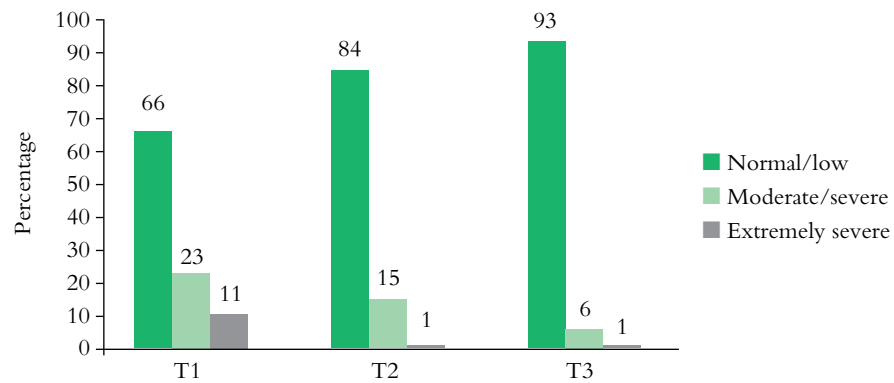
At T1, 11 parents (13 per cent) were classified as having severe depression on the DASS21. This had dropped to only one parent (1 per cent) by T2, and at T3 no parents reported symptoms that indicated they were severely depressed (although 12 were still moderately depressed). The Marginal Homogeneity test confirmed that among the 85 parents, PuP resulted in a significant change between T1 and T2 in the proportion of parents falling into each of the three categories ($T = 53, p < .001$). It appears that most of the change was a reduction in the number of parents reporting severe depression (rather than an increase in the number reporting low levels). Due to the zero value in the ‘extremely severe’ category at T3, it is not possible to compare the T2 and T3 classifications. Given the descriptive statistics above, however, it can be assumed that the change mainly occurred between T1 and T2, and was maintained at T3.

Figure 2b: Change in anxiety (n=85)



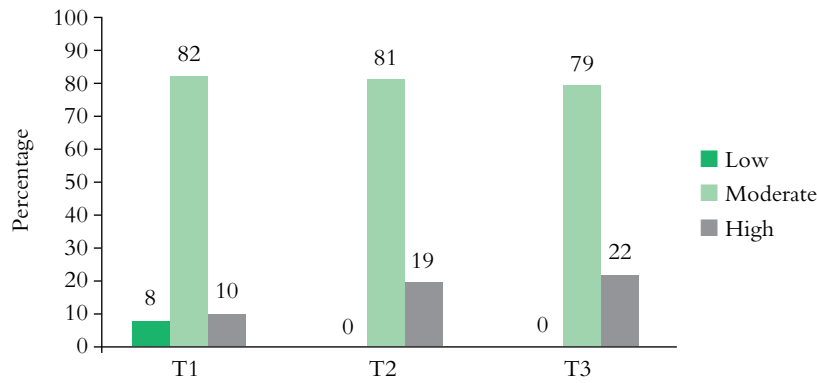
At T1, 17 parents (20 per cent) were classified as having extremely severe anxiety on the DASS21. This had dropped to only four parents (5 per cent) by T2, and this was the same at T3. By T3, 83.5 per cent of the parents were within the normal/low range for anxiety (compared with 65 per cent at T1). The Marginal Homogeneity test confirmed that among the 85 parents, PuP resulted in a significant change between T1 and T2 in the proportion of parents falling into each of the three categories ($T = 59, p = .002$). It appears most of the change was a reduction in the severity of anxiety. There was no significant change between T2 and T3 ($T = 28, p = .071$).

Figure 2c: Change in stress (n=85)



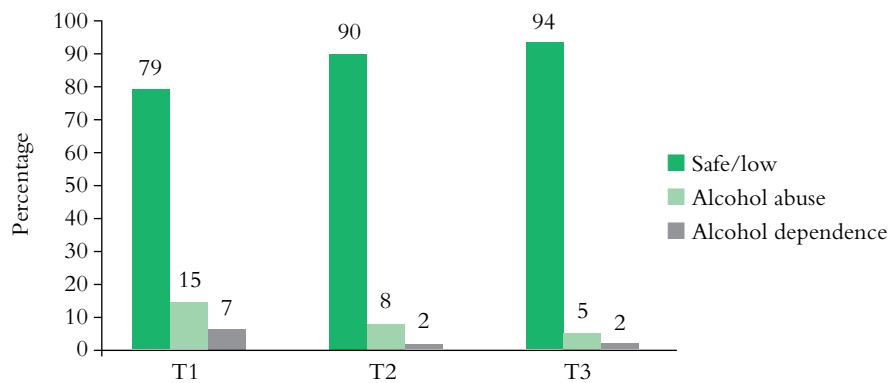
At T1, nine parents (11 per cent) were classified as having extremely severe stress on the DASS21. This had dropped to only one parent (1 per cent) by T2, and this was the same at T3. By T3, almost all of the parents (93 per cent) were within the normal/mild range for stress (compared with 66 per cent at T1). The Marginal Homogeneity test confirmed that among the 85 parents, PuP resulted in a significant change between T1 and T2 in the proportion of parents falling into each of the three categories ($T = 51, p < .001$). It appears most of the change was a reduction in the severity of stress. There was also a significant change between T2 and T3 ($T = 27, p = .033$), which from the descriptive statistics seemed mainly to be due to a reduction in the numbers of parents classified as moderately/severely stressed.

Figure 2d: Change in IMP-I (n=79)



At T1, six parents (8 per cent) were classified as having low mindfulness, while eight (10 per cent) were classified as having high mindfulness. At T2, the number of parents in the high category had almost doubled to 15 (19 per cent) and this only increased slightly by T3. At T2 and T3 there were no parents who were classified as having low mindfulness. In order to test these changes statistically, it was necessary to collapse the T1 scores so that low and moderate become one category, given that there were only two categories (moderate and high) at T2 and T3. McNemar tests were then used to explore change. These tests confirmed that among the 79 parents, there was no statistically significant change in the parents' mindfulness between T1 and T2 ($\chi^2(1) = 2.77, p = .092$) or between T2 and T3 ($\chi^2(1) = 0.10, p = .754$).

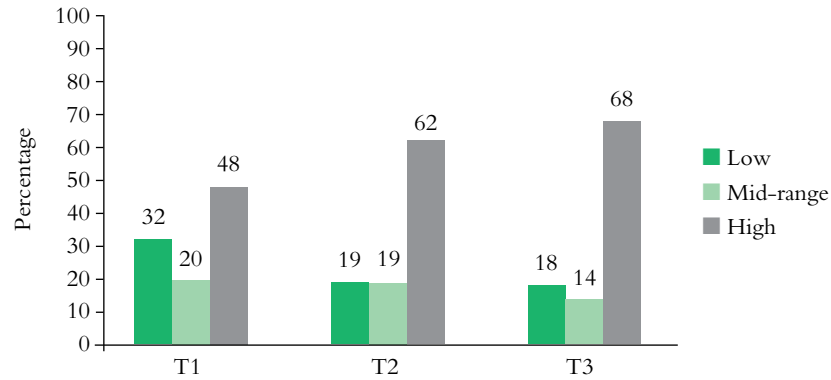
Figure 2e: Change in AUDIT-C (n=62)



At T1, only four parents (7 per cent) were classified as having alcohol dependence, while nine were categorised as abusing alcohol (14.5 per cent). At T2, the number of parents considered to be dependent (n=1) and abusing alcohol (n=5) had reduced by around half. By T3, 94 per cent of the parents were considered to have safe or low levels of alcohol use. The Marginal Homogeneity test confirmed that among the 62 parents, there was a significant change between T1 and T2 in the proportion of parents falling into each of the three categories ($T = 21, p = .004$). It appears most of the change was a reduction in

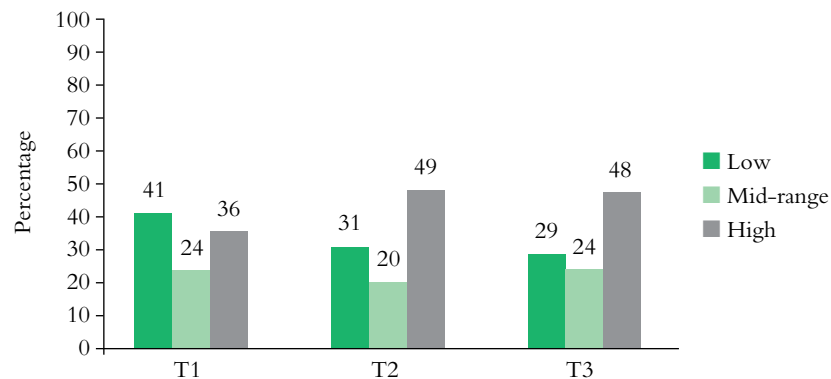
the severity of alcohol use. There was no significant change between T2 and T3 ($T = 4, p = .157$), indicating that the initial changes were maintained over time.

Figure 2f: Change in MSPSS (n=84)



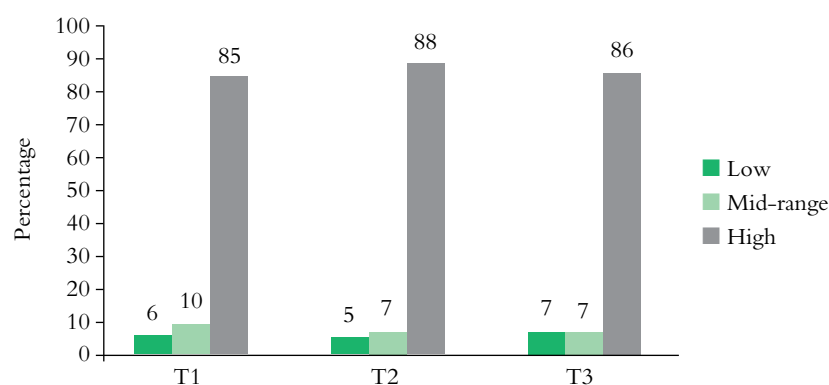
At T1, 40 parents (48 per cent) reported high levels of family support, while almost a third (n=27; 32 per cent) reported having low family support. At T2, these numbers had improved, with 62 per cent reporting high levels of support from family members and only 19 per cent reporting low levels of support. There was a very slight improvement on these numbers at T3. The Marginal Homogeneity test confirmed that among the 84 parents, PuP resulted in a significant change between T1 and T2 in the proportion of parents falling into each of the three categories ($T = 42, p = .002$). From the previous tables, it appears most of the change was an improvement in the level of family support. There was no significant change between T2 and T3 ($T = 44, p = .396$), indicating that the initial changes were maintained over time.

Figure 2g: Change in friend support (n=84)



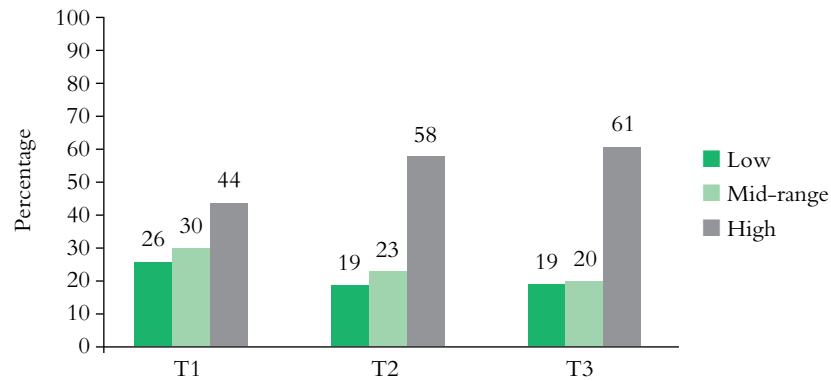
At T1, 30 parents (36 per cent) reported high levels of support from friends, while a similar number (n=34, 41 per cent) reported having low levels of friend support. At T2, these numbers had improved, with 49 per cent reporting high levels of support from friends and only 31 per cent reporting low levels of support. There was little change in these numbers at T3. The Marginal Homogeneity test confirmed that among the 84 parents, there was a significant change between T1 and T2 in the proportion of parents falling into each of the three categories ($T = 56, p = .030$). From the previous tables, it appears most of the change was an improvement in the level of support from friends. There was no significant change between T2 and T3 ($T = 65, p = .903$), indicating that the initial changes were maintained over time.

Figure 2h: Change in significant other support (n=84)



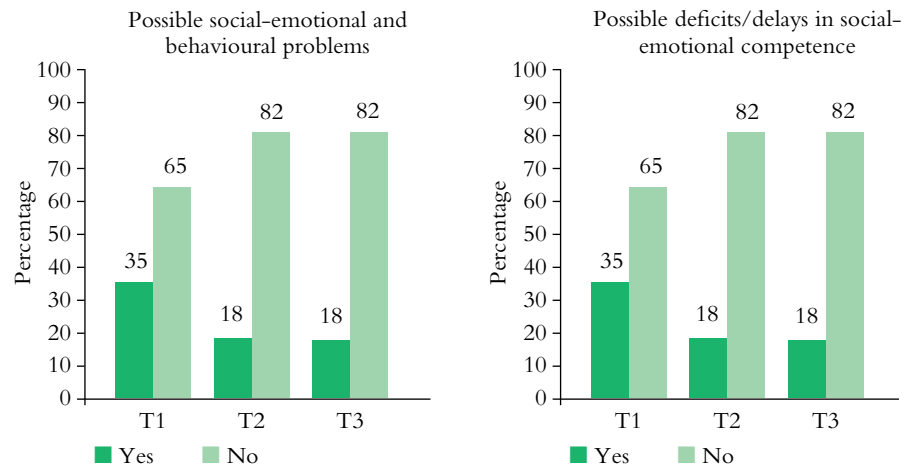
At T1, the majority of parents (n=71, 85 per cent) reported high levels of support from a significant other. At T2, there was a very slight increase (by three parents) in the number of parents reporting high levels of support from a significant other, and this dropped by two parents at T3. So, in essence, there was minimal change over the three time-points (perhaps due to a ceiling effect – most people already felt well supported before they began the intervention). The Marginal Homogeneity tests confirmed that among the 84 parents, there were no significant changes in levels of support from a significant other reported by parents between T1 and T2 ($T = 24, p = .371$) or between T2 and T3 ($T = 36, p = .465$).

Figure 2i: Change in total support (n=84)



At T1, 37 parents (44 per cent) reported high levels of total support, while just over a quarter (n=22; 26 per cent) reported having low levels of total support. At T2, these numbers had improved, with 58 per cent reporting high levels of total support and only 19 per cent reporting low levels of support. There was virtually no change in these numbers at T3. The Marginal Homogeneity test confirmed that among the 84 parents, there was a significant change between T1 and T2 in the proportion of parents falling into each of the three categories ($T = 65, p = .020$). From the previous tables, it appears most of the change was an improvement in the level of total support. There was no significant change between T2 and T3 ($T = 46, p = .763$), indicating that the initial changes were maintained over time.

Figure 2j: Change in BITSEA (n=17)



At T1, six children (35 per cent) were classified as having a possible problem on the BITSEA. This roughly halved by T2 (18 per cent) and stayed stable at T3. The Cochran's Q test found that among the 17 parents, PuP did not result in a significant change in the proportion of their children considered to have a possible problem across the three time-points ($\chi^2(2) = 2.57, p = .276$). Repeating the analysis using a related samples McNemar test comparing T1 and T2 did not reveal any significant differences either ($\chi^2(1) = 0.57, p = .453$).

At T1, six children (35 per cent) were classified as having a possible delay/deficit on the BITSEA. This roughly halved by T2 (18 per cent) and stayed stable at T3. The Cochran's Q test found that among the 17 parents, there was a borderline significant reduction in the proportion of their children considered to have a possible delay/deficit across the three time-points ($\chi^2(2) = 6.00, p = .050$). However, repeating the analysis using a related samples McNemar test comparing T1 and T2 did not reveal a statistically significant change in those classified as having a possible delay/deficit between these two time-points ($\chi^2(1) = 1.33, p = .250$).

Individual reliable change across the primary outcome measures (DASS21, IMP-I, and BITSEA)

As no scale measures a concept 100 per cent accurately, it is important to check that the changes seen in the scores on a measure still occur if the error in the measuring tool is taken into account (for example, a parent's change score is statistically significantly greater than a difference that could have occurred due to random measurement error alone). This test is called the reliable change index (RCI). We were able to calculate the RCI for our primary outcomes (the DASS21, IMP-I, and BITSEA) as the scores on these measures could be categorised and we had information from other published papers on the measure's reliability. This then allows us to identify which parents showed reliable change during the course of PuP (at T1 vs T2, T2 vs T3 and T1 vs T3).

Change in an individual score is considered to be reliable/clinically significant if the RCI is greater than 1.96 (Jacobson & Truax, 1991). To calculate the RCI for each PuP evaluation measure, several steps were undertaken:

1. The standard deviation (SD) for the T1 variable was calculated.
2. The absolute values of the change scores were obtained to remove the minus signs (which indicate a reduction in scores from one time-point to the other).
3. The standard error of measurement (SEM) was calculated:
 $SEM = SD \cdot \sqrt{(1-\alpha)}$ where alpha is the internal reliability estimate (using Cronbach's alpha) associated with the measure that is available from the existing literature.
4. The standard error of difference (SEdiff) was calculated:
 $SEdiff = \sqrt{2 \cdot (SEM)^2}$.
5. The RCI was calculated by dividing the individual change scores by the SEdiff.
6. The RCI was recoded to obtain the number of parents with an RCI >1.96.

Depression, Anxiety and Stress Scale (DASS21)

In accordance with Harnett and Dawe (2008), the internal reliabilities (Cronbach's alphas) used to calculate the RCI for the Depression, Anxiety, and Stress subscales of DASS21 were 0.94, 0.87, and 0.91, respectively, as previously reported by Antony et al (1998). For the DASS21 total score, the conservative value of 0.90 was used, as proposed by Harnett and Dawe (2008). The proportion of parents who demonstrated a reliable change on each of the DASS21 subscales between each of the time-points is provided in Table 16. As can be seen, over a third of parents experienced significant reliable change in their levels of depression and stress during PuP, and around a quarter of parents made significant change in their anxiety and total emotional wellbeing.

Table 16: Parents' reliable change on the DASS21

DASS21 subscale (N=85)	Reliable change achieved for T1 vs T2 n (%)	Reliable change achieved for T2 vs T3 n (%)	Reliable change achieved for T1 vs T3 n (%)
Depression	25 (29.4)	19 (22.4)	33 (38.8)
Anxiety	11 (12.9)	5 (5.9)	19 (22.4)
Stress	22 (25.9)	19 (22.4)	33 (38.8)
Total	20 (23.5)	9 (10.6)	22 (25.9)

DASS21: Depression, anxiety and stress scale.

Interpersonal Mindfulness in Parenting (IMP-I) measure

Based on the Laurent et al (2017) paper on the IMP-I, the internal reliability (Cronbach's alpha) used to calculate the RCI for the total parental mindfulness score was 0.81. The proportion of parents who demonstrated a reliable change on the IMP-I between each of the time-points is provided in Table 17. In total, over a quarter of parents experienced reliable change in their levels of mindful parenting.

Table 17: Parents' reliable change on the IMP-I

N=79	Reliable change achieved for T1 vs T2 n (%)	Reliable change achieved for T2 vs T3 n (%)	Reliable change achieved for T1 vs T3 n (%)
Total	13 (16.5)	11 (13.9)	22 (27.8)

IMP-I: Interpersonal mindfulness in parenting scale- infant version.

BITSEA

Based on the Karabekiroglu et al (2010) paper on the BITSEA, the internal reliability (Cronbach’s alpha) used to calculate the RCI for the problem subscale was 0.80, and for the competence subscale it was 0.69. The proportion of children who demonstrated a reliable change on the BITSEA between each of the time-points is provided in Table 18. Around a quarter of parents reported significant reliable change in their child’s social-emotional and behavioural problems and competencies during PuP.

Table 18: Parents’ reliable change on the BITSEA

BITSEA subscale N=17	Reliable change achieved for T1 vs T2 n (%)	Reliable change achieved for T2 vs T3 n (%)	Reliable change achieved for T1 vs T3 n (%)
Problem	4 (23.5)	0 (0)	5 (29.4)
Competence	2 (11.8)	0 (0)	4 (23.5)

BITSEA: Brief Infant-Toddler Social and Emotional Assessment.

For the DASS21 and the BITSEA, we were able to assess whether the reliable change was clinically meaningful between T1 and T3 by looking at whether the change in their scores meant the parent or child had moved from a clinical to a sub-clinical level of need (referred to here as ‘recovered’ for ease of interpretation), whether they clinically ‘deteriorated’ (their needs rose to a clinical level from a sub-clinical level over the duration of the programme), or whether they stayed in the same clinical category (despite the change in their scores this did not place them into a different clinical/sub-clinical bracket). This data can be seen in Table 19 and shows that the majority of parents who made reliable changes in their emotional wellbeing made significant improvement and often moved from a clinical to a sub-clinical range.

However, there were a small number of parents whose emotional wellbeing worsened during PuP and some of these went from a sub-clinical to clinical level of need. For the BITSEA, most of the change represented an improvement in social-emotional and behavioural functioning, but there was one child for whom the parent reported clinical deterioration in their social-emotional and behavioural problems.

Table 19: Reliable and clinical change in DASS21 and BITSEA scores from the start to the end of PuP

	DASS21			BITSEA ¹		
	Depression n (%)	Anxiety n (%)	Stress n (%)	Problems n (%)	Competencies n (%)	
Made reliable and clinically meaningful change from T1 to T3	33 (39)	19 (22)	33 (39)	5 (29)	4 (24)	
Recovered						
Improved by 1 category	11 (33)	5 (27)	18 (55)	3	2	
Improved by 2 categories	8 (24)	12 (63)	5 (15)			
Improved by 1 category but still in clinical range	3 (9)	1 (5)	3 (9)			
Statistically significant, but not clinically meaningful, change in scores						
Started in a clinical (moderate/severe) range	2 (6)		0	1	0	
Worsened	1 (3)			0	0	
Started in sub-clinical (normal/mild) range	4 (12)	0	6 (18)	0	2	
Worsened	2 (6)		1 (3)	0	0	
Deteriorated						
Deteriorated by 1 category	2 (6)	1 (5)	0	1	0	

¹Percentages are not reported for the proportion of parents reporting reliable and clinically meaningful change for their children on the BITSEA due to the small numbers involved.

DASS21: Depression, anxiety and stress scale. BITSEA: Brief Infant-Toddler Social and Emotional Assessment.

Recovered: made statistically significant reliable change and clinical classification improved from clinically moderate/severe/extremely severe on the DASS21 or from being in a clinical to a sub-clinical range on the BITSEA.

Deteriorated: made statistically significant reliable change that indicated a clinically relevant worsening of emotional symptoms or went from a sub-clinical to a clinical range of problems or deficits/delays on the BITSEA.

Statistical, but not clinically meaningful, change: scores showed the parent made statistically significant reliable change but the reduction/increase in scores did not move them into the next clinical/sub-clinical bracket.

For the IMP-I, we are able to explore whether the parent moved between a low/moderate/high bracket to show the extent of change in their levels of mindful parenting. However, we are not able to consider clinically relevant levels of mindful parenting as there are no population norms for comparison. These findings can be seen in Table 20 and show that all but one of the parents who evidenced reliable change in their levels of mindful parenting improved in this area – more than half of whom made an increase that placed them in a higher category of scores. However, one parent showed deterioration in their levels of mindful parenting.

Table 20: Reliable change in categorical classification of mindful parenting during the PuP programme

		Mindful parenting n (%)	
Made reliable and clinically meaningful change from T1 to T3		22 (28)	
Improve to moderate or high levels of support	Improved by 1 category	11 (50)	
	Improved by 2 categories	0	
Statistically significant change but do not change in category	Started as low	Improved	0
		Deteriorated	0
	Started as moderate	Improved	10 (45)
		Deteriorated	1 (5)
	Started as high	Improved	0
		Deteriorated	0
Deteriorated	Deteriorated by 1 category	0	
	Deteriorated by 2 categories	0	

Question 4. What factors are associated with the change experienced by parents and children during PuP?

Change scores were calculated to reflect the amount of change made by parents on each measure from T1 to T2, T2 to T3, and T1 to T3. We hoped to use these change scores to explore predictors of change on the measures over time. However, the change scores were significantly skewed, and we were unable to transform them using square root, logarithmic and reciprocal transformation (scores were reversed prior to transformation given that many fell below zero when the scores reduced on measures over time). This prevented us from exploring these variables as predictors of change given the data did not meet the assumptions for parametric tests. We therefore carried out non-parametric tests (Mann-Whitney) to explore whether certain variables were associated with parents making a greater amount of change during PuP (T1–T3). For example, we compared the amount of change made by parents who misused drugs versus those who abused alcohol or drugs and alcohol to see if one group made more change than the other. We did this for each of our primary outcomes (each DASS21 subscale and the IMP-I) and included only the parents who were found to have made significant reliable change on that outcome within the analysis. The numbers of parents included in each test therefore varied and are given in Tables 21 and 22 below.

In order to categorise the predictors that had interval data (for example, the child's age at the start of PuP), we grouped the children/parents into meaningful categories as described in Tables 21 and 22. Mann-Whitney tests were carried out for variables that had two categories (for example, drug abuse) and we used Kruskal-Wallis with post-hoc Mann-Whitney tests for those with three categories (for example, child's age). Median change scores were explored to determine which groups made the most significant change.

Note that a negative change score denotes a decrease in scores between T1 and T3.

Table 21: Factors associated with change in DASS21 subscale scores between T1 and T3 for the parents who showed reliable change for each sub-section

Variable	Depression			Anxiety			Stress						
	Median change score (IQR)	U	Z	p	Median change score (IQR)	U	Z	p	Median change score (IQR)	U	Z	p	
Parent characteristics													
Drug abuse (n = 33; 19; 33)	Drugs only	-12 (20)			-22 (5)				-13 (14)				
	Alcohol or drug and alcohol abuse	-14 (19)	130.50	-0.09	.927	-16 (8)	12.50	-2.65	.008	-18 (14)	104.00	-1.13	.258
Psychopathology (based on severe or extremely severe symptoms in 1+ area of the DASS21 at assessment; n = 33; 19; 33)	Yes	-18 (17)	25.00	-3.81	<.001	-20 (8)	5.00	-1.62	.105	-20 (14)	25.00	-3.56	<.001
	No	-8 (18)			-2 (4)					-10 (2)			
Diagnosed mental health difficulty (n = 29; 17; 30)	Yes	-14 (20)	75.00	-0.71	.476	-20 (6)	28.00	-0.51	.610	-16 (13)	81.00	-0.33	.740
	No	-10 (26)			-18 (10)					-15 (18)			
Domestic violence (perpetrator and/or victim; n = 26; 16; 28)	Yes	-14 (20)	63.50	-1.00	.320	-20 (7)	28.50	-0.32	.717	-14 (9)	76.00	-0.84	.404
	No	-10 (8)			-22 (10)					-12 (14)			
Criminality (n = 26; 14; 25)	Yes	-14 (20)	80.00	0.00	1.000	-18 (7)	16.50	-0.51	.614	-18 (14)	65.50	-0.15	.883
	No	-12 (27)			-20 (22)					-16 (17)			
Single parenting (n = 33; 19; 33)	Yes	-13 (19)	125.00	-0.19	.853	-20 (8)	38.00	-0.09	.929	-14 (15)	122.50	-0.13	.895
	No	-12 (28)			-19 (10)					-19 (14)			
Parent has other children who have been removed from their care at assessment (n = 21; 12; 20)	Yes	0 (29)	19.50	-1.31	.192	-12 (4)	5.00	-1.61	.108	-19 (32)	30.00	-0.191	.849
	No	-14 (17)			-20 (6)					-13 (15)			

Median scores are presented as the data was not normally distributed. IQR = Interquartile Range. DASS21: Depression, anxiety and stress scale.

n = number of parents in the analysis for depression, anxiety and stress, respectively.

^aNo IQR could be calculated as there were just two people in this category. ^bNo IQR could be calculated as there were just three people in this category

Variable	Depression			Anxiety			Stress					
	Median (IQR)	χ^2	df	p	Median (IQR)	χ^2	df	p	Median (IQR)	χ^2	df	p
Child characteristics												
Age of child (<i>n</i> = 33; 19; 33)	Infant (0–12 months)	-8 (26)			-14 (11)**				-16 (11)			
	Toddler (13–24 months)	-12 (13)	2	.412	-22 (10)	6.23	2	.044	-12 (10)	5.71	2	.058
	25+ months	-22 (19)			-21 (4)**				-26 (18)			
Child was on a child protection plan at assessment (<i>n</i> = 33; 19; 33)	Yes	-10 (21)	-1.29	.198	-20 (6)	43.50	-0.13	.901	-14 (9)	118.00	-0.65	.513
	No	-15 (20)			-19 (11)				-20 (17)			
Programme-level factors												
Number of weeks open to NSPCC (<i>n</i> = 31; 18; 32)	20 weeks or less	-9 (a)			b				-18 (c)			
	21–30 weeks	-14 (20)	2	.560	-20 (10)	1.679	2	.432	-16 (13)	0.379	2	.827
	31+ weeks	-9 (20)			-20 (9)				-18 (14)			
Number of PuP sessions received ¹ (<i>n</i> = 33; 19; 33)	Below average	-			-				-			
	Average	-12 (20)	1	1.00	-16 (10)	1.490	1	.222	-18 (15)	0.065	1	.799
	Above average	-14 (18)			-21 (2)				-16 (8)			

Median scores are presented as the data was not normally distributed. IQR = Interquartile Range.

n = number of parents in the analysis for depression, anxiety and stress, respectively.

¹Below average number of sessions = less than the mean minus the SD; <7, average = mean +/- SD; 8–24, and above average = greater than the mean plus the SD; >25.

^aNo IQR could be calculated as there was just one person in this category. ^bThere was only one person in this category. ^cNo IQR could be calculated as there were just two people in this category.

******The significant difference was between the scores for parents of children aged 0–12 months and 25+ months: *U* = 2.00, *Z* = -2.59, *p* = .009.

Table 22: Factors associated with change in IMP-I scores between T1 and T3 for the parents who showed reliable change for each sub-section

Variable	Mindful parenting (IMP-I)				
	Median change score (IQR)	<i>U</i>	<i>Z</i>	<i>p</i>	
Parent characteristics					
Drug abuse (<i>n</i> =22)	Drugs only	22 (16)	49.50	-0.61	.545
	Alcohol or drug and alcohol abuse	21 (22)			
Psychopathology (based on severe or extremely severe symptoms in 1+ area of the DASS21 at assessment; (<i>n</i> =22)	Yes	28 (20)	48.50	-0.67	.501
	No	22 (17)			
Diagnosed mental health difficulty (<i>n</i> =18)	Yes	21 (14)	38.00	-0.22	.823
	No	28 (23)			
Domestic violence (perpetrator and/or victim; (<i>n</i> =17)	Yes	22 (15)	25.00	-1.07	.286
	No	30 (21)			
Criminality (<i>n</i> =19)	Yes	28 (21)	39.50	-0.38	.707
	No	22 (8)			
Single parenting (<i>n</i> =22)	Yes	22 (22)	54.00	-0.30	.762
	No	21 (16)			
Parent has other children who have been removed from their care at assessment (<i>n</i> =12)	Yes	22 (*)	8.50	-0.93	.352
	No	28 (22)			

IQR = Interquartile Range. *IMP-I*: Interpersonal mindfulness in parenting scale- infant version.

*No IQR could be calculated as there were just three people in this category.

Variable	Mindful parenting (IMP-I)				
	Median (IQR)	χ^2	df	p	
Child characteristics					
Age of child (n=22)	Infant (0–12 months)	22 (20)	2.54	2	.282
	Toddler (13–24 months)	30 (22)			
	25+ months	20 (12)			
Child was on a child protection plan at assessment (n=22)	Yes	22 (21)	56.50	-0.232	.816
	No	23 (14)			
Programme-level factors					
Number of weeks open to NSPCC (n=22)	20 weeks or less	^a	0.071	1	.790
	21–30 weeks	21 (19)			
	31+ weeks	25 (22)			
Number of PuP sessions received ¹ (n=22)	Below average	^a	2.279	2	.320
	Average	23 (21)			
	Above average	20 (^b)			

IQR = Interquartile Range. *IMP-I*: Interpersonal mindfulness in parenting scale - infant version.


¹Below average number of sessions = less than the mean minus the SD; <7, average = mean +/- SD; 8–24, and above average = greater than the mean plus the SD; >25. ^aThere was only one person in this category. ^bNo IQR could be calculated as there were just three people in this category.

As can be seen in Tables 21 and 22, drug abuse (as opposed to alcohol abuse or alcohol *and* drug abuse), psychopathology and the age of the child were all significantly associated with the extent of change in parents' emotional wellbeing during PuP. However, none of the programme-level factors that were explored predicted change in this area and we did not identify any predictors of change in levels of mindful parenting.

Question 5. Does the safeguarding status of children change during PuP and are there any predictors of change?

In order to classify change in safeguarding status, a coding system was developed to determine the ranking of children’s services involvement with a family, from the least intrusive to the most intrusive (ie, increased surveillance on the family). This is described in Table 23. An assumption was made that when a parent ticked ‘parenting assessment order’, this was in reference to a legal order and therefore greater in severity than a child protection plan. Where a parent ticked more than one box regarding current safeguarding status, we coded the intervention that was most intrusive/had the greatest level of surveillance in our analysis. When the data was taken from the NSPCC’s case recording system at the end of PuP (for parents that were not part of the RCT), we recorded which of these children’s services interventions were present and again recorded that which was the most intrusive. There are limitations with our approach to collecting safeguarding status information for our evaluation, which is discussed in greater detail in the discussion of the main [service evaluation report](#) (Page 50).

Table 23: Ranking of safeguarding status

Children’s services intervention ranked from least to most intrusive	
	Team around the child
	Child in Need (s17)
	Child Protection Plan (s47)
	Parenting Assessment Order
	Pre-legal proceedings
Legal Proceedings	Care order (no PR) interim
	Supervision order (PR remains with parent)
	Special guardianship order (child lives with family member; parents still have contact)
	Care order (no PR) long term

PR = Parental Responsibility

We explored whether any factors associated with the parent or their relationship were predictive of change in safeguarding status from the start to the end of PuP. We first explored whether five factors hypothesised to influence safeguarding status were significantly associated with safeguarding status. This was done for all of the parents in the sample who we have information on regarding change in safeguarding status. The findings from this analysis are outlined in Table 24.

Table 24: Factors associated with change in safeguarding status during PuP

Predictor		Reduced surveillance		Increased surveillance		No change		Significant difference		
		n	%	n	%	n	%	χ^2	df	p
Psychopathology at T1 (n=144)	Yes	12	22	12	22	30	56	1.43	2	.489
	No	28	31	16	18	46	51			
Opiate replacement treatment (n=131)	Yes	23	28	21	26	38	46	4.98	2	.083
	No	14	29	5	10	30	61			
Parent has other children who are not currently in their care (n=101)	Yes	12	44	3	11	12	44	5.50	2	.064
	No	17	23	20	27	37	50			
Single parenting (n=139)	Yes	22	31	15	21	34	48	0.891	2	.640
	No	18	27	12	18	38	56			
Having experienced or perpetrated domestic violence (n=120)	Yes	20	30	11	16	36	54	0.011	2	.994
	No	16	30	9	17	28	53			

None of these variables were found to be significantly associated with change in safeguarding status.

Table 25 outlines the change in safeguarding status for parents who did and did not complete PuP. The figures suggest that more of the parents who did not complete PuP experienced increased surveillance from children's services by the point at which they disengaged with the programme than those who did complete PuP. Fewer of them experienced a reduction in surveillance.

Table 25: Change in safeguarding status for the parents who did (n=96) and did not (n=47) complete PuP and for whom we could assess change

	Reduced surveillance/ child returned to their care		No change		Increased surveillance/ child removed from their care	
	n	%	n	%	n	%
Parent completed PuP	34	35	49	51	13	14
Parent did not complete PuP	5	11	27	57	15	32

Question 6: What do we know about the fathers enrolled in PuP?

We know very little about the fathers who take part in PuP, and the inclusion of 10 fathers within our evaluation of the programme therefore permitted a case study of the profile of these fathers when the programme began. While this provides us with some insight into their needs, it is important to stress the small number of fathers included here and how this limits our ability to generalise the findings beyond this small sample.

Table 26 shows the levels of emotional wellbeing, alcohol use and support reported by fathers at the start of PuP. More of the fathers reported normal/mild levels of stress, anxiety and depression than moderate to extremely severe levels, similar to the pattern identified in the full sample of parents. Five of nine fathers reported safe/low risk alcohol consumption and most reported high levels of support from their family and a significant other. Support from friends appears to be mixed.

Table 26: Categorisation of fathers' scores for emotional wellbeing, alcohol consumption and social support at T1 (n=10)

Time 1 (assessment)				
DASS21	N	Normal/mild (N)	Moderate/ severe (N)	Extremely severe (N)
Stress	10	8	2	0
Depress	10	7	3	0
Anxiety	10	8	2	0

AUDIT	N	Safe/low risk (N)	Alcohol abuse (N)	Alcohol dependence (N)
Total score	9	5	3	1

MSPSS	N	Low support (N)	Mid-range (N)	High support (N)
Family support	10	3	1	6
Friends support	10	5	1	4
Significant other support	10	1	1	8
Total score	10	2	4	4

DASS21: Depression, anxiety and stress scale. *AUDIT-C*: Alcohol Use Disorders Identification. *MSPSS*: Multidimensional Scale of Perceived Social Support.

Conclusion

The evaluation findings highlight that parents who misuse substances and access a parenting programme like PuP are likely to have complex needs and experience a range of multiple adversities. With support from this parenting programme, however, the parents were able to demonstrate change in all of the primary and secondary outcomes assessed (with the exception of child social-emotional and behavioural difficulties).

References

- Antony, M.M., Bieling, P.J., Cox, B.J., Enns, M.W. & Swinson, R.P. (1998) Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment*, 10, pp. 176–181.
- Briggs-Gowan, M.J., Carter, A.S., Irwin, J.R., Wachtel, K. & Cicchetti, D.V. (2004) The Brief Infant -Toddler Social and Emotional Assessment: screening for social-emotional problems and delays in competence. *Journal of Pediatric Psychology*, 29, pp. 143–55.
- Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, USA.
- Crníc, K.A. & Greenberg, M.T. (1990) Minor Parenting Stresses with Young Children. *Child Development*, 61, pp. 1,628–1,637.
- Department of Health (2000) *Framework for the Assessment of Children in Need and their Families*. The Stationery Office, London.
- Early Intervention Foundation (2017) *EIF Evidence Standards*. Retrieved from <http://guidebook.eif.org.uk/eif-evidence-standards>. Accessed on 15/11/2017.
- Gossop, M., Darke, S., Griffiths, P., Hando, J., Powis, B., Hall, W. & Strang, J. (1995) The Severity of Dependence Scale (SDS): psychometric properties of the SDS in English and Australian samples of heroin, cocaine and amphetamine users. *Addiction*, 90, pp. 607–614.
- Harnett, P.H. & Dawe, S. (2008) Reducing child abuse potential in families identified by Social Services: Implications for assessment and treatment. *Brief Treatment & Crisis Intervention*, 8, pp. 226–235.
- Jacobson, N. & Truax, P.J. (1991) Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting & Clinical Psychology*, 59, pp. 12–19.
- Karabekiroglu, K., Briggs-Gowan, M.J., Carter, A.S., Rodopman-Arman, A. & Akbas, S. (2010) The clinical validity and reliability of the Brief Infant–Toddler Social and Emotional Assessment (BITSEA). *Infant Behavior & Development*, 33, pp. 503–509.
- Laurent, H.K., Duncan, L.G., Lightcap, A. & Khan, F. (2017) Mindful Parenting Predicts Mothers' and Infants' Hypothalamic-Pituitary-Adrenal Activity During a Dyadic Stressor. *Developmental Psychology*, 53, pp. 417–424.

- Lovibond, S.H. & Lovibond, P.F. (1995) *Manual for the Depression Anxiety Stress Scales*. (2nd edition). Psychology Foundation, Sydney.
- Meneses-Gaya, C., Zuardi, A.W., Loureiro, S.R., Hallack, J.E.C., et al. (2010) Is the full version of the AUDIT really necessary? Study of the validity and internal construct of its abbreviated versions. *Alcoholism: Clinical & Experimental Research*, 34, pp. 1,417–1,424.
- Power, M.J., Champion, L.A. & Aris, S.J. (1988) The development of a measure of social support: the Significant Others (SOS) Scale. *British Journal of Clinical Psychology*, 27, pp. 349–358.
- Sherin, K.M., Sinacore, J.M., Li, X., Zitter, R.E. & Shakil, A. (1998) HITS: A Short Domestic Violence Screening Tool for Use in a Family Practice Setting. *Family Medicine*, 30, pp. 508–12.
- Zimet, G.D., Dahlem, N.W., Zimet, S.G. & Farley, G.K. (1988) The Multidimensional Scale of Perceived Social Support. *Journal of Personality Assessment*, 52, pp. 30–41.