July 2025 | **Report**

SELECTING AND VALIDATING OUTCOME MEASURES FOR THE DOMESTIC VIOLENCE AND ABUSE CORE OUTCOME SET (DVA-COS)

Work Package 2: Psychometric assessment of the Short Warwick-Edinburgh Mental Wellbeing Scale

Foundations
What Works Centre for Children & Families
foundations.org.uk

Authors

Jenna Harewell, University of Sussex

Elizabeth Dunk, University of Sussex

Shivi Bains, University of Sussex

Emma Howarth, University of Sussex

Claire Powell, University College London

Lazaros Gonidis, University of Sussex

Acknowledgements

We would like to thank members of the Expert Advisory Group: Amanda McIntyre, Professor Christine Barter, Elaine Fulton, Dr Estela Capelas Barbosa, Dr Eszter Szilassy, Professor Gene Feder, Dr Melissa Kimber, Dr Simona Skripkauskaite, Professor Yo Jackson, and Dr Trina Moseley. We also say thank you to representatives from VOICES, for their role as the Lived Experiences Advisory group. We'd also like to thank SafeLives' Changemakers, Jasmine and Latoya, as well as everyone with lived experience of domestic abuse who took part in the study – we really appreciate you giving your time.

We gratefully acknowledge the OxWell study team[1], the many students who contributed their time and experiences, and the OxWell study's funders, including the NIHR Applied Research Collaboration Oxford and Thames Valley, for supporting this research.

We would additionally like to thank the domestic abuse service and its servicer users, who wished to remain anonymous, for their time and contribution to this work.

Further thanks to Professor Sally McManus for offering time and insight into this report. Finally thank you to colleagues from Foundations – What Works Centre for Children & Families for their support.

Funding and competing interests

Foundations provided funding to the University of Sussex to deliver the research study. Foundations is funded by the Department for Education, England.

About Foundations, the national What Works Centre for Children & Families

Foundations, the national What Works Centre for Children & Families, believes all children should have the foundational relationships they need to thrive in life. By researching and evaluating the effectiveness of family support services and interventions, we're generating the actionable evidence

[1] See: https://osf.io/sekhr

needed to improve them, so more vulnerable children can live safely and happily at home with the foundations they need to reach their full potential.

About University of Sussex

For over 60 years the aim of University of Sussex's courses, research, culture, and campus has been to stimulate, excite, and challenge. From scientific discovery to global policy, from student welfare to career development, the University of Sussex innovates and takes a lead. In every part of society and across the world, you will find someone from the University of Sussex making an original and valuable contribution. Visit www.sussex.ac.uk

To cite this report:

Harewell, J., Dunk, E., Bains, S., Howarth, E., Powell, C. & Gonidis, L. (2025) *Selecting and Validating Outcome Measures for the DVA-COS. Work Package 2: Psychometric Assessment of the Short Warwick–Edinburgh Mental Wellbeing Scale*. Foundations, UK.

ABBREVIATIONS & ACRONYMS

Abbreviation/acronym	Description
APMS	Adult Psychiatric Morbidity Survey
CFA	Confirmatory factor analysis
CFI	Comparative Fit Index
CIS-R	Clinical Interview Schedule – Revised
COS	Core outcome set
COSMIN	COnsensus-based Standards for the selection of health Measurement INstruments
DVA	Domestic violence and abuse
OMI	Outcome measurement instrument
RCADS	Revised Children's Anxiety and Depression Scale
RMSEA	Root mean square error of approximation
SPSS	Statistical Package for the Social Sciences
SRMR	Standardised root mean square residual
SWEMWBS	Short Warwick–Edinburgh Mental Wellbeing Scale
TLI	Tucker-Lewis Index

WEMWBS	Warwick–Edinburgh Mental Wellbeing Scale
WLSMV	Weighted least squares mean and variance adjusted

CONTENTS

EXECUTIVE SUMMARY

Background

The domestic abuse core outcome set (DVA-COS) is an agreed set of five outcomes intended for use in evaluations of interventions or services for children and families with experience of domestic violence and abuse (DVA, hereafter referred to as domestic abuse). A COS is a minimum standard for measurement in intervention studies, the purpose of which is to overcome heterogeneity in outcome selection and measurement. The aim of a COS is to maximise the value of a body of evidence by facilitating comparison between and synthesis across studies, thus reducing research wastage. Since the development of the DVA-COS, work has been undertaken to identify, select, and validate outcome measurement instruments (OMIs) to measure the core outcomes. The Warwick–Edinburgh Mental Wellbeing Scale (WEMWBS) was previously identified as acceptable by stakeholders to capture two outcomes: *child and caregiver emotional health and wellbeing*. This work seeks to extend those findings by validating the measure for use with domestic abuse-experienced populations.

Aims

Foundations, the national What Works Centre for Children & Families, commissioned two work packages to develop and integrate previous work to outline and validate OMIs for use to assess outcomes comprising the DVA-COS. Work package 1 seeks to identify three OMIs, and this report focuses on work package 2, which aimed to validate the Short WEMWBS (SWEMWBS) for use with children and young people (aged 11 to 18) who have experienced domestic abuse. The studies that make up this work package used mixed methods to examine the acceptability, content validity, structural validity, internal consistency, and measurement invariance for the scale in children and young people experiencing domestic abuse. We also report a validation study of the WEMWBS for adults with experience of domestic abuse.

Methods

The above aims were addressed across four individual studies: two planned and two supplementary. First, a qualitative 'think aloud' study assessed the acceptability of the SWEMWBS with children and young people who had experienced domestic abuse. The remaining three studies were quantitative analyses of secondary data on using the SWEMWBS and WEMWBS with children and young people and adult samples.

- Study A: a qualitative think aloud study that involved interviews and a focus group to gather feedback from children and young people with domestic abuse experience on use of the SWEMWBS.
- Study B: examined cross-sectional data collected by the OxWell Student Survey to validate the SWEMWBS with children and young people affected by domestic abuse.
- Study C: examined anonymised longitudinal service data to validate the SWEMWBS with children and young people affected by domestic abuse.

- Study D: validated the WEMWBS with adults who have experienced domestic abuse using cross-sectional data from the Adult Psychiatric Morbidity Survey (APMS).

Key findings

Our findings demonstrate the validity and acceptability of the SWEMWBS and WEMWBS in domestic abuse-experienced child and adult populations respectively. Study A indicated that the SWEMWBS is broadly acceptable for use with children and young people, while raising important considerations regarding respondents' interpretation of the measure's items as well as the emotional impact of the measure on this population. Studies B and C demonstrated robust psychometric validity[2] of the SWEMWBS with children and young people affected by domestic abuse, and Study D showed robust psychometric validity of the WEMWBS with adult victims of domestic abuse.

These are significant findings given the limited number of measures that have been evaluated for use with this population across practice and research contexts. Moreover, this represents an important step forward in the implementation of the DVA-COS, which we hope will help to unify outcome measurement in domestic abuse research and evaluation, as well as service monitoring.

Recommendations

We recommend that the SWEMWBS and WEMWBS be used to measure wellbeing in the context of evaluation studies (of any quantitative design) seeking to assess the impact of child-focused domestic abuse interventions. To enhance the acceptability of the measure to children and adults we suggest minor adaptations for use in the domestic abuse context. Finally, we recommend the development of guidelines for practitioners and researchers about how to use the tools in a 'care-first' way and how to guard against the tools being used for screening or triaging, or rationing care, as well as guidance for commissioners on how to interpret and use evidence, generated by the completion of the SWEMWBS and WEMWBS, for the basis of decision making. This guidance needs to reflect the balance between the benefits of data-driven decision making and the risk of unduly narrowing the breadth of services or thwarting innovation in the sector. The OMI's implementation (including the use of guidance) should be closely monitored and evaluated, to inform any associated refinements and to develop an in-depth understanding of the process and outcomes associated with embedding routine measurement in practice. Further work is also required to identify an alternative OMI or adapt the SWEMWBS for appropriate use with children under the age of 11.

[2] Psychometric validity refers to whether a psychometric scale (such as the SWEMWBS) measures what it purports to measure. See Kimberlin and Winterstein (2008) for more detail on validity and reliability of measurement instruments.

BACKGROUND

It is widely recognised that domestic violence and abuse (hereafter referred to as domestic abuse) is common and can have long-term health and wellbeing consequences for children and their families (Evans et al., 2008; Vu et al., 2016; Walker-Descartes et al., 2021). In 2023, over 800,000 children and over 450,000 adults in England and Wales experienced domestic abuse (Foundations, 2023).

Children can be deeply affected by domestic abuse even without direct physical harm or witnessing abuse firsthand. Simply knowing that a trusted caregiver is experiencing domestic abuse can cause significant stress. This exposure, whether through its aftermath, a sibling's account, or changes in parenting, is increasingly recognised as a form of maltreatment, such as a form of emotional abuse or within the wider umbrella of adverse childhood experiences (Callaghan et al., 2018; Holden, 2003; Katz et al., 2020; Lawson, 2019; Macmillan et al., 2009).

Children exposed to domestic abuse are two to four times more likely to experience significant mental health issues, including anxiety, depression, aggression, and trauma symptoms, compared to children who have not been exposed (Kitzmann et al., 2003). Even where difficulties do not meet diagnostic criteria, they can cause substantial distress and impairment. Early adjustment difficulties, particularly behaviour problems, partly mediate the link between childhood domestic abuse exposure and negative adult outcomes (Dargis and Koenigs, 2017; Springer et al., 2003).

In 2021, the UK introduced a landmark Domestic Abuse Act (Domestic Abuse Act, 2021) that recognised children as primary victims of domestic abuse who may require support (Carlisle et al., 2024), thus increasing the policy imperative to offer acceptable and effective interventions. However, the evidence for effective interventions that aim to address the impact on children and their families is limited (Allen et al., 2022). The usefulness of existing evidence is in part restricted by the variety of outcomes measured and the range of outcome measurement instruments (OMI) used in the context of evaluative research, which makes it difficult to compare interventions or to synthesise findings across studies. Systematic reviews repeatedly highlight these challenges and recommend greater consistency in outcome measurement and reporting (Hameed et al., 2020; Livings et al., 2023; Weeks et al., 2024).

The role of core outcome sets

One way to address the challenge of outcome priority and diversity is to develop a core outcome set (COS) – a small number of outcomes that service users/survivors, practitioners/service providers, commissioners, policy makers, and researchers agree are the most important to be measured in academic research and programme evaluation (Williamson et al., 2012, 2017). Widespread use of a COS can improve the quality of evidence by increasing consistent measurement of outcomes and reducing reporting bias (Kirkham et al., 2013).

Funded by the National Institute for Health and Care Research via the Children and Families Policy Research Unit, in 2019 an adapted core outcome methodology was used to develop a COS for use in evaluating targeted psychosocial interventions aimed at improving outcomes for children exposed to domestic abuse. Following a two-year consensus process involving over 300 survivors

of domestic abuse, practitioners, and researchers, we identified five outcomes to be included in the COS: 1) *child emotional health and wellbeing*; 2) *feelings of safety*; 3) *caregiver emotional health and wellbeing*; 4) *family relationships*; 5) *freedom to go about daily life* (Howarth et al., 2021; Powell, Feder, et al., 2022; Powell et al., 2023, 2025). The outcome set represents a minimum measurement standard for quantitative evaluation of child-focused domestic abuse interventions (Krause et al., 2021). The expectation is that these outcomes would be reported in trials and practice-based evaluations, and where certain outcomes are not considered relevant to a particular intervention, the rationale for not measuring them would be reported.

Despite the promise of COSs for improving evidence quality, much of their potential impact has not yet been realised; studies show use in trials and systematic reviews to be low (Hughes et al., 2022; Williamson et al., 2022). One of the key barriers to COS uptake is a lack of guidance on *how* to measure outcomes. Therefore, for the DVA-COS, or any other, to make a material impact on the quality of evidence on effectiveness, it is critical to identify OMIs that can be used to assess outcomes in the context of research and evaluation. Published guidance provides a standardised process by which OMIs for outcomes included in a COS should be selected (Prinsen et al., 2016).

A number of reviews have highlighted a lack of well-validated tools designed or validated for use in the domestic abuse field (O'Doherty et al., 2014), with the greatest number of tools tending to focus on the measurement of domestic abuse itself (e.g. types of abusive behaviour, severity, cessation), rather than on broader outcomes such as wellbeing (Carlisle et al., 2024, 2025). The same can also be said of adjacent fields and associated literature, such as child maltreatment (Fallon et al., 2010; Georgieva et al., 2023; Saini et al., 2019). Moreover, many of the tools used in the domestic abuse field have been widely criticised for having limited evidence with respect to psychometric properties, acceptability, and feasibility (O'Doherty et al., 2014).

In 2021 the Home Office funded a rapid review of OMIs used in practice settings that mapped to the outcomes of the DVA-COS (Clark et al., 2023; Powell, Feder, et al., 2022). Using an adapted consensus process involving survivors and practitioners, the study identified the Short Warwick–Edinburgh Mental Wellbeing Scale (SWEMWBS) and the WEMWBS as acceptable tools to measure child and caregiver emotional health and wellbeing, respectively (Powell, Feder, et al., 2022). These tools were recommended for further exploration for three key reasons. First, they were well validated and being used in intervention trials. Second, both tools were already being used by two national domestic abuse organisations, Women's Aid and SafeLives (Stanley et al., 2021), and therefore, frontline services were familiar with them. Third, Powell, Clark, et al (2022) also found that survivors appreciated that the WEMWBS and SWEMWBS were brief, as well as clearly and positively worded (Clark et al., 2023). While the work of Powell and colleagues (2022) identified some minor concerns among participants, these were mainly over its acceptability with younger age groups due to inaccessible language. Collectively, the measures were promising as OMIs for the DVA-COS, with minor concerns deemed surmountable through the addition of a free text box and careful trauma-informed guidance for the use of the tool in research and practice settings.

Overview of the WEMWBS

The WEMWBS was developed in 2007 as a positive measure of mental wellbeing that could be used with the general population (Tennant et al., 2007). Positive mental health and wellbeing is

comprised of two distinct aspects: the hedonic (also known as subjective wellbeing), which focuses on feeling good, and the eudemonic (also known as psychological functioning), which focuses on functioning well (Ryan and Deci, 2001). The WEMWBS scale is comprised of 14 items that cover both facets of mental wellbeing – hedonic (e.g. feelings of optimism, cheerfulness, relaxation) and eudemonic (e.g. energy, clear thinking, self-acceptance, personal development, competence, and autonomy) (Tennant et al., 2007). Although the scale was initially validated with an adult population, it was later validated for children and young people aged 13+ (Clarke et al., 2011). Moreover, it was identified in a review by Waite and Atkinson (2021) as one of the few suitable measures for universal mental health and wellbeing screening among secondary-school-aged children and young people.

To provide a more accurate interval scale for mental wellbeing, a shorter, seven-item version (SWEMWBS) was later developed using the Rasch measurement model and, although it is a more restrictive view because it focuses mostly on eudemonic wellbeing, it presents stronger psychometric properties (Stewart-Brown et al., 2009). Its brevity makes it less arduous to complete and it is also validated for use with younger children, aged 11+ (McKay and Andretta, 2017; Melendez-Torres et al., 2019).

The WEMWBS was initially created to measure wellbeing at the population level, but it has been shown to be responsive to change at both the group and individual level, so it is now used in a wide variety of healthcare and non-clinical settings (Maheswaran et al., 2012; Shah et al., 2021). Large-scale population surveys, such as the Health Survey for England, have incorporated the WEMWBS to track mental wellbeing trends at a national level (Ng Fat et al., 2017). Additionally, the SWEMWBS has been used in randomised controlled trials evaluating a wide range of interventions and programmes aimed at improving mental health, including evaluating a wide range of interventions (Blodgett et al., 2022; Shah et al., 2021). The measure is also commonly used in the context of research with vulnerable groups such as looked-after children (Anthony et al., 2022). Moreover, through our work with domestic abuse services, we know that SWEMWBS is already partially implemented in some domestic abuse service settings as a means of gathering outcome data from adults and children (Powell, Feder, et al., 2022).

The WEMWBS and SWEMWBS have several key strengths that highlight their potential as OMIs for the DVA-COS. Most notable is the positive framing of items. As our previous work highlights, OMIs that are positively worded are favoured by individuals with domestic abuse experiences, with negatively worded items identified as potentially re-traumatising (Clark et al., 2023; Powell, Feder, et al., 2022; Shah et al., 2021). Furthermore, the growing use of the scales internationally and across disciplines highlights a general acceptance of the scales (e.g. Arnull and Stewart, 2021; Bartram et al., 2011, 2013; Cilar et al., 2020; Dong et al., 2016; Fung, 2019; Konaszewski et al., 2021; Koushede et al., 2019; Paton et al., 2023; Shah et al., 2021; Soraci et al., 2024; Trousselard et al., 2016; Yadav et al., 2025; Zayed et al., 2023). The scales are also very concise, making them a feasible addition to intervention evaluations, even where they are not the primary outcome. The SWEMWBS has demonstrated sensitivity to change, particularly in individuals with lower baseline wellbeing, making it suitable for evaluating the efficacy of interventions for individuals that have experienced domestic abuse, who are known to experience mental health difficulties at a higher rate than the general population (Blodgett et al., 2022; Evans et al., 2024; Oram et al., 2022; Trevillion et al., 2016).

Despite its strong psychometric properties and widespread use, the WEMWBS has some limitations. One critique is the psychometric impact of its positive wording, which, although it is beneficial in many contexts, may lead to ceiling effects – where individuals with already high wellbeing scores have limited room to show improvement (Melin et al., 2022). However, when used to evaluate interventions, ceiling effects are argued to be minimal, probably due to the nature of interventions targeting individuals with lower levels of wellbeing (Maheswaran et al., 2012). Additionally, the SWEMWBS, despite its brevity, places greater emphasis on functioning rather than subjective wellbeing, meaning it may not fully capture the emotional aspects of mental health (Stewart-Brown et al., 2009). Finally, and potentially linked to the functioning focus of the SWEMWBS, some service providers involved in our programme of work raised concerns that they observed defensive reporting when the SWEMWBS is initially completed by service users at the point of entry to a service. Defensive reporting is a form of social-desirability bias in which respondents respond in a way that is 'socially favoured', usually overestimating positive self-descriptions (Tourangeau et al., 2000). This can be an unconscious positive self-bias or an intentional attempt to conform to what is desirable. In the context of families who have experienced domestic abuse, it is possible that survivors may feel the need to indicate they are functioning well, owing to fear of the possible consequences of less optimal functioning, such as social service involvement and child protection proceedings (Rose et al., 2011). Research has demonstrated that social-desirability bias does not significantly impact reports of relationship satisfaction in intimate-partner violence contexts (Visschers et al., 2017) or in self-report wellbeing measures more generally (Caputo, 2017). However, there has not been empirical research (to our knowledge) exploring these wellbeing measurement concerns in the context of families who have experienced domestic abuse.

COnsensus-based Standards for the selection of health Measurement INstruments (COSMIN) guidance suggests that an OMI can be provisionally included in a COS if there is at least high-quality evidence for good content validity and for good internal consistency (if applicable), and if the OMI seems feasible (Prinsen, Vohra, Rose, et al., 2016). As we outline above, the WEMWBS and SWEMWBS meet these criteria for inclusion in the DVA-COS (Clarke et al., 2011; Koushede et al., 2019; Stewart-Brown et al., 2011; Yadav et al., 2025). However, studies that properly evaluate the measurement properties of the WEMWBS and SWEMWBS with domestic abuse populations are lacking.

It is important that an OMI is validated for the intended population to ensure there is no risk of measurement bias. Individuals with domestic abuse experience are a vulnerable and unique population. The social stigma that can be experienced among this population could result in specific sensitivity to the language used in the scale, as well as social desirability in responding. These potential differences in the interpretation and completion of the scale could lead to systematic error in data collection and assessment (Hays, 2008). Following from this, analysis of reliability is sample-dependent and, therefore, must also be assessed with the domestic abuse population to fully adhere to COSMIN guidance.

Current study

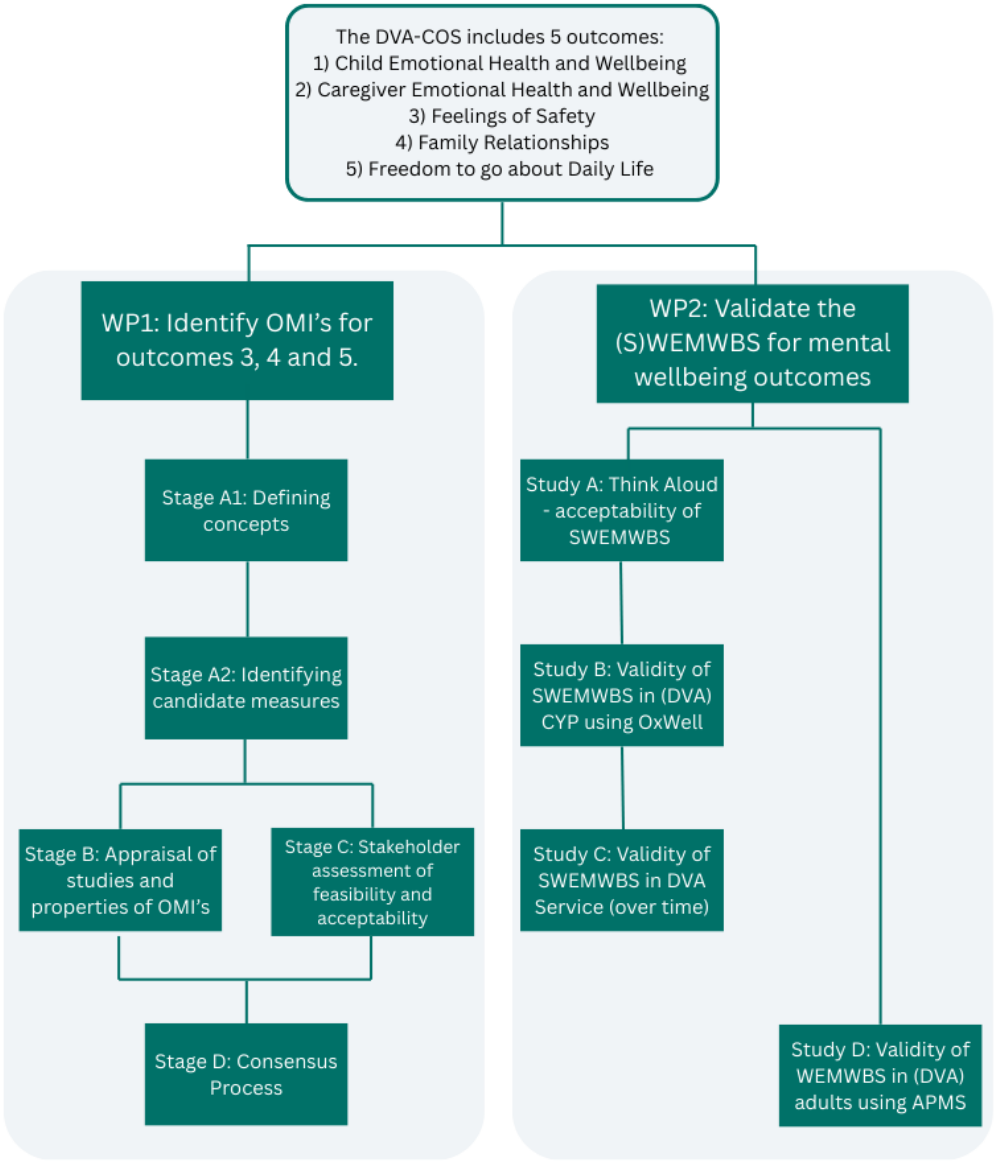
This study, funded by Foundations, the national What Works Centre for Children & Families, sought to consolidate and build on our previous work (outlined above), and was undertaken to identify and validate OMIs with which to measure the DVA-COS (Prinsen et al., 2016; Powell,

Clark, et al., 2022; Powell, Feder, et al., 2022). The current study was formed of two work packages:

Work package 1 aimed to identify OMIs to measure three of the five outcomes that comprise the COS: *feelings of safety, family relationships, freedom to go about daily life* (Bains et al., forthcoming).

Work package 2, reported here, aimed to validate the SWEMWBS for use with children and young people who have experienced domestic abuse. The studies comprising this work package sought to examine content validity, structural validity, internal consistency, and measurement invariance for the scale in children and young people experiencing domestic abuse. We used a mixed-methods design because this approach increases validity (McKim, 2017).

Figure 1: Schematic of full study (go to accessibility text)



The present research involved four individual studies: two planned and two supplementary (see Howarth et al. (2025) for the research protocol). These studies included: a qualitative think aloud study and a quantitative analysis of large-scale OxWell Student Survey data, both undertaken to evaluate the SWEMWBS with children and young people with experience of domestic abuse. The protocol identified two additional avenues for supplementary analyses: (1) to validate the Stirling Children's Wellbeing Scale as an additional OMI (for use with younger children) for the core outcome of child emotional wellbeing, and (2) to validate the WEMWBS with adults who have experienced domestic abuse. During this project, an opportunity was also presented to analyse the validity of the SWEMWBS with children and young people within domestic abuse services. As a result, the decision was made to include supplementary analyses validating the SWEMWBS with children and young people within domestic abuse services and validating the WEMWBS with adults who have experienced domestic abuse. Figure 2 gives an overview of each of the four studies.

In this programme of study, domestic abuse experience in children and young people refers to witnessing domestic abuse, while domestic abuse experience in adults refers to direct victimisation.

Figure 2: Overview of methods (go to accessibility text)

A – Think aloud study	• Focus group and interviews with young people with lived experience of domestic abuse • Framework and thematic analysis of transcripts
B – OxWell: Evaluating the psychometric properties of the SWEMWBS in children who have experienced domestic abuse	• Sample: 17,801 CYP aged 11–18 (1,215 had experienced domestic abuse) • Comparison of demographics and SWEMWBS scores across samples • Reliability: internal consistency, split-half reliability • Validity: convergent validity, factor analysis, measure invariance
C – Supplementary analyses: Evaluating the psychometric properties of the SWEMWBS with children in domestic abuse service settings	• Sample: 268 CYP who have partaken in a domestic abuse service • Comparison of SWEMWBS scores across service timepoints • Reliability: internal consistency, split-half reliability • Validity: factor analysis, measure invariance
D – Supplementary analyses: Evaluating the psychometric properties of the WEMWBS in adults who have experienced domestic abuse	• Rapid review for secondary data source • Sample: 6,621 adults aged 18–95+ (1,625 domestic abuse victims) • Comparison of demographics and WEMWBS scores across samples • Reliability: internal consistency, split-half reliability • Validity: convergent validity, factor analysis, measure invariance

Objectives

Study A: To explore the feasibility and acceptability of the SWEMWBS among children and young people who have experienced domestic abuse using think aloud methodology.

Study B: To evaluate the construct validity, internal consistency, measurement invariance, and convergent validity of the SWEMWBS in a cross-sectional sample of children and young people who have experienced domestic abuse.

Study C: To examine the construct validity, internal consistency, measurement invariance (across time), and responsiveness of the SWEMWBS with children and young people who have experienced domestic abuse through secondary analysis of data from a domestic abuse service provider.

Study D: To investigate the construct validity, internal consistency, measurement invariance, and convergent validity of the WEMWBS in a cross-sectional sample of adults who have experienced domestic abuse.

A. THINK ALOUD STUDY

Aim

Think aloud methodology asks individuals to verbalise their interpretation of an OMI (Boateng et al., 2018). This method was employed to assess the feasibility and acceptability of the SWEMWBS among young people with lived experience of domestic abuse. The study aimed to understand how respondents interpret and respond to the SWEMWBS items, identify potential challenges in comprehension, and explore emotional responses to the scale.

Method

Participants

Nine female participants aged 13–24 were recruited through previously established networks with domestic abuse organisations, whereby participants were members of lived experience groups or current service users. Participants represented diverse regions across the UK, including individuals from greater London, south-west England, and north-east England. Demographic data was collected for eight participants, with one participant not providing demographic data. A small sample size is acceptable for think aloud interviews, because informative data can be generated with as few as five participants (Patel-Syed et al., 2024).

Table 1. Age and ethnicity of participants

Characteristic	N	%
Age		
13–15 years	3	33%
16–18 years	3	33%
19–24 years	2	22%
Missing data	1	11%
Ethnicity		
White: English/Welsh/Scottish/Northern Irish/British	5	56%
Mixed/multiple ethnic groups: White and Asian	1	11%

Characteristic	N	%
Asian/Asian British: Indian	1	11%
Black/African/Caribbean background	1	11%
Missing data	1	11%

Procedure and materials

The study was initially advertised to young people through gatekeeper organisations. Each interested participant who made contact was offered an information sheet and an online briefing session with the researcher to ask any questions before consenting to participate. Participants were given the option of a focus group or an individual interview. One domestic abuse service arranged for an in-person focus group of five participants. The remaining four interviews were facilitated online via Microsoft Teams. Informed consent was given at the beginning of each session.

Each session began with participants completing the Short Warwick–Edinburgh Mental Wellbeing Scale (SWEMWBS) without intervention from the researcher. The time taken to complete the scale was recorded for each individual participant. Following completion of the SWEMWBS, the participant was asked a series of questions (see Appendix A) to identify difficulties they may have experienced when completing the scale. This approach has been used in similar studies and with a domestic abuse population (Evans et al., 2015; Tourangeau et al., 2000). The interview schedule addressed each item in turn and sought to understand how responder interpretation may differ from the scale developer's intention for the tool. The same questions were used in the focus group, and opportunity was given to each participant to provide their response. To conclude the session, participants were asked to rate each item on a scale of 0 to 10 to indicate how upset each item made them, 0 being not upset and 10 being extremely upset. The interviews and focus group ranged from 31 to 50 minutes in duration.

Data analysis

All interviews were audio recorded, either with a Dictaphone or on Microsoft Teams, and transcribed. Framework analysis was used to analyse the transcripts and was completed by two researchers using an Excel template. The framework included seven broad themes, with four themes reflecting the categories of difficulties as set out in Evans et al. (2015). These difficulties were:

1. Comprehension: the participant did not understand or was not certain how to interpret the meaning of the item
2. Recall: the participant was not confident on their memory of the scenario that fitted the corresponding item
3. Judgement: the participant was unable to assess the frequency within the timeframe
4. Response mapping: the participant was not able to respond accurately using the given response scale or they used this incorrectly.

The remaining three themes in the framework were contextual information, recommendations, and direct comments regarding the response scale. Contextual information included any relevant information that could not be categorised using one of the above difficulties – this was thematically analysed separately. After both researchers had individually analysed the transcripts, their analyses were compared, and any discrepancies addressed (see Appendix B for overview of analysis).

An acceptability score was calculated per item by averaging participants' rating for each item.

Ethics

Ethical approval was granted by the University of Sussex's Sciences and Technology Cross-School Research Ethics Committee (reference ER/EHH24/3 and ER/EHH24/6).

Informed consent was obtained from all participants as well as from parents of participants under the age of 15 years old, which is in keeping with Gillick competence[3] (Griffith, 2016). Consent forms, demographic questionnaires, and recordings were all stored separately to ensure anonymity. Audio recordings were deleted after transcripts were received and checked for accuracy. The transcripts were anonymised before analysis. Each participant was given time to debrief with the researcher or, where applicable, a representative of the domestic abuse service. They were also given a following support sheet with signposting to support services.

Results

Acceptability

The SWEMWBS took between one to three minutes for participants to complete, demonstrating it is acceptable because it was not overly burdensome. Largely, the SWEMWBS was acceptable, as indicated by low mean acceptability scores, meaning that the items broadly did not cause upset (see Table 2). The highest mean score was attributed to the item "I've been dealing with problems well". Participants who were not active service users ranked items as less upsetting than participants currently using services. Furthermore, active service users reported "I've been feeling close to other people" to be the most upsetting item.

[3] Children under the age of 16 can demonstrate they have the competence to understand the consequences of their decisions and therefore are able to provide their own consent.

Table 2. Item-level acceptability score per item*

Item	P1	P2	P3	P4	P5	P6	P7	P8	P9	Mean
I've been feeling optimistic about the future	7	0	0	0	2	0	0	0	6	**1.67**
I've been feeling useful	1	0	0	3	1	3	-	2	6	**1.78****
I've been feeling relaxed	2	0	1	3	0	0	0	0	4	**1.11**
I've been dealing with problems well	3	0	3	5	2	4	0	3	7	**3.00**
I've been thinking clearly	2	0	3	5	1	6	1	0	6	**2.67**
I've been feeling close to other people	3	0	1	2	1	3	0	0	8	**2.00**
I've been able to make up my own mind about things	3	0	0	4	3	1	0	1	5	**1.89**

* Acceptability score: 0 (not upset at all) to 10 (very upset).

** Mean was calculated with only eight participants because P7 provided a rating of 1 or 2 for "I've been feeling useful".

Cognitive difficulties

Participants demonstrated 'comprehension' and 'response mapping' difficulties; however, no 'recall' or 'judgement' difficulties were identified. Every participant experienced 'comprehension' difficulties on one to four items (M=2.6 items), although there were no 'comprehension' difficulties on the items "I've been feeling relaxed" and "I've been feeling close to other people". Seven participants experienced 'response mapping' difficulties on one to two items (M=0.8 items). Response mapping was only reported on three items ("I've been feeling optimistic about the future", "I've been feeling useful", and "I've been thinking clearly").

Comprehension difficulties

Use of unfamiliar words

The first item contained the word "optimistic", which was unfamiliar to younger participants. This resulted in them being unable to understand and thus respond to the statement.

Ambiguity of wording and interpretation

The wording of five out of seven items was perceived as ambiguous and therefore participants' interpretations of these items varied. For the item "I've been feeling optimistic about the future" respondents were not clear on the parameters of "future"; this led to some considering their immediate future, whereas others defined future as looking further ahead at later adulthood and career prospects.

There was also uncertainty over the interpretation of "I've been feeling useful" and "I've been thinking clearly". Participants varied in their understanding of which contexts and scenarios this might apply to. For example, should "thinking clearly" be associated with a specific problem and, if not, how do you determine thinking clearly on a day-to-day basis? Similarly, participants were not sure to which context usefulness might apply:

> "I wasn't quite sure of what this one might cover. So, like, in general I think useful means you're kind of adding value to something or you're doing something for someone else … "
> – Participant 8

As the above quote outlines, participants rationalised usefulness by the degree to which they supported others or were useful within a situation. An added complexity participants documented when answering this item was whose perception of "useful" they would hold in mind, with different responses being attributed to family, friends, or work/school.

For two items, it was not clear how severe the participants' example should be. For example, participants were uncertain of the scale of the problems inferred in "I've been dealing with problems well":

> " … I think day-by-day that one can alter, and it depends on what problem it is."
> – Participant 6

This interpretation would change the response to the item – if a problem was understood as small and therefore surmountable, then the response was more likely to be positive compared with more existential or more enduring problems. This difficulty was also experienced with "I've been able to make up my own mind about things"; this item was reported to be ambiguous over whether this statement applied for daily decisions, " … what you wanted for breakfast … " (Participant 7), compared with more complex decisions.

Subjective interpretation

Participants also found that "I've been dealing with problems well" was subjective and therefore could be misinterpreted. Participants felt that a person's strategy for dealing with a problem might be considered maladaptive by others. For example:

> "Your 'well' might be, I don't know, doing yoga or something, and my 'well' might be, I don't know, drinking a pint or something … "
> – Participant 9

This question was perceived as implicitly judgemental, due to the perception that there are good and bad ways to respond to a problem. Therefore, responses could be shaped by social desirability, particularly in the context of early help seeking, before trust and rapport have been established.

Response mapping difficulties

No applicable response option

There were two items where participants felt they would have preferred an "I don't know" option, and participants responded with a 3 on the Likert scale in its absence, because this was considered the most neutral option. The need for this response option might be resolved in part if the 'comprehension' difficulties are addressed. For example, a definition could be provided for "optimistic". However, regarding "I've been feeling optimistic about the future" there were participants who reported not thinking about or considering themselves to have future:

> "That one's a bit difficult because I don't, I don't see a future if that makes sense? … I don't see myself reaching to like that future point."
> – Participant 9

In these instances, there was no option for the participant to accurately reflect this feeling. This may be more common in a population who have experienced domestic abuse, and therefore their future feels less certain, especially for those entering a service.

Contextual information

Previously weaponised by person who harms

Participants reported that three items elicited discomfort in the context of experiences of domestic abuse, because these concepts could be weaponised by the person who harms. For example, regarding the item "I've been feeling useful", it is a common experience to be called "useless" in an abusive relationship. Similarly, if a person's value in the abusive relationship was based on their usefulness this might be upsetting:

> "It's hard to characterise a person [as] useful. It seems almost a, almost a touch of dehumanising in the question."
> – Participant 6

If a person was forced to do chores in the abusive relationship, they might want to distance themselves from this experience and the notion of "usefulness":

> "I was kind of like a slave basically … I always like had to make myself useful in order to … be liked at that point or like not turn against."
> – Participant 9

This might lead to defensive reporting, where a person completing the outcome measure might not respond truthfully out of fear of ramifications (Rose et al., 2011), particularly when they do not

trust the service provider. Similar concerns were raised with respect to "I've been thinking clearly". Gaslighting, a form of psychological abuse where the person that harms causes the victim to distrust their perception of reality (Solace Women's Aid, 2025), is common in abusive relationships and so this item might also upset some individuals. One participant highlighted that in the context of help seeking because of domestic abuse, you might not want it on record that you are thinking unclearly – for example:

> " … you never want to … report that you're not thinking clearly because you can see how that can be used against you … maybe in a court perspective … "
> – Participant 6

Finally, although some felt the item "I've been able to make up my own mind about things" might also be weaponised, one participant felt that this item can demonstrate the success of an intervention responding to domestic abuse:

> " … like a nice, privileged thing to say that you are able to have freedom of thought and opinion."
> – Participant 6

Inappropriate to ask in context of domestic abuse

Participants questioned the appropriateness of two items in a domestic abuse context. The first item, "I've been feeling optimistic about the future", was perceived as upsetting to those in the early stages of help seeking. This was because they were accessing support while still having difficult experiences. For those in the process of separating from the person that harms, the future may feel particularly uncertain. With respect to the item "I've been dealing with problems well", it was felt that the complexity of the situation and the lack of control over the person who harms may mean that problems could not be easily dealt with by the individual:

> "Coming from an outside force something that's kind of not in your control … how can you deal with a problem well that maybe you didn't create or you had no effect on?"
> – Participant 6

The context for completing SWEMWBS

Many participants reported that their response would vary depending on their experiences directly before the completion of the scale:

> " … it's dependent on situation factors in your day maybe, but I think it depends also on your state … "
> – Participant 6

Similarly:

> "Yeah, I think so. I think deadlines and external pressures or expectations would be the things that would change my mind or answer on that one."
> – Participant 8

This is not controlled or altered by the wording of the scale, but perhaps something that should be considered in the presenting of the scale or the interpretation of the answers.

Discussion

This study assessed the feasibility and acceptability of the SWEMWBS for young people affected by domestic abuse. It demonstrated that the measure was not burdensome to complete, taking no more than three minutes. This supports previous literature that found the SWEMWBS to be brief enough for use with children and young people (Clark et al., 2023; McKay and Andretta, 2017). This is of particular importance for domestic abuse settings because many services are reluctant to complete lengthy measures, particularly in the current funding climate where they are under-resourced (Carlisle et al., 2024).

Results also suggest that the SWEMWBS is broadly acceptable to this population. The average acceptability score was 3 or below out of 10, indicating that items did not cause any major distress to individuals. This is supported by Children Affected by Domestic Abuse[4] (CADA) evaluation findings that demonstrated a high completion rate for the SWEMWBS for young people aged 13 to 17 (Barter et al., forthcoming). This also resonates with our previous work, and that of others, suggesting this measure is largely acceptable to children and young people with experience of domestic abuse (Clark et al., 2023; Powell, Feder, et al., 2022; Shah et al., 2021). The one participant who found the OMI to be most distressing was the individual still actively accessing the service. This could suggest that individuals new to domestic abuse services require additional support when first being introduced to the measure. This strengthens work package 1's call for trauma-informed guidance to supplement the implementation of the DVA-COS (Bains et al., forthcoming).

No participants demonstrated 'recall' or 'judgement' difficulties, suggesting that they were confident in their recollection of instances to respond to all items as well as the frequency in the given two-week timeframe. Participants reported that the addition of an "I don't know" response would make the OMI more acceptable. In the measure's current form, participants tended to use the middle option, "some of the time", as indicating neutral response, which participants did not feel accurately reflected their experience. The developers of the SWEMWBS stipulate that no changes should be made to the measure's response categories[5], so this is not an adaptation that can be easily made.

The main difficulty experienced by participants was 'comprehension', primarily due to ambiguous wording and concerns from participants that their interpretation of the items might differ from others'. This highlights the need to validate the measure within populations affected by domestic abuse, particularly through the evaluation of measurement invariance. Another difficulty included the definition of "optimistic", which was unknown to the younger participants in the sample. Previous studies have similarly identified young people's lack of familiarity with the word "optimistic", which may result in difficulties completing the measure (Clarke et al., 2011). While

[4] Kimberlin and Winterstein, 2008.

[5] See: https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/using/faq

previous studies have identified the wording of the SWEMWBS as clear (Clarke et al., 2011), participants found five out of seven items ambiguous, and identified one as being very subjective, which could result in differing interpretations. Another factor raised by participants, which could alter the response to the OMI, was the influence of events preceding completion of the measure, which could include a stressful journey or difficult day at school; therefore, consideration should be given to how this can be countered in the implementation of the measures via guidance given to practitioners.

There was limited research into the use of the SWEMWBS with a domestic abuse population; therefore, this study expands this evidence base. The participants commented on the ways in which their experience of domestic abuse impacted their interpretation of the OMI, which centred on concepts that had previously been weaponised by the person who harms. There were two items, "I've been feeling optimistic about the future" and "I've been dealing with problems well", that participants felt were less appropriate in the context of domestic abuse. Given the limited scope to adapt the measure, this feedback needs to be considered when developing trauma-informed guidance to support the implementation of the measure. This could simply take the form of practitioners or researchers acknowledging that completion of the measure may have felt difficult and offering the opportunity to voice concerns or distress, as well as offering subsequent support to respond to the distress caused. This consideration may be most important when individuals are new to a service, or even a study, to ensure they do not experience distress as a result of the measure that causes them to disengage. It is important that developed guidance should be evaluated after implementation; this should inform future refinement and improvement to the guidance.

Limitations

The sample size for this study was small, and although this is acceptable for cognitive interviews, the participants were not wholly representative of the population. Although attempts were made to recruit participants from, by, and for services, the majority of the participants were of White ethnicity; thus, future research should seek to recruit more ethnically diverse participants. Future studies should recruit young people from Black, south Asian, and east Asian ethnicities as well as other minoritised groups such as Gypsy, Roma, and Irish Traveller and individuals with physical and mental disabilities. To properly reflect the perspectives of any one group, separate think aloud studies may be required for each minoritised group of interest. All participants in this study were female; therefore, results are limited to this perspective and further efforts should seek to engage young males and other gender identities to further test the measure. Many of the participants were engaging with services in an advisory capacity, meaning their experience of domestic abuse was more historical. This is a limitation, because the SWEMWBS in practice would be used with both this population and those newly engaging with domestic abuse services.

Participants' responses to the interview questions might have been impacted by social desirability. This may include them providing answers that they perceived to be helpful, causing them to find fault or comment on the measure where they would not have had a problem when completing the measure alone. Additionally, five participants took part in a focus group; this might have caused some individuals not to be as honest or vocal as if they speaking to the researcher one on one.

Conclusion

Overall, this study indicates that the SWEMWB is broadly acceptable for use with young people who have experienced domestic abuse. Some relatively minor adaptations could enhance the acceptability of this measure for use in the domestic abuse context; this must be weighed against the potential to undermine the validity of the tool. Moreover, these findings highlight the need to assess the acceptability of measures with a domestic abuse population because their experience are unique and tools designed for the general population may not be appropriate for use in this context. Similarly, there is a clear need for careful trauma-informed guidance that supports the OMI's implementation in research and service delivery contexts. Guidance must support those with domestic abuse experiences when completing the tools to prevent disengagement or any negative consequences that can result from completing the tool.

B. OXWELL: EVALUATING THE PSYCHOMETRIC PROPERTIES OF THE SWEMWBS IN CHILDREN WHO HAVE EXPERIENCED DOMESTIC ABUSE

Aim

The primary objective was to evaluate the psychometric properties of the SWEMWBS, including its internal consistency, convergent validity, and construct validity among children with domestic abuse experiences. This study also aimed to determine whether SWEMWBS functions equivalently across children with and without domestic abuse experiences and, therefore, effectively measures wellbeing in children and young people affected by domestic abuse.

Methodology

Design

This study was a psychometric validation of the SWEMWBS with children aged 11–18 using secondary data from the OxWell Student Survey 2023. It compared psychometric properties of the scale across children who have experienced domestic abuse (the DVA group) and those who have not (the Non-DVA group).[6] This included reliability analyses (internal consistency and split-half reliability) and validity analyses (construct validity, convergent validity, and measurement invariance).

The OxWell Student Survey

The OxWell Student Survey is a large-scale, anonymous survey designed to assess the wellbeing of children and young people aged 8 to 18 in England. It is conducted annually or every other year and uses a repeated cross-sectional design to capture a broad range of factors influencing mental health and happiness, including mental wellbeing, anxiety, loneliness, bullying, substance use, sleep patterns, online safety, vulnerability, and access to mental health support. The sampling strategy includes three key components: (1) targeting schools in areas with varying levels of deprivation, (2) using the opt-out consent model to increase participation rates, and (3) ensuring

[6] We use the acronym 'DVA' when distinguishing between samples that have experienced domestic abuse (DVA) and those that have not (non-DVA) to ease interpretation of findings.

anonymity to promote accurate and open responses (Mansfield et al., 2021). The survey is also conducted during school hours to ensure consistency in data collection.

In 2023, the survey collected 42,000 responses from students in 105 primary schools and 80 secondary schools and further education colleges, representing a diverse cross-section of the student population. Participation came from several key regions, including Liverpool (16,886 students from 73 schools), Berkshire (12,511 students from 60 schools), Oxfordshire (7,826 students from 31 schools), Milton Keynes (2,789 students from 9 schools), and Sandwell, Surrey, and Buckinghamshire (2,194 students from 9 schools).

Sample

The full OxWell dataset was 'cleaned' by excluding data entries from primary-school-aged children, children who were not aged 11–18 or did not specify their age, entries with missing SWEMWBS scores, and ineligible DVA scores (e.g. contradictory responses that included both DVA experience and 'never' or 'prefer not to say'; see Appendix C). Respondents who 'preferred not to say' whether they had experienced DVA were also not included in the analysis sample. This resulted in a total sample of 17,801 participants and a sample of 1,215 children and young people who have experienced domestic abuse (either in their lifetime or the past year; i.e. DVA sample). Within the DVA sample, 29 participants reported experiencing domestic abuse both in the past year and across their lifetime, while a further 153 participants reported domestic abuse experience in the past year and 1,033 children and young people reported domestic abuse experience within their lifetime.

The key demographics of the sample are reported in Table 3. Compared with national averages for this age group, the sample includes more ethnic minorities compared with White participants. The national average of White persons in England and Wales, aged 10–19, is an estimated 82% according to the ONS census 2021, while in the present sample only 55% of the participants who disclosed their ethnicity were White. Furthermore, the latest official government statistics show that 28% of children in the UK are living in material deprivation and 31% live in a household with relative low income (Department for Work and Pensions, 2025). Poverty was indicated by a slightly larger proportion of this sample (32%).

Table 3. The key demographics of the included sample

Demographic		DVA sample *N(%)*	Non-DVA sample *N(%)*	Total sample *N(%)*
Gender	Boy	379 (31%)*	7,301 (44%)*	7,680 (43%)

Demographic		DVA sample *N(%)*	Non-DVA sample *N(%)*	Total sample *N(%)*
	Girl	729 (60%)*	8,532 (51%)*	9,261 (52%)
	Gender diverse	48 (4%)*	216 (1%)*	264 (2%)
	Not specified	59 (5%)	537 (3%)	741 (4%)
Age	11	66 (5%)*	1,451 (9%)*	1,517 (9%)
	12	189 (16%)*	3,325 (20%)*	3,514 (20%)
	13	232 (19%)	3,340 (20%)	3,572 (20%)
	14	220 (18%)	2,787 (17%)	3,007 (17%)
	15	189 (16%)	2,309 (14%)	2,498 (14%)
	16	171 (14%)*	1,867 (11%)*	2,038 (11%)
	17	98 (8%)*	1,039 (6%)*	1,137 (6%)
	18	50 (4%)*	468 (3%)*	518 (3%)
Ethnicity	White	639 (53%)	9,180 (55%)	9,819 (55%)
	Mixed/multiple ethnic groups	99 (8%)*	897 (5%)*	996 (6%)
	Asian	188 (15%)	2,429 (15%)	2,617 (15%)
	Black/African/Caribbean	65 (5%)*	684 (4%)*	750 (4%)

Demographic		DVA sample *N(%)*	Non-DVA sample *N(%)*	Total sample *N(%)*
	Arab	12 (1%)	268 (2%)	280 (2%)
	Another ethnic group	25 (2%)	327 (2.0%)	352 (2%)
	No response	187 (15%)	2,801 (17%)	2,988 (17%)
Poverty	No poverty indicated	455 (37%)*	11,398 (68%)*	11,853 (67%)
	Poverty indicated	738 (60%)*	4,992 (30%)*	5,730 (32%)
	No response	22 (2%)	196 (1%)	218 (1%)

* Significant differences between the DVA and non-DVA samples.

Measures

For the present research, measures included DVA experience; mental wellbeing (SWEMWBS); a measure of anxiety and depression; and demographics.

Domestic abuse experience was measured as "having seen or heard a parent/carer be slapped, kicked, punched, beaten, or deliberately hurt by a partner or ex-partner". The present analysis will therefore underestimate the full extent of domestic abuse experience because it only captures physical abuse. Participant responses could include "No, never"; "Yes, it has happened within the past 12 months"; "Yes, it has happened in my life"; or "Prefer not to answer". As *lifetime* and *past year* response options were delivered simultaneously, it is possible that children and young people who had experienced domestic abuse in the past year as well as previously in their lifetime may have selected only one of these options without indication for which should be prioritised. The DVA sample in this study (unless otherwise specified) included all children and young people that indicated domestic abuse experience, whether in their lifetime or within the past year.

The OxWell survey uses the SWEMWBS to measure wellbeing (see Appendix D). To test convergent validity, we also retained the 11-item version of the Revised Children's Anxiety and Depression Scale (RCADS-11; Radez et al., 2021). The RCADS-11 is a shortened version of the original RCADS designed to screen for symptoms of anxiety and depression in children and adolescents aged 8 to 18. It includes 11 items selected from the full 47-item RCADS, maintaining

strong psychometric properties while reducing the time burden on respondents. It uses a four-point Likert scale from Never to Always to measure how often things happen (e.g. "I feel sad or empty"). The items are averaged to present an overall total internalising anxiety and depression score. The RCADS has been shown to relate to other wellbeing measures in the literature (Piqueras et al., 2017).

Key demographics retained for the present study included gender, age, ethnicity, and indication of poverty (see Table 3). Poverty was measured dichotomously with participants who responded affirmatively that "they worry about not having enough money for things their family needs e.g. food, bills, electric or gas" or that "their family uses foodbanks" recorded as poverty indicated. The OxWell survey includes additional measures of poverty relating to home and school life but for the purpose of this study the focus was on these two items because they were used to measure financial-based poverty in the Wales's Young People's Survey on Child and Family Poverty 2019 (McFarlane, 2021).

Analysis plan

First, the full OxWell dataset was 'cleaned' by excluding data entries from primary-school-aged children, children that were not aged 11–18 or did not specify their age, entries with missing SWEMWBS scores on any item, and ineligible DVA scores (for a summary of excluded data see Appendix C). Missing data was explored to assess issues with underreporting domestic abuse experience but excluded from subsequent analyses. Data cleaning was performed using SPSS, and analyses were conducted in R.

Descriptive statistics were examined to summarise the sample characteristics and assess the distribution of SWEMWBS scores. The raw SWEMWBS scores were used in all analyses unless otherwise stated.

To assess the reliability of the SWEMWBS scale the internal consistency was evaluated using Cronbach's alpha, with values of ≥0.7 considered acceptable. Item-total correlations were also reviewed to identify any weak or problematic items. Additionally, split-half reliability was assessed to further evaluate internal consistency. However, because the data is cross-sectional and was collected at a single timepoint, test-retest reliability could not be assessed.

Convergent validity was examined by analysing the relationship between the SWEMWBS and the RCADS-11 measure of anxiety and depression. A negative correlation was expected, with the anticipation that higher wellbeing scores on the SWEMWBS would correspond with lower depression scores on the RCADS-11.

To explore the factor structure of the SWEMWBS within children and young people who have and have not experienced domestic abuse, a confirmatory factor analysis (CFA) was conducted for each subsample using a weighted least squares mean, and variance-adjusted (WLSMV) estimator, which provides robust standard errors for ordinal data. Model fit was evaluated using the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardised root mean square residual (SRMR). An acceptable fit was defined as CFI and TLI >.90 and RMSEA and SRMR <.08 (Hu and Bentler, 1999).

Finally, measurement invariance was tested across individuals with and without DVA experience to determine whether the WEMWBS functions consistently across these groups. This also used

WLSMV estimation. Invariance was confirmed when there were minimal changes in fit indices between nested models for configural, metric, and scalar invariance (ΔCFI and Δ TLI <-.01, ΔRMSEA <.015, and Δ SRMR <.01; Cheung and Rensvold, 2002; Chen, 2007).

Results

Descriptives

Data from the 'non-responding' sample was explored initially to identify whether it was more likely to represent those with DVA experience and therefore suggest that domestic abuse is underreported. It was identified that participants who chose not to report whether they had experienced domestic abuse (by selecting a "prefer not to say" response option) had significantly lower wellbeing than those that had never experienced domestic abuse (see Appendix E). Exploration of item-level missingness on the SWEMWBS did not indicate that participants were particularly deterred from any one item. There was also no indication that any of the items were avoided more for participants with DVA experience compared with those without (see Appendix E).

The distribution of SWEMWBS scores can be seen in Appendix F. The prevalence of domestic abuse exposure (lifetime or past year) in the included sample was 6.8%. Of those reporting exposure to domestic abuse, 17.4% had experienced domestic abuse within the past year (1.0% of the total sample). Research conducted by the NSPCC identified that 17.5% of children aged 11–17 have been exposed to domestic abuse (involving threats, physical abuse, and property damage in the context of an argument) in their lifetime and 2.5% of 11–17 year olds have been exposed within the past year (Radford et al., 2011). When specifically referring to witnessing a parent be pushed, slapped, hit, punched, or beaten up by the other parent or their partner, 7.1% of 11–17 year olds have been exposed in their lifetime and 0.8% within the past year.

The average SWEMWBS scores transformed in line with guidance (Stewart-Brown et al., 2009) were significantly different between the DVA (M=17.76, SD=4.93) and non-DVA samples (M=20.85, SD=4.78), t(17,799)=21.688, p<.001. The average raw SWEMWBS score (used in all subsequent analyses) was also significantly different between the DVA (M=18.31, SD=6.24) and non-DVA samples (M=22.40, SD=5.68), t(17,799)=22.145, p<.001. See Table 4.

Table 4. The differences in mean SWEMWBS scores between the DVA and non-DVA sample

	Non-DVA[a]	DVA[b]		
	M(SD)	***M(SD)***	***t***	***p***
Transformed SWEMWBS	20.85 (4.78)	17.76 (4.93)	21.13***	<.001

	Non-DVA[a]	DVA[b]		
	M(SD)	*M(SD)*	*t*	*p*
Raw SWEMWBS	22.40 (5.82)	18.31 (6.24)	22.145***	<.001

[a] N=16,586, [b] N=1,215, *** p<.001.

Raw SWEMWBS scores were also significantly different between those that experienced DVA within the past year (M= 16.77, SD=6.17) and lifetime DVA exposure (M=17.89, SD=4.67), $t(178.64)=2.152$, p=.033.

In line with expectations, wellbeing is lower in the DVA-experienced group compared with those who have never experienced (physical) DVA. Furthermore, wellbeing is lower in those that have experienced DVA within the past year compared with those who have experienced DVA in their lifetime.

Reliability

Internal consistency

The Cronbach's alpha of the SWEMWBS scale for the non-DVA sample (α=.86; 95% confidence intervals=.86–.86) and the DVA sample (α=.87; 95% confidence intervals=.85–.88) were similar in strength and both greater than the required standard of .7, indicating good reliability of the scale for both samples.

In both samples the individual item scores correlated with each other, demonstrating medium to strong relationships (r=.302–.673). This shows that the items are consistently related, suggesting that they align well with a single construct.

Corrected item-total correlations for the seven items ranged from 0.46 to 0.74, indicating that all items sufficiently correlated with the overall construct in both samples. The values of the Cronbach's alpha if items were removed ranged from 0.83 to 0.86, demonstrating that no single item substantially detracted from the scale's reliability. This further confirms that all items contribute meaningfully to the overall construct. Overall, the findings demonstrate that the internal consistency of the scale is supported (see Appendix G for supporting tables).

Split-half reliability

As testing only took place at a single timepoint, evaluation of test-retest reliability was not possible. Instead, we examined the split-half reliability of the test, assessed using Guttman's lambda coefficients (see Appendix G: Table G3). The average split-half reliability was 0.84 for the non-DVA sample and 0.85 for the DVA sample, indicating good internal consistency above the generally accepted threshold of 0.7. The minimum split-half reliability was 0.78 for the non-DVA sample and

0.79 for the DVA sample, indicating that even in the least favourable split, reliability remained above the accepted threshold.

Validity

Convergent validity

The relationship between the SWEMWBS and the RCADS-11 was explored with the expectation that they would be negatively correlated due to the constructs measured by the scales being inversely related: mental wellbeing (measured by SWEMWBS) versus symptoms of depression and anxiety (measured by RCADS-11). Both the raw SWEMWBS and transformed SWEMWBS scores strongly negatively correlated with the RCADS-11 in both samples (see Table 5). The consistency of these findings, across both the non-DVA and DVA samples, underscores the validity of the SWEMWBS as an indicator of mental wellbeing inversely related to psychological distress.

Table 5. Correlation between RCADS-11 and SWEMWBS

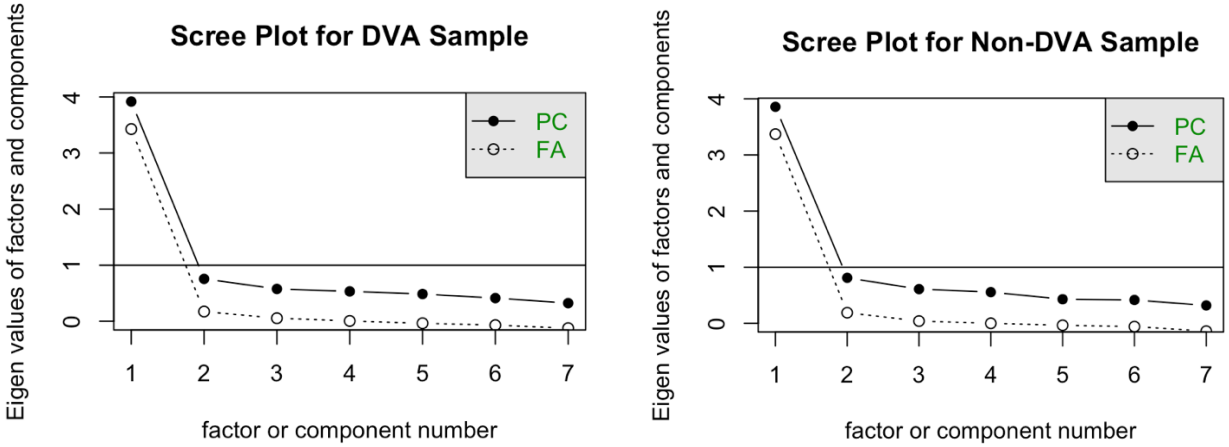
	Raw SWEMWBS	Transformed SWEMWBS
Non-DVA sample[a]	-.618***	-.576***
DVA sample[b]	-.625***	-.586***

[a] N=15,864, [b] N=1,151, *** p<.001.

Factor analysis

Scree plots were generated for both the DVA sample and the non-DVA sample to assess the number of factors underlying the SWEMWBS and ensure that confirmatory factor analysis with a single factor model was appropriate. In both samples, the scree plots showed a clear “elbow” after the first factor, suggesting that a single factor explained the majority of the variance in the data (see Figure 3).

Figure 3. Scree plots of number of factors measured by the SWEMWBS scale for the DVA sample and the non-DVA sample (go to accessibility text)



A CFA using the WLSMV estimator was conducted to assess the one-factor structure of the SWEMWBS in each sample. The chi-square test was significant for both the non-DVA sample ($\chi2(14)=3{,}356.97$, $p<.001$) and the DVA sample ($\chi2(14)=183.68$, $p<.001$), likely due to the large sample sizes, as the chi-square test is sensitive to sample size. Although the chi-square was significant, the model demonstrated adequate fit to the data in all other fit indices (see Appendix H). The CFI and TLI exceed the commonly accepted threshold of 0.90 in both samples, indicating excellent fit. The SRMR values of 0.03 and 0.04 also indicate excellent fit (threshold: < 0.08). Both samples had RMSEA values slightly above the threshold, indicating a mediocre fit. However, RMSEA is a less reliable fit statistic because it can be distorted in models using the WLSMV estimator with ordinal data of 5 or more categories (the SWEMWBS is a 5-item scale) and models with low degrees of freedom (Beauducel and Herzberg, 2006; Kenny et al., 2015). On balance, the strong fit indicated by the CFI, TLI, and SRMR and the mediocre fit indicated by the borderline RMSEA suggest acceptable model fit.

All items demonstrate moderate to strong factor loadings to the single factor in both samples (see Appendix H: Table H2). This indicates that each item contributes meaningfully to the measurement of the overall construct of mental wellbeing. Items 4 and 5 consistently showed stronger loadings, suggesting they are particularly representative of the construct.

The CFA results support the one-factor structure of the SWEMWBS in both the non-DVA and DVA samples. Despite significant chi-square values, the model demonstrated strong or adequate fit according to all other fit indices (CFI, TLI, RMSEA, and SRMR). The strong factor loadings provide further evidence of the unidimensional nature of the scale and its effectiveness in capturing the construct of mental wellbeing across diverse populations.

Measurement invariance

Measurement invariance was assessed to determine whether the SWEMWBS is interpreted and functions equivalently across those with and without DVA experiences. The sequence of testing

followed the steps for configural, metric, and scalar invariance using the WLSMV estimator (see Appendix I).

Configural invariance

The configural invariance model showed acceptable fit. Although the chi-square was significant, $\chi^2(28)=3{,}510.87$, $p<.001$, other fit indices show a strong fit, CFI=.97, TLI=.96, SRMR=.04, indicating that the factor structure was equivalent across groups, meaning that both groups conceptualise mental wellbeing similarly.

Metric invariance

The metric invariance model showed acceptable fit, $\chi^2(34)=2{,}106.56$, $p<.001$, CFI=.98, TLI=.98, RMSEA=.08, 90% CI [.08, .09], SRMR=.04. Although the χ^2 difference test comparing the configural and metric models was significant, $\Delta\chi^2(6)=36.01$, $p<.001$, there was improvement in other fit indices, suggesting that the factor loadings were equivalent across groups. This means that the relationship between SWEMWBS items and the latent factor (mental wellbeing) is consistent between DVA and non-DVA groups.

Scalar invariance

The scalar invariance model showed acceptable fit, $\chi^2(54)=2{,}249.51$, $p<.001$, CFI=.98, TLI=.99, RMSEA=.07, 90% CI [.07, .07], SRMR=.03. Although the χ^2 difference test comparing the metric and scalar models was significant, $\Delta\chi^2(20)=58.63$, $p<.001$, there was almost no change in other fit indices, indicating that the item intercepts were equivalent across groups. This means that group differences in item scores reflect true differences in the latent construct rather than measurement bias.

Discussion

This study provided robust psychometric validation of the SWEMWBS in a large sample of children and young people (age 11–18), confirming its strong internal consistency, convergent validity, and unidimensional structure. Measurement invariance testing indicated that the scale functioned equivalently across those with and without domestic abuse experiences, supporting its utility in this population. Wellbeing scores were significantly lower among those who reported domestic abuse experience, particularly among those who had experienced it within the past year. Additionally, children who chose not to disclose their domestic abuse experience had similar SWEMWBS scores to those who reported domestic abuse, raising potential concerns about underreporting. These findings validate the use of the SWEMWBS as a measure of wellbeing in children aged 11+ and specifically in children and young people who have experienced domestic abuse.

Limitations

Although the study benefited from a large and diverse sample, this specific methodology also presents limitations in the context of our research. The school-based research setting may not generalise to the completion of measures in a domestic abuse service context. First, the school

experience for children who have experienced domestic abuse is mixed: although some children and young people report school as a safe place where they feel supported, others face adversity in the school environment and feel more supported in service settings (Bracewell et al., 2020; Roy et al., 2022). Additionally, within school settings children are assisted by familiar teachers in completing the SWEMWBS but completion of scales in a service context could be impacted by the absence of this assistance and support, particularly in early stages of service provision. It is important to consider the increased likelihood that children and young people within services are within the throes of, or have more recent and/or severe experience with, domestic abuse that could impact the completion of wellbeing scales (McGarry and Ali, 2016). As highlighted by the previous study, this could be addressed with trauma-informed guidance.

On the other hand, the anonymity of the OxWell survey was a key strength because it encouraged honest and candid responding, which is particularly beneficial when addressing sensitive topics such as domestic abuse experiences and wellbeing. However, this level of openness may not be replicable in service settings, in which individuals could report defensively due to perceived risks or lack of anonymity.

The study included only those attending schools that opted into the survey, potentially excluding more vulnerable children who may have disengaged from education or were in alternative care settings, which could impact the representativeness of the sample. Furthermore, the measurement of domestic abuse exposure in the OxWell study was based solely on reports of witnessing physical abuse between parents/partners, which does not capture the full spectrum of domestic abuse experiences, such as coercive control or emotional abuse, or abuse between other adults in a household. This underestimation of domestic abuse prevalence may have influenced the findings. However, research has shown that exposure to physical forms of domestic abuse is closely associated with other forms of domestic abuse and family violence. The prevalence of poly-victimisation and shared impacts of different forms of domestic abuse suggest that these findings are likely to be generalisable to children and young people who have been exposed to non-physical forms of domestic abuse (UNICEF., 2006; Radford et al., 2011; Radford, Stanley and Elwen, 2021).

Additionally, comparisons between lifetime and past-year domestic abuse experiences may be limited by potential recall biases or underreporting in self-reported measures. Within the OxWell survey, domestic abuse experience was a single measure that included lifetime and past-year response options simultaneously. It is therefore ambiguous as to which response option to prioritise for children and young people who have experienced domestic abuse both within the past year and previously. As a result, the domestic abuse experience group had to be combined for analyses, and measurement invariance between past-year experience and experience previously in lifetime could not be explored.

Moreover, reliance on cross-sectional data limited the ability to assess the longitudinal validity of the scale. In particular, it meant that the SWEMWBS' sensitivity to change could not be established. The lack of repeated measures also prevented the analysis of test-retest reliability; thus, stability of the scales over time remains uncertain. Further limitations on the implementation of the SWEMWBS as an OMI are outlined in the general discussion.

Conclusion

In summary, these findings indicate that the SWEMWBS is a valid OMI for assessing emotional health and wellbeing in children and young people aged 11+ affected by domestic abuse, offering an efficient measure for research and intervention evaluations.

C. EVALUATING THE PSYCHOMETRIC PROPERTIES OF THE SWEMWBS WITH CHILDREN IN DOMESTIC ABUSE SERVICE SETTINGS

Aim

We aimed to validate the use of the SWEMWBS within service settings that support children affected by domestic abuse. Recognising that the scale is already implemented in some service contexts, this research sought to assess the validity of the scale for use in these settings and identify whether it reliably measures wellbeing over time. The study analysed existing data from a service provider to examine how the SWEMWBS functions when used within intervention programmes. Specifically, analyses included internal consistency testing, factor analysis, and measurement invariance assessments to determine whether SWEMWBS scores can detect meaningful differences when used in practice. This allowed us to expand on the previous study and determine the ability of the SWEMWBS to detect changes in wellbeing throughout the course of interventions in domestic abuse services.

Methodology

Sample

We conducted a secondary analysis of data collected by a service provider that offers a range of services to families affected by domestic abuse (children and young people aged 4–18). The provider uses the SWEMWBS to record wellbeing at an initial timepoint when the children and young people enter the service, an interim timepoint during service provision, and a final timepoint after completion of service. The services include short-term first-response interventions (for age 11+), mental health and wellbeing services, counselling, and group work. Interventions varied in length from 2-week first-response interventions to between 10 weeks and 1 year (although rarely this long) for longer-term interventions.

There were 676 individual timepoint entries provided by the service and there was no item-level missingness in the completion of the SWEMWBS. Once entries from the adult programme were removed, this left 580 timepoint entries representative of children and young people. Some individuals had re-entered services on multiple occasions. Where this was the case, only the most recent and completed data was included in analysis. This resulted in a total sample of 268 participants: 45 who had completed all 3 timepoints and 159 who had completed at least the initial and final timepoints. These participants had been referred between March 2020 and October

2024. To protect privacy and anonymity, demographic information of participants was not provided by the service.

Analysis plan

To validate the use of the SWEMWBS with children and young people in the provision of domestic abuse services, analyses similar to those conducted with the OxWell sample were carried out. All analyses used raw SWEMWBS scores so that comparisons could be made at the item level. First, descriptive statistics were explored to assess the distribution of SWEMWBS scores and compare them across the three timepoints. All subsequent analyses were carried out on data from initial timepoints and final timepoints to establish the reliability and validity of the scale for use throughout the course of interventions.

The internal consistency of the scale was evaluated using Cronbach's alpha, with a threshold of ≥0.7 indicating acceptable reliability. Additionally, item-total correlations were examined to identify any weak or problematic items. The split-half reliability of the scale was also assessed to further evaluate the internal consistency of the scale. Although the sample includes multiple timepoints, these are pre-post scores that are expected to change, and thus test-retest reliability could not be explored.

To explore the factor structure of SWEMWBS in this population, CFA was performed for the initial and final timepoints. Fit indices including the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardised root mean square residual (SRMR were used to evaluate model adequacy. Acceptable fit is indicated by CFI and TLI >.90 and RMSEA and SRMR <.08.

Finally, measurement invariance was tested across initial and final timepoints to determine whether the SWEMWBS functions consistently across time during service provision. It must be highlighted that there is not consistency in the interval of time between initial and final timepoints across participants and thus there may be individual differences (associated with time and intervention type) that will not be accounted for in the model. Measurement invariance (configural, metric, and scalar) was confirmed where there were minimal changes in fit indices between nested models (ΔCFI and Δ TLI <-.01, <.015, and Δ SRMR <.01; Cheung and Rensvold, 2002; Chen, 2007).

Results

Descriptives

The distributions of SWEMWBS scores are presented in Appendix K (UNICEF., 2006; Radford et al., 2011; Radford, Stanley and Elwen, 2021). The average SWEMWBS scores at the final timepoint (M=23.86) were significantly higher than at the initial timepoint (M=20.1), $t(158)=17.61$, $p<.001$ (see Table 6). The mean difference of 5.21 (95% confidence interval = 4.62, 5.79) underscores the substantial improvement in participant wellbeing (see Table 6). The effect size for the change in SWEMWBS scores, measured by Cohen's D, was 1.17. This large effect size indicates that the interventions may have considerably impacted wellbeing, although without an experimental design the driver of improvements in wellbeing cannot be established. These findings also highlight the

responsiveness of the SWEMWBS; although we cannot determine whether this change is due to the intervention or external influences, the scale was able to detect improvements in wellbeing between timepoints. For comparisons between SWEMWBS scores over time at the item level see Appendix K.

Table 6. Mean SWEMWBS scores

		Raw scores			Transformed scores[a]		
	N	***M***	***SD***	***Range***	***M***	***SD***	***Range***
Initial timepoint	252	20.1	4.92	9–35	19.09	3.69	11.25–25
Interim timepoint	69	23.86	5.61	10–35	22.01	4.64	12.40–35
Final timepoint	172	25.07	4.19	11–35	22.81	3.79	13.33–35
Initial–final score change	159	5.21	3.73	-6–19	3.89	2.92	-3.85–14.05

[a] Scores are transformed in line with official SWEMWBS guidance (Stewart-Brown et al., 2009).

While the SWEMWBS was not developed as a measure of individual differences, research has shown that a change in transformed SWEMWBS scores of either 1.03 points or 2.87 points exceed the minimum threshold for important levels of change (there is a lower and higher figure due to varying statistical methods used to derive this; Shah et al., 2021). Using the lower minimum threshold, 86.8% of participants demonstrated a clinically significant improvement in wellbeing at the end of the service. Using the higher minimum threshold, 63.5% of participants demonstrated a clinically significant improvement in wellbeing at the end of the service.

On average, there were significant improvements recorded between initial and interim timepoints, and interim and final timepoints ($F(2,487)=7.58$, $p<.001$, pairwise comparisons with Bonferroni corrections demonstrated significant improvements across all timepoints, all $p<.001$). Of those that completed all timepoints, 17.8% of respondents reported a reduction in wellbeing between initial and interim timepoints, and 20% reported a reduction in wellbeing between interim and final timepoints. When comparing transformed SWEMWBS scores, 40.0%–71.1% of participants between initial and interim, and 33.3%–57.8% of participants between interim and final timepoints, presented clinically important improvements. Overall, these findings do not suggest an issue with defensive reporting at initial timepoints.

Reliability

Internal consistency

The SWEMWBS demonstrated high internal consistency, with Cronbach's alpha values of 0.84 at both initial and final assessments (95% confidence interval = .81–87). Both are greater than the required standard of .7, indicating good reliability of the scale for both samples.

Corrected item-total correlations for the 7 items ranged from .56 to .75, indicating that all items sufficiently correlated with the overall construct in both samples. The values of the Cronbach's alpha if items were removed ranged from 0.80 to 0.84, demonstrating that no single item substantially detracted from the scale's reliability (see Appendix L: Table L1).

Split-half reliability

Additionally, we report the split-half reliability of the SWEMWBS, assessed using Guttman's lambda coefficients (see Appendix L: Table L2). The average split-half reliability was .83 at the initial timepoint and .82 at the final timepoint, indicating good internal consistency at both timepoints, above the generally accepted threshold of .7. The minimum split-half reliability scores were .76 and .80, indicating that even in the least favourable split, reliability remained above the accepted threshold.

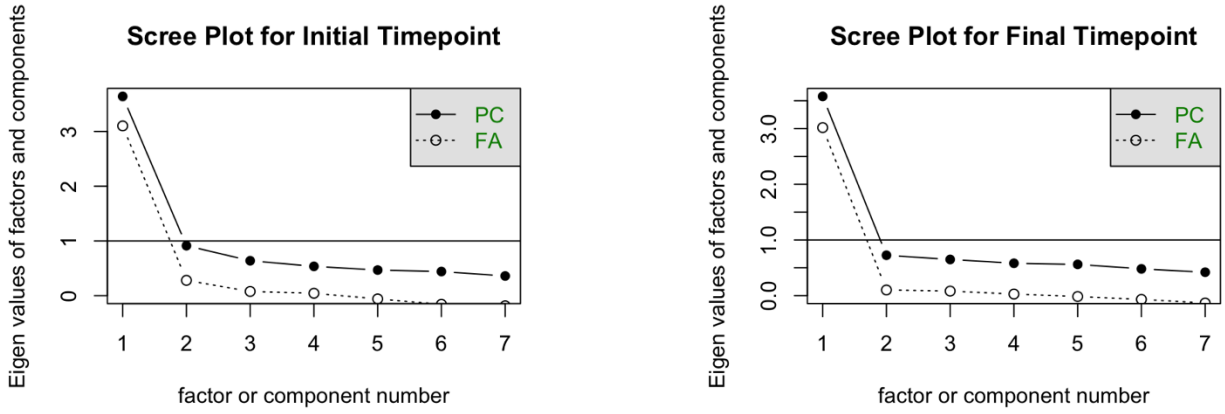
The results of the split-half reliability and Cronbach's alpha testing support the internal consistency of the SWEMWBS scale, indicating that the scale performs robustly when used within service settings.

Validity

Factor analysis

Scree plots were generated for both the initial timepoint and the final timepoint to assess the homogeneity of the SWEMWBS and ensure that confirmatory factor analysis with a single factor model is appropriate. In both samples, the scree plots show a clear "elbow" after the first factor, suggesting that a single factor explains most of the variance in the data (see Figure 4).

Figure 4. Scree plots of number of factors measured by the SWEMWBS scale at initial and final timepoints (go to accessibility text)



A CFA was conducted to assess the one-factor structure of the SWEMWBS at each timepoint. The chi-square test was significant at the initial timepoint ($\chi 2(14)=52.48$, $p<.001$) and the RMSEA was 0.104 (90% confidence interval: [0.075, 0.135]). However, other fit indices indicated an acceptable model fit (see Appendix M: Table M1). The model demonstrated excellent fit for the final timepoint, with all goodness of fit indices exceeding thresholds and a non-significant chi-square, $\chi 2(14)=11.40$, $p=.655$ (see Appendix M: Table M1).

Factor loadings for individual items ranged from 0.56 to 0.75, demonstrating that each item contributed meaningfully to the measurement of the latent construct (Appendix M: Table M2). These results support the unidimensional nature of the SWEMWBS.

The CFA results support the one-factor structure of the SWEMWBS in both timepoints. The final timepoint demonstrated stronger model fit. The strong factor loadings provide further evidence of the unidimensional nature of the scale and its effectiveness in capturing the construct of mental wellbeing.

Measurement invariance

Measurement invariance was assessed to determine whether the SWEMWBS is interpreted and functions equivalently before and after interventions. The sequence of testing followed the steps for configural, metric, and scalar invariance (see Appendix N).

The configural invariance model showed acceptable fit, $\chi^2(70)=119.0$, $p<.001$, CFI=.94, TLI=.92, RMSEA=.07, 90% CI [.05, .09], SRMR=.051. Although the chi-square was significant, other fit indices show acceptable fit, indicating that the factor structure was equivalent across time.

Metric invariance is supported (see Appendix N). The non-significant chi-square change between the configural and metric model, along with minimal change in fit indices, indicates that factor loadings are equivalent across timepoints, $\Delta\chi^2(6)=7.21$, $p<.301$, ΔCFI=-.001, ΔTLI=.004, ΔRMSEA=-.002, ΔSRMR=.01. Although the change in CFI from metric to scalar invariance slightly exceeds the recommended threshold (ΔCFI=-.011), the change in RMSEA is minimal

(ΔRMSEA=.003), and other absolute fit indices remain acceptable(ΔTLI=-.007, ΔSRMR=.013). Given the theoretical rationale and the consistency of the results across fit indices, scalar invariance is deemed to hold with caution.

Discussion

This study extended the validation of SWEMWBS by evaluating its effectiveness within real-world service settings for children impacted by domestic abuse. The results demonstrated that the SWEMWBS maintained strong internal consistency and a unidimensional structure across intervention timepoints. Wellbeing scores significantly improved from the initial to final assessment, indicating that SWEMWBS is sensitive to change and can effectively capture wellbeing improvements following intervention. Measurement invariance analysis confirmed that the scale functions consistently across pre- and post-intervention assessments. This further supports the SWEMWBS as a viable OMI for tracking wellbeing changes in children and young people receiving domestic abuse-related services.

Limitations

This study was invaluable in supplementing our findings with data on the SWEMWBS in practice but is not without its limitations.

First, we relied on data collected from a single service provider, which may limit generalisability and is also constrained by data availability and potential selection bias. The data represents children and young people actively engaged in services, meaning those who did not access or dropped out of services were not represented. This raises concerns about whether the findings reflect all service users or are biased towards those who remained engaged in interventions. Additionally, demographic data was unavailable due to anonymity constraints, preventing an assessment of whether findings were influenced by factors such as age, gender, or ethnicity and making any participant biases unknown. The data was also collected between 2020 and 2024, so there is the possibility that the COVID-19 pandemic could have exacerbated struggles with mental wellbeing during this time. Research has highlighted that wellbeing worsened during the COVID-19 pandemic for children and young people who reported feeling unsafe or very unsafe at home compared with reported improvements in wellbeing for children who feel very safe at home (Soneson et al., 2023).The nature of the impact of the COVID-19 pandemic and associated public health measures on the wellbeing of children and young people was probably dependent on whether the child was living with the parent that harms during this time.

Furthermore, the dataset compiled different intervention programmes provided by the service with varying and unspecified durations. This lack of control over data collection timepoints does pose issues in interpreting the meaningfulness of the change in SWEMWBS scores. However, even in first-response programmes, which were the shortest, significant score changes were demonstrated, indicating that the SWEMWBS is sensitive to change. Nevertheless, in a controlled study it would be preferable for the SWEMWBS to be completed consistently with equal intervals between timepoints for all participants to best minimise individual covariates. Therefore, inconsistency in intervention type and intervals between initial and final timepoints are to be considered here. That said, the concept of time is abstract for longitudinal research and is to be determined in the context

of the research goal: here we are interested in the scale functioning before and after service use rather than a specified duration of time (Wang et al., 2017).

Another important consideration for implementation identified by this study is the potential impacts of participant attrition. Only a small proportion of participants completed measures at all three timepoints. Although it is unclear whether this was due to dropout, administrative issues, or participant refusal, this will be something to consider when implementing OMIs during service delivery. Measurement burden and responsivity are important to consider when interpreting defensive reporting. While our findings did not suggest that defensive reporting was an issue, it is based on interim data which was only reported by a limited number of participants. Defensive reporting is addressed further in the general discussion below.

Conclusion

In summary, this study supports the use of the SWEMWBS as a measure for evaluating wellbeing in children and young people accessing domestic abuse services. The scale demonstrated strong psychometric properties, consistent with the findings using OxWell data, and indicated that the SWEMWBS is sensitive to change in wellbeing, highlighting its suitability as a valid OMI for use within this population.

D. EVALUATING THE PSYCHOMETRIC PROPERTIES OF THE WEMWBS IN ADULTS WHO HAVE EXPERIENCED DOMESTIC ABUSE

Aim

The last study sought to validate the WEMWBS for use with adults who have experienced domestic abuse. A rapid review was conducted to identify secondary datasets containing both WEMWBS and domestic abuse measures, ultimately selecting the Adult Psychiatric Morbidity Survey (APMS) 2014. The study replicated the analyses conducted with the children and young people sample from the OxWell survey testing the internal consistency, convergent validity, factor structure, and measurement invariance between groups that have and have not experienced domestic abuse. When we refer to adults who have experienced domestic abuse, we refer to direct victimisation.

Methodology

Rapid review for a secondary data source

A rapid review was conducted to identify existing datasets suitable for further validation of the WEMWBS within domestic abuse-affected populations. Following the methodological framework outlined by Arksey and O'Malley (2005), three complementary searches were carried out between August and November 2024 to map available datasets that included the WEMWBS or SWEMWBS and domestic abuse-related variables.

The first was a broad search aimed at identifying longitudinal studies administering SWEMWBS at two timepoints within UK-based adult populations. Although domestic abuse was not a specific inclusion criterion, studies mentioning domestic abuse-affected populations were flagged. This search was conducted using Google Scholar and the official WEMWBS website. Studies were included if they used SWEMWBS longitudinally, involved adult UK populations, and were published in English. Cross-sectional studies and those with only a single wave of SWEMWBS were excluded. This search identified the Avon Longitudinal Study of Parents and Children (ALSPAC, University Of Bristol, 2009) as a potential dataset for further validation analyses.

The second search was a focused publication search that explicitly included both the WEMWBS and domestic abuse-affected populations, with at least two waves of data collection. It was conducted across six databases: Medline, Scopus, PsycInfo, Web of Science, Cochrane Library, and Google Scholar, using Boolean search strings that combined WEMWBS and domestic abuse-related terms. Although no eligible references were found through initial screening, two additional

studies that met eligibility criteria were identified from Blodgett and colleagues' (2022) systematic review. The first dataset identified was from the Roadmap Programme (Stanley et al., 2021), a collaborative initiative by Women's Aid Federation England (WAFE) and SafeLives (SL). Over five years, the programme conducted two interventions aimed at the prevalence, impact, and tolerance of domestic abuse for female survivors of domestic abuse. Although the WEMWBS was administrated at 3 timepoints, only 35 participants completed 2 waves. The second potential dataset identified was from the Barnardo's Opening Closed Doors Programme to support children and families who experienced domestic abuse (Institute of Public Care, 2020). WEMWBS was completed by 154 parents at 2 timepoints.

Finally, a dataset catalogue search was conducted to identify large-scale, UK-based datasets including WEMWBS and domestic abuse measures. Within the Catalogue of Mental Health Measures (2025), the term "Warwick Edinburgh Mental Wellbeing Scale" was searched, identifying 22 datasets that administered WEMWBS. On reviewing the documentation of each dataset, two datasets were found to include populations affected by domestic abuse: the Adult Psychiatric Morbidity Survey 2014 (APMS, McManus et al., 2016) and the Avon Longitudinal Study of Parents and Children (ALSPAC) was re-identified.

The APMS is a repeated cross-sectional study series that is completed every 7–10 years to monitor mental health and treatment among adults living in private households in England (McManus et al., 2016). It is conducted by the National Centre for Social Research, in collaboration with the University of Leicester, commissioned by NHS Digital, and funded by the Department of Health and Social Care (DHSC). The survey had a 57% response rate, resulting in 7,546 participants aged 16 and over. For full details on the methodology of the APMS 2014 data collection cycle, see McManus et al., 2016.

The APMS 2014 was selected due to its suitable sample size of 7,546 participants. The dataset uses multiple indicators to measure domestic abuse and administers several mental health measures that allow convergent validity to be tested. In contrast to other datasets, the APMS dataset could be proficiently accessed via the UK Data Service alongside the accompanying documentation.

Measures

For the purposes of this validation study, only a small number of measures were used from the APMS. These included demographics, the WEMWBS, a measure of depression, and measures of domestic abuse.

Wellbeing was measured using the 14-item WEMWBS (see Appendix O). Each item is measured using a 5-item Likert scale ranging from "None of the Time" (1) to "All of the Tim" (5). Total WEMWBS scores were calculated as the sum of the 14 items.

The APMS measured domestic abuse victimisation comprehensively, including items focusing on abuse perpetrated by partners, family members, and other cohabiting adults. For this study, domestic abuse was measured with five items pertaining to financial abuse, emotional abuse, harassment, less severe physical force, and more severe physical force perpetrated by a partner (see Appendix P). For initial analyses, physical and non-physical domestic abuse was also assessed. The financial abuse, emotional abuse, and harassment items indicated non-physical abuse, while the remaining two items indicated physical abuse. Participants were recorded as having DVA

experience if they reported experience of any of these behaviours.[7] Follow-up questions also differentiated between DVA experiences within the past 12 months and previously in their lifetime. The initial analyses explored differences across categories of DVA experience (type and recency). For the psychometric analyses validating the WEMWBS scale, the DVA group included those who had experienced any form of DVA at any moment in their lifetime.

To explore convergent validity, the Clinical Interview Schedule – Revised (CIS-R) measure of depression symptoms was retained (Lewis et al., 1992). The sub-scale of depression symptoms includes four items that ask participants to indicate whether they have experienced a particular symptom within a specified timeframe (responses are binary, "yes/no"). Participants score one for each of (a) "Unable to enjoy or take an interest in things as much as usual in past week"; (b) "Felt sad, miserable or depressed/unable to enjoy or take an interest in things on four days or more in the past week"; (c) "Felt, sad, miserable or depressed/unable to enjoy or take an interest in things for more than three hours in total on any day in past week"; and (d) "When sad, miserable or depressed you did not become happier when something nice happened, or when in company." Total scores are the sum of these items, ranging from zero to four, with higher scores indicating a greater presence of depression.

The single item "Do you (and your family or partner) have enough money to make regular savings of £10 a month or more for rainy days or retirement?" was used as an indicator of poverty. Response options included "we have this" (indicating no poverty (for the purpose of this analysis)), "we would like to have this but cannot afford this at the moment" (indicating poverty (for the purpose of this analysis)) and "we do not want/need this at the moment or does not apply", which was recorded as not applicable. This item is used, among others, to measure material deprivation of the UK Household Longitudinal Study (Cribb et al, 2024)Other demographics assessed included gender, age and ethnicity (see Table 7).

Sample

The present study included 6,621 participants (who were aged over 18 and had completed all WEMWBS and DVA items); their key demographics are shown in Table 7. Lifetime prevalence of DVA was significantly more likely to be reported by women, 25–54-year-olds, and individuals experiencing poverty.

[7] As above, we use the acronym 'DVA' when distinguishing between samples that have experienced domestic abuse (DVA) and those that have not (non-DVA) to ease interpretation of findings.

Table 7. Key demographics of the sample

Demographic		Non-DVA sample N (%)	DVA sample N (%)	Total sample N (%)
Gender	Man	2,207 (44.2%)*	446 (27.4%)*	2,653 (40.1%)
	Woman	2,789 (55.8%)*	1,179 (72.6%)*	3,968 (59.9%)
Age	18–24	284 (5.7%)	110 (6.8%)	394 (6%)
	25–34	679 (13.6%)*	280 (17.2%)*	959 (14.5%)
	35–44	758 (15.2%)*	320 (19.7%)*	1,078 (16.3%)
	45–54	823 (16.5%)*	370 (22.8%)*	1,193 (18%)
	55–64	810 (16.2%)	296 (18.2%)	1,106 (16.7%)
	65–74	895 (17.9%)*	172 (10.6%)*	1,067 (16.1%)
	75+	747 (15.0%)*	77 (4.7%)*	824 (12.4%)
Ethnicity	White British	4,261 (85.3%)*	1,418 (87.3%)*	5,679 (85.8%)
	White other	297 (5.9%)*	74 (4.6%)*	371 (5.6%)
	Asian	220 (4.4%)*	49 (3.0%)*	269 (4.1%)
	Black/African/Caribbean	122 (2.4%)	46 (2.8%)	168 (2.5%)
	Mixed/multiple/other	88 (1.8%)	30 (1.8%)	118 (1.8%)
	No response	8 (0.2%)	8 (0.5%)	16 (0.2%)
Poverty	No poverty indicated	3,989 (79.8%)*	1,026 (63.1%)*	5,015 (75.7%)

Demographic		Non-DVA sample N (%)	DVA sample N (%)	Total sample N (%)
	Poverty indicated	738 (14.8%)*	516 (31.8%)*	1,254 (18.9%)
	No response/not applicable	269 (5.5%)	83 (5.1%)	442 (5.3%)

* Significant differences between the DVA and non-DVA samples.

Data processing and analysis plan

Missing data was reviewed first to explore any item-level missingness and the scope of potential underreporting. The full dataset was 'cleaned' by excluding data entries from participants who were under the age of 18, entries with missing WEMWBS scores, and entries with missing DVA scores (*N*=925 excluded). This resulted in a total sample of 6,621 participants and a sample of 1,625 adults that have experienced DVA (either in their lifetime or the past year). Data cleaning was completed in SPSS. Analysis was then conducted using R. To validate the use of the WEMWBS in individuals who have experienced domestic abuse, several statistical analyses were conducted. These mirrored the analyses conducted with the OxWell sample. First, descriptive statistics were explored to summarise the sample characteristics and assess the distribution of WEMWBS scores.

The internal consistency of the scale was evaluated using Cronbach's alpha, with a threshold of ≥0.7 indicating acceptable reliability. Additionally, item-total correlations were examined to identify any weak or problematic items. The split-half reliability of the scale was also assessed to further evaluate the internal consistency of the scale. As the sample only includes cross-sectional data at a single timepoint, test-retest reliability could not be explored.

Convergent validity was explored by analysing the relationship between the WEMWBS and the Clinical Interview Schedule – Revised (CIS-R) measure of depression. A negative correlation is expected to determine that the WEMWBS measure of wellbeing is inversely related to the CIS-R measure of depression.

To explore the factor structure of WEMWBS in this population, CFA was performed using the WLSMV estimator, which provides robust standard errors for ordinal data. Fit indices including the Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), root mean square error of approximation (RMSEA), and standardised root mean square residual (SRMR) were used to evaluate model adequacy. Acceptable fit is indicated by CFI and TLI>.90 and RMSEA and SRMR <.08 (Hu and Bentler, 1999).

Finally, measurement invariance was assessed across individuals with and without DVA experience to determine whether WEMWBS functions consistently in those who have and have not experienced DVA. Measurement invariance (configural, metric, and scalar) was confirmed where there were minimal changes in fit indices between nested models (ΔCFI and ΔTLI <-.01, ΔRMSEA

<.015, and Δ SRMR <.01; Cheung and Rensvold, 2002; Chen, 2007). The WLSMV estimator was also used in measurement invariance testing.

Key findings

Descriptives

Exploration of missing data

There were 6,737 participants (aged 18 and over) who responded to DVA measures but 58 of these did not complete the WEMWBS; 54 were missing because they selected "I don't know" in response to one or more WEMWBS item (40 from the non-DVA group and 14 from the DVA group), and 4 participants refused to respond to one or more WEMWBS item (all from the non-DVA group). Item-level missingness on the WEMWBS ranged between 0 (Item 14: feeling cheerful) and 18 (Item 1: feeling optimistic about the future).

On the other hand, there were 7,462 participants who completed the WEMWBS, but 783 of these did not complete domestic abuse measures; 196 were missing because they had never been in a relationship. Further item-level missingness showed between 17 and 19 participants refused to respond to DVA items, between 6 and 79 participants selected they "did not know" and between 469 and 541 selected "not applicable" to DVA items. An Analysis of Variance (ANOVA) with Bonferroni post-hoc analyses demonstrated that the wellbeing scores of participants who were missing DVA scores (for any of the reasons outlined) were significantly lower than those who have never experienced DVA (p<.001) but were not significantly different from the DVA group (p=1.00), F (2, 7459)=186.98, p<.001.

Prevalence of DVA

The dataset consisted of complete responses from a total of 6,621 participants. Of these, 1,625 individuals (24.5%) reported having experienced at least one form of DVA (as measured in this study) within their lifetime, while 4,996 individuals (75.5%) reported no history of DVA. Of those reporting DVA victimisation, 17.04% had experienced DVA within the past year (4.18% of the total sample).

A breakdown of the DVA-experienced group showed that 21.7% experienced only physical DVA, 31% experienced only non-physical DVA, and 47.2% experienced both physical DVA and non-physical DVA. The prevalence of each category of DVA within the total sample is shown in Table 8.

Table 8. DVA prevalence by type, method, and recency

DVA victimisation item		*N* (%)
Type	Physical	1,120 (16. 9%)
	Non-physical	1,272 (19.2%)

DVA victimisation item		*N* (%)
Method*	Financial	524 (7.9%)
	Belittling	969 (14.6%)
	Pushing	983 (14.8%)
	Harassing	470 (7.1%)
	Kicking	796 (12.0%)
Recency	Within lifetime (inc. past year)	1,625 (24.5%)
	Within past year	277 (4.2%)

Total *N*=6,621; * methods correspond to the scale items for DVA measurement presented in Appendix P; percentages represent proportion of the full sample.

Mean differences in WEMWBS Scores

The distribution of WEMWBS scores in the DVA and non-DVA samples can be found in Appendix Q. A statistical comparison of the mean WEMWBS scores indicated a significant difference in mental wellbeing between individuals who had experienced DVA and those who had not, $t(2{,}444.6)=22.145$, $p<.001$. Participants who had experienced DVA ($M=49.03$, $SD=10.14$) reported significantly lower WEMWBS scores on average, suggesting reduced overall wellbeing compared with those who had not experienced DVA ($M=53.53$, $SD=8.68$).

Scores were also significantly different between those that experienced DVA within the past year ($M=47.57$, $SD=10.80$), compared with previously in their lifetime ($M=49.33$, $SD=9.98$), $t(379.09)=2.51$, $p=.01$. See Table 9 for mean scores across subsamples.

Table 9. Mean WEMWBS score in each subsample

	N	*M*	*SD*
Non-DVA sample	4,996	53.53	8.68
Any DVA sample	1,625	49.03	10.14
Lifetime DVA sample	1,348	49.33	9.98
Past year DVA sample	277	47.57	10.80

	N	*M*	*SD*
Physical DVA only sample	353	51.88	9.25
Non-physical DVA only sample	505	49.40	9.54
Both physical & non-physical DVA sample	767	47.48	10.62

Total *N*=6,621.

Reliability

Internal consistency

The internal consistency of the WEMWBS was assessed using Cronbach's alpha to determine the scale's reliability. The Cronbach's alpha of the WEMWBS scale for the non-DVA sample (α=.89; 95% confidence intervals = .89–.90) and the DVA sample (α=.92; 95% confidence intervals = .91–.93) were similar in strength and both estimates were greater than the required standard of .70, indicating good reliability of the scale in both samples.

Furthermore, corrected item-total correlations ranged from .53 to .77, confirming that all items were sufficiently correlated with the overall scale. These findings demonstrate that the WEMWBS is a reliable measure of mental wellbeing in both DVA and non-DVA samples.

The values of the Cronbach's alpha, if items were removed, ranged from .89 to .92, demonstrating that no single item substantially detracted from the scale's reliability. These alpha ranges further confirm that all items contribute meaningfully to the overall construct. Overall, findings demonstrate that the internal consistency of the scale is supported (see Appendix R for supporting tables).

Split-half reliability

As testing only took place at a single timepoint, test-retest reliability is not possible. Instead, we report the split-half reliability of the test, assessed using Guttman's lambda coefficients (see Appendix R: Table R3). The average split-half reliability was .89 for the non-DVA sample and .92 for the DVA sample, indicating good internal consistency above the generally accepted threshold of .70. The minimum split-half reliability was .81 for the non-DVA sample and .88 for the DVA sample, indicating that even in the least favourable split, reliability remained above the accepted threshold.

The results of the split-half reliability and Cronbach's alpha testing support the internal consistency of the WEMWBS scale, indicating that the scale performs robustly within these populations.

Validity

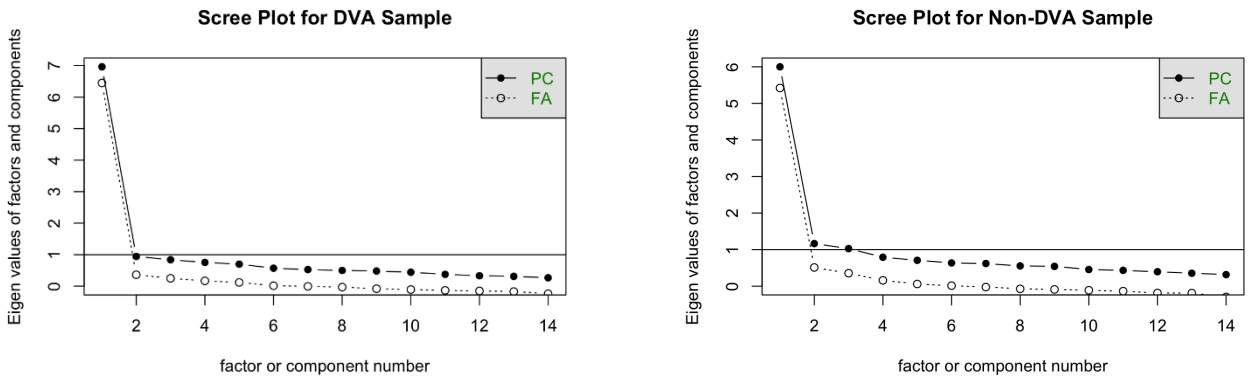
Convergent validity

The relationship between the WEMWBS and the Clinical Interview Schedule – Revised (CIS-R) was explored with the expectation that they would be negatively correlated due to the constructs measured by the scales being inversely related: mental wellbeing (measured by WEMWBS) versus symptoms of depression (measured by the CIS-R depression symptom score). The WEMWBS was strongly negatively correlated with the depression symptom measure of the CIS-R (r=-.51, p<.001) across the full sample. This finding was consistent across the non-DVA (r=-.61, p<.001) and DVA (r=-.42, p<.001) samples, underscoring the validity of the WEMWBS as an indicator of mental wellbeing inversely related to depression.

Factor analysis

Scree plots were generated for both the DVA sample and the non-DVA sample to assess the number of factors underlying the WEMWBS and ensure that confirmatory factor analysis (CFA) with a single factor model is appropriate. In both samples, the scree plots show a clear "elbow" after the first factor, suggesting that a single factor explains most of the variance in the data (see Figure 5).

Figure 5. Scree plots of number of factors measured by the WEMWBS scale for the DVA sample and the non-DVA sample (go to accessibility text)



CFAs with the WLSMV estimator were conducted to evaluate whether the WEMWBS adheres to a one-factor structure in each of the samples. In the initial model, model fit indices did not present the strongest fit, with fit indices slightly below thresholds of acceptability, particularly in the non-DVA sample (see Appendix S: Table 1). Modification indices were inspected, and the recommended correlated residuals were added to the model in a stepwise fashion.

The error of Items 9 and 12 were correlated, followed by Items 6 and 7. These items demonstrate additional similarities focusing on relationships and on decision making, which theoretically justify these correlated residuals. It is commonplace in the literature that error covariance parameters are

released in one-factor WEMWBS models, with these items repeatedly demonstrating additional relationships (e.g. Smith et al., 2017; Sarasjärvi et al., 2023).

In the final model, the chi-square test was significant for both the non-DVA sample ($\chi2(75)=3,331.29$, $p<.001$) and the DVA sample ($\chi2(75)=912.89$, $p<.001$), probably due to the large sample sizes, because the chi-square test is sensitive to sample size (Shi et al., 2018). Although the chi-square was significant, the model demonstrated adequate fit to the data in all other fit indices in the full sample model and in the DVA sample (with the exception of RMSEA, but as outlined above this is a less reliable index; see Appendix S: Table S1). The model for the non-DVA sample presented slightly mixed findings; the CFI exceeds the commonly accepted threshold of .90, indicating acceptable fit. However, the SRMR and TLI exceeded their respective thresholds by .01.

Items demonstrated moderate to strong factor loadings to the single factor in both samples, ranging from .46 to .83 (see Appendix S : Table 2). This indicates that each item contributes meaningfully to the measurement of the overall construct of mental wellbeing. Items 8 ("I've been feeling good about myself") and 10 ("I've been feeling confident") consistently showed stronger loadings, suggesting they are particularly representative of the construct.

The CFA results support the one-factor structure of the WEMWBS in both the full sample and the DVA sample. Despite significant chi-square values, the adjusted model demonstrated adequate fit according to all other fit indices (CFI, TLI, RMSEA, and SRMR). The strong factor loadings provide further evidence of the unidimensional nature of the scale and its effectiveness in capturing the construct of mental wellbeing. Although the non-DVA model presented borderline fit, it was deemed appropriate to move forward with measurement invariance testing due to the strong factor loadings and existing research justification for the single factor model in the general population (Tennant et al., 2007).

Measurement invariance

Measurement invariance was assessed to determine whether the WEMWBS is interpreted and functions equivalently across those with and without DVA experiences. The sequence of testing followed the steps for configural, metric, and scalar invariance using the WSLMV estimator. Each of the configural, metric, scalar, and residual invariance models demonstrated acceptable fit; although chi-square tests were significant, all other fit indices demonstrated good fit (see Appendix T: Table T1).

Comparison of nested models demonstrated minimal changes in fit indices, indicating full measurement invariance (see Appendix T: Table T2). This confirms that both groups conceptualise mental wellbeing similarly, the relationship between the items and the latent factor (wellbeing) is consistent, group differences reflect true differences in wellbeing instead of measurement bias, and the measurement error of items is similar across groups.

Discussion

This study provided strong evidence that the WEMWBS scale reliably measures wellbeing in adults with experiences of domestic abuse. CFA confirmed its unidimensional structure, and measurement invariance testing indicated that the WEMWBS functioned equivalently across those with and without domestic abuse experiences. As expected, WEMWBS scores were significantly

lower among those with a history of domestic abuse, and a strong negative correlation with depression scores further supported its validity. These findings reinforce that the WEMWBS is a valid tool for assessing emotional health and wellbeing in adults who experience domestic abuse.

Limitations

Due to almost identical design, this study presented many of the same limitations as the OxWell study, particularly in relation to the reliance on a single wave of data collection, which prevented the examination of sensitivity to change and the test-retest reliability of the WEMWBS. Furthermore, the measure of domestic abuse in this dataset is also retrospective, introducing potential recall bias.

A limitation unique to this dataset is that it only includes individuals living in private households, excluding those in refuges or temporary accommodation, who may be more likely to have experienced domestic abuse and to report more severe impacts on their wellbeing (Stulz et al., 2024).

Furthermore, although this study demonstrates the validity of the scale with adults who have experienced domestic abuse, the participants are not limited to parents and caregivers. Within the domestic abuse victim population there may be differences that are specific to parents and caregivers. For example, mothers are more likely to be held responsible for violence and abuse and can therefore be more susceptible to defensively reporting (Arnull and Stewart, 2021).

Conclusion

In summary, the results of this study provide strong evidence that the WEMWBS is a valid OMI that can be used to measure emotional health and wellbeing in adults who have experienced domestic abuse, with robust psychometric properties.

DISCUSSION

To our knowledge, this suite of studies is the first to systematically explore the validity of a well-used OMI specifically with a domestic abuse population. Our findings are strengthened by the use of mixed methods and datasets that reflect the use of the measure in practice and research contexts.

Overall, our findings demonstrate the validity and acceptability of the SWEMWBS and WEMWBS in domestic abuse-experienced child and adult populations respectively. This is a significant finding given the limited number of measures that have been evaluated for use with this population across practice and research contexts in the UK and internationally. Moreover, it represents an important step forward in the implementation of the DVA-COS, which we hope will help to unify outcome measurement in domestic abuse research and evaluation, as well as service monitoring. However, there are important considerations which emerged, primarily from the think aloud study, but also from our consultation with service users and providers, regarding interpretation, emotional impact, and contextual relevance. These issues are discussed in more detail below.

Validity and reliability of the (S)WEMWBS

Studies B and C demonstrated robust psychometric validity of the SWEMWBS. Both datasets confirmed the OMI's strong internal consistency and unidimensional structure. Study B also highlighted the convergent validity of the SWEMWBS, with scale responses strongly negatively correlating with the RCADS-11. The RCADS-11 is another reputable measure of mental health but is negatively framed, with items focusing on anxiety and depression (mental illness rather than mental wellbeing; Radez et al., 2021). The divergent relationship between these measures highlights the utility of using a positively framed OMI to capture emotional health and wellbeing; the SWEMWBS still effectively captures the intended outcome while being more trauma-informed (i.e. strength-based) and sensitive. Additionally, the test of measurement invariance with the non-DVA comparison group (see Study B) showed that children and young people who had experienced domestic abuse interpreted the SWEMWBS in the same way as those who had not. This further demonstrates the robustness of the SWEMWBS as a measure of wellbeing and allows for scores to be meaningfully compared.

The supplementary analyses in Study C extended these findings by evaluating the SWEMWBS within real-world domestic abuse intervention settings. Wellbeing scores significantly improved between initial and final timepoints, suggesting that SWEMWBS is sensitive to change and can effectively capture wellbeing improvements following intervention. Measurement invariance testing confirmed that the scale operated consistently across pre- and post-intervention assessments. This is in line with the literature that has demonstrated the responsiveness of the SWEMWBS to change at both the individual and group level (Maheswaran et al., 2012; Shah et al., 2021). Furthermore, our findings did not indicate defensive reporting (a concern of practitioners participating in our study) within this sample, with average SWEMWBS scores improving at each timepoint. That said, concerns about defensive reporting were initially raised in relation to adults, whereas these analyses relate to children and young people. Nevertheless, the main takeaway of these findings is that the SWEMWBS is a useful tool for tracking wellbeing changes in service

settings, though further research is needed to explore potential response biases or external influences on scoring.

Taken together, these two studies conclude that the SWEMWBS functions equivalently and is psychometrically valid in the domestic abuse population and is responsive to change in wellbeing. Therefore, the SWEMWBS is a valid tool for the measurement of the *child emotional health and wellbeing* core outcome.

In addition, this project validated the use of the WEMWBS with adults who have experienced domestic abuse. Study D showed the WEMWEBS demonstrates strong reliability and validity in populations who have experienced domestic abuse. The scale functioned equivalently across samples of adults with and without domestic abuse experience, demonstrating consistency in the way wellbeing is captured. This allows for comparison across populations and, as expected, those with a history of domestic abuse reported noticeably lower wellbeing scores. These findings highlight the psychometric suitability of the WEMWBS as an OMI for assessing the *caregiver emotional health and wellbeing* outcome. It is important to note that this study used data from a single timepoint so could not assess the responsiveness to change in this specific population. However, research has demonstrated the ability of the WEMWBS to measure changes in wellbeing over time in both the general population and clinical samples (Chanfreau et al., 2014; Freeman et al., 2015; Maheswaran et al., 2012). As our findings demonstrated equivalence in the functioning of the WEMWBS across populations that had and had not experienced domestic abuse, it could be inferred that it would also detect change in wellbeing in the context of domestic abuse service settings.

In summary, our findings show the SWEMWBS and WEMWBS are appropriate for measuring emotional health and wellbeing in child and adult DVA experienced populations respectively, and therefore meets standards for inclusion in a COS (Prinsen et al., 2016).

Acceptability considerations

The findings about the psychometric robustness of the SWEMWBS/WEMWBS, while significant, must be balanced with acceptability considerations. Previous studies, along with input from our advisory group and stakeholders, emphasise that the acceptability of a measure is the highest priority when determining its use in research or practice settings (Bains et al., forthcoming; Clark et al., 2023; Powell, Clark, et al., 2022). There is a clear consensus that although robust psychometrics are considered necessary when using OMIs with vulnerable groups, they are not sufficient on their own – whereas acceptability is seen as essential and non-negotiable (Prinsen, Vohra, Rose, et al., 2016).

The think aloud study (Study A) suggests that further thought is warranted on how to maximise acceptability of the short-form measure for children and young people with experience of domestic abuse. A potential adaptation – which has been requested in previous work as well as during work package 1 of the current study (Bains et al., forthcoming) – is for the introduction of a free text box in the SWEMWBS/WEMWBS measure (Clark et al., 2023; Powell, Clark, et al., 2022). Feedback suggests this would allow service users to provide additional context to their responses and what might be impacting them at that time. It is possible that a text box can be added without impacting the psychometric functioning of the tool, though this would need careful implementation and

testing. However, clear direction on how free text responses would or would not be used should be given as part of the measure. For example, for use in service contexts, it should be made clear that although text may aid the understanding of the practitioner, it will not be used to augment the score from the measure.

Other concerns relating to the terms and concepts addressed by specific questions of the measure highlight areas for potential refinement; however, this must be balanced against the pressing need for validated measures in this field of research and practice. Material changes to the substance of the measure would require further evaluation and are prohibited by the measure developers. Additionally, the item-level missingness analysis of the OxWell data did not present any undue deterrence from specific items of the SWEMWBS in the DVA group nor the wider population, though this is a school sample and so may not fully reflect item concerns in the context of service provision. We suggest that some of the concerns might be addressed through the development of trauma-informed guidance to supplement the implementation of the OMI. This was also a request of participants in work package 1 (Bains et al., forthcoming); trauma-informed guidelines would be needed for practitioners and researchers to support the use of the tools in a 'care-first' approach, rather than as a tool for screening, triaging, and rationing care. These guidelines might support scaling the implementation of the OMI, serving to address some concerns without disrupting the integrity and functioning of the measure. The implementation of guidance should be closely monitored and evaluated, to inform further refinement to the developed guidance.

Feasibility considerations

It is worth noting that since beginning this work, policy regarding access to the WEMWBS has changed. Up until 1 December 2024 the measure could be freely accessed by researchers and service providers using it in a not-for-profit capacity. However, since 1 December 2024 a charge was introduced for NHS services to use the measure, which may represent a significant barrier to services that are funded or partially funded by the NHS. It remains available free of charge for "organisations and individuals whose main purpose is not directed towards commercial advantage or monetary compensation ('Non-Commercial Organisations'), including Public Sector Organisations (e.g. Universities, Schools, Public Health, Social Services, NGOs), Registered Charities, Registered Community Interest Companies and Registered Social Enterprises only." We recommend exploring the possibility of negotiating reduced-cost or free access when the measure is being used by NHS-commissioned services seeking to implement the DVA-COS.

In response to the findings and recommendations of prior work we have validated and recommended the 14-item WEMWBS to measure caregiver emotional health and wellbeing (Powell, Feder, et al., 2022). It is noted that the short version is nested within the full WEMWBS. If implementation highlights any undue feasibility considerations based on the length of the WEMWBS further work may be conducted to revisit and validate the SWEMWBS for use in adults. However, because the WEMWBS is still a relatively short and quick measure, and captures a wider scope of wellbeing, we recommend its use with adults to measure *caregiver emotional health and wellbeing*.

Age considerations

A notable consideration when implementing the SWEMWBS with children and young people is the age limitation. The official recommendations provided are that the WEMWBS has been validated for adult populations and young people aged 13 years and older (Clark et al., 2011). The official user guidance publication from May 2015 specifies that "there is no evidence for the use of WEMWBS under the age of 13 and it is therefore not recommended to do so". The WEMWBS official website more recently, however, does acknowledge that SWEMWBS has been shown to be validated in children and young people aged 11 years and older in a large-scale study in Welsh secondary schools (Melendez-Torres et al. 2019).

The quantitative research outlined in this report has reiterated the validity of the scale for children and young people aged 11–18, but the think aloud study raises concern over younger participants not being familiar with some wording, particularly "optimistic". Due to the fact that there is not an "I don't know" response, individuals were uncertain how to respond. Therefore, future research must identify or validate an appropriate measure of child emotional health and wellbeing for children under the age of 11. Observation can be used to measure mental wellbeing in younger children to overcome difficulties in completing questionnaires; however, this may not be practical in either practice or research contexts. When considering outcome measurement within service contexts it is paramount that feasibility in terms of financial cost and time taken are at the forefront for implementation to be successful (Deighton et al., 2014). Observation-based measurement is likely to burden resources and could raise concerns over reliability when making comparisons across services and interventions. Parent-proxy completion of scales can also be used for younger children, and has been shown to be reliable and valid for measuring quality of life in children as young as 2 years old (Varni et al., 2007). However, the WEMWBS or SWEMWBS has not been validated for use via parent-proxy reporting. Furthermore, research indicates that parent-proxy measurement is less reliable for measures of psychological wellbeing, particularly when the reporting parent is experiencing depression or anxiety (Oltean and Ferro, 2019). This is especially important in the context of domestic abuse services in which parent-survivors often experience high levels of depression, anxiety, and post-traumatic stress disorder (Ferrari et al., 2016). Although it is often assumed that young children cannot accurately self-report, research has demonstrated that children as young as 5 reliably report their health (Selwyn and Wood, 2015). Previous studies which used the DVA-COS, such as CADA (Barter et al., forthcoming), used the Stirling Children's Wellbeing Scale (Liddle and Carter, 2015) with children 8–12 years old. However, the SWEMWBS is shorter, and therefore less burdensome, as well as being used in many clinical and non-clinical settings, along with being available and validated in many more languages.[8]

Diversity and inclusion

As highlighted in Studies A, B, and D, the samples in this programme of research were majority White. It is important to consider the applicability of these findings to ethnically minoritised

[8] See: https://warwick.ac.uk/fac/sci/med/research/platform/wemwbs/about/use

groups, because they can be more vulnerable to the impacts of domestic abuse (Femi-Ajao et al., 2020; Istratii et al., 2024). The WEMWBS has previously been cross-culturally validated, demonstrating a high level of consistency in measurement across minority ethnic groups (Pakpour et al., 2024; Taggart et al., 2013). The WEMWBS is also available and validated in over 30 languages, and therefore can be confidently used with individuals who speak English as a second language as well as by and for services who support service users in their first language.

Adaptation and implementation

The evidence outlined in this report demonstrates that the SWEMWBS is validated for use with individuals aged 11+ who have experience of domestic abuse. We recommend that the tool can be used with confidence to measure changes in wellbeing in an individual being offered an intervention. Therefore, implementing the SWEMWBS and WEMWBS will start to build an evidence base of how services impact child and caregiver mental wellbeing. However, we do make the above recommendations in conjunction with the following suggested next steps to support its implementation:

1. Further work is needed to identify an alternative OMI or adapt the SWEMWBS for appropriate use by a child under the age of 11.
2. We recommend that a free text box is supplied following the administration of the SWEMWBS and WEMWBS. This is in line with the recommendation in previous work (Clark et al., 2023; Powell, Clark, et al., 2022) and work package 1 (Bains et al., forthcoming). This also addresses the need outlined in this report, that the circumstances which precede the completion of a measure may provide context to an individual's response.
3. We recommend the development of trauma-informed guidelines for practitioners and researchers on how to use the tools in a 'care-first' approach, rather than as a tool for screening, triaging, and rationing care; for commissioners, guidance about how to interpret the data collected is also needed.

BIBLIOGRAPHY

Allen K, Melendez-Torres GJ, Ford T, et al. (2022) Family focused interventions that address parental domestic violence and abuse, mental ill-health, and substance misuse in combination: A systematic review. *PloS One* 17(7): e0270894.

Anthony R, Moore G, Page N, et al. (2022) Measurement invariance of the short Warwick-Edinburgh Mental Wellbeing Scale and latent mean differences (SWEMWBS) in young people by current care status. *Quality of Life Research* 31(1): 205–213.

Arksey H and O'Malley L (2005) Scoping studies: Towards a methodological framework. *International Journal of Social Research Methodology* 8(1): 19–32.

Arnull E and Stewart S (2021) Developing a theoretical framework to discuss mothers experiencing domestic violence and being subject to interventions: A cross-national perspective. *International Journal for Crime, Justice and Social Democracy* 10(2): 113–126.

Bains S, Dunk E, Gonidis L, et al. (forthcoming) *Selecting and Validating Outcome Measures for the Domestic Violence And Abuse Core Outcome Set (DVA-COS) – Work Package 1: Identifying Outcome Measures [manuscript in preparation].*

Barter CA, Richardson Foster H, Howarth E, et al. (forthcoming) *CADA Evaluation – Children Affected by Domestic Abuse [manuscript in preparation].*

Bartram DJ, Sinclair JM and Baldwin DS (2013) Further validation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) in the UK veterinary profession: Rasch analysis. *Quality of Life Research* 22(2): 379–391.

Bartram DJ, Yadegarfar G, Sinclair JMA, et al. (2011) Validation of the Warwick–Edinburgh Mental Well-being Scale (WEMWBS) as an overall indicator of population mental health and well-being in the UK veterinary profession. *The Veterinary Journal* 187(3): 397–398.

Beauducel A and Herzberg PY (2006) On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal* 13(2): 186–203.

Blodgett JM, Birch JM, Musella M, et al. (2022) What works to improve wellbeing? A rapid systematic review of 223 interventions evaluated with the Warwick-Edinburgh Mental Well-Being Scales. *International Journal of Environmental Research and Public Health* 19(23): 15845.

Boateng GO, Neilands TB, Frongillo EA, et al. (2018) Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health* 6: 149.

Bracewell K, Larkins C, Radford L, et al. (2020) Educational opportunities and obstacles for teenagers living in domestic violence refuges. *Child Abuse Review* 29(2): 130–143.

Callaghan JEM, Alexander JH, Sixsmith J, et al. (2018) Beyond "witnessing": Children's experiences of coercive control in domestic violence and abuse. *Journal of Interpersonal Violence* 33(10): 1551–1581.

Caputo A (2017) Social desirability bias in self-reported well-being measures: Evidence from an online survey. *Universitas Psychologica* 16(2).

Carlisle S, Bunce A, Prina M, et al. (2024) Trends in outcomes used to measure the effectiveness of UK-based support interventions and services targeted at adults with experience of domestic and sexual violence and abuse: A scoping review. British Medical Journal Publishing Group. Epub ahead of print 1 April 2024. DOI: 10.1136/bmjopen-2023-074452.

Carlisle S, Bunce A, Prina M, et al. (2025) Effectiveness of UK-based support interventions and services aimed at adults who have experienced or used domestic and sexual violence and abuse: A systematic review and meta-analysis. *BMC Public Health*. BioMed Central. Epub ahead of print 10 February 2025.

Chanfreau J, Lloyd C, Byron C, et al. (2014) *Predicting Wellbeing*. NatCen Social Research. Available at: http://rgdoi.net/10.13140/2.1.3267.4564 (accessed 28 March 2025).

Chen FF (2007) Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal* 14(3): 464–504.

Cheung GW and Rensvold RB (2002) Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal* 9(2): 233–255.

Cilar L, Pajnkihar M and Štiglic G (2020) Validation of the Warwick-Edinburgh Mental Well-being Scale among nursing students in Slovenia. *Journal of Nursing Management*: jonm.13087.

Clark SE, Kimber M, Downes L, et al. (2023) Engaging domestic abuse practitioners and survivors in a review of outcome tools: Reflections on differing priorities. *Research for All* 7(1).

Clarke A, Friede T, Putz R, et al. (2011) Warwick-Edinburgh Mental Well-being Scale (WEMWBS): Validated for teenage school students in England and Scotland. A mixed methods assessment. *BMC Public Health* 11(1): 487.

Cribb, J., Henry, A., Karjalainen, H., Ray-Chaudhuri, S., Waters, T. and Wernham, T. (2024). Living standards, Poverty and Inequality in the UK: 2024. *Living standards, Poverty and Inequality in the UK: 2024*. Available at: https://ifs.org.uk/publications/living-standards-poverty-and-inequality-uk-2024.

Dargis M and Koenigs M (2017) Witnessing domestic violence during childhood is associated with psychopathic traits in adult male criminal offenders. *Law and Human Behavior* 41(2): 173–179.

Deighton J, Croudace T, Fonagy P, et al. (2014) Measuring mental health and wellbeing outcomes for children and adolescents to inform practice and policy: A review of child self-report measures. *Child and Adolescent Psychiatry and Mental Health* 8(1): 14.

Department for Work and Pensions (2025) *Households Below Average Income: An analysis of the UK income distribution: FYE 1995 to FYE 2024*. Available at: https://www.gov.uk/government/statistics/households-below-average-income-for-financial-

years-ending-1995-to-2024/households-below-average-income-an-analysis-of-the-uk-income-distribution-fye-1995-to-fye-2024.

Domestic Abuse Act (2021) Domestic Abuse Act 2021. King's Printer of Acts of Parliament. Available at: https://www.legislation.gov.uk/ukpga/2021/17/contents/enacted (accessed 10 August 2023).

Dong A, Chen X, Zhu L, et al. (2016) Translation and validation of a Chinese version of the Warwick–Edinburgh Mental Well-being Scale with undergraduate nursing trainees. *Journal of Psychiatric and Mental Health Nursing* 23(9–10): 554–560.

Evans M, Gregory A, Feder G, et al. (2015) "Even 'daily' is not enough": How well do we measure domestic violence and abuse? A think-aloud study of a commonly used self-report scale. *Violence and victims* 31.

Evans R, Trubey R, MacDonald S, et al. (2024) What mental health and wellbeing interventions work for which children and young people in care? Systematic review of potential outcome inequities. *Child and Adolescent Social Work Journal*. Epub ahead of print 28 January 2024. DOI: 10.1007/s10560-023-00956-7.

Evans SE, Davies C and DiLillo D (2008) Exposure to domestic violence: A meta-analysis of child and adolescent outcomes. *Aggression and Violent Behavior* 13(2): 131–140.

Fallon B, Trocmé N, Fluke J, et al. (2010) Methodological challenges in measuring child maltreatment. *Child Abuse & Neglect* 34(1): 70–79.

Femi-Ajao O, Kendal S and Lovell K (2020) A qualitative systematic review of published work on disclosure and help-seeking for domestic violence and abuse among women from ethnic minority populations in the UK. *Ethnicity & Health* 25(5): 732–746.

Ferrari G, Agnew-Davies R, Bailey J, et al. (2016) Domestic violence and mental health: A cross-sectional survey of women seeking help from domestic violence support services. *Global Health Action* 9(1): 29890.

Foundations (2023) Over 827,000 children estimated to have experienced domestic abuse this year, with concerns of increase at Christmas underscoring urgent need for stronger support. Available at: https://foundations.org.uk/press-release/over-827000-children-estimated-to-have-experienced-domestic-abuse-this-year-with-concerns-of-increase-at-christmas-underscoring-urgent-need-for-stronger-support (accessed 6 March 2025).

Freeman D, Dunn G, Startup H, et al. (2015) Effects of cognitive behaviour therapy for worry on persecutory delusions in patients with psychosis (WIT): A parallel, single-blind, randomised controlled trial with a mediation analysis. *The Lancet Psychiatry* 2(4): 305–313.

Fung S (2019) Psychometric evaluation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS) with Chinese University Students. *Health and Quality of Life Outcomes* 17(1): 46.

Georgieva S, Tomás JM, Navarro-Pérez JJ, et al. (2023) Systematic review and critical appraisal of five of the most recurrently validated child maltreatment assessment instruments from 2010 to 2020. *Trauma, Violence & Abuse* 24(4): 2448–2465.

Griffith R (2016) What is Gillick competence? *Human Vaccines & Immunotherapeutics* 12(1): 244–247.

Hameed M, O'Doherty L, Gilchrist G, et al. (2020) Psychological therapies for women who experience intimate partner violence. *The Cochrane Database of Systematic Reviews* 7(7): CD013017.

Hays PA (2008) *Addressing Cultural Complexities in Practice: Assessment, Diagnosis, and Therapy*. 2nd ed. Washington, DC: American Psychological Association.

Holden GW (2003) Children exposed to domestic violence and child abuse: Terminology and taxonomy. *Clinical Child and Family Psychology Review* 6(3): 151–160.

Howarth E, Powell C, Dunk E, et al. (2025) *Developing Outcome Measures: Domestic Abuse Core Outcome Set Research Protocol*. Foundations – What Works Centre for Children & Families. Available at: https://foundations.org.uk/wp-content/uploads/2025/01/developing-outcome-measures-domestic-abuse-core-outcome-set-protocol.pdf.

Howarth E, Powell C, Woodman J, et al. (2021) Protocol for developing core outcome sets for evaluation of psychosocial interventions for children and families with experience or at risk of child maltreatment or domestic abuse. *BMJ Open* 11(8). British Medical Journal Publishing Group: e044431.Hu L and Bentler PM (1999) Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal* 6(1): 1–55.

Hughes KL, Williamson PR, and Young B (2022) In-depth qualitative interviews identified barriers and facilitators that influenced chief investigators' use of core outcome sets in randomised controlled trials. *Journal of Clinical Epidemiology* 144: 111–120.

Institute of Public Care (2020) *Barnardo's Cymru Opening Closed Doors Programme Evaluation Final Report*. Oxford Brookes University.

Istratii R, Ali P, and Feder G (2024) Integration of religious beliefs and faith-based resources in domestic violence services to migrant and ethnic minority communities: A scoping review. *Violence: An International Journal* 5(1): 94–122.

Katz E, Nikupeteri A, and Laitinen M (2020) When coercive control continues to harm children: Post-separation fathering, stalking and domestic violence. *Child Abuse Review* 29(4): 310–324.

Kenny DA, Kaniskan B, and McCoach DB (2015) The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research* 44(3): 486–507.

Kimberlin CL and Winterstein AG (2008) Validity and reliability of measurement instruments used in research. *American Journal of Health-System Pharmacy* 65(23): 2276–2284.

Kirkham JJ, Gargon E, Clarke M, et al. (2013) Can a core outcome set improve the quality of systematic reviews? A survey of the Co-ordinating Editors of Cochrane review groups. *Trials* 14(1): 21.

Kitzmann KM, Gaylord NK, Holt AR, et al. (2003) Child witnesses to domestic violence: A meta-analytic review. *Journal of Consulting and Clinical Psychology* 71(2): 339–352.

Konaszewski K, Niesiobędzka M, and Surzykiewicz J (2021) Factor structure and psychometric properties of a Polish adaptation of the Warwick–Edinburgh Mental Wellbeing Scale. *Health and Quality of Life Outcomes* 19(1): 70.

Koushede V, Lasgaard M, Hinrichsen C, et al. (2019) Measuring mental well-being in Denmark: Validation of the original and short version of the Warwick-Edinburgh mental well-being scale (WEMWBS and SWEMWBS) and cross-cultural comparison across four European settings. *Psychiatry Research* 271: 502–509.

Krause KR, Chung S, Adewuya AO, et al. (2021) International consensus on a standard set of outcome measures for child and youth anxiety, depression, obsessive-compulsive disorder, and post-traumatic stress disorder. *The Lancet. Psychiatry* 8(1): 76–86.

Lawson J (2019) Domestic violence as child maltreatment: Differential risks and outcomes among cases referred to child welfare agencies for domestic violence exposure. *Children and Youth Services Review* 98: 32–41.

Lewis G, Pelosi AJ, Araya R, et al. (1992) Measuring psychiatric disorder in the community: A standardized assessment for use by lay interviewers. *Psychological Medicine* 22(2): 465–486.

Liddle I and Carter GFA (2015) Emotional and psychological well-being in children: The development and validation of the Stirling Children's Well-being Scale. *Educational Psychology in Practice*. Routledge. Epub ahead of print 3 April 2015.

Livings MS, Hsiao V, and Withers M (2023) Breaking the cycle of family violence: A critique of family violence interventions. *Trauma, Violence & Abuse* 24(4): 2544–2559.

Macmillan HL, Wathen CN, Barlow J, et al. (2009) Interventions to prevent child maltreatment and associated impairment. *Lancet* 373(9659): 250–266.

Maheswaran H, Weich S, Powell J, et al. (2012) Evaluating the responsiveness of the Warwick Edinburgh Mental Well-Being Scale (WEMWBS): Group and individual level analysis. *Health and Quality of Life Outcomes* 10(1): 156.

Mansfield KL, Puntis S, Soneson E, et al. (2021) Study protocol: The OxWell school survey investigating social, emotional and behavioural factors associated with mental health and well-being. *BMJ Open* 11(12): e052717.

McGarry J and Ali P (2016) Researching domestic violence and abuse in healthcare settings: Challenges and issues. *Journal of Research in Nursing* 21(5–6): 465–476.

McFarlane, K. (2021). *Report on the 5th Annual Child and Family Poverty Surveys 2021 'Poverty affects every aspect of their lives' (10-13yrs)*. Available at: https://www.childreninwales.org.uk/application/files/5816/3663/4886/Child_Poverty_Survey_Report_2021_English_v7.pdf.

McKay MT and Andretta JR (2017) Evidence for the psychometric validity, internal consistency and measurement invariance of Warwick Edinburgh Mental Well-being scale scores in Scottish and Irish adolescents. *Psychiatry Research* 255: 382–386.

McKim, C.A. (2017) The value of mixed methods research: A mixed methods study. *Journal of mixed methods research*, *11*(2): 202-222.

McManus S, Bebbington P, Jenkins R, et al. (2016) *Mental health and wellbeing in England: Adult Psychiatric Morbidity Survey*. Leeds: NHS Digital. Available at: https://discovery.ucl.ac.uk/id/eprint/1518055/1/APMS%202014-full-rpt.pdf.

Melendez-Torres GJ, Hewitt G, Hallingberg B, et al. (2019) Measurement invariance properties and external construct validity of the short Warwick-Edinburgh mental wellbeing scale in a large national sample of secondary school students in Wales. *Health and Quality of Life Outcomes* 17(1): 139.

Melin J, Lundin A, and Johansson M (2022) An off-target scale limits the utility of Short Warwick-Edinburgh Mental Well-Being Scale (SWEMWBS) as a measure of well-being in public health surveys. *Public Health* 202: 43–48.

Ng Fat L, Scholes S, Boniface S, et al. (2017) Evaluating and establishing national norms for mental wellbeing using the short Warwick–Edinburgh Mental Well-being Scale (SWEMWBS): Findings from the Health Survey for England. *Quality of Life Research* 26(5): 1129–1144.

O'Doherty LJ, Taft A, Hegarty K, et al. (2014) Screening women for intimate partner violence in healthcare settings: Abridged Cochrane systematic review and meta-analysis. *BMJ* 348: g2913.

Oltean II and Ferro MA (2019) Agreement of child and parent-proxy reported health-related quality of life in children with mental disorder. *Quality of Life Research* 28(3): 703–712.

Oram S, Fisher HL, Minnis H, et al. (2022) The Lancet Psychiatry Commission on intimate partner violence and mental health: Advancing mental health services, research, and policy. *The Lancet Psychiatry* 9(6): 487–524.

Pakpour AH, Eriksson M, Erixon I, et al. (2024) The Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): A psychometric evaluation of adolescents in Sweden during the COVID-19 pandemic. *Heliyon* 10(6): e27620.

Patel-Syed Z, Becker S, Olson M, et al. (2024) What do you think it means? Using cognitive interviewing to improve measurement in implementation science: Description and case example. *Implementation Science Communications* 5(1): 14.

Paton LW, Tiffin PA, Barkham M, et al. (2023) Mental health trajectories in university students across the COVID-19 pandemic: Findings from the Student Wellbeing at Northern England Universities prospective cohort study. *Frontiers in Public Health* 11: 1188690.

Piqueras JA, Pineda D, Martin-Vivar M, et al. (2017) Confirmatory factor analysis and psychometric properties of the Revised Child Anxiety and Depression Scale–30 (RCADS-30) in clinical and non-clinical samples. *Revista de Psicopatología y Psicología Clínica* 22(3): 183.

Powell C, Clark SE, Downes L, et al. (2022) A rapid review of outcome measurement tools related to the DVA Core Outcome Set. Epub ahead of print March 2022.

Powell C, Feder G, Gilbert R, et al. (2022) Child and family-focused interventions for child maltreatment and domestic abuse: Development of core outcome sets. *BMJ Open* 12(9): e064397.

Powell C, Peeren S, Ostrowska A, et al. (2023) The Importance of Grey and Qualitative Literature in Developing Domestic Violence and Abuse and Child Maltreatment Core Outcome Sets: A Brief Report. *Journal of Family Violence*. Epub ahead of print 31 October 2023.

Powell C, Szilassy E, Cowan K, et al. (2025) Adapting a consensus process for survivors of domestic abuse and child maltreatment: a brief report about adopting a trauma-informed approach in multistakeholder workshops. British Medical Journal Publishing Group. Epub ahead of print 1 January 2025.

Prinsen CAC, Vohra S, Rose Michael R, et al. (2016) Guideline for selecting outcome measurement instruments for outcomes included in a core outcome set. Epub ahead of print September 2016.

Prinsen CAC, Vohra S, Rose Michael R., et al. (2016) How to select outcome measurement instruments for outcomes included in a "core outcome set": A practical guideline. *Trials* 17(1): 449.

Radez J, Waite P, Chorpita B, et al. (2021) Using the 11-item version of the RCADS to identify anxiety and depressive disorders in adolescents. *Research on Child and Adolescent Psychopathology* 49(9): 1241–1257.

Radford L, Corral S, Bradley C, et al. (2011) *Child Abuse and Neglect in the UK Today*.

Radford, L., Stanley, N. and Elwen, A. (2021) Safeguarding Children and Young People Exposed to Domestic Violence and Abuse. *Working with Domestic Violence and Abuse Across the Lifecourse: Understanding Good Practice*: 22.

Rose D, Trevillion K, Woodall A, et al. (2011) Barriers and facilitators of disclosures of domestic violence by mental health service users: Qualitative study. *British Journal of Psychiatry* 198(3): 189–194.

Roy J, Williamson E, Pitt K, et al. (2022) "It felt like there was always someone there for us": Supporting children affected by domestic violence and abuse who are identified by general practice. *Health & Social Care in the Community* 30(1): 165–174.

Ryan RM and Deci EL (2001) On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology* 52(1): 141–166.

Saini SM, Hoffmann CR, Pantelis C, et al. (2019) Systematic review and critical appraisal of child abuse measurement instruments. *Psychiatry Research* 272: 106–113.

Sarasjärvi KK, Elovainio M, Appelqvist-Schmidlechner K, et al. (2023) Exploring the structure and psychometric properties of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) in a representative adult population sample. *Psychiatry Research* 328: 115465.

Selwyn J and Wood M (2015) *Measuring Well-Being: A Literature Review*. University of Bristol. Available at: http://www.coramvoice.org.uk/professional-zone/news/launch-literature-reviews-support-bright-spots-project.

Shah N, Cader M, Andrews B, et al. (2021) Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): Performance in a clinical sample in relation to PHQ-9 and GAD-7. *Health and Quality of Life Outcomes* 19(1): 260.

Shi D, DiStefano C, McDaniel HL, et al. (2018) Examining chi-square test statistics under conditions of large model size and ordinal data. *Structural Equation Modeling: A Multidisciplinary Journal* 25(6): 924–945.

Smith ORF, Alves DE, Knapstad M, et al. (2017) Measuring mental well-being in Norway: Validation of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS). *BMC Psychiatry* 17(1): 182.

Solace Womens Aid. (2023). Have you heard about gaslighting? - Solace Womens Aid. [online] Available at: https://www.solacewomensaid.org/have-you-heard-about-gaslighting/.

Soneson E, Puntis S, Chapman N, et al. (2023) Happier during lockdown: A descriptive analysis of self-reported wellbeing in 17,000 UK school students during Covid-19 lockdown. *European Child & Adolescent Psychiatry* 32(6): 1131–1146.

Soraci P, Bevan N, Griffiths MD, et al. (2024) The Short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS): An Italian validation using confirmatory factor analysis and Rasch analysis. *BMC Psychology* 12(1): 680.

Springer KW, Sheridan J, Kuo D, et al. (2003) The long-term health outcomes of childhood abuse. *Journal of General Internal Medicine* 18(10): 864–870.

Stanley N, Barter CA, Bracewell K, et al. (2021) *Roadmap Evaluation Final Report*. Project Report. UCLan. Available at: https://clok.uclan.ac.uk/39447/2/Roadmap_Report_280921.pdf.

Stewart-Brown S, Platt S, Tennant A, et al. (2011) The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): A valid and reliable tool for measuring mental well-being in diverse populations and projects. *Journal of Epidemiology & Community Health* 65(Suppl 2): A38–A39.

Stewart-Brown S, Tennant A, Tennant R, et al. (2009) Internal construct validity of the Warwick-Edinburgh Mental Well-being Scale (WEMWBS): A Rasch analysis using data from the Scottish Health Education Population Survey. *Health and Quality of Life Outcomes* 7(1): 15.

Stulz V, Francis L, Naidu A, et al. (2024) Women escaping domestic violence to achieve safe housing: An integrative review. *BMC Women's Health* 24(1): 314.

Taggart F, Friede T, Weich S, et al. (2013) Cross cultural evaluation of the Warwick-Edinburgh mental well-being scale (WEMWBS): A mixed methods study. *Health and Quality of Life Outcomes* 11(1): 27.

Tennant R, Hiller L, Fishwick R, et al. (2007) The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): Development and UK validation. *Health and Quality of Life Outcomes* 5(1): 63.

Tourangeau R, Rips LJ, and Rasinski K (2000) *The Psychology of Survey Response*. 1st ed. Cambridge University Press. Available at: https://www.cambridge.org/core/product/identifier/9780511819322/type/book (accessed 28 March 2025).

Trevillion K, Corker E, Capron LE, et al. (2016) Improving mental health service responses to domestic violence and abuse. *International Review of Psychiatry* 28(5): 423–432.

Trousselard M, Steiler D, Dutheil F, et al. (2016) Validation of the Warwick-Edinburgh Mental Well-Being Scale (WEMWBS) in French psychiatric and general populations. *Psychiatry Research* 245: 282–290.

UNICEF (2006) *Behind Closed Doors: The Impact of Domestic Violence on Children*. Available at: https://books.google.co.uk/books?id=QNubtQEACAAJ.

University Of Bristol DOSM (2009) *ALSPAC; Children of the 90s Avon Longitudinal Study of Parents and Children, 1990–2003: Social Science Sampler Datasets*. UK Data Service. Available at: https://beta.ukdataservice.ac.uk/datacatalogue/doi/?id=6147#1 (accessed 14 April 2025).

Varni JW, Limbers CA, and Burwinkle TM (2007) Parent proxy-report of their children's health-related quality of life: An analysis of 13,878 parents' reliability and validity across age subgroups using the PedsQL™ 4.0 Generic Core Scales. *Health and Quality of Life Outcomes* 5(1): 2.

Visschers J, Jaspaert E, and Vervaeke G (2017) Social desirability in intimate partner violence and relationship satisfaction reports: An exploratory analysis. *Journal of Interpersonal Violence* 32(9): 1401–1420.

Vu NL, Jouriles EN, McDonald R, et al. (2016) Children's exposure to intimate partner violence: A meta-analysis of longitudinal associations with child adjustment problems. *Clinical Psychology Review* 46: 25–33.

Waite M and Atkinson C (2021) A review of mental health and wellbeing measures for universal secondary school screening. *Emotional and Behavioural Difficulties* 26(2): 206–222.

Walker-Descartes I, Mineo M, Condado LV, et al. (2021) Domestic violence and its effects on women, children, and families. *Pediatric Clinics* 68(2): 455–464.

Wang M, Beal DJ, Chan D, et al. (2017) Longitudinal research: A panel discussion on conceptual issues, research design, and statistical techniques. *Work, Aging and Retirement* 3(1): 1–24.

Weeks LE, Stilwell C, Rothfus M, et al. (2024) A review of intimate partner violence interventions relevant to women during the COVID-19 pandemic. *Violence Against Women* 30(3–4): 981–1021.

Williamson PR, Altman DG, Bagley H, et al. (2017) The COMET Handbook: Version 1.0. *Trials* 18(3): 280.

Williamson PR, Altman DG, Blazeby JM, et al. (2012) Developing core outcome sets for clinical trials: Issues to consider. *Trials* 13(1): 132.

Williamson PR, Barrington H, Blazeby JM, et al. (2022) Review finds core outcome set uptake in new studies and systematic reviews needs improvement. *Journal of Clinical Epidemiology* 150: 154–164.

Yadav P, Arias De La Torre J, Bakolis I, et al. (2025) Internal structure, reliability and cross-cultural validity of the Warwick-Edinburgh Mental Wellbeing Scale in three European populations. *BMJ Mental Health* 28(1): e301433.

Zayed K, Omara E, Al-Shamli A, et al. (2023) A validation study of the Arabic version of the Warwick-Edinburgh Mental Well-being scale among undergraduate students. *BMC Psychology* 11(1): 399.

APPENDICES

Appendix A. Think aloud – focus group/interview schedule: young person

The option of focus group and interview will be given to each young person, and the decision of which to facilitate will be made based on their preference. Young person will have received questions in advance.

Introduction

Introduce yourself and the aims of the interview (as per the information sheet). Review their consent form, discuss limits to confidentiality and give a brief outline of what will be covered in the interview/focus group (provide ground rules if focus group). Remind the young person that they can ask questions about things they don't understand, and they can ask for a break whenever they need one.

1. Completing the SWEWBS

Participants to complete the SWEWBS with no intervention from the researcher.

2. Think-aloud interview/focus group: item by item

Researcher to introduce the think-aloud interview, advising that they will ask a series of questions for each item on the measure.

- **a.** What does this question mean to you?
- **b.** Was the question easy or difficult for you to answer?
    - Prompt: In what way? Were there any words you didn't understand?
- **c.** Could you give me an example of that kind of behaviour?
- **d.** How did it feel to be asked that question?
    - Prompt: Did it make you think of anything difficult?
- **e.** How easy or difficult was it to select an answer from the options provided?
    - Prompt: Why?
- **f.** How sure do you feel about your answer?

Participant are asked to rate each item against at scale of how upset they were by each question: 0 (not upset) to 10 (very upset).

3. General questions about the measure

Researcher will then ask more general questions about the measure as a whole.

- **a.** What was your overall opinion of the measure?
- **b.** How would you feel coming into a service, *[name associated DA service],* and being asked to complete this measure? (On first meeting and completing again)
- **c.** Are there things that would be important when introducing the measure?
    - Prompt: Change to wording?

d. Is there anything else that would have helped you to complete the measure?

4. Close and debrief

a. Is there anything else that you think is important for us to know?

- How are you feeling? Has it been ok?

Appendix B. Key difficulties** and findings for each item of the SWEMWBS

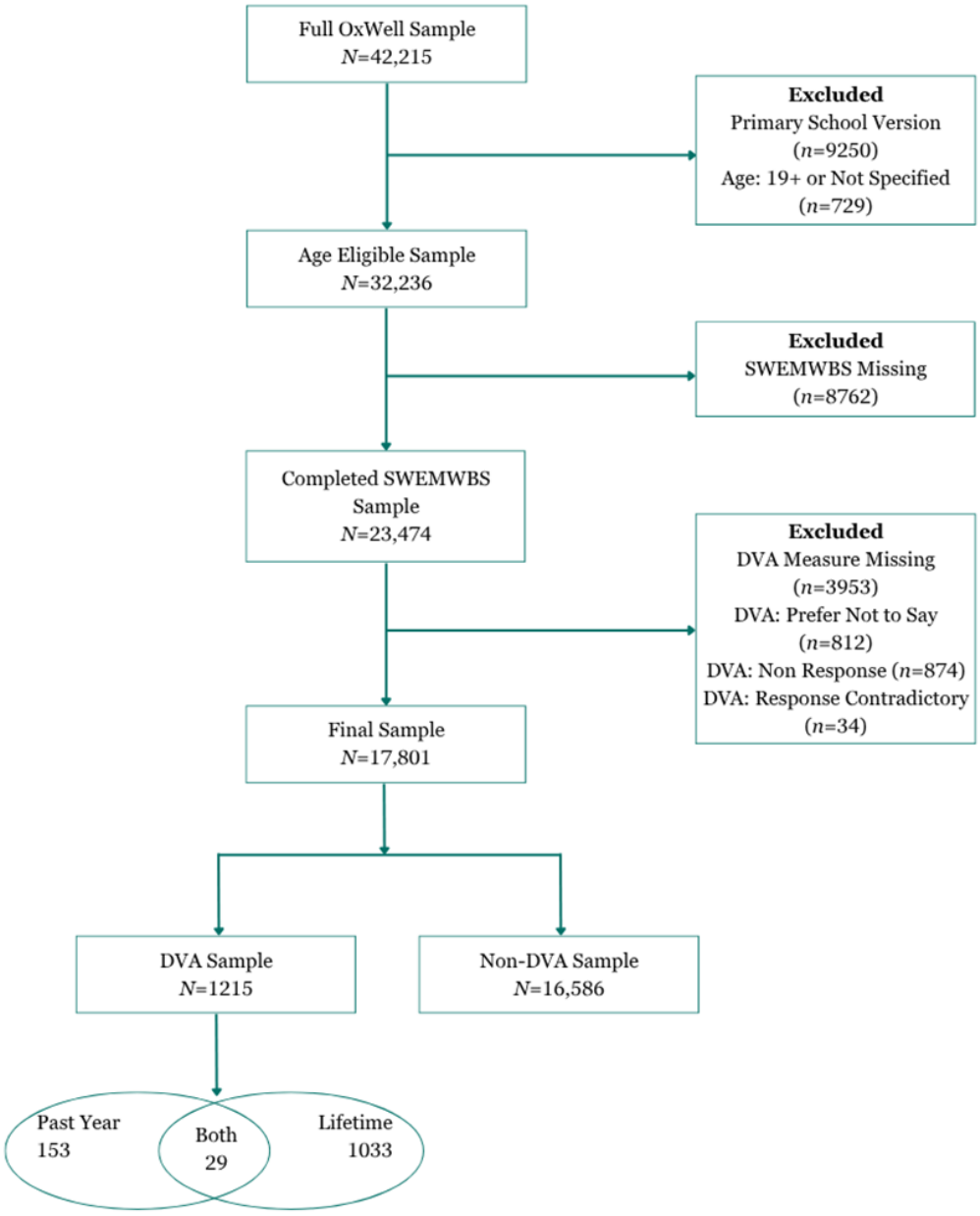
Item	Comprehension difficulty	Response mapping difficulty	Contextual information	Recommendation	Scale
I've been feeling optimistic about the future	Younger participants were not comfortable with the meaning of "optimistic". Ambiguity over the intended time scale of "future": is this the immediate future or adulthood?	Participants requested "I don't know" option due to not understanding the word "optimistic". Participants who did not consider themselves to have a future did not feel they were able to respond using the scale. This was particularly pertinent in a domestic abuse context	Participants felt that their answers were dependant on contextual factors preceding the completion of the measure. If a person was new to a domestic abuse service, this item might cause them to feel sad.	Would like a definition for "optimistic" as well as a timeframe for "future".	Some participants requested a larger Likert scale. The accompanying prompts for the Likert scale were too disparate.
I've been feeling useful	Uncertainty over how to interpret "useful" – to whom and in what capacity?	Unlikely that you would respond "all the time" to this item.	In a domestic abuse context, a person may have been called useless or exploited for chores. This is interpreted as the perception of other people and could be weaponised.		"I don't know" was requested by those who did not know how to interpret "useful".

Item	Comprehension difficulty	Response mapping difficulty	Contextual information	Recommendation	Scale
I've been feeling relaxed			Context-specific; may have other factors making them unrelaxed, not to do with service.	This item is not appropriate to be asked at beginning of service delivery.	
I've been dealing with problems well	Uncertainty on severity of the problem this is referring to. "Well" feels subjective, and can be interpreted as judgemental.		In a DVA context external factors may cause the problem – therefore, is it appropriate to suggest the responsibility of resolution is the respondent's? Also whether there is a resolution.		Preference for a middle "neutral" option.
I've been thinking clearly	Uncertainty on what this means and which situations it would apply to.	Request for an "I don't know" option.	This might result in defensive reporting in a DVA context due to courts or statutory agencies using this against you. Similarly this might have been weaponised by the person that harms.		Would prefer to agree/disagree for this item.
I've been feeling close to other people			If a person was not feeling close to people, this might make them feel upset		

Item	Comprehension difficulty	Response mapping difficulty	Contextual information	Recommendation	Scale
I've been able to make up my own mind about things	Uncertainty of level of decision making.		Decisiveness weaponised by person that harms.		

**Table does not display "recall" or "judgement" difficulties as these were not applicable to any transcripts.

Appendix C. Flow diagram demonstrating the process of data exclusion



Note. *Within Past Year* and *Within Lifetime* response options were delivered simultaneously. Therefore, although one would assume past year also means lifetime, the responses are presented in line with the raw survey data. Within analyses in this report the DVA group includes participants that selected either lifetime or past year as well as those that selected both.

Appendix D. The Short Warwick–Edinburgh Mental Wellbeing Scale (SWEMWBS)

Below are some statements about feelings and thoughts.

Please select the answer that best describes your experience of each over the last 2 weeks.

	None of the Time / Rarely		Some of the Time / Often		All of the Time
I've been feeling optimistic about the future	**1**	**2**	**3**	**4**	**5**
I've been feeling useful	1	2	3	4	5
I've been feeling relaxed	1	2	3	4	5
I've been dealing with problems well	1	2	3	4	5
I've been thinking clearly	1	2	3	4	5
I've been feeling close to other people	1	2	3	4	5
I've been able to make up my own mind about things	1	2	3	4	5

Appendix E. The 'non-responding' sample in OxWell analyses

Of the 8,762 that did not respond to the SWEMWBS, 6,300 had withdrawn from the questionnaire before this point. There were 1,832 non-responders that did continue the questionnaire and were presented with the DVA measure. Of these, 23 responded in a contradictory way (selecting they both had and had never experienced DVA) and 533 did not respond to the measure. Of those that responded to the DVA measure, but not the SWEMWBS, 1,104 had never experienced DVA (87%), 99 had experienced DVA (7.8%), and 73 preferred not to say (5.7%). The prevalence rate of DVA is higher in those that did not complete the SWEMWBS than in the full sample who did complete the SWEMWBS (6.8%).

Exploring SWEMWBS missingness at the item level did not indicate that participants were particularly deterred from any one item. Of the 25,936 participants that were presented with the SWEMWBS, item-level missingness ranged from 1,339 (Item 3) and 1,554 (Item 6). When exploring item missingness by domestic abuse experience similar patterns emerged, the same item (Item 6) had the most missing responses across all three DVA categories (see Table E1). Those with DVA experience had the smallest percentage of missing responses across all items compared to those without DVA experience and those who did not disclose DVA experience.

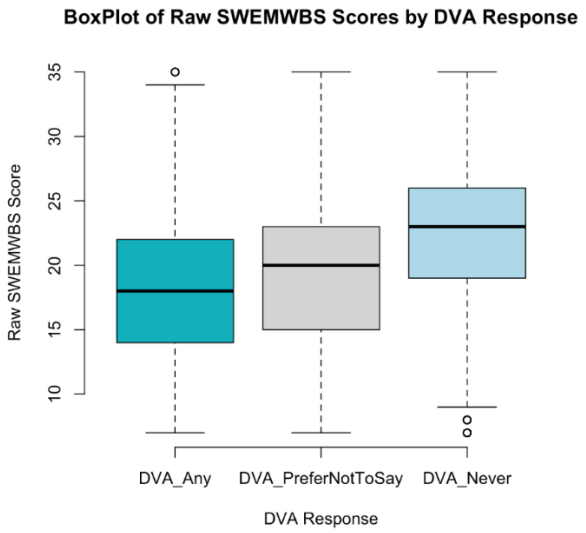
Table E1. Item-level missingness on the SWEMWBS by domestic abuse experience

	Item-level missingness		
	DVA[a]	Non-DVA[b]	DVA missing[c]
Item 1 – Optimistic about future	38 (2.9%)	966 (4.8%)	370 (8.1%)
Item 2 – Feeling useful	40 (3%)	939 (4.7%)	372 (8.1%)
Item 3 – Feeling relaxed	39 (3%)	922 (4.6%)	378 (8.2%)
Item 4 – Dealing with problems well	36 (2.7%)	972 (4.9%)	382 (8.3%)
Item 5 – Thinking clearly	41 (3.1%)	1,012 (5.1%)	412 (9%)
Item 6 – Feeling close to others	46 (3.5%)	1,076 (5.4%)	432 (10.61%)
Item 7 – Make up own mind	34 (2.6%)	948 (4.7%)	390 (8.5%)

N=25,936 [a] *n*=1,215, [b] *n*=16,586, [c] *n*=4,583.

Upon assessing the SWEMWBS scores of those who chose “prefer not to say” as a response option to the DVA measure (*M*=19.37, *SD*=6.19), they are more synonymous with those with DVA experience (*M*=18.31, *SD*=6.24) than those without (*M*=22.4, *SD*=5.82). A one-way ANOVA followed by TukeyHSD post-hoc analyses identified that mean SWEMWBS scores were significantly different between all groups: those who preferred not to say, those who have never experienced DVA, and those that have, F(2,18408)= 340.8, *p*<.001.

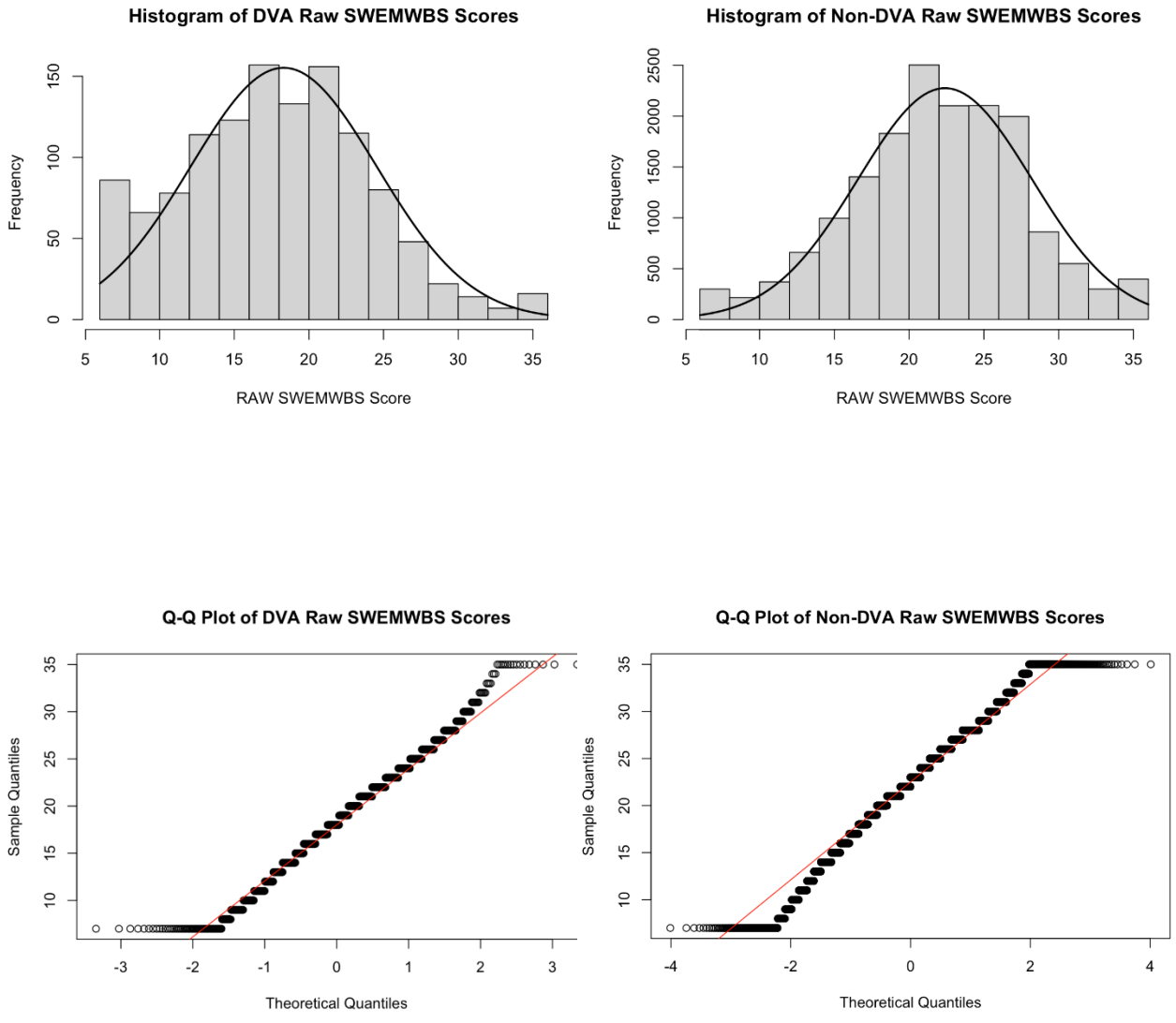
Figure E1. Box plot of raw SWEMWBS scores for those that have no DVA experience, those with DVA experience, and those that chose not to disclose



Appendix F. Distribution of SWEMWBS scores (OxWell)

The distribution of SWEMWBS scores across the DVA and non-DVA sample are shown in the figures below (see figure 2). The raw SWEMWBS scores in the DVA sample are slightly positively skewed (0.16) and have a slightly flatter distribution (-.29) than a normal distribution. The raw SWEMWBS scores in the non-DVA sample are slightly negatively skewed (-.22) and kurtosis presents an approximately normal distribution (-.05).

Figure F1. Histograms and Q-Q plots of the raw SWEMWBS scores for the DVA and non-DVA samples



Appendix G. Reliability analysis of the SWEMWBS (OxWell)

Table G1. Cronbach's alpha of the SWEMWBS scale in each sample

	α	95% confidence intervals
Non-DVA sample[a]	.86	.86–.86
DVA sample[b]	.87	.85–.88

[a] *N*=16,586, [b] *N*=1,215.

Table G2. Item statistics for Cronbach's alpha analysis of the SWEMWBS scales

	Corrected item-total correlation		Cronbach's α if item removed	
	DVA[a]	Non-DVA[b]	DVA[a]	Non-DVA[b]
Item 1	.52	.46	.86	.86
Item 2	.64	.67	.85	.84
Item 3	.63	.65	.85	.84
Item 4	.70	.71	.84	.83
Item 5	.74	.73	.83	.83
Item 6	.60	.57	.85	.85
Item 7	.65	.61	.85	.84

[a] *N*=1,215, [b] *N*= 16,586. The corrected item-total correlation assesses how well an item correlates with the total scale score, excluding that item.

Table G3. Statistics for split-half analysis of the SWEMWBS scale in each sample

	DVA[a]	Non-DVA[b]
Average split-half reliability	.85	.84
Minimum split-half reliability (β)	.79	.78
Maximum split-half reliability (λ4)	.87	.88
Guttman's split-half coefficient (λ6)	.86	.85
Average inter-item correlation	.48	.47

[a] *N*=1,215, [b] *N*=16,586.

Appendix H. Confirmatory factor analysis of the SWEMWBS (OxWell)

Table H1. Fit indices of CFA of the one-factor models of the SWEMWBS in each sample

Model	χ^2 (df)	p-value	CFI	TLI	RMSEA (90% CI)	SRMR
Non-DVA sample[a]	3,356.97 (14)	<.001***	.96	.94	.10 (.09, .10)	.04
DVA sample[b]	183.68 (14)	<.001***	.97	.96	.09 (.08, .10)	.03

[a] N=16,586, [b] N=1,215, *** p<.001

Table H2. Factor loadings of SWEMWBS items onto a single factor for each sample

Item	Factor loading (standardised)	
	DVA[a]	Non-DVA[b]
Item 1	.60	.54
Item 2	.74	.75
Item 3	.74	.75
Item 4	.82	.83
Item 5	.86	.85
Item 6	.68	.65
Item 7	.74	.70

[a] N=1,215, [b] N= 16,586.

Appendix I. Measurement invariance of the SWEMWBS across children and young people who have experienced DVA and those who have not (OxWell)

Table I1. Fit indices of models testing for measurement invariance between DVA and non-DVA samples

Model	χ^2 (df)	*p*	CFI	TLI	RMSEA (90% CI)	SRMR
CFA (one-factor model)	3,496.36 (14)	<.001	.98	.96	.12 (.12, .12)	.03
Configural invariance	3,510.87 (28)	<.001	.97	.96	.12 (.12, .12)	.04
Metric invariance	2,106.56 (34)	<.001	.98	.98	.08 (.08, .09)	.04
Scalar invariance	2,249.21 (54)	<.001	.98	.99	.07 (.07, .07)	.04

Note. Fit statistics reported are scaled values because robust fit indices were not available for the scalar invariance model due to estimation constraints under the WLSMV estimator.

Table I2. Differences in fit indices between models testing for measurement invariance between DVA and non-DVA samples

Comparison	$\Delta\chi^2$	Δdf	*p*	ΔCFI	ΔTLI	ΔRMSEA	ΔSRMR
Configural vs. metric	36.01	6	<.001	.01	.020	-.035	.001
Metric vs. scalar	58.63	20	<.001	-.001	.006	-.015	-.001

Note. Changes <-.01 to the CFI and TLI, <.015 to RMSEA, and<.01 to the SRMR, indicate invariance.

Appendix J. Distribution of SWEMWBS scores (children and young people in domestic abuse service)

The distribution of initial and final SWEMWBS scores was assessed using histograms, Q-Q plots, and normality tests. The initial SWEMWBS scores showed a slightly positively skewed distribution. The Shapiro-Wilk test indicated deviations from normality (W=0.982, p=.004). Skewness was 0.37, and kurtosis was 0.35, suggesting a distribution close to normal but with a slight tail towards higher scores. The final SWEMWBS scores indicated a more symmetrical distribution compared with initial scores. Skewness was -0.07, and kurtosis was 0.57, suggesting an approximately normal distribution with minor flattening. The Shapiro-Wilk test also showed deviations from normality (W = 0.980, p = .015).

Figure J1. Histograms of the raw SWEMWBS scores at initial and final timepoints

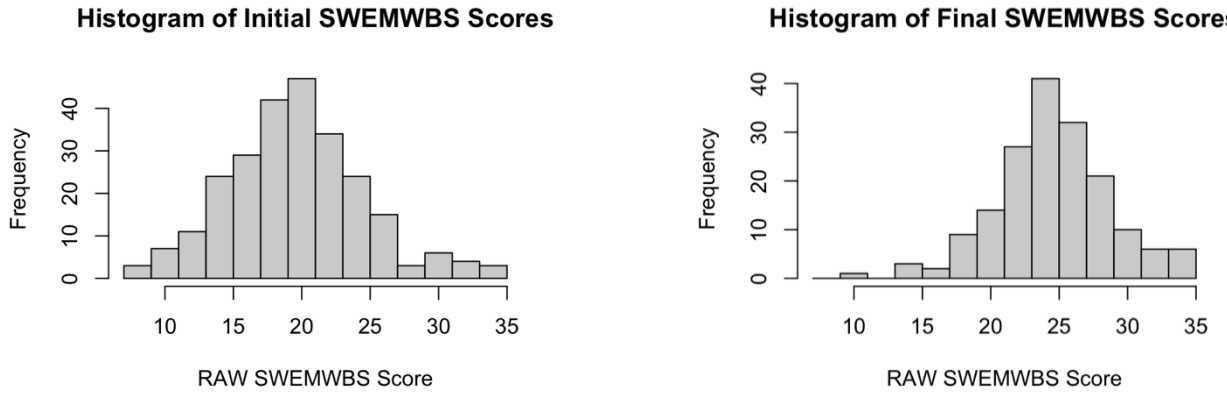
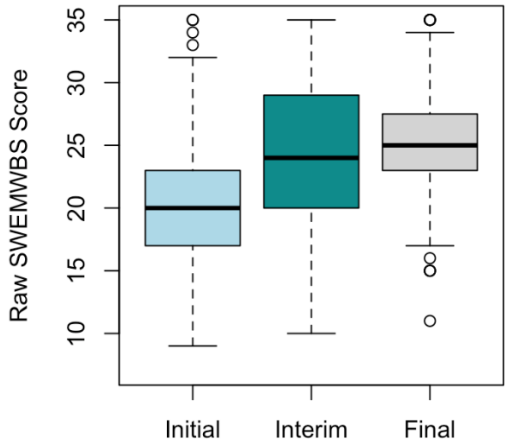


Figure J2. Boxplot of the raw SWEMWBS scores at initial, interim, and final timepoints



Appendix K. Item-level mean differences in SWEMWBS scores (children and young people in domestic abuse service)

The SWEMWBS scores were explored by item. The largest improvement between initial and final timepoints was made in item 4 ("I've been dealing with problems well") while item 7 ("I've been able to make up my mind about things") had the smallest mean change. These correspond to the smallest and largest initial scores, with the lowest-scoring item seeing most improvement and the highest scoring item seeing least improvement.

Table K1. Mean SWEMWBS scores across timepoints at the item level

	Mean raw score – M(SD)				Mean score change – M(SD)		
	Initial N=252	Interim N=69	Final N=172		Initial to interim N=64	Interim to final N=47	Initial to final N=159
Item 1	2.86(.92)	3.41(9.3)	3.58(.77)		.31(.85)*	.47(.86)*	.74(.89)*
Item 2	2.93(.90)	3.38(.84)	3.53(.73)		.55(.97)*	.19(.74)	.62(.83)*
Item 3	2.61(.94)	3.46(1.02)	3.34(.85)		.72(1.12)*	-.09(1.00)	.75(.98)*
Item 4	2.45(1.00)	3.23(1.11)	3.35(.83)		.67(1.18)*	.47(1.02)*	.96(.94)*
Item 5	2.78(1.00)	3.10(1.19)	3.60(.88)		.24(1.08)	.68(.78)*	.86(1.02)*
Item 6	3.17(1.10)	3.74(1.02)	3.83(.95)		.47(1.13)*	.19(.92)	.68(1.05)*
Item 7	3.29(.99)	3.55(1.08)	3.84(.87)		.31(.89)*	.26(.79)*	.59(.96)*

* Change in score is significant $p<.05$.

Interestingly, although item 5 ("I've been thinking clearly"*)* had the highest improvement between interim and final reports, it had the lowest mean change between initial and interim timepoints. This may be due to defensive reporting for that item or an indicator that "thinking clearly" has an initially slower progress rate. On the other hand, item 3 ("I've been feeling relaxed") presented the highest mean change between initial and interim timepoints but the lowest change between interim and final timepoints. It was the only item to demonstrate a significant difference in the

average score change between *initial to interim* and *interim to final* timepoints, $t(44)=3.82$, $p<.001$ – i.e. substantial improvements in feeling relaxed are made by interim timepoints, but this then plateaued.

All the items demonstrated significant improvement in scores between the initial timepoint at the start of the project and the final timepoint at the end of project ($p<.001$).

Appendix L. Reliability analysis of the SWEMWBS (children and young people in domestic abuse service)

Table L1. Item statistics for Cronbach's alpha analysis of the SWEMWBS scales

	Corrected Item-total correlation		Cronbach's α if item removed	
	Initial[a]	Final[b]	Initial[a]	Final[b]
Item 1	.75	.63	.81	.82
Item 2	.64	.70	.82	.81
Item 3	.64	.63	.82	.82
Item 4	.61	.69	.83	.81
Item 5	.75	.71	.81	.80
Item 6	.66	.58	.82	.83
Item 7	.56	.59	.84	.82

[a] N=252, [b] N=172. The corrected item-total correlation assesses how well an item correlates with the total scale score, excluding that item.

Table L2. Statistics for split-half analysis of the SWEMWBS scale across timepoints

	Initial[a]	Final[b]
Average split-half reliability	.83	.82
Minimum split-half reliability (β)	.76	.80
Maximum split-half reliability (λ4)	.88	.86
Guttman's split-half coefficient (λ6)	.84	.82
Average inter-item correlation	.44	.43

[a] N=252, [b] N=172.

Appendix M. Confirmatory factor analysis of the SWEMWBS (children and young people in domestic abuse service)

Table M1. Fit indices of CFA of the one-factor models of the SWEMWBS in each sample

Model	χ^2 (df)	*p*-value	CFI	TLI	RMSEA (90% CI)	SRMR
Initial[a]	52.48(14)	<.001	.94	.90	.10 (.08, .14)	.052
Final[b]	11.40(14)	.655	1.00	1.00	.00 (.00, .06)	.027

[a] N=252, [b] N=172.

Table M2. Factor loadings of SWEMWBS items onto a single factor for each sample

	Factor loading (standardised)	
Item	**Initial timepoint**[a]	**Final timepoint**[b]
Item 1	.75	.64
Item 2	.63	.70
Item 3	.66	.64
Item 4	.63	.70
Item 5	.75	.71
Item 6	.67	.59
Item 7	.56	.60

[a] N=252, [b] N=172.

Appendix N. Measurement invariance of the SWEMWBS across time: before and after service provision

Table N1. Fit indices of models testing for measurement invariance between initial and final timepoints

Model	χ^2 (df)	*p*	CFI	TLI	RMSEA (90% CI)	SRMR
Configural invariance	119.0 (70)	<.001	.94	.92	.07 (.05, .09)	.051
Metric invariance	126.2 (76)	<.001	.94	.93	.06 (.04, .08)	.061
Scalar invariance	141.2 (82)	<.001	.93	.92	.07 (.05, .09)	.074

Note. Thresholds of acceptability include >.90 for CFI and TLI, and <.08 for RMSEA and SRMR.

Table N2. Differences in fit indices between models testing for measurement invariance between initial and final timepoints

Comparison	Δ χ^2	Δ df	*p*	Δ CFI	Δ TLI	Δ RMSEA	Δ SRMR
Configural vs. metric	7.21	6	.301	-.001	.004	-.002	.010
Metric vs. scalar	14.95	6	.021	-.011	-.007	.003	.013

Note. Changes <-.01 to the CFI and TLI, <.015 to RMSEA, and<.01 to the SRMR, indicate invariance.

Appendix O. The Warwick–Edinburgh Mental Wellbeing Scale

Below are some statements about feelings and thoughts.

Please select the answer that best describes your experience of each over the last 2 weeks.

	None of the Time	Rarely	Some of the Time	Often	All of the Time
I've been feeling optimistic about the future	**1**	**2**	**3**	**4**	**5**
I've been feeling useful	1	2	3	4	5
I've been feeling relaxed	1	2	3	4	5
I've been feeling interested in other people	1	2	3	4	5
I've had energy to spare	1	2	3	4	5
I've been dealing with problems well	1	2	3	4	5
I've been thinking clearly	1	2	3	4	5
I've been feeling good about myself	1	2	3	4	5
I've been feeling close to other people	1	2	3	4	5
I've been feeling confident	1	2	3	4	5
I've been able to make up my own mind about things	1	2	3	4	5
I've been feeling loved	1	2	3	4	5
I've been interested in new things	1	2	3	4	5
I've been feeling cheerful	1	2	3	4	5

Appendix P. The DVA measures used from the APMS

Financial
Has a partner or ex-partner ever prevented you from having your fair share of the household money? (By partner we mean any boyfriend or girlfriend, as well as a husband, wife, or civil partner)
1 Yes
2 No
3 Never been in a relationship (*SKIP rest of DVA questions*)

Belittle
Has a partner or ex-partner ever repeatedly belittled you to the extent that you felt worthless?

1 Yes
2 No
9 Don't understand/does not apply

Harass
Has a partner or ex-partner ever sent you more than one unwanted letter, email, text message or card that was either obscene or threatening and which caused you fear, alarm or distress?
1 Yes
2 No

Pushed
Has a partner or ex-partner ever pushed you, held or pinned you down or slapped you?

1 Yes
2 No

Kicked
Has a current or ex-partner ever kicked you, bit you, or hit you with a fist or something else, or threw something at you that hurt you?
1 Yes
2 No

For each question a participant indicated “Yes” they were also asked:
Has this happened within the past 12 months?

1 Yes
2 No

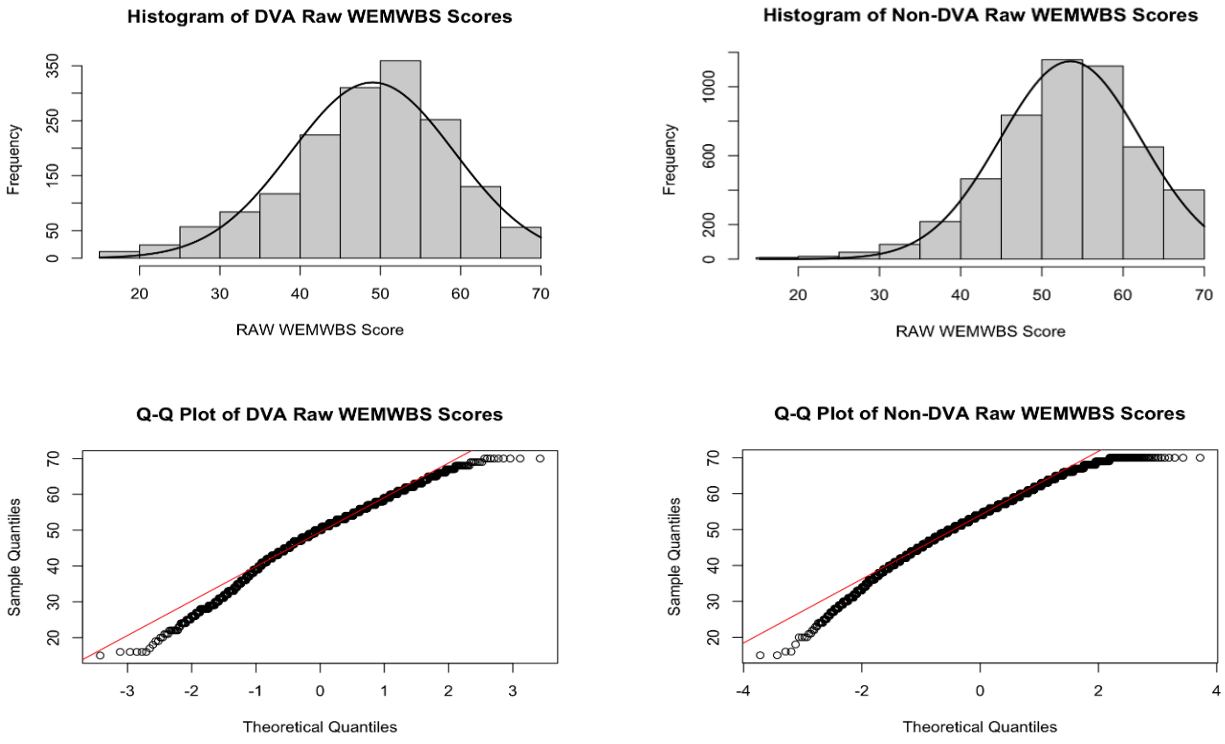
Where a participant indicated that “Yes” it has happened in the past 12 months they were then asked:
How often has this happened in the past year?

1 At least once a week
2 At least once a fortnight
3 At least once a month
4 Less often than once a month

Appendix Q. Distribution of WEMWBS scores (adults)

The distributions of WEMWBS scores across the DVA and non-DVA samples are shown in the figures below. The WEMWBS scores in the DVA sample are slightly negatively skewed (-0.57) and have an approximately normal distribution (kurtosis excess=0.23). The WEMWBS scores in the non-DVA sample are also slightly negatively skewed (-0.55) and kurtosis presents an almost normal distribution (kurtosis excess=0.57). The Shapiro-Wilk tests of normality are significant for both DVA (W=.98, p<.001) and non-DVA samples (W=.98, p<.001). Accounting for test statistics (W) close to 1, the visual distributions in figure Q1, and minimal skewness and kurtosis, it can be concluded that the distribution of WEMWBS scores is approximately normal with no concerning deviation from normality.

Figure Q1. Histograms and Q-Q plots of the raw WEMWBS scores for the DVA and non-DVA samples



Appendix R. Reliability analysis of the WEMWBS (adults)

Table R1. Cronbach's alpha of the WEMWBS scale in each sample

	n	α	95% confidence intervals
Non-DVA sample[a]	4,996	.89	.89–.90
DVA sample[b]	1,625	.92	.91–.93

Table R2. Item statistics for Cronbach's alpha analysis of the WEMWBS scales

	Corrected item-total correlation		Cronbach's α if item removed	
	Non-DVA[a]	DVA[b]	Non-DVA[a]	DVA[b]
Item 1	.45	.53	.89	.92
Item 2	.58	.63.	.88	.91
Item 3	.55	.60	.89	.92
Item 4	.50	.58	.89	.92
Item 5	.53	.57	.89	.92
Item 6	.59	.70	.88	.91
Item 7	.61	.67	.88	.91
Item 8	.72	.77	.88	.91
Item 9	.59	.64	.88	.91
Item 10	.72	.76	.88	.91
Item 11	.49	.58	.89	.92

Item 12	.51	.55	.89	.92
Item 13	.57	.68	.89	.91
Item 14	.71	.75	.88	.91

[a] N=4,996, [b] N=1,625. The corrected item-total correlation assesses how well an item correlates with the total scale score, excluding that item.

Table R3. Statistics for split-half analysis of the WEMWBS scale in each sample

	Non-DVA sample[a]	DVA sample[b]
Average split-half reliability	.89	.92
Minimum split-half reliability (β)	.81	.88
Maximum split-half reliability (λ4)	.93	.94
Guttman's split-half coefficient (λ6)	.90	.92
Average inter-item correlation	.38	.44

[a] N=4,996, [b] N=1,625.

Appendix S. Confirmatory factor analysis of the WEMWBS (adults)

Table S1. Fit indices of CFA of the one-factor models of the WEMWBS in each sample

Model	χ^2 (df)	p-value	CFI	TLI	RMSEA (90% CI)	SRMR
Model 1						
Full sample	7,209.06 (77)	<.001	.87	.84	.114(.11,.12)	.058
DVA sample[a]	1,563.25 (77)	<.001	.90	.88	.105(.10,.11)	.048
Non-DVA sample[b]	5,499.78 (77)	<.001	.85	.83	.117(.11,.12)	.061
Model 2						
Full sample	4,184.85 (150)	<.001	.92	.90	.091(.09,.09)	.046
DVA sample[a]	912.89 (75)	<.001	.94	.92	.084(.08,.09)	.038
Non-DVA sample[b]	3,331.29 (75)	<.001	.91	.89	.095(.09,.10)	.051

[a] N=1,625, [b] N=4,996.

Table S2. Factor loadings of WEMWBS items onto a single factor for each sample

Item	Factor loading (standardised)	
	Non-DVA[a]	DVA[b]
Item 1	.51	.58
Item 2	.65	.68
Item 3	.65	.67
Item 4	.56	.64
Item 5	.60	.64
Item 6	.67	.75
Item 7	.71	.74
Item 8	.83	.86
Item 9	.65	.68
Item 10	.83	.85
Item 11	.63	.67
Item 12	.57	.60
Item 13	.63	.73
Item 14	.80	.83

[a] N=4,996, [b] N= 1,625.

Appendix T. Measurement invariance of the WEMWBS across adults who have and have not experienced DVA

Table T1. Fit indices of models testing for measurement invariance between DVA and non-DVA samples

Model	χ^2 (df)	p	CFI	TLI	RMSEA (90% CI)	SRMR
Configural invariance	4,338.47(150)	<.001	.95	.95	.092 (.09, .09)	.048
Metric invariance	2,565.88 (163)	<.001	.97	.97	.067 (.06, .07)	.049
Scalar invariance	3,415.01 (204)	<.001	.97	.97	.069 (.07, .07)	.048

Note. Fit statistics reported are scaled values because robust fit indices were not available for the scalar invariance model due to estimation constraints under the WLSMV estimator.

Table T2. Differences in fit indices between models testing for measurement invariance between DVA and non-DVA samples

	Δ	Δ		Δ	Δ	Δ	Δ
Configural vs. metric	56.51	13	<.001	.020	.026	-.0025	.002
Metric vs. scalar	121.25	41	<.001	-.009	-.002	.002	-.001

Note. Changes <-.01 to the CFI and TLI, <.015 to RMSEA, and<.01 to the SRMR, indicate invariance.

Appendix U. Accessibility text

Figure 1. Schematic of full study

A flowchart illustrating the project's two work packages (WP1 and WP2). The first overarching box lists the five DVA-COS outcomes:

4. Child Emotional Health and Wellbeing
5. Caregiver Emotional Health and Wellbeing
6. Feelings of Safety
7. Family Relationships
8. Freedom to go about Daily Life

The left branch describes **WP1** which addresses outcomes 3–5 and includes the following stages:

- **A1:** Defining concepts
- **A2:** Identifying candidate measures
- **B:** Appraisal of studies and properties of outcome measure indicators (OMIs)
- **C:** Stakeholder assessment of feasibility and acceptability
- **D:** Consensus process

The right branch describes **WP2** which focuses on validating the SWEMWBS/WEMWBS tools for mental wellbeing through four studies:

- **Study A:** Think Aloud – acceptability of SWEMWBS
- **Study B:** Validity of SWEMWBS in DVA children and young people using OxWell
- **Study C:** Validity of SWEMWBS in DVA service users over time
- **Study D:** Validity of WEMWBS in DVA adults using APMS

Click here to return to main report.

Figure 2. Overview of methods

Figure 2 summarises WP2's four studies which validate the SWEMWBS and WEMWBS for mental wellbeing in domestic abuse contexts.

- **A:** Think Aloud interviews to gather feedback from young people with domestic abuse experience on use of the SWEMWBS.
- **B:** OxWell study of 17,801 children and young people (1,215 with experience of domestic abuse) assessing SWEMWBS reliability and validity.
- **C:** Analysis of 268 children in domestic abuse services, assessing SWEMWBS reliability and validity over time.
- **D:** Secondary data analysis on dataset of 6,621 adults (1,625 with experience of domestic abuse) assessing WEMWBS psychometrics.

Click here to return to main report.

Figure 3. Scree plots of number of factors measured by the SWEMWBS scale for the DVA sample and the non-DVA sample

Figure 3 shows two side-by-side scree plots illustrating the number of factors measured by the SWEMWBS scale.

- **Left plot:** "Scree Plot for DVA Sample"
- **Right plot:** "Scree Plot for Non-DVA Sample"

Both plots have:

- **X-axis:** Factor or component number (1 to 7)
- **Y-axis:** Eigenvalues (0 to 4)
- Two lines: a **solid line** for Principal Component (PC) and a **dashed line** for Factor Analysis (FA)

In both plots, the first factor/component has the highest eigen value, followed by a sharp decline and then a plateau, indicating a strong first factor.

Click here to return to main report.

Figure 6. Scree plots of number of factors measured by the SWEMWBS scale at initial and final timepoints

Figure 4 displays two scree plots comparing the number of factors measured by the SWEMWBS scale at two timepoints: initial and final.

- **Left plot:** "Scree Plot for Initial Timepoint"
- **Right plot:** "Scree Plot for Final Timepoint"

Both plots have:

- **X-axis:** Factor or component number
- **Y-axis:** Eigenvalues of factors and components
- A visible drop in eigen values after the first factor/component, followed by a levelling off, suggesting a dominant first factor at both timepoints.

Click here to return to main report.

Figure 7. Scree plots of number of factors measured by the WEMWBS scale for the DVA sample and the non-DVA sample

Figure 5 presents two scree plots comparing the number of factors measured by the WEMWBS scale for DVA and non-DVA samples.

- **Left plot:** "Scree Plot for DVA Sample"
- **Right plot:** "Scree Plot for Non-DVA Sample"

Both plots feature:

- **X-axis:** Factor or component number (1 to 14)
- **Y-axis:** Eigenvalues (0 to 7 on the left plot and 0 to 6 on the right plot)
- Two lines:
    - **Solid line** representing Principal Component (PC)
    - **Dotted line** representing Factor Analysis (FA)

In both plots, eigen values drop sharply after the first factor and then plateaus, indicating a dominant initial factor.

Click here to return to main report.