

Bangor University

PROFESSIONAL DOCTORATES

Child and adolescent mental health Intervention and appropriate Measurement

Connolly, John

Award date:
2017

Awarding institution:
Bangor University

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

**Child and adolescent mental health: Intervention and appropriate
Measurement.**

John Francis Connolly

North Wales Clinical Psychology Programme
Bangor University

June 2017

Thesis submitted in partial fulfilment of the requirements for the degree of
Doctorate in Clinical Psychology

Acknowledgements

This thesis would not have been possible without the love and support of family and friends. It was you who made me who I am today. Thank you to my parents, who equipped me with the personal resources to take on challenges and chase dreams. I could not have asked for a more committed and loving upbringing. Thank you from the bottom of my heart.

Thank you, Hannah, for all your support and flexibility throughout the process.

A massive thank you to my research supervisor, Dr Chris Saville, without whom this project would not have been possible and whose enthusiasm and reassurance was invaluable at all stages of the project. His accessibility, combined with the wisdom to provide both enough support, and independence, was greatly appreciated.

Thank you also to Dr Mike Jackson for advice and feedback throughout the process.

Declarations

This work has not been previously accepted in substance for any degree and is not being concurrently submitted in candidature for any degree.

Signed

Date

Statement 1

This thesis is the result of my own investigations, except where otherwise stated. Other sources are acknowledged by footnotes giving explicit references. A list of references is appended.

Signed

Date

Statement 2

I agree to deposit an electronic copy of my thesis (the Work) in the Bangor University (BU) Institutional Digital Repository, the British Library ETHOS system, and /or in any other repository authorized for use by Bangor University and where necessary have gained the required permissions for the use of third party material.

Signed

Date

Word Count Statement

Title Page: 37

Thesis Abstract: 194

Chapter 1: Literature Review

Positive psychology interventions (PPI's) to improve the mental health of children and adolescents: a systematic review

Word count without references and Appendices: 5, 079

Word count with references: 7, 069

Tables and Figures: 3, 441

Chapter 2: Empirical Study

The measurement of child emotional health and norm development in a representative national sample.

Word count without references: 4, 913

Word count with references: 7, 681

Tables and Figures: 768

Paper 3: Contributions to Theory and Clinical Practice

Word count without references: 4, 986

Word count with references: 6, 269

Appendices (excluding Ethics Submission appendix)

Word Count: 2, 035

Total without References: 14, 978

Total with References: 21, 019

Total of Tables, Figures and Appendices: 6, 244

Overall Thesis Word Count: 32, 149

Table of Contents

Thesis Abstract	6
Chapter 1. Literature Review.....	7
Author guidelines for the Review of Education journal.....	8
Abstract.....	18
Introduction.....	19
Method.....	23
Results.....	27
Discussion.....	39
References.....	45
Chapter 2. Empirical Paper.....	54
Author guidelines for the Journal of Clinical Child and Adolescent Psychology.....	55
Abstract.....	62
Introduction.....	63
Method.....	69
Results.....	72
Discussion.....	83
References.....	88
Chapter 3. Contributions to theory and clinical practice.....	97
Overview.....	9
Implications for theory development and future research.....	97
Implications for mental health measurement and the SDQ.....	108
References	115
Appendices.....	121

Thesis Abstract

This thesis explores issues relating to intervention in, and measurement of, child mental health and well-being. The first chapter was a systematic review of Positive psychology interventions used to improve mental health and well-being outcomes. This identified 12 randomised controlled trials (n= 1668) in children and adolescents identifying a growing but generally weak evidence base. The array of outcomes and confounding variables assessed were identified alongside implementation issues. The second chapter explored psychometric properties of a ubiquitous tool for measuring these same outcomes, namely the Strengths and Difficulties Questionnaire. Factor structure, reliability, and measurement invariance were investigated and normative data developed. A 5-factor model was found to best fit the data in a large nationally representative sample. Omega (but not alpha) reliability coefficients revealed acceptable reliability and measurement invariance (configural, metric, scalar) was found for gender, parent status, and socio-economic status. Findings suggested that the method of analysis employed greatly affected subsequent interpretation. Normative analysis revealed current norms underestimate at risk children (by 1-2 points) and new norms were created. The clinical, theoretical, and research implications of these findings are presented in the final part of the thesis, alongside a series of recommendations.

Chapter 1.
Literature Review

Positive psychology interventions to improve the mental health of children and adolescents: a
systematic review.¹

¹ An article (Connolly, May, Jackson, & Saville, 2017) based on this chapter was submitted to *Review of Education*, the guidelines of which follow.

Author guidelines

Review of Education: An International Journal of Major Studies in Education is a new online journal launched by the British Educational Research Association and Wiley-Blackwell.

This significant new journal in educational research specialises in publishing reports on major studies, substantial reviews and research syntheses, with a genuinely international reach and orientation. Supplementary materials, such as video abstracts, interviews and teaching resources are welcomed and can be published alongside articles, allowing authors to maximise the impact of their work.

Article Types

The journal is interdisciplinary in approach and specialises in publishing substantial papers of the highest quality from across the field of education that have a genuinely international reach and orientation. In particular, the journal publishes:

- Reports of major studies, including detailed discussion of findings of major research projects. A substantial part of each article should put the study into the context of its field. The design and methodology should also be explained and any limitations discussed.
- Substantial research syntheses, integrating critically results from qualitative, quantitative, or mixed-method empirical studies. Such syntheses may include systematic reviews of studies relevant to particular research questions, thematic narrative reviews, and quantitative meta-analyses.
- Original state-of-the-art reviews that assess the state of knowledge in a field of education research. Thematic, historical, conceptual or theoretical reviews of cutting-edge fields of research and reviews that connect creatively different fields of research are especially welcome.
- Overviews of a number of studies or a field; or any paper that requires substantially more

space than a standard article. The length must be justified by the breadth or detail of the presentation.

Submission

All submissions should be made online at the journal's ScholarOne Manuscript Site [here](#). New users should first create an account. Once a user is logged on to the site submissions should be made via the Author Centre. Full submission instructions can be found on the site. If you have any queries please contact the Editorial office at reviewofeducation@gmail.com

Reviewing

All submissions will be initially screened by the Editors. If a paper falls outside the scope of the journal, fails to meet the style/length requirements, or does not meet basic publication standards it may be rejected outright. All articles that pass the initial screening will be peer reviewed by at least two appropriately qualified referees. The Editors will use these reviews and their own judgment when deciding which papers will go forward for publication. The Editors' decision is final.

Preparing your Paper

Authors should prepare and upload two versions of their manuscript. One should be a complete text, while in the second all information identifying the author should be removed from files to allow them to be sent anonymously to referees. When uploading files authors will then be able to define the non-anonymous version as "File not for review".

Articles should be between 8,000 and 20,000 words, including references. Manuscripts should be double spaced, with ample margins, and bear the title of the contribution, name(s) of the author(s) and the address where the work was carried out. Each article should be accompanied by an abstract of 200-250 words and four keywords. Abstracts should be

accurate summaries including the rationale for the article, methods employed and conclusions drawn. The full postal and email address of the author who will check proofs and receive correspondence should also be included. Footnotes to the text should be avoided. Sponsorship of the research reported (e.g. by research councils, government departments and agencies, etc.) should be declared.

Please visit Wiley's Author Services (<http://www.wiley.com/WileyCDA/Section/id-301854.html>) page for more information about writing an article, preparing your manuscript and general guidance for authors.

Tables and captions to illustrations

Tables must be in separate files and not included as part of the text. Captions to illustrations should also be sent in a separate file. Tables and figures should be numbered by Arabic numerals. The approximate position of tables and figures should be indicated in the manuscript. Captions should include keys to symbols.

Figures

Please supply one set of artwork in a finished form, suitable for reproduction. Figures will not normally be redrawn by the publisher. Please click [here](#) for more detailed information on the submission of electronic artwork.

As an author, you are required to secure permission if you want to reproduce any figure, table, or extract from the text of another source. This applies to direct reproduction as well as "derivative reproduction" (where you have created a new figure or table which derives substantially from a copyrighted source).

References

References should be indicated in the typescript by giving the author's name, with the year of

publication in parentheses. If several papers by the same author and from the same year are cited, a, b, c, etc. should be put after the year of publication. The references should be listed in full at the end of the paper in the following standard form:

Journal

Smyth, J. & Hattam, R. (2002) Early school leaving and the cultural geography of high schools, *British Educational Research Journal* , 28(3), 375-398.

Book

Whitty, G. (2002) *Making sense of education policy: studies in the sociology and politics of education* (London, Paul Chapman).

Chapters in a book

Zukas, M. & Malcolm, J. (2002) Pedagogies for lifelong learning: building bridges or building walls?, in: R.Harrison, F. Reeve, A. Hanson & J. Clarke (Eds) *Supporting lifelong learning. Vol. One: Perspectives on learning* (London, Routledge).

For online documents: Standler, R. (2000) *Plagiarism in colleges in the USA*. Available online at: www.rbs2.com/plag.htm (accessed 6 August 2004).

For online articles with DOI: Smith, A. (2006) A paper on statistics, *Statistics Journal*
doi:10.1111/j.1234-4321.2006.01234.x

Titles of journals should not be abbreviated.

Supplementary materials

The journal welcomes supplementary materials, such as video abstracts, podcast interviews and teaching resources to sit alongside articles. We will provide guidance in how to create these at article acceptance; the items should not be sent with the initial submission. Please contact the Associate Editor for further information.

Context and Implications Document

All authors are required, upon acceptance, to provide a short Context and Implications document, which will be published alongside their article online. These documents allow authors to succinctly demonstrate the relevance of their article to Policy and/or Practice and/or Teaching & Learning in Higher Education.

The document will be reviewed by the Associate Editor for Supplementary Materials before publication but will not be formally copyedited or proofread, so authors will be expected to thoroughly check the content for accuracy before submitting it. The Context and Implications Document template is available for download [here](#).

The Editors may request author amendments to any Context and Implications document before publication.

Please do not submit a Context and Implications document with your initial submission – these will be requested at article acceptance and instructions supplied.

Please address any queries about Context and Implications to the Associate Editor, Alison Kington: a.kington@worc.ac.uk

Video Abstracts

Authors may also, if they wish, provide a short video abstract to accompany their article. This should be a video recording of 2-5 minutes in length, which provides background to the article. You might like to include information such as research rationale, research design/methodology, key findings, relevance to policy, research or practice, etc. It should not simply be a verbatim reading of the abstract.

Video abstracts will be reviewed by the Associate Editor for Supplementary Materials before publication but will not be edited, so please ensure you are happy with the quality and content of the recording before submitting it.

The Editors may request author edits, or decline any video abstract they feel does not meet the required standards of quality or relevance.

Please do not submit a video abstract with your initial submission – these will be requested at article acceptance stage and instructions supplied.

Please address any queries about video abstracts to the Editorial Assistant, Sarah Newman:
reviewofeducation@gmail.com

Non-discriminatory writing

Please ensure that writing is free from gender or ethnic bias. Authors might wish to note the BERA Ethical Guidelines for Educational Research.

Originality and Copyright

Articles submitted to the journal must be original and not previously published elsewhere.

Drafts of papers which can be accessed on individual/personal websites will be regarded as unpublished. However 'working papers' or 'formal reports' on institutional or quasi-institutional websites which are specifically designed to publicise findings and act as dissemination vehicles will be regarded as 'previously published'. Authors should ensure that papers submitted to the journal are substantially different from such working papers and reports, summarising and referring to them as appropriate, but not repeating them verbatim.

It is a condition of publication that authors assign copyright in their articles to the British Educational Research Association (BERA). This ensures full copyright protection and allows dissemination of the article, and the journal, to the widest possible readership.

If your paper is accepted, the author identified as the formal corresponding author for the paper will receive an email prompting them to login into Author Services; where via the Wiley Author Licensing Service (WALS) they will be able to complete the license agreement on behalf of all authors on the paper.

For authors signing the copyright transfer agreement:

If the OnlineOpen option is not selected the corresponding author will be presented with the copyright transfer agreement (CTA) to sign. The terms and conditions of the CTA can be previewed in the samples associated with the Copyright FAQs below:

CTA Terms and Conditions http://authorservices.wiley.com/bauthor/faqs_copyright.asp

For authors choosing OnlineOpen:

If the OnlineOpen option is selected the corresponding author will have a choice of the following Creative Commons License Open Access Agreements (OAA):

Creative Commons Attribution Non-Commercial License OAA

Creative Commons Attribution Non-Commercial -NoDerivs License OAA

To preview the terms and conditions of these open access agreements please visit the Copyright FAQs hosted on Wiley Author Services

http://authorservices.wiley.com/bauthor/faqs_copyright.asp and visit

<http://www.wileyopenaccess.com/details/content/12f25db4c87/Copyright--License.html>.

If you select the OnlineOpen option and your research is funded by The Wellcome Trust and members of the Research Councils UK (RCUK) you will be given the opportunity to publish your article under a CC-BY license supporting you in complying with Wellcome Trust and Research Councils UK requirements. For more information on this policy and the Journal's compliant self-archiving policy please visit: <http://www.wiley.com/go/funderstatement>.

For RCUK and Wellcome Trust authors click on the link below to preview the terms and conditions of this license:

Creative Commons Attribution License OAA

To preview the terms and conditions of these open access agreements please visit the Copyright FAQs hosted on Wiley Author Services

http://authorservices.wiley.com/bauthor/faqs_copyright.asp and visit

<http://www.wileyopenaccess.com/details/content/12f25db4c87/Copyright--License.html>.

Online Open

OnlineOpen is available to authors of primary research articles who wish to make their article available to non-subscribers on publication, or whose funding agency requires grantees to archive the final version of their article. With OnlineOpen, the author, the author's funding agency, or the author's institution, pays a fee to ensure that the article is made available to non-subscribers upon publication via Wiley Online Library, as well as deposited in the funding agency's preferred archive. For the full list of terms and conditions, see Wiley-Blackwell's [OnlineOpen Guidelines](#).

Any authors wishing to send their paper OnlineOpen will be required to complete the payment form available from Wiley-Blackwell's website [here](#).

Prior to acceptance there is no requirement to inform an Editorial Office that you intend to publish your paper OnlineOpen if you do not wish to. All OnlineOpen articles are treated in the same way as any other article. They go through the journal's standard peer-review process and will be accepted or rejected based on their own merit.

Authorship

Authors should be listed in order of contribution to the paper and all authors take responsibility for their own contributions. Only include those authors who have made a

substantive contribution either to the writing and/or to the underpinning research; those who have made marginal contributions (for example, colleagues or supervisors who have reviewed drafts of the work) should be named in an Acknowledgments paragraph after the conclusions.

Plagiarism

Review of Education is a member of CrossCheck by iThenticate. iThenticate is a plagiarism detection service that verifies the originality of submitted materials by screening against millions of published research papers, documents on the web, and other relevant sources. Authors, researchers and freelancers can also use the iThenticate system to screen their work before submission by visiting [iThenticate](#). Papers submitted to *Review of Education* may be screened using iThenticate at any stage between submission and publication.

Author Offprints

Authors will receive electronic offprints of their article through the Author Services system. Registration details will be provided to authors after acceptance. By registering for Author Services, you will be able to: track the progress of your article through the production process, receive publication status alerts, enable free access to your published article online for all co-authors and up to ten nominated colleagues, and enjoy a 25% discount on Wiley books. Future benefits will enable you to discover the number of times your article has been downloaded and which articles have cited yours.

Submission to the Journal of Child Psychology and Psychiatry

Positive psychology interventions to improve the mental health of children and adolescents: a systematic review.

Connolly^{1,2}, J. F., May³, J., Jackson^{1,2}, M., & Saville¹, C.W. N.

¹North Wales Clinical Psychology Programme (NWCPP), School of Psychology, Bangor University, Wales, UK;

²Betsi Cadwaladr University Health Board, Wales, UK

³MerseyCare National Health Service Foundation Trust

Word count (exc. reference list, tables & Appendices): 5, 079

*Correspondence should be addressed to Dr John F. Connolly, North Wales Clinical Psychology Programme (NWCPP), School of Psychology, Brigantia Building, Bangor, Gwynedd, North Wales, United Kingdom, LL57 2DG (e-mail: John.Connolly@wales.nhs.uk)

Abstract

Background: There is limited understanding of the efficacy of Positive Psychology Interventions (PPIs) in children and adolescents. Three randomised controlled trials (RCTs) have been identified in previous reviews, which have been methodologically criticised. This systematic review addresses these methodological challenges. It is one of the first attempts to systematically identify and summarise the related RCT literature.

Method: A systematic search for PPIs targeting children and adolescents was conducted. Databases searched included the Cochrane Library, EBSCOhost, PsycINFO, Wiley online library, Web of Science, and PubMed. Empirical peer reviewed studies in English that were randomised and had a control condition were included.

Results: Twelve studies (n=1668) were identified. Four found no significant effect compared to control while eight found significant effects. All but one study was classified as poor according to Jadad scores/method appraisal. That study involved group intervention (gratitude/hope based) and resulted in significant improvements (medium effect sizes for depression, life satisfaction, gratitude, and hope) compared to controls. Of the remaining studies, effect sizes tended to be small, except in one study. This found a large effect in positive affect for a multi-component PPI at follow up.

Discussion: The evidence base has grown from three to twelve RCTs. There is weak evidence that PPIs are effective in improving mental health in children and adolescents. The array of outcomes, and confounding variables accounted for were identified. Considerable methodological limitations mean that it remains to be seen if effects can generalise outside of study settings and whether small effect sizes are practically meaningful.

Key words: Positive Psychology; systematic review; child and adolescent; mental health

Introduction

Spurred on by the contribution of Seligman (2002; 2009) during and since his presidency of the American Psychological Association, the past 15 years has seen burgeoning research into Positive Psychology and Positive Psychology Interventions (PPIs). Positive psychology approaches report a focus on the promotion of well-being, happiness and ‘flourishing’, a fact that is often contrasted with traditional interventions in mental health, which have historically focused on psychopathology and the reduction of distress (Boiler, Haverman, Westerhof, Riper, Smit, & Bohlmeijer, 2013).

Within the child and adolescent literature (and particularly the education literature), there are strong claims to incorporate PPIs to improve well-being and the justification for this is the robust evidence base authors suggest is underlying these approaches (Bird & Markle, 2012; Waters, 2012). Despite these claims, there is mixed evidence for such incorporation (Chodkiewicz & Boyle, 2017).

Collectively, previous systematic reviews and meta-analyses (Sin & Lyubomirsky, 2009; Bolier et al., 2013) have identified only three randomised controlled trials (RCTs) concerning PPIs in child and adolescent populations. Despite this, the findings are cited frequently to justify PPIs in child and adolescent populations (for example, Jaser et al., 2014). Further, the mechanisms by which PPIs might work are largely unknown in youth populations and there is scant information about the process of implementation.

Preliminary evidence that positive interventions are a viable tool for improving well-being and mental health is highlighted in the aforementioned reviews and meta-analyses which address mixed age (and clinical / non-clinical) populations (Sin & Lyubomirsky, 2009; Bolier et al., 2013). The former review found a significant PPI effect on well-being ($r = .29$) and depression ($r = .31$) and notably that age moderated outcome (greater PPI effect for older

participants). Only three studies (n= 274) concerned children and adolescents (Froh et al., 2008; Rashid, Anjum, & Lennox, 2006; Ruini, Belaise, Brombin, Caffo, & Fava, 2006). A re-analysis by White (2016) shows a significant small effect on well-being ($r = .10$), and no statistically significant effect on depression ($r = .00$), highlighting a lack of reproducibility. The review by Bolier et al (2013) aimed to update Sin's 2009 review (to 2012) and to add to it by focusing only on RCTs and including clear inclusion criteria. No child or adolescent findings were identified.

Background to this review – Why is it important?

There has been limited summary and review of the area and a greater understanding of PPI within the child and adolescent setting is required. Previous reviews focused primarily on the Positive Psychology literature, neglect a wide variety of research. Importantly, intervention findings from adult populations cannot be presumed to extrapolate to child populations. This literature review aims to develop an understanding of which PPIs have been evaluated, are effective among this population, and therefore provide related guidance to clinicians and researchers.

Challenges faced by this review

This review was faced with several challenges associated with methodological issues in the field. Previous reviews of PPIs have had several limitations. First, search methodology has focussed primarily on the Positive Psychology literature (e.g. Sin, 2009) and / or used the term 'positive psychology' as a limiter. This is problematic because interventions and study authors may not associate themselves with this term.

Second, positive psychology, while not a new concept, is a relatively new term and relevant interventions from before its advent would be missed in reviews. It is widely accepted that positive psychology has existed long before the term was coined (Vella-Broderick, 2009).

Finally, and a fact that contributes to the previous points, it is increasingly recognised that a major challenge faced by Positive Psychology is the ongoing challenge of definition. Previous reviews have broadly defined PPIs as “psychological intervention (training, exercise, therapy) primarily aimed at raising positive feelings, positive cognitions or positive behaviour as opposed to interventions aiming to reduce symptoms, problems, or disorders” (Sin & Lyubomirsky, 2009; Bolger et al., 2013). These reviews contradict their definition by excluding studies within the definition's remit (exercise, coping etc.), as almost any intervention (such as exercise, meditation, acupuncture, CBT) could be argued to be a ‘positive psychology’ intervention by this definition.

Further, it is difficult to draw conclusions from literature reviews because of the disparity in methodology and lack of coherence or continuity in search terms. Thus different reviews include vague terms with potentially little connection to the underlying philosophy behind Positive Psychology, or that identify PPIs. Reviews have used arbitrary search terms associated with outcomes or emotional states rather than intervention (for example, Macaskillet et al., 2016; Bolger et al., 2013; Schueller, Kashdan, & Parks, 2014) and found it impossible to apply their inclusion criteria (Macaskillet et al., 2016). It is necessary to improve the transparency of search methodology. Without this, findings cannot be replicated, and scientific quality is questionable.

This review suggests that in order to identify PPIs it is necessary to name specific interventions in *search terms*, particularly when considering the difficulties associated with defining Positive Psychology.

This review defines PPIs as follows:

'interventions which aim to increase well-being, often through invoking positive internal experiences and feelings (for example, gratitude, hope, happiness) that are

embodied in a range of interventions which have been specified in the literature and that are commonly considered to come under the Positive Psychology umbrella. PPI techniques often rely on internal production (cultivating feelings of gratitude through gratitude diaries; recalling positive autobiographical memories; imagining best possible selves).

Aims of this review

The primary aim of this review is to identify and assess systematically the effects of PPI interventions in children and adolescents. Because there is a lack of knowledge about the process underlying intervention and implementation (Ciarrochi et al., 2016) secondary aims are to identify process and implementation issues, outcome measures, and confounding variables typically measured.

Method

Criteria for considering studies for this review

Types of studies

Interventions using PPIs to improve outcomes were considered in this review. Only empirical peer reviewed studies in English that were randomised and had a control condition were included.

As Bolier et al. (2013) highlight, interventions should explicitly fit the theoretical tradition of Positive Psychology. Examples include: gratitude interventions (diaries/ visits/ letters, e.g. Froh et al., 2008); Best Possible Selves intervention (e.g. Owens & Patterson, 2013); character strengths intervention or strengths based therapy (e.g. Proyer, Gander, Wellenzohn, & Ruch, 2015); specific interventions to develop hope and optimism (e.g. Pedrotti, 2000); group PPI interventions (e.g. Harrison et al., 2016); broad minded affective coping interventions (e.g. Johnson et al., 2013).

Exclusion criteria

We excluded interventions without evaluative pre-and post-outcome measures, medical or drug interventions, and studies in adult populations (>18). We excluded interventions that did not include a PPI and those not typically considered to be a PPI. These included a primary focus on Cognitive Behavioural Therapy, Solution Focussed Therapy, Dialectical Behaviour Therapy, Acceptance and Commitment Therapy, parenting programmes (for example, Incredible Years, Triple P), self-management disease programmes, social skills training, complementary and alternative therapies, massage, physiotherapeutic intervention, yoga, exercise and dietary interventions.

Types of outcome measures

Because the Child and adolescent PPI literature is in its infancy, this review was interested in understanding the variety of applicable measures and change therein. Information was collected on validated quantitative intervention outcome measures. This included measures of psychopathology (e.g. anxiety and depression) as well as measures relating to well-being, satisfaction or positive emotion (happiness, hope, gratitude etc.).

Search strategy for identifying studies

A scoping review identified search strategies from previous reviews and meta-analyses (Sin & Lyubomirsky, 2009; Bolier et al., 2013; Meyers, van Woerkom, & Bakker, 2013). These then informed search strategy and exclusion criteria. Terms unnecessary to identifying PPIs in our context were removed (for example, search terms used by Meyers et al. [2013] specifically relating to organisational research or ‘depression’), to lead to the below search term combination. As referred to in the literature discussion we did not limit our search to positive psychology fields or terms (‘positive psychology’), or by date the Positive Psychology movement is purported to originate (a key criticism of the area; Schueller, Kashdan, & Parks, 2014). We were primarily interested in specific treatments that have been identified as PPIs in the child and adolescent literature. Feedback on our search criteria from from previously mentioned systematic review authors did not identify any additional PPIs, although one reviewer questioned whether a CBT based program constituted a PPI. Because certain programs that are heavily CBT- based have been referred to as PPIs (e.g. The Penn Resiliency Program), we included them in search criteria to allow description in discussion but these were not included in the final analysis. The effectiveness of the search strategy was tested by ensuring it produced known child and adolescent interventions.

The following electronic sources were searched: The Cochrane Library; EBSCOHOST (including Cumulative Index to Nursing and Allied Health [CINAHL]); PSYCHINFO; the

Wiley online library; and the Web of Science (including PubMed). Papers were limited to English language peer-reviewed empirical investigations.

Search terms were as follows: “Gratitude diary” “OR” “positive psychotherapy” OR “well-being therapy” OR “benefit finding” OR “positive reappraisal coping” OR “savouring” OR “kindness” OR “QOL therapy” OR “meaning making” OR “hope therapy” OR “hope intervention” OR “strength centred therapy” OR “strength-based counselling” OR “positive psychotherapy” OR “meaningful life therapy” OR “appreciative inquiry” OR “strengths coaching” OR “strengths approach” OR “strengths-based approach” OR “Optimal functioning therapy for adolescents” OR “best possible self” OR “best possible selves” OR “broad minded affective coping” or “broad minded coping” OR “Positive Psychotherapy Intervention” OR “well-being therapy” OR “Acts of Kindness” or “blessing” or “count your blessings” OR “the Penn Resiliency Program” OR “the Strath Haven Positive Psychology Curriculum” OR “the positive psychology programme” OR “positive reminiscence” OR “positive reminiscing” or “happiness” AND (“child” OR “children” OR “adolescent” OR “adolescence” OR “teen” OR “teenager”) AND (“rct” OR “randomised trial” OR “randomized trial” OR “intervention”).

Study evaluation and summary

Studies were included based on the above criteria and evaluated using validated measures (Appendix 1), including the Jadad rating scale (Jadad, Moore, & Carroll, 1996) and an adapted method appraisal measure (MAM, Connolly et al., 2013). Because it is unlikely that an RCT study will be retrospective we replaced one item of the MAM (item 1, ‘Is the study prospective?’) with a validity question identified as important in systematic review guidance (CASP, 2016).

A dual rating system assessed inter-rater reliability and measure performance.

Because the RCT literature for this population is small and in the initial stages, we extract a range of information that may guide clinicians and future research (Appendix 2; 7).

Specifically, the array of confounding variables and outcome measures presented in the RCT literature is summarised. Information relating to implementation was also extracted but is presented in a separate paper to follow.

Results

A total of 3494 studies were identified in the initial search. An additional 8 were identified following communication with authors and in reference searches. A breakdown of the process is presented in Figure 1. Excluded studies and reasons for exclusion are available on request.

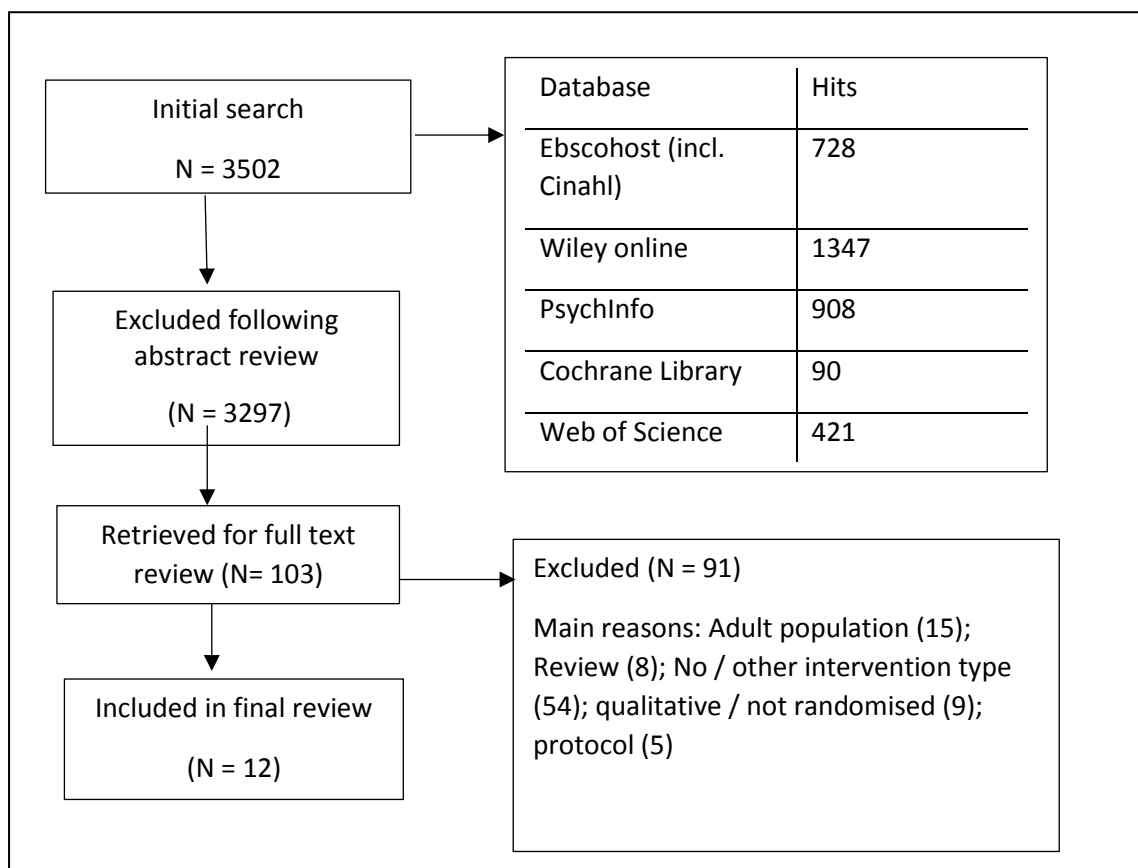


Figure 1 Breakdown of search results

Included studies

Twelve studies involving a total sample of 1668 children and adolescents were identified (mean age range 10.3 to 15.6). These took place in Australia (n = 2), Hong Kong (n = 1), U.S.A. (n = 5), and Italy (n = 3) and were published between 2006 and 2017. Most studies took place in school settings (n = 9), one took place in an inpatient psychiatric unit, another in

an outpatient diabetes clinic, and one study was a self-directed online program. Studies are summarised in Table 1.

Types of interventions

Seven studies concerned PPIs relating to gratitude (Froh et al., 2008), gratitude and best possible selves (Froh et al., 2009), gratitude and hope (Kwok et al., 2016), character strengths (Toback et al., 2016), happiness (Jaser et al., 2014a), or multi component PPI programs (Suldo et al., 2014; Roth et al., 2017). A further 5 studies concerned PPIs combined with other therapeutic approaches (e.g. healthy lifestyle activities; mindfulness, cognitive therapy). Two of these involved an online program (Burkhardt et al., 2015; Manicavasagar et al., 2014) and 3 studies evaluated a well-being therapy protocol within school settings (Ruini et al., 2006; Ruini et al., 2009; Tomba et al., 2010).

Types of Outcomes and measures

In the RCT literature, PPI studies most commonly measure depression ($n = 4$), affect ($n = 4$), life satisfaction ($n = 5$), and well-being ($n = 5$). The most common measures used are the Positive and Negative Affect Scale (Laurent et al., 1999) and the Student Life Satisfaction Scale (Huebner, 1991). Measures vary in length from 7 (Brief Multidimensional Students Life Satisfaction Scale, BMSLSS; Huebner & Valois, 2003) to 92 items (Symptoms Questionnaire, Kellner, 1987). While not used as an outcome, two studies have used the VIA-Youth measure (Park & Peterson, 2009) to inform subsequent character strengths interventions, and the BMLSS has been used as a pre-screening measure, also in two studies. A summary of all outcome measures is provided in the study summary that follows (Table 1).

Summary of effectiveness of PPIs not combined with other therapies ($n = 7$)

The largest effect ($d = .81$). was found for Roth et al. (2017) which found significant improvement in Positive Affect at 7 week follow up for a multi-component PPI group

compared to control. One study of a gratitude / hope program (Kwok et al., 2016) found significant improvements in intervention compared to control on hope (partial eta squared $[\eta^2] = .21$), gratitude (partial $\eta^2 = .28$), life satisfaction (partial $\eta^2 = .10$), and depression (partial $\eta^2 = .21$), representing medium effects at post-test. Changes in gratitude (but not hope) mediated the PPI effect on depression and satisfaction and hope mediated effects on life satisfaction (Kowk et al., 2016).

The remaining studies found small effect sizes. One study (Froh et al., 2008) found a brief gratitude intervention group (writing down 5 gratitudes daily) significantly improved satisfaction with school, and lowered negative affect compared to a treatment as usual ($d = .33$) and active control group ($d = .27$) at 3 week follow up. A psychiatric inpatient study involving a brief (44 minutes) character strengths intervention (Toback et al., 2016) found improvements maintained only for intervention on self-esteem ($d = .10$) and self-efficacy ($d = .20$) at 3 months.

Compared to control no significant improvements were identified in three studies at follow up (Jaser et al., 2014; Suldo et al., 2014; Froh et al., 2009). Notably, Froh, Kashdab, Ozimkowski & Miller (2009) found no intervention effect but that those low in positive affect at baseline made greater gains (although this could be a regression to the mean). Jaser, Russell, Rothman, Choi, & Whittemore's (2014) PPI (gratitude, self-affirmation, parent affirmations, small gifts) did not improve on any outcome measure compared to active control except self-reported adherence.

Table 1 Included studies table

Study	Design, aim, methods appraisal, & analyses	Setting and participants	Intervention	Psychosocial outcomes and results ↑ = improvement (in PPI vs control) ↔ = no change. ↓ = worsening.
Kwok et al., 2016	<p><i>Design:</i> Intervention v control; Within each school, half of children randomised to each condition.</p> <p><i>Methods appraisal score:</i> 6 / 10 [Individual criteria met: 1; 3; 5; 6; 7; 10] Jadad score: 10 / 13</p> <p><i>Aim:</i> To investigate the effects of a gratitude / hope intervention on Depression and life satisfaction.</p> <p><i>Analyses:</i> Baseline, post intervention; t-test & χ^2 tests of group baseline differences; 2 (intervention v control) x 2 (pre-post) repeated measures ANOVA; mediation analyses.</p>	<p><i>Setting:</i> Hong Kong, 5 primary schools; convenience sample.</p> <p><i>Participants:</i> N = 447; 'marginal' cases on HADS (those scoring 9-11; n = 77 / 17%) invited for interview; 68 recruited.</p> <p><i>Intervention:</i> N = 34; 6-8 in each condition; mean age = 10.5; 57% male.</p> <p><i>Control:</i> N = 34; 6-8 in each condition; mean age = 10.3; 50% male.</p>	<p><i>Details:</i> <i>Intervention:</i> 'Live a positive life' Group sessions of a gratitude /hope intervention (90 minutes) after class schedule. Variety of exercises (promoting expression of gratitude; appreciating self and others; cultivating agency towards goals) and related homework; Therapist facilitators.</p> <p><i>Control:</i> School as usual.</p> <p>Attrition: 0%</p>	<p><i>Measures:</i> ↑Depression (HADS) ↑Hope (CHS) ↑Gratitude (GQ) ↑Life satisfaction (SLS)</p> <p><i>Details:</i> Improved scores in intervention group on all outcomes compared to control at post-test. Changes in hope and gratitude mediated life satisfaction. Gratitude mediated depression symptoms.</p> <p><i>Effect size:</i> $h^2 = .21$ for depression; $h^2 = .10$ for life satisfaction; $h^2 = .21$ for hope; $h^2 = .28$ gratitude.</p>
Jaser et al., 2014	<p><i>Design:</i> Intervention v active control; Block randomised</p> <p><i>Methods appraisal score:</i> 6 / 10 [Individual criteria met: 3; 5; 6; 8; 9;10] Jadad score: 9 / 13</p> <p><i>Aim:</i> To investigate the effect of a PPI on a variety of outcomes (adherence, family conflict, glycaemic control) in adolescents with Diabetes.</p> <p><i>Analyses:</i> Baseline, 3 and 6 months; mixed effects models tested intervention effects; Bivariate correlations between affect and adherence; Analyses controlled for group differences in depression, but not ethnicity or income and no results provided.</p>	<p><i>Setting:</i> U.S.A; outpatient clinic</p> <p><i>Participants:</i> Mainly white, aged between 13-17 years (mean = 15), of moderate income; Recruited during clinic (no further detail); good adherence generally;</p> <p><i>Intervention:</i> N = 20; 40% male; mean age = 15.3 (1.4) For 37% income < \$40,00</p> <p><i>Control:</i> N = 19; 58% male; mean age =15 (1.6) For 6% income < \$40,00 <i>Attrition:</i> no drop-outs</p>	<p><i>Details:</i> Duration was 8 weeks</p> <p><i>Intervention:</i> PPI exercises tailored based on initial interview ("e.g. what are some of the things that make you happy...?" p. 3); parents sent weekly positive affirming messages; instructed to spend time remembering things of which led to them being happy and proud when checking blood levels. Phone call reminders (5 minutes) fortnightly; preferred gifts (\$3 items) by mail every 2 weeks</p> <p><i>Control:</i> Received educational information about Diabetes and management every 2 weeks</p>	<p><i>Measures:</i> ↔ Self-care /adherence (SCI) (but ↑ in control) ↔ Family conflict (DFC) ↔Quality of Life (PQLI) ↔Adherence (Mean Blood Glucose (BG) previous week) ↔HBA1C (average BG over 2-3 months)</p> <p><i>Details:</i> No significant improvement in blood glucose monitoring over time; But increased adherence (SCI) in control. Positive affect associated with greater monitoring in intervention but not control at 3 & 6 months.</p> <p><i>Effect size (d):</i> Not available</p>

Study	Design, aim, methods appraisal, & analyses	Setting and participants	Intervention	Psychosocial outcomes and results ↑ = improvement (in PPI vs control) ↔ = no change. ↓ = worsening.
Froh et al., 2009	<p><i>Design:</i> Intervention v control; Matched by grade and randomised. No detail reported as to initial sampling; likely convenience sample.</p> <p><i>Methods appraisal score:</i> 7 / 10 [Individual criteria met:10; 9; 8; 6; 5; 4; 3] Jadad score: 8 / 13</p> <p><i>Aim:</i> To investigate a gratitude intervention and affect as a moderator.</p> <p><i>Analyses:</i> Baseline, immediately post intervention, and at 1 and 2 months. Phi correlations; 2 (condition) x 3 (t1-t3) repeated measures ANCOVA (baseline as covariate); Hierarchical regressions</p>	<p><i>Setting:</i> U.S.A school</p> <p><i>Participants:</i> N = 89; mean age = 12.74 (3.48); 49.4% male, 67% Caucasian; 12.4% Asian American, 9% Hispanic.</p> <p><i>Intervention:</i> 3rd grade intervention: N =14, control = 15 9th grade intervention N= 20, control = 19 12th grade intervention N = 10, control = 11</p>	<p><i>Details:</i> 10-15 minutes daily for 5 days; in class time; spread over 10 day period. One author supported process (ensure pupils stayed focussed and facilitated pupils reflective discussions- which took place in corner of the same classroom as controls). Parents completed a form to check implementation.</p> <p><i>Intervention:</i> Writing a letter of gratitude and reading it to the person they are thankful to.</p> <p><i>Control:</i> N = Write about yesterday, the things you did and what if felt like; teacher implemented</p>	<p><i>Measures:</i> ↔Gratitude (GAC) ↔Positive affect (PANAS-C)</p> <p><i>Details:</i> No group differences in Gratitude, Positive or negative effect at any time point when compared to control. A sub group analysis showed those low in Positive Affect reported more gratitude and Positive affect at T2 and T4 compared to control.</p> <p><i>Effect size:</i> Gratitude $n^2 = 0.03$ Positive affect $n^2 = 0.01$ Negative affect $n^2 = 0.01$</p>
Toback et al., 2016	<p><i>Design:</i> Intervention v active control;</p> <p><i>Methods appraisal score:</i> 4 / 10 [Individual criteria met: 1; 7; 9; 10] Jadad score: 8 / 13</p> <p><i>Aim:</i> To test the effects of a character strengths intervention on self-esteem and efficacy.</p> <p><i>Analyses:</i> Baseline at admission, 2 days post intervention, 2 weeks, 3 months. Two-tailed paired t-tests for esteem and efficacy changes; two tailed independent t-tests for other analyses.</p>	<p><i>Setting:</i> Inpatient psychiatric unit, Michigan</p> <p><i>Participants:</i> 81 hospitalised adolescents, 49 female; age range 12-17; Mean age = 15.3 (1.4); variety of diagnoses (95% mood disorder; 31% anxiety disorder); 81% white; affluent area.</p> <p><i>Intervention:</i> N = 40; Breakdown NR</p> <p><i>Control:</i> N = 41; Breakdown NR</p>	<p><i>Details:</i> Recruited on admission, initial sample prior to exclusion criteria not reported;</p> <p><i>Intervention:</i> Character strengths identified (VIA-Youth) on day 2 of admission (20 minutes); coping skills (2-4 per strength) then identified with researcher based on strengths on index cards (24 minutes); participants asked to use skills from then on.</p> <p><i>Control:</i> One to one activities / games (24 minutes) Attrition:15 (37%); 60% missing at follow-up</p>	<p><i>Measures:</i> ↑Self-efficacy (GSE) ↑Self-esteem (SES)</p> <p><i>Details:</i> Improvements in esteem and efficacy for both groups immediately post intervention but only maintained in intervention at 3months.</p> <p><i>Effect size (d):</i> Self-efficacy: $d = .20$ Self-esteem $d = .10$</p>

Study	Design, aim, methods appraisal, & analyses	Setting and participants	Intervention	Psychosocial outcomes and results ↑ = improvement (in PPI vs control) ↔ = no change. ↓ = worsening.
Froh et al., 2008	<p><i>Design:</i> 11 classes randomly assigned to 3 conditions; Intervention v active v no treatment control</p> <p><i>Methods appraisal score:</i> 4 / 10 [Individual criteria met: 1; 3; 7; 10] Jadad score: 5 / 13</p> <p><i>Aim:</i> To partially replicate a previous adult study (Emmons et al., 2003) in testing a gratitude intervention.</p> <p><i>Analyses:</i> Immediately post intervention & 3 week follow up Baseline group equivalence stated for outcomes but data not provided;</p> <p>ANCOVA's; regression analyses; Bivariate correlations between composites</p>	<p><i>Setting:</i> U.S. Middle school</p> <p><i>Participants:</i> 221 6th and 7th graders; Mean age 12.17 (.67); 50% male vs 41% female; 69% Caucasian.</p> <p>Demographic breakdown NR for conditions</p>	<p><i>Details:</i> 2 week intervention period; scripts provided.</p> <p><i>Gratitude intervention:</i> 4 classes; N = 76; ask to write down up to 5 things they were grateful /thankful for (past day); No examples of gratitude given.</p> <p><i>Hassles:</i> 4 classes; N =80; asked to write down up to 5 hassles (past day). Examples of hassles given</p> <p><i>TAU Control:</i> 3 classes; N = 65; TAU</p> <p>On day 8, one class failed to complete measures (teacher forgot);</p>	<p><i>Measures:</i> ↑Life satisfaction school (BMSLSS) ↑ / ↔ Gratitude (study composite measure) ↔Positive Affect (study composite measure) ↑ / ↔ Negative Affect ↔Life satisfaction item (how feel about life as a whole) ↑ / ↔Life satisfaction item (how expect to feel about life) ↔Physical symptoms (list) ↑ / ↔Reactions to aid (sympathy; help/advice; talked to someone and how it felt) in solving problems. ↔ Prosocial items (x2; have you helped someone or offered emotional support?)</p> <p><i>Details:</i> The gratitude group reported significantly higher school satisfaction compared to both hassles and control. Mixed findings as above: better scores than active but not TAU control (↑ / ↔).</p> <p><i>Effect size (d):</i> Small to medium (.01-.06); School satisfaction (d =.27 v control; d = .33 v TAU)</p>
Roth et al., 2017	<p><i>Design:</i> Intervention v wait-list control; Participants allocated alternately (based on life satisfaction scores)</p> <p><i>Methods appraisal score:</i> 6 / 10 [Individual criteria met: 1; 5; 6; 7; 9; 10] Jadad score: 6 / 13</p> <p><i>Aim:</i> To investigate whether a modified version of a multi-component PPI was effective.</p> <p><i>Analyses:</i> Baseline (Jan 2014), postintervention (April 2014) and two follow-up points (5 & 7 weeks post intervention); Piecewise growth modelling tested change pre to post; then post to 7 week follow up. Groups equivalent on outcome at baseline but demographic tests NR</p>	<p><i>Setting:</i> U.S. middle school</p> <p><i>Participants:</i> All 274 students completed a Life satisfaction scale (BMSLSS); low scorers (<6) invited to participate (n = 111). N = 42 7th graders; 40% male; primarily white (n = 35). Age range = 11-13.</p> <p><i>Intervention:</i> N = 21; 43% male; 24% eligible for free lunch; 24% Hispanic, Latino or Spanish; 76% not Hispanic; 45% parents married.</p> <p><i>Control:</i> N = 21; 57% male; 19% eligible for free lunch; 19% Hispanic, Latino or Spanish; 76% non-Hispanic; 43% parents married.</p>	<p><i>Details:</i> Manualised program with hand-outs; expanded Suldo et al.'s (2013) PPI program by including 2 follow-ups and parents.</p> <p><i>Intervention:</i> 10 x 50 minute group (n=7) meetings as in Suldo et al (2014); Two follow up booster sessions (50 minutes) at 5 and 7 weeks post; Parent information session (60 minutes) included PowerPoint and discussion; weekly email handouts (x 10) to parents overviewed child sessions, homework and how they can support child.</p> <p><i>Control:</i> Wait list assigned</p>	<p><i>Measures:</i> ↑Negative Affect (PANAS-C) ↑Positive Affect (PANAS-C) ↑Life Satisfaction (SLSS)</p> <p><i>Details:</i> Significant gains in Life satisfaction, PA & NA compared to control post intervention and in PA at 7 weeks.</p> <p><i>Effect size (d):</i> Medium to large effect size (0.81) for Positive Affect at 7 week follow up.</p>

Study	Design, aim, methods appraisal, & analyses	Setting and participants	Intervention	Psychosocial outcomes and results ↑ = improvement (in PPI vs control) ↔ = no change. ↓ = worsening.
Suldo et al., 2014	<p><i>Design:</i> Intervention v wait list control;</p> <p><i>Methods appraisal score:</i> 4 / 10 [Individual criteria met: 1; 3; 7; 10] Jadad score: 9 / 13</p> <p><i>Aim:</i> To develop and test a manualised PPI that intends to increase well-being.</p> <p><i>Analyses:</i> Baseline, immediately after, & 6 months post intervention. No baseline differences according to gender, ethnicity, parental marital status, and guardianship but intervention had lower satisfaction and higher externalising symptoms so propensity score matching was used; ANOVAs primarily analysis method.</p>	<p><i>Setting:</i> U.S. middle school</p> <p><i>Participants:</i> N = 55; Age range = 10-12 years; Mean age = 11.43 (.55);</p> <p><i>Intervention:</i> N = 20; 40% male; Hispanic (30%) Caucasian (25%); Asian (25%); African American (15%); Native American (5%)</p> <p><i>Control:</i> N = 20; 35% male; Hispanic (30%) Caucasian (40%); Asian (5%); Native American (10%) <i>Attrition:</i> 11 dropped out (16%) during intervention (Final analyses: 40 / 67 [60%]).</p>	<p><i>Details:</i> 10 weekly 50 minute group sessions; Manualised program with hand-outs.</p> <p><i>Intervention:</i> 5 groups (n=7); each had a leader and co-facilitator; content included range of tasks, discussion relating to topics including gratitude journaling (5 things), a gratitude visit, character strengths (VIA-youth questionnaire informed), future oriented emotions of hope and optimism (including best possible selves task), performing acts of kindness and recording feelings.</p> <p><i>Control:</i> Wait list assigned; completed intervention the following school year.</p>	<p><i>Measures:</i> ↔Negative Affect (PANAS-C) ↔Positive Affect (PANAS-C) ↑Life Satisfaction (SLSS) (not maintained at follow up) ↔Child Behaviour Checklist (YSR Internalising, externalising)</p> <p><i>Details:</i> Gains in Life satisfaction for intervention at post-test but this was matched by similar control group gain at follow up. No other group differences.</p> <p><i>Effect size (d):</i> Small effect sizes (.01-.32)</p>
Ruini et al., 2006	<p><i>Design:</i> Intervention v active control;</p> <p><i>Methods appraisal score:</i> 4 / 10 [Individual criteria met: 3; 5; 9; 10] Jadad score: 7 / 13</p> <p><i>Aim:</i> To test well-being based therapy protocol in comparison to standard CBT.</p> <p><i>Analyses:</i> Paired t tests; ANCOVA (statistics NR)</p>	<p><i>Setting:</i> Italian middle school</p> <p><i>Participants:</i> N =111; 6 classes</p> <p><i>Intervention:</i> N = 57; 31 male; mean age = 13.04 (0.76)</p> <p><i>Active Control:</i> N = 54; 29 male; mean age =13.22 (0.69) Protocol based solely on CBT techniques (Beck, 1979) focussed on negative emotion.</p>	<p><i>Details:</i> 4 x 2 hour sessions fortnightly; in the first 2 sessions both conditions included games, role play and discussion and were thought to recognise emotions and link to behaviour, e.g. through diary keeping.</p> <p><i>Intervention:</i> Protocol based on well-being therapy (Fava, 1999); attention focussed on recognising, expressing and sharing positive emotions, according to (Rhyff's 1995 model). Session 3: e.g. name a positive thing aspect about a friend and pay a compliment, record in diary. session 4: identify and share with class positive factors relating to personality and specific positive life memories; identify potential daily positive moments and record.</p>	<p><i>Measures:</i> ↔ Psychological Well-Being Scales (PWB) ↔ Symptom Questionnaire (SQ); ↑ Physical well-being (SQ subscale)</p> <p><i>Details:</i> No overall differences in PWB or SQ, but significant improvement in physical well-being (SQ subscale) in intervention compared to control.</p> <p><i>Effect size (d):</i> Physical well-being = -0.29</p>

Study	Design, aim, methods appraisal, & analyses	Setting and participants	Intervention	Psychosocial outcomes and results ↑ = improvement (in PPI vs control) ↔ = no change. ↓ = worsening.
Ruini et al., 2009	<p><i>Design:</i> Intervention v attention-placebo; classes randomised.</p> <p><i>Methods appraisal score:</i> 3 / 10 [Individual criteria met: 1; 5; 10] Jadad score: 7 / 13</p> <p><i>Aim:</i> To compare a well-being program adapted for high school adolescents against an attention-control.</p> <p><i>Analyses:</i> Baseline; post and at 6 months Students t tests for baseline difference analysis; x2; repeated measure ANOVA for efficacy analyses.</p>	<p><i>Setting:</i> Italian high school</p> <p><i>Participants:</i> <i>Initial sample unclear ('various' schools)</i> 227; 139 female; mean age = 14.4 (0.673); Eight 9th grade & one 10th grade class; unequal groups at baseline. Attrition: 1 class (n = 25) assigned to intervention was lost to follow up (graduated);</p> <p><i>Intervention:</i> N = 5 classes (129) but only 4 classes (n=104) followed up; 42 (32%) male; mean age = 14.47 (.076).</p> <p><i>Control:</i> 4 classes (n = 98); 46 (47%) male; mean age = 14.32 (0.54).</p>	<p><i>Details:</i> Adapted well-being therapy (Ruini et al., 2006) to high school adolescents;</p> <p><i>Intervention:</i> <i>6 x 2 hour weekly classes in school; As per Ruini et al 2006 but an extra 2 sessions added (e.g. cognitive restructuring).</i> Sessions targeted Rhff's (1989) well-being dimensions of autonomy, environmental mastery, positive interpersonal relationships, personal growth, purpose in life and self-acceptance.</p> <p><i>Control:</i> <i>6 x 2 hour weekly classes in school</i></p>	<p><i>Measures:</i> ↔Symptom Questionnaire (SQ) total ↑SQ physical well-being subscale ↑SQ somatisation subscale ↑Psychological Well-Being Scales (PWB) ↔ ↑PWB Personal growth subscale ↔ Children's Anxiety Scale (RCMAS) total ↑RCMAS physiological subscale</p> <p><i>Details:</i> At 6 months, the only effects remaining favouring control were for SQ somatisation (d = -.21) and SQ anxiety (d = .36).</p> <p><i>Effect size (d):</i> Small effect sizes (d= -.42-.36 in intervention); unexpected negative effect sizes for emotional mastery in both groups.</p>
Tomba et al., 2010	<p><i>Design:</i> Intervention v active control;</p> <p><i>Methods appraisal score:</i> 8 / 10 [Individual criteria met: 1; 3; 5; 6; 7; 8; 9; 10] Jadad score: 8 / 13</p> <p><i>Aim:</i> To investigate differential effects of well-being therapy versus anxiety management.</p> <p><i>Analyses:</i> Baseline, immediately post, & 6 months Repeated measures ANCOVA with baseline scores as covariates. ANOVA's for within group testing. Higher well-being in intervention group at baseline.</p>	<p><i>Setting:</i> Italian middle schools</p> <p><i>Participants:</i> 8 classes (N=162) across 4 schools; 94 female; mean age = 11.41 (.56)</p> <p><i>Intervention:</i> 4 classes (N = 82); 40 males; mean age = 11.44 (.56)</p> <p><i>Control:</i> 4 classes (N = 80); 28 males; mean age = 11.39 (.53)</p>	<p><i>Details:</i> Manualised protocols developed for both groups; 6 x 2 hour sessions.</p> <p><i>Intervention:</i> Well-being therapy protocol; as in Ruini et al., 2010; based on Rhff's (1989) well-being dimensions</p> <p><i>Control:</i> Anxiety management (AM) CBT based protocol including education; restructuring; communication skills; breathing; self- talk; relaxation; Discussion of positive emotion or Rhff's (1989) well-being dimensions omitted.</p>	<p><i>Measures:</i> ↔Symptom Questionnaire (SQ) total ↔Psychological Well-Being Scales (PWB) ↔ Children's Anxiety Scale (RCMAS) total</p> <p><i>Subscales:</i> ↔ SQ anxiety (↑ for control not intervention) ↑SQ friendliness</p> <p><i>Details:</i> At post-test or 6 months, there were no significant differences for the overall measures between the groups. The AM control improved on SQ anxiety subscale compared to intervention. The intervention improved on SQ friendliness compared to the AM control.</p> <p><i>Effect size (d):</i> Friendliness (d = .26)</p>

Study	Design, aim, methods appraisal, & analyses	Setting and participants	Intervention	Psychosocial outcomes and results ↑ = improvement (in PPI vs control) ↔ = no change. ↓ = worsening.
Burckhardt et al., 2015	<p><i>Design:</i> Intervention v active control; Block randomised</p> <p><i>Methods appraisal score:</i> 1/10 [Individual criteria met:10] Jadad score: 8/13</p> <p><i>Aim:</i> To adapt and evaluate a structured online PPI.</p> <p><i>Analyses:</i> T-tests of baseline differences; Multilevel linear modelling.</p>	<p><i>Setting:</i> 5 Australian high schools</p> <p><i>Participants:</i> N=338 single sex and co-ed schools; Grades 7-12; mean age range 12.6-15.6 (0.3-1.6)</p> <p>Religious non-governmental schools chosen due to proximity to researchers; gender breakdown not calculatable.</p> <p><i>Attrition:</i> High attrition (59%);</p>	<p><i>Details:</i> Both conditions delivered / supervised by teachers in class time (6 hours over 6 weeks). Website posts moderated and clinical psychologist advised when necessary (e.g. if users present in need of support).</p> <p><i>Intervention:</i> n = 313; 'Bite-back' online interactive activities with supporting workbook; activities relating to gratitude entries, hope, photo's, personal stories, mindfulness. Sample mean ages not calculatable.</p> <p><i>Control:</i> n = 259; similar format but without the PPI content.</p>	<p><i>Measures:</i> ↔Depression, Anxiety, and Stress Scale (DASS-21) ↔Student Life Satisfaction Scale (SLSS) ↔ The short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS)</p> <p><i>Details:</i> Improvements in both conditions but no benefit for intervention group when compared to control. Increased well-being /flourishing for control but not intervention.</p> <p><i>Effect size (d):</i> n/a Significant time and condition interaction (F1,202.9=5.88, P=.02.). Controls improved well-being from pre (mean 16.68) to post intervention (18.19) but intervention did not (18.18 to 18.37).</p>
Manicavasagar et al., 2014	<p><i>Design:</i> Intervention v active control; Block randomised</p> <p><i>Methods appraisal score:</i> 4 / 10 [Individual criteria met: 3; 7; 8; 10] Jadad score: 9 / 13</p> <p><i>Aim:</i> To investigate the feasibility and implementation of an online PPI package ('Biteback')</p> <p><i>Analyses:</i> Two tailed Wilcoxon signed rank tests</p>	<p><i>Setting:</i> Australian schools and youth groups</p> <p><i>Participants:</i> N = 235; 104 (67%) female; Aged 12-16; mean age = 15.4 (1.7).</p> <p>Recruited by mail, flyer, email, letter, & organisation. No further demographic information (e.g. SES) reported.</p>	<p><i>Details:</i> 1 hour on website per week; self-directed (over 6 weeks). Website posts moderated.</p> <p><i>Intervention:</i> N = 120. Online interactive activities relating to gratitude entries, hope, photo's, personal stories, mindfulness. Mean ages not calculatable.</p> <p><i>Control:</i> N = 115; similar online format but without the PPI content.</p>	<p><i>Measures:</i> ↑ Well-being (SWEMWBS) ↑ Depression and Stress (DASS-21)</p> <p><i>Details:</i> Overall improvements in PPI but not control for all measures.</p> <p>However, effects only observed for high frequency / time users suggesting a dose effect. Note: These users also improved well-being in the control condition.</p> <p><i>Effect size (d):</i> Depression d = .30; Anxiety d=.001; Stress d = .16; Well-being d = .20</p>
<p>Methods appraisal: 1 = valid; 2 = representative sample; 3 = appropriate comparison group; 4 = baseline response .60%; 5 = follow-up .80% in cohort, .60% in cross-section; 6 = adjustment for non-response and drop-out; 7 = conclusions substantiated by data; 8 = adjustment for confounders; 9 = all intervention group exposed, non-contaminated comparison group; 10 = appropriate statistical tests. NR = not reported; where effect sizes were not reported we computed the difference between effect sizes in the two groups; Positive and Negative Effects scale (PANAS, Watson & Clark, 1988); Depression, Anxiety, and Stress Scale (DASS-21; Lovibond & Lovibond, 1995); Student Life Satisfaction Scale (SLSS, Huebner, 1991); The short Warwick-Edinburgh Mental Well-being Scale (SWEMWBS, Stewart-Brown et al., 2009; Children's Depression Inventory, Kovacs (CDI, 1985); Self-care Inventory (SCL, La Greca, 2004); Diabetes Family conflict scale (DFC, Rubin et al., 1989); Paediatric Quality of Life Inventory Type 1 modules (PQLI, Varni et al., 2003); Hospital Anxiety and Depression Scale (HADS, Chinese version, Leung et al., 1993); Childrens Hope Scale (CHS, Chinese version, Snyder et al., 1997); The Gratitude Questionnaire (GQ, Mc Cullough et al., 2002); Satisfaction with Life Scale (SLS, Diener et al., 1985); The Gratitude Adjective Checklist (GAC, Mc Cullough et al., 2002); The Positive and Negative Effect Scale for Children (PANAS-C, Laurent et al., 1999); General Self Efficacy Scale (GSE; Schwarzer & Jerusalem, 1995); Rosenberg Self-esteem Scale (SES; Rosenberg, 1965); Values in Action Inventory of Strengths for Youth (VIA-Youth, Park & Peterson, 2009); Brief Multidimensional Students' Life Satisfaction Scale (BMSLSS; Seligson, Huebner, & Valois, 2003); The Students' Life Satisfaction Scale (SLSS; Huebner, 1991); The Positive and Negative Affect Scale for Children (PANAS-C; Laurent et al., 1999); The Brief Problem Monitor-Youth (BPM-Y; Achenbach, McConaughy, Ivanova, & Rescorla, 2011); Child behaviour checklist- Youth self-report (YSR; Achenbach & Rescorla, 2001); Kellner's Symptom Questionnaire (SQ, Kellner, 1987) ; Ryff's Psychological Well-Being Scales (PWB, Ryff, 1989)</p>				

Summary of effectiveness of PPIs combined with other therapies (n = 5)

One study of an online PPI program (Manicavasagar et al., 2014) found significant improvements in intervention but no improvement in controls. This represented small effects for depression ($d = .30$), anxiety ($d = .017$), stress ($d = .16$) and well-being ($d = .20$). This same program was not replicated in the Burkhardt (2015) study which found no significant effects. Three studies evaluated well-being therapy in Italian schools with several small but significant subscale effects. One study (Ruini et al., 2006) found significant improvements in a physical well-being subscale compared to a CBT control ($d = .29$). Another study (Ruini et al., 2009) found that the only significant improvement remaining at 6-month follow-up compared to a treatment as usual (TAU) control was for SQ somatisation ($d = .052$) and SQ anxiety ($d = .047$). The final well-being study (Tomba et al., 2010) found the SQ friendliness subscale improved compared to an anxiety management control ($d = .26$). A consistent finding from the three studies of well-being therapy is that it is associated with improvements in physical indicators (e.g. physiological anxiety; somatisation; physical well-being).

Quality and methodological limitations of studies

Jadad scores classified all but one study (Kwok et al., 2016) in the poor range² (Table 1). This pattern was supported by low method appraisal scores (seven studies scored below 5)³. RCT design is problematic with this group, particularly within school settings. It is unlikely that many studies (e.g. Burkhardt et al., 2015; Kwok et al., 2016; Froh et al., 2008) could control for condition-control interaction and blinding was not possible. In Froh et al. (2009)

²Poor ≤ 9 ; Good >9 (Tuech et al., 2005).

³ The agreement between raters was high (Intra-class correlation coefficients [ICC] = .916) on the method appraisal scores. In fact, agreement was higher for the adapted method appraisal measure than for the Jadad ratings (ICC=.864).

contamination can be considered highly likely given that the intervention group discussed their experiences in the same classroom as the control. There was also difficulty with implementing interventions according to pre-specified plans due to the school schedule. Because these studies randomised within each school rather than for the overall sample, context or school effects could have influenced findings.

Baseline response was poor and attrition was high in five studies (Burckhardt et al., 2015; Manicavasagar et al., 2014; Suldo et al., 2014; Toback et al., 2016). Most studies comprise unrepresentative convenience samples and the small sample size of several studies prohibits statistical power to analyse complex multivariate relationships. The array of confounding variables considered by studies (Appendix 2) highlights a general lack of consideration here.

In Suldo et al (2014) randomisation was ineffective in balancing group differences. In Manicavasagar et al (2014) treatment dose was not maintained. Few studies have used active and non-treatment control groups. The one study that did (Froh et al., 2008) had mixed findings. That is, on all but school satisfaction, they found that the gratitude group improved more than the active control group but did not improve more than the TAU control on several variables (including gratitude, life satisfaction, reactions to aid, physical illness, positive affect, negative affect). Indeed, the control demonstrated more satisfaction over the past few weeks than the active control, whilst no such relationship existed between the gratitude and active control. Thus, it can be interpreted that a TAU control had greater effect on this measure of life satisfaction. Similarly, both TAU and gratitude groups had significantly less negative effect than the hassles group, but not each other.

Excluded Studies

A full list of studies included for full text review (n = 91) but that were subsequently excluded is available from the authors on request. Rashid, Anjum & Lennox (2006) was

excluded despite previous inclusion in the meta-analysis by Sin & Lyubomirsky (2009). We emailed the author, previous reviewers, and the University of Toronto school board. None could provide the original unpublished manuscript and as we had inadequate statistical information to work with, we excluded this study.

Discussion

This is one of the first studies to identify and summarise the child and adolescent RCT literature relating to Positive Psychology, to identify the array of outcome measures used, and confounding variables accounted for. Previously, systematic reviews of PPIs had collectively identified three studies. A total of 12 RCTs with a total sample of 1668 children and adolescents were identified by this review. Four of these found no significant effect compared to control while eight generally found small significant effects of PPIs.

All but one study was classified as poor. That one study (Kwok et al., 2016) involved a group intervention (gratitude / hope based) in mildly depressed adolescents and resulted in significant improvements (medium effect sizes for depression, life satisfaction, gratitude, and hope) compared to controls.

While it is unremarkable to suggest that the philosophy behind PPIs (to promote well-being, flourishing etc) is compatible with clinical groups (and likely regularly applied by clinicians already), surprisingly little evidence exists as to which interventions are effective clinically. Related to this point this review identified the first RCT relating to inpatient adolescents (Toback et al., 2016) which demonstrated (albeit tentatively, due to considerable flaws in condition-dose equivalency) that a brief character strengths intervention significantly increased self-esteem and self-efficacy in an inpatient psychiatric setting. More research is needed in clinical populations and future research could incorporate additional outcome measures likely important in this group (e.g. depression and anxiety measures). Kwok et al.'s (2016) study discussed above also is suggestive of clinical applicability.

Of the remaining studies, PPIs tended to be small except for one study. The finding of a large effect size in Positive Affect for Roth et al.'s (2017) multi-component PPI that remained at 7 week follow up (0.81) seems promising. Its use of techniques to encourage participant and

facilitator adherence, and in involving parents and booster sessions were commendable. Few PPIs have addressed implementation fidelity so comprehensively. However, research that addresses the methodological difficulties identified is warranted to replicate the large effect. Findings suggest that compared to controls, well-being therapy impacts physical symptoms more so than psychological or well-being indicators (Ruini et al., 2006; Ruini et al., 2009), a finding that may be relevant to clinical health populations. Ruini et al.'s (2006) study is increasingly cited as finding that Well-being therapy and CBT showed 'significant and comparable improvements' relating to psychological well-being and symptoms (Akhtar & Boniwell, 2010; Ruini et al., 2006). We would like to point out that this is not the case (see results).

The review identified that PPIs effected outcome differently. For example, improvement on subjective well-being measures such as life satisfaction did not necessarily correspond to improvement in psychopathology (e.g. Suldo, Savage & Mercer, 2014). This provides further evidence that these are distinguishable constructs (Suldo & Shaffer, 2008). Clinicians should therefore be mindful that targeting one domain may not lead to improvements in the other.

Tobach et al. (2016) found that those unaware of their character strengths prior to the intervention showed greater improvement than those who were aware. Similarly, those low in Positive Affect initially see greater improvements over time (Froh et al., 2009). This suggests that targeting PPIs accordingly could lead to greater benefits. However, further study is warranted to ensure this is not a 'regression to the mean' phenomenon.

Relatedly, there can be ceiling effects on measures of well-being, that is, a proportion of children will score at, or close to, the maximum score, particularly on shorter measures and in non-clinical community samples (as in Suldo, Mercer & Savage, 2014; Roth et al., 2017). These studies pre-screened participants to remove the highest scorers because there

would be no room for improvement. This is questionable because it may bias results towards positive findings and neglect potential adverse effects. Future research could avoid screening in this way or use measures with a greater scoring range to avoid removing such large proportions of the sample. Targeting those with greatest need (or room for improvement) may be appropriate and necessary, particularly if resource constraints are an issue, as they were in some studies (and likely would be if adopted generally or in public health programmes). It remains to be seen if the small or negligible effects seen in several studies are practically meaningful or important and whether a targeted or a broad public health or community approach is recommended. Those considering them are well advised to consider the cost-benefit as well as the context (for discussion see Ciarrochi et al., 2016).

PPIs have the potential to directly or indirectly create adverse outcomes (for example, Mauss, Tamir, Anderson, & Savino, 2011) and these should be considered in more detail. For example, Burckhardt et al., (2015) implement admirable safety protocols for their online intervention. However, no data is reported for adverse outcomes alluded to (encountered by the moderators or supporting clinical psychologist). Similarly, that effect sizes in the unexpected negative direction on environmental mastery were found for both groups in the Ruini et al. (2009) well-being study received limited attention (in fact these were the largest effect sizes found; $d = -.42$ in intervention; $-.40$ for control).

The time period for follow up is important as effects of PPIs (such as gratitude and character strength interventions) may take time to manifest and grow (e.g. Froh et al., 2008; Tobach et al., 2016). Based on Tobach et al., (2016) a minimum follow up of 3 months may be needed to detect effects. The question remains as to how long the effects of PPIs remain. There is evidence that they disappear compared to waitlist control at 6 months (as in Suldo, Mercer &

Savage, 2014). However, time may introduce potential confounds in this group, i.e. because developmental changes can be rapid and variable during adolescence. This point underscores the logic for exploring and adjusting for demographic factors (such as age) not just at baseline but also at follow up points. Future research should consider the delivery format and setting. The acceptability of content and format of PPI's to different age ranges is likely relevant given the Manicavasagar et al., 2014 study and previous findings that those who self-select do better (Sin & Lyubovitsky, 2009). Most identified studies were group based. Research is needed to assess whether individual versus group format leads to differential outcomes.

This review has several limitations. As with any review, it is limited by the interventions and terms specified. Ours was a deliberate strategy, chosen to address difficulties with the definition of what constitutes a PPI. Further, we reviewed previous systematic reviews and studies in the area and consulted review authors to confirm that we were not omitting PPI interventions. However, it is likely that this review does not encompass all PPI interventions. We hope that as understanding develops, more interventions can be added to our protocol and future reviews updated.

This review focussed on interventions aimed at the individual child. Future reviews could assess interventions aimed at broader context (for example: community, family, parents) and consider indirect or cascade effects. Additionally, the review only concerned RCTs, and neglects other types of evidence, particularly qualitative methods that can add depth in meaning (Abushaba & Woelfel, 2003). Evaluating methods that access the perspective and experience of young people (for example, Coverdale & Long, 2015) may overcome criticisms relating to the neglect of context within the PPI literature (Ciarrochi et al., 2016).

Despite these limitations our review has considerable strengths, including its non-reliance on limiting terms (i.e. positive psychology), the use of a dual reviewers to evaluate studies and use of an established appraisal tool (which demonstrated high levels of inter-rater reliability).

What variables are, and should be, measured and a general lack of consistency in this regard is a considerable criticism levelled at the PPI literature (Chodkiewicz & Boyle, 2017). As a first step in addressing this, this review provides a summary of the variables currently measured and controlled for in RCTs (Appendix 2) that should be considered when implementing and evaluating interventions. Further, this is the first time the array of measures used within the PPI literature has been identified (Table 1). It is difficult to isolate the mechanisms of change in multi-component interventions (Boulton, 2016), and more research to understand relative contributions is required.

When designing, implementing and evaluating PPIs in the future several points warrant consideration. In summary, the format (e.g. structured or unstructured; choice as to how participants can engage with content); duration of intervention (particularly when online) and of follow up (minimum 3-6 month follow up is suggested); whether participants receive a reward or not; if parental consent is required (reduces participation and creates a sampling issue); whether content is adjusted to different age ranges (e.g. early, middle or late adolescents); whether an intervention is elective (as opposed to part of a mandated school program); how treatment dose will be maintained and managed (particularly in online interventions).

Additionally, the resource intensity and sustainability (e.g. reliance on school or teachers; time demand on the participants increases attrition; and/or decreases adherence), and the method of data collection (the social desirability effect is likely to be particularly relevant in

classroom settings where the facilitator is collecting data) were specifically identified as important to the process.

Conclusion

The evidence base has grown from 3 to 12 RCTs. Considerable methodological limitations and poor study quality means that it remains to be seen if effects can generalise outside of study settings. To overcome its 'bad science' label (Frawley, 2015, p. 66), PPI literature needs to address limitations identified in this review or, in contrast to its underlying philosophy, will fail to flourish.

References

- Abushaba, R., Woelfel, M. L. (2003). Qualitative vs quantitative methods: Two opposites that make a perfect match. *The American Dietetic Association*, 103(5), 566-569.
- Akhtar, M., & Boniwell, I. (2010). Applying positive psychology to alcohol-misusing adolescents. *Groupwork*, 20(3), 6-31.
- Amini, S., Namdari, K., & Kooshki, H. M. (2016). The effectiveness of positive psychotherapy on happiness and gratitude of female students. *International Journal of Educational and Psychological Researches*, 2(3), 163.
- Beck, A. T. (Ed.). (1979). *Cognitive therapy of depression*. Guilford press.
- Bird, J. M., & Markle, R. S. (2012). Subjective Well- Being in School Environments: Promoting Positive Youth Development Through Evidence- Based Assessment and Intervention. *American Journal of Orthopsychiatry*, 82(1), 61-66.
- Boiler L., Haverman M., Westerhof G. J., Riper H., Smit F., Bohlmeijer E., et al. (2013). Positive psychology interventions: a meta-analysis of randomised controlled studies. *BMC Public Health*, 13, 119.
- Burckhardt, R., Manicavasagar, V., Batterham, P. J., & Hadzi-Pavlovic, D. (2016). A randomized controlled trial of strong minds: A school-based mental health program combining acceptance and commitment therapy and positive psychology. *Journal of school psychology*, 57, 41-52.
- Burckhardt, R., Manicavasagar, V., Batterham, P. J., Miller, L. M., Talbot, E., & Lum, A. (2015). A web-based adolescent positive psychology program in schools: randomized controlled trial. *Journal of medical Internet research*, 17(7), e187.

- Burns, D. S., Robb, S. L., & Haase, J. E. (2009). Exploring the feasibility of a therapeutic music video intervention in adolescents and young adults during stem-cell transplantation. *Cancer nursing*, 32(5), E8-E16.
- CASP Critical Appraisal skills Programme. (2016). Critical Appraisal Skills Systematic Review Checklist. Available from, <http://www.casp-uk.net/casp-tools-checklists>.
- Chodkiewicz, A. R., & Boyle, C. (2017). Positive psychology school- based interventions: A reflection on current success and future directions. *Review of Education*, 5(1), 60-86.
- Ciarrochi, J., Atkins, P. W., Hayes, L. L., Sahlra, B. K., & Parker, P. (2016). Contextual positive psychology: policy recommendations for implementing positive psychology into schools. *Frontiers in psychology*, 7.
- Connolly, J. F. & Willock, J. (2009). *Coping predicts Quality of Life in a group of Irish women with iatrogenic Hepatitis c infection*. Retrieved 23-02-2013 from, <http://eresearch.qmu.ac.uk/606/1/606.PDF>
- Connolly, J. F., Connolly, P. S, & Chisholm, V. (2013). *Stress interventions in universities: A systematic review*. Report. Retrieved 20-04-2017 from, <http://eresearch.qmu.ac.uk/3257/>
- Coverdale, G. E., & Long, A. F. (2015). Emotional well-being and mental health: an exploration into health promotion in young people and families. *Perspectives in public health*, 135(1), 27-36.
- Frawley, A. (2015) Happiness research: a review of critiques, *Sociology Compass*, 9(1), 62–77. doi:10.1111/soc4.12236
- Freire, T., Teixeira, A., Silva, E., & Matias, G. P. (2015). Interventions for youth depression: From symptom reduction to well-being and optimal functioning. *Journal of Behavior, Health & Social Issues*, 6(2), 9-19.

- Froh, J. J., Kashdan, T. B., Ozimkowski, K. M., & Miller, N. (2009). Who benefits the most from a gratitude intervention in children and adolescents? Examining positive affect as a moderator. *The Journal of Positive Psychology, 4*(5), 408-422.
- Froh, J. J., Sefick, W. J., & Emmons, R. A. (2008). Counting blessings in early adolescents: An experimental study of gratitude and subjective well-being. *Journal of school psychology, 46*(2), 213-233.
- Gillham, J. E., Hamilton, J., Freres, D. R., Patton, K., & Gallop, R. (2006). Preventing depression among early adolescents in the primary care setting: A randomized controlled study of the Penn Resiliency Program. *Journal of abnormal child psychology, 34*(2), 195-211.
- Harrison, A., Al-Khairulla, H., & Kikoler, M. (2016). The feasibility, acceptability and possible benefit of a positive psychology intervention group in an adolescent inpatient eating disorder service. *The Journal of Positive Psychology, 11*(5), 449-459.
- Ho, H. C., Mui, M., Wan, A., Ng, Y. L., Stewart, S. M., Yew, C., ... & Chan, S. S. (2016). Happy Family Kitchen: A community-based research for enhancing family communication and well-being in Hong Kong. *Journal of Family Psychology, 30*(6), 752.
- Ho, H. C., Mui, M., Wan, A., Ng, Y. L., Stewart, S. M., Yew, C., ... & Chan, S. S. (2016). Happy Family Kitchen II: a cluster randomized controlled trial of a community-based positive psychology family intervention for subjective happiness and health-related quality of life in Hong Kong. *Trials, 17*(1), 367.
- Jadad AR, Moore RA, Carroll D, et al. (1996). Assessing the quality of reports of randomized clinical trials: is blinding necessary? *Control Clin Trials, 17*, 1–12.

- Jaser, S. S., Patel, N., Linsky, R., & Whittemore, R. (2014b). Development of a positive psychology intervention to improve adherence in adolescents with type 1 diabetes. *Journal of Pediatric Health Care*, 28(6), 478-485.
- Jaser, S. S., Patel, N., Rothman, R. L., Choi, L., & Whittemore, R. (2014a). Check it! A randomized pilot of a positive psychology intervention to improve adherence in adolescents with type 1 diabetes. *The Diabetes Educator*, 40(5), 659-667.
- Johnson, J., Gooding, P. A., Wood, A. M., Fair, K. L., & Tarrier, N. (2013). A therapeutic tool for boosting mood: the broad-minded affective coping procedure (BMAC). *Cognitive therapy and research*, 37(1), 61-70. doi:10.1007/s10608-012-9453-8
- Kellner, R. (1987). A symptom questionnaire. *J Clin Psychiatry*, 48(7), 268-274.
- Kim, S., Kverno, K., Lee, E. M., Park, J. H., Lee, H. H., & Kim, H. L. (2006). Development of a music group psychotherapy intervention for the primary prevention of adjustment difficulties in Korean adolescent girls. *Journal of Child and Adolescent Psychiatric Nursing*, 19(3), 103-111.
- Kini, P., Wong, J., McInnis, S., Gabana, N., & Brown, J. W. (2016). The effects of gratitude expression on neural activity. *NeuroImage*, 128, 1-10.
- Krabbenborg, M. A., Boersma, S. N., & Wolf, J. R. (2013). A strengths based method for homeless youth: Effectiveness and fidelity of Houvast. *BMC public health*, 13(1), 359.
- Kwok, S. Y., Gu, M., & Cheung, A. P. (2017). A longitudinal study of the role of children's altruism and forgiveness in the relation between parental aggressive discipline and anxiety of preschoolers in China. *Child Abuse & Neglect*, 65, 236-247.
- Kwok, S. Y., Gu, M., & Kit, K. T. K. (2016). Positive psychology intervention to alleviate child depression and increase life satisfaction: A randomized clinical trial. *Research on social work practice*, 26(4), 350-361.

- Lazarus, R. (2003b). The Lazarus manifesto for positive psychology and psychology in general. *Psychol. Inq.* 13, 173–189. Doi:10.1207/S15327965PLI1
- Macaskill, A. (2016, September). Review of Positive Psychology Applications in Clinical Medical Populations. In *Healthcare* (Vol. 4, No. 3, p. 66). Multidisciplinary Digital Publishing Institute.
- Manicavasagar, V., Horswood, D., Burckhardt, R., Lum, A., Hadzi-Pavlovic, D., & Parker, G. (2014). Feasibility and effectiveness of a web-based positive psychology program for youth mental health: randomized controlled trial. *Journal of medical Internet research*, 16(6), e140.
- Martínez-Martí, M. L., Avia, M. D., & Hernández-Lloreda, M. J. (2010). The effects of counting blessings on subjective well-being: A gratitude intervention in a Spanish sample. *The Spanish Journal of Psychology*, 13(02), 886-896.
- Mauss, I. B., Tamir, M., Anderson, C. L., & Savino, N. S. (2011). Can seeking happiness make people unhappy? Paradoxical effects of valuing happiness. *Emotion*, 11(4), 807.
- Meyers, M. C., van Woerkom, M., & Bakker, A. B. (2013). The added value of the positive: A literature review of positive psychology interventions in organizations. *European Journal of Work and Organizational Psychology*, 22(5), 618-632.
- Owens, R. L. & Patterson, M. M. (2013) Positive Psychological Interventions for Children: A Comparison of Gratitude and Best Possible Selves Approaches, *The Journal of Genetic Psychology*, 174:4, 403-428, DOI: 10.1080/00221325.2012.697496
- Pawelski, J. O. (2016). Defining the ‘positive’ in positive psychology: Part II. A normative analysis. *The Journal of Positive Psychology*, 11(4), 357-365.
- Proyer, R. T., Gander, F., Wellenzohn, S., & Ruch, W. (2015). Strengths-based positive psychology interventions: a randomized placebo-controlled online trial on long-term

- effects for a signature strengths-vs. a lesser strengths-intervention. *Frontiers in psychology*, 6, 456.
- Rashid, T., Anjum, A., & Lennox, C. (2006). Positive psychotherapy for middle school children. *Unpublished manuscript, Toronto District School Board. Results presented in: Rashid, T., & Anjum, A.(2008). Positive psychotherapy for young adults and children. In JRZ Abela & BL Hankin (Eds.), Handbook of depression in children and adolescents, 250-287.*
- Roth, R. A., Suldo, S. M., & Ferron, J. M. (2017). Improving Middle School Students' Subjective Well-Being: Efficacy of a Multicomponent Positive Psychology Intervention Targeting Small Groups of Youth. *School Psychology Review*.
- Ruini, C., Belaise, C., Brombin, C., Caffo, E., & Fava, G. A. (2006). Well-being therapy in school settings: a pilot study. *Psychotherapy and Psychosomatics*, 75(6), 331-336.
- Ruini, C., & Fava, G. A. (2009). Well- being therapy for generalized anxiety disorder. *Journal of Clinical Psychology*, 65(5), 510-519.
- Ruini, C., & Fava, G. A. (2012). Role of Well-Being Therapy in Achieving a Balanced and Individualized Path to Optimal Functioning. *Clinical Psychology & Psychotherapy*, 19(4), 291–304. <https://doi.org/10.1002/cpp.1796>
- Ryff, C. D. (1989). Happiness is everything, or is it? Explorations on the meaning of psychological well-being. *Journal of personality and social psychology*, 57(6), 1069.
- Ryff, C. D. (1995). Psychological well-being in adult life. *Current directions in psychological science*, 4(4), 99-104.
- Saltzman, W. R. (2016). The FOCUS family resilience program: an innovative family intervention for trauma and loss. *Family process*, 55(4), 647-659.

- Schueller, S. M., Kashdan, T. B., & Parks, A. C., (2014). Synthesizing positive psychological interventions: Suggestions for conducting and interpreting meta-analyses. *International Journal of Well-being*, 4(1), 91-98. doi:10.5502/ijw.v4i1.5
- Seligman, M. E. P. Randal M. Ernst , Jane Gillham , Karen Reivich & Mark. Linkins (2009) Positive education: positive psychology and classroom interventions, *Oxford Review of Education*, 35:3, 293-311, DOI: 10.1080/03054980902934563
- Seligman, M. E. P. (2002). *Authentic happiness* . New York, NY: Simon & Schuster.
- Sezici, E., Ocakci, A. F., & Kadioglu, H. (2017). Use of Play Therapy in Nursing Process: A Prospective Randomized Controlled Study. *Journal of Nursing Scholarship*, 49(2), 162-169.
- Shoshani, A., Mifano, K., & Czamanski-Cohen, J. (2016). The effects of the make a wish intervention on psychiatric symptoms and health-related quality of life of children with cancer: a randomised controlled trial. *Quality of Life Research*, 25(5), 1209-1218.
- Sin, N. L., & Lyubomirsky, S. (2009). Enhancing well- being and alleviating depressive symptoms with positive psychology interventions: A practice- friendly meta-analysis. *Journal of clinical psychology*, 65(5), 467-487.
- Stewart-Brown S, Tennant A, Tennant R, Platt S, Parkinson J, Weich S. (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 Qual Life Outcomes*, 7:15.
- Suldo, S. M., & Shaffer, E. J. (2008). Looking beyond psychopathology: The dual-factor model of mental health in youth. *School Psychology Review*, 37(1), 52.
- Suldo, S. M., Hearon, B. V., Bander, B., McCullough, M., Garofano, J., Roth, R. A., & Tan, S. Y. (2015). Increasing elementary school students' subjective well-being through a

- classwide positive psychology intervention: results of a pilot study. *Contemporary School Psychology*, 19(4), 300-311.
- Suldo, S. M., Savage, J. A., & Mercer, S. H. (2014). Increasing middle school students' life satisfaction: Efficacy of a positive psychology group intervention. *Journal of happiness studies*, 15(1), 19-42.
- Sundar, S., Qureshi, A., & Galiatsatos, P. (2016). A positive psychology intervention in a Hindu community: the pilot study of the Hero Lab Curriculum. *Journal of religion and health*, 55(6), 2189-2198.
- Tak, Y. R., Kleinjan, M., Lichtwarck-Aschoff, A., & Engels, R. C. (2014). Secondary outcomes of a school-based universal resiliency training for adolescents: a cluster randomized controlled trial. *BMC public health*, 14(1), 1171.
- Tak, Y. R., Lichtwarck-Aschoff, A., Gillham, J. E., Zundert, R. M., & Engels, R. C. (2016). Universal school-based depression prevention 'Op Volle Kracht': a longitudinal cluster randomized controlled trial. *Journal of abnormal child psychology*, 44(5), 949-961.
- Toback, R. L., Graham-Bermann, S. A., & Patel, P. D. (2016). Outcomes of a Character Strengths-Based Intervention on Self-Esteem and Self-Efficacy of Psychiatrically Hospitalized Youths. *Psychiatric Services*, 67(5), 574-577.
- Tomba, E., Belaise, C., Ottolini, F., Ruini, C., Bravi, A., Albieri, E., ... & Fava, G. A. (2010). Differential effects of well-being promoting and anxiety-management strategies in a non-clinical school setting. *Journal of Anxiety Disorders*, 24(3), 326-333.
- White, C. A. (2016). *Meta-analyses of positive psychology interventions on well-being and depression: reanalyses and replication* (M. A. dissertation, University of British Columbia).

Wood, L., Theron, L., & Mayaba, N. (2012). 'Read me to resilience': Exploring the use of cultural stories to boost the positive adjustment of children orphaned by AIDS.

African Journal of AIDS Research, 11(3), 225-239.

Chapter 2.

Empirical Paper

Measurement of child emotional health using the SDQ and norm development in a representative national sample.⁴

⁴ An article (Connolly, Jackson, & Saville, 2017) based on this chapter was submitted to the Journal of Clinical Child and Adolescent Psychology, the guidelines of which follow.

Author guidelines

Journal of Clinical Child and Adolescent Psychology

Instructions for authors

Thank you for choosing to submit your paper to us. These instructions will ensure we have everything required so your paper can move through peer review, production and publication smoothly. Please take the time to read and follow them as closely as possible, as doing so will ensure your paper matches the journal's requirements. For general guidance on the publication process at Taylor & Francis please visit our [Author Services website](#).

This journal uses ScholarOne Manuscripts (previously Manuscript Central) to peer review manuscript submissions. Please read the [guide for ScholarOne authors](#) before making a submission. Complete guidelines for preparing and submitting your manuscript to this journal are provided below.

Aims and Scope. The *Journal of Clinical Child and Adolescent Psychology* (*JCCAP*) is the official journal for the Society of Clinical Child and Adolescent Psychology, American Psychological Association, Division 53. It publishes original contributions on the following topics: (1) development and evaluation of assessment and intervention techniques for use with clinical child and adolescent populations; (2) development and maintenance of clinical child and adolescent problems; (3) cross-cultural and sociodemographic issues that have a clear bearing on clinical child and adolescent psychology theory, research, or practice; and (4) training and professional practice in clinical child and adolescent psychology as well as child advocacy. Manuscripts that discuss theoretical and/or methodological issues on topics pertinent to clinical child and adolescent psychology also are considered. Authors need not be members of Division 53 to submit articles to *JCCAP*.

There are several criteria that increase the likelihood that a manuscript will be favorably evaluated in *JCCAP*: (1) The paper reflects a substantive advance in our understanding of clinical child and adolescent psychology. (2) The paper is of such importance that it likely will influence an area of research. (3) The paper presents new ideas or creative methods. (4) The paper offers theoretically-driven hypotheses. (5) Multiple measures, informants, or procedures are used to collect data. (6) Sophisticated methodologies are carefully employed. (7) Longitudinal methods are used. (8) Data are rigorously and appropriately analyzed. (9) The implications of the findings for clinical child and adolescent psychology are well articulated.

Style of Manuscripts. Manuscripts should be prepared according to the guidelines in the *Publication Manual of the American Psychological Association* (6th edition; see www.apastyle.com). Typing instructions, including format, organization, and the preparation of figures, tables, and references appear in the Manual. Manuscripts may be submitted as Regular Articles, Brief Reports, or Future Directions. A Regular Article may not exceed 11,000 words (i.e., 35 pages), including references, footnotes, figures, and tables. Brief Reports include empirical research that is soundly designed, but may be of specialized interest or narrow focus. Brief Reports may not be submitted in part or whole to another journal of general circulation. Brief Reports may not exceed 4,500 words for text and references. These limits do not include the title page, abstract, author note, footnotes, tables, and figures. Manuscripts that exceed these page limits and that are not prepared according to the guidelines in the Manual will be returned to authors without review. Future Directions submissions are written by leading scholars within the field. These articles provide a brief summary of important advances that are needed within a specific research or practice area pertinent to clinical child and adolescent psychology. Future Directions submissions are by invitation only and undergo peer review.

All Regular Article and Brief Report submissions must include a title of 15 words or less that identifies the developmental level of the study participants (e.g., children, adolescents, etc.). *JCCAP* uses a structured abstract format. For studies that report randomized clinical trials or meta-analyses, the abstract also must be consistent with the guidelines set forth by CONSORT or MARS, respectively. The Abstract should include up to 250 words, presented in paragraph form. The Abstract should be typed on a separate page (page 2 of the manuscript), and must include each of the following label sections: 1) Objective (i.e., a brief statement of the purpose of the study); 2) Method (i.e., a detailed summary of the participants, N, age, gender, ethnicity, as well as a summary of the study design, measures, and procedures; 3) Results (i.e., a detailed summary of the primary findings that clearly articulate comparison groups (if relevant); 4) Conclusions (i.e., a description of the research and clinical implications of the findings). Avoid abbreviations, diagrams, and reference to the text in the abstract. A list of up to five keywords that describe the central themes of the manuscript should be included below the abstract on page 2. *JCCAP* will scrutinize manuscripts for a clear theoretical framework that supports central study hypotheses.

In addition, a clear developmental rationale is required for the selection of participants at a specific age. The Journal is making diligent efforts to insure that there is an appropriately detailed description of the sample, including a) the population from which the sample was drawn; b) the number of participants; c) age, gender, ethnicity, and SES of participants; d) location of sample, including country and community type (rural/urban), e) sample identification/selection; f) how participants were contacted; g) incentives/rewards; h) parent consent/child assent procedures and rates; i) inclusion and exclusion criteria; j) attrition rate. The Discussion section should include a comment regarding the diversity and generality (or lack thereof) of the sample. The Measures section should include details regarding item content and scoring as well as evidence of reliability and validity in similar populations.

All manuscripts must include a discussion of the clinical significance of findings, both in terms of statistical reporting and in the discussion of the meaningfulness and clinical relevance of results. Manuscripts should a) report means and standard deviations for all variables, b) report effect sizes for analyses, and c) provide confidence intervals wherever appropriate (e.g., on figures, in tables), particularly for effect sizes on primary study findings. In addition, when reporting the results of interventions, authors should include indicators of clinically significant change. Authors may use one of several approaches that have been recommended for capturing clinical significance, including (but not limited to) the reliable change index (i.e., whether the amount of change displayed by a treated individual is large enough to be meaningful, the extent to which dysfunctional individuals show movement to the functional distribution).

All manuscripts should conform to the criteria listed in Table 1 of the 2008 APA Publications and Communications Board Working Group on Journal Article Reporting Standards (published in *American Psychologist*). These reporting standards apply to all empirical papers. In addition, *JCCAP* requires that reports of randomized clinical trials conform to CONSORT reporting standards (<http://www.consort-statement.org/index.aspx?o=2965>), including the submission of a flow diagram and checklist. Nonrandomized clinical trials must conform to TREND criteria (see http://www.cdc.gov/trendstatement/docs/AJPH_Mar2004_Trendstatement.pdf) and meta-analyses should conform to MARS standards (see Table 4 in 2008 *American Psychologist* article).

Peer Review Process. *JCCAP* uses a two-tiered peer review process. All manuscripts are evaluated by the Editor or at least one Associate Editor to determine whether the manuscript is likely to make a significant impact to the scientific literature in clinical child and adolescent psychology. A significant proportion of manuscripts submitted to *JCCAP* are not

selected for peer review, and this decision is typically made within 2–4 days after submission.

Manuscripts selected for peer review will undergo a masked review procedure. To prepare manuscripts for masked review, authors' names and affiliations should not appear on the title page or elsewhere in the manuscript file (they can be entered into the system and placed on a separate page in the cover letter file). Footnotes identifying the authors should be typed on a separate page and submitted in the cover letter file. Authors should make every effort to ensure that the manuscript file itself contains no clues to their identities. Manuscripts that do not comply with these instructions will be returned to the authors without review.

Publication Policies. Authors are responsible for all statements made in their work and for obtaining permission from copyright owners to use a lengthy quotation (500 words or more) or to reprint or adapt a table or figure published elsewhere. Authors should write to both author(s) and publisher of such material to request nonexclusive world rights in all language for use in print and nonprint forms of the article and in future editions. This applies to direct reproduction as well as "derivative reproduction" (where you have created a new figure or table which derives substantially from a copyrighted source). Authors are required to sign an agreement for the transfer of copyright to the publisher. All accepted manuscripts, artwork, and photographs become the property of the publisher.

Submitting Manuscripts. *JCCAP* receives all manuscript submissions electronically via its ScholarOne Manuscripts site located at: <http://mc.manuscriptcentral.com/jccap>. ScholarOne Manuscripts allows for rapid submission of original and revised manuscripts, and facilitates the review process and internal communication between authors, editors, and reviewers via a web-based platform. ScholarOne technical support can be accessed at

<http://scholarone.com/services/support/>. If you have any other requests, please contact the journal's editor, Andres Des Los Reyes, at adlr@umd.edu.

Color Reproduction. Color art will be reproduced in color in the online publication at no additional cost to the author. Color illustrations will also be considered for print publication; however, the author will be required to bear the full cost involved in color art reproduction. Please note that color reprints can only be ordered if print reproduction costs are paid. **Print Rates:** \$900 for the first page of color; \$450 per page for the next three pages of color. A custom quote will be provided for articles with more than four pages of color. Art not supplied at a minimum of 300 dpi will not be considered for print.

Proofs and Reprints. Page proofs are sent to the designated author using Taylor & Francis' Central Article Tracking System (CATS). They must be carefully checked and returned within 48 hours of receipt. Authors for whom we receive a valid email address will be provided an opportunity to purchase reprints of individual articles, or copies of the complete print issue. These authors will also be given complimentary access to their final article on *Taylor & Francis Online*.

Open Access. Taylor & Francis Open Select provides authors or their research sponsors and funders with the option of paying a publishing fee and thereby making an article fully and permanently available for free online access – *open access* – immediately on publication to anyone, anywhere, at any time. This option is made available once an article has been accepted in peer review. [Full details of our Open Access programme.](#)

Search Engine Optimization. Search Engine Optimization (SEO) is a means of making your article more visible to anyone who might be looking for it. Please consult our guide [here](#).

Submission to the Journal of Clinical Child and Adolescent Psychology

Measurement of child emotional health using the SDQ and norm development in a representative national sample.

Connolly¹, J. F, Jackson², M., & Saville³, C. W. N

^{1,2,3}North Wales Clinical Psychology Programme (NWCPP), School of Psychology, Bangor University, Wales, UK;

Word count (exc. abstract, reference list, figures & tables): 6,217

*Correspondence should be addressed to Dr John F. Connolly, North Wales Clinical Psychology Programme (NWCPP), School of Psychology, Brigantia Building, Bangor, Gwynedd, North Wales, United Kingdom, LL57 2DG (e-mail: john.connolly@wales.nhs.uk)

Abstract

Objective: Despite the ubiquity of the Strengths and Difficulties Questionnaire (SDQ) as a screening measure of child mental health/psychopathology, the literature identifies several ambiguities which this study investigates, including a requirement for further psychometric analysis, analysis of functioning between groups, and provision of up to date norms.

Method: The performance of the parent SDQ (4 to 12 year olds version) was analysed in a stratified national sample (n=4828). Confirmatory factor analysis assessed competing models of the SDQ's structure, and measurement invariance was examined for important demographic variables. Alternate psychometric indicators (Omega versus alpha) and differential multigroup confirmatory factor analyses (categorical versus continuous) were evaluated. Updated normative bandings were created.

Results: Confirmatory factor analysis best supported a five-factor model. Subscale internal reliability ranged from $\alpha = 0.64$ (conduct problems) to $\alpha = 0.79$ (hyperactivity). The method of analyses influenced differences within reliability and model acceptance considerations. The SDQ demonstrated invariance (configural, metric, scalar) for gender, parent status, and SES. Differences were identified compared to previous norms.

Conclusions: This study provides one of the largest and most up to date normative data samples identifying psychometric properties of the SDQ. Contrary to some previous suggestion, findings recommend that it is appropriate to 1) use the subscales of the SDQ in screening and 2) interpret the 5-factor rather than the 3-factor structure. Measurement invariance further supports its use. Current normative bands were updated as they underestimate at risk children (by 1-2 points on total or subscale score). Contributions to theory and practice are considered.

Keywords: Strengths and Difficulties Questionnaire; factor structure; invariance; norms

Introduction

Countries around the world have identified a need for collecting and monitoring population level data relating to child mental health (Bayer, Ukomunne, Lucas, Wake, Scalzo, & Nicholson, 2011; Statistics Canada, 2000). The importance of up to date and culturally calibrated norms has been established by the fact that norms vary widely across cultures (Moriwaki & Kamio, 2014) and because prevalence rates of certain mental health problems can change over a reasonably short period of time (Green, McGinnity, Meltzer, Ford, & Goodman, 2004). In the UK, the available data suggest that 10% of those between 5-16 years old have a diagnosable mental health disorder (Green et al., *ibid.*). Elsewhere, there are more recent estimates that 20% of children are affected by internalised (emotional) or externalised (behavioural) clinical level problems (Bayer et al., 2011).

Over the past two decades one instrument that has been increasingly used to screen for childhood mental health problems at population, community, and clinical level is the Strengths and Difficulties Questionnaire (SDQ, Goodman, 1997). This 25-item screening questionnaire consists of 5 subscales (conduct problems, emotional symptoms, inattention /hyperactivity, peer problems, and prosocial behaviour). It was designed to map on to DSM-IV and ICD-10 classifications and can be completed by parents and / or teachers (for 3-16 year olds) or in self-report format (by 11-16 year olds). There are several advantages in using the SDQ, including that it is brief, freely available, demonstrates good user acceptability, discriminant validity and performs well against other longer measures such as the Child Behaviour Checklist (Goodman & Scott, 1999; Goodman, 2001; Achenbach, Becker, Dopfner, Heiervang, Steinhausen, & Rothenberger, 2008).

Despite the large body of research on the SDQ, several important questions remain to be answered and these are discussed in more detail in the following section. In summary, there are few studies relating to the psychometrics of the parent version in 4-12 year olds, findings

relating to reliability and construct validity are conflicting, and the SDQ literature has been affected by several methodological issues. Additionally, the norms provided by the publisher (available at <http://www.sdqinfo.org/>) for the UK are based on data from the year 2000 (Meltzer, Gatward, Goodman, & Ford, 2000). Because up to date normative bandings are lacking, it is possible that inaccurate cut-off points are currently being used to screen for those at risk. This study seeks to investigate these issues.

Psychometric properties of the Strengths and Difficulties Questionnaire

Reliability

A considerable amount of research has assessed the reliability of the SDQ and this consistently shows inadequate reliability of subscales, especially for the peer problems and conduct problems.

While a 2010 review (Stone, Otten, Rutger, Engels, Vermulst, & Jansens, 2010) concluded sufficient psychometric properties of the 5-factor model there were few studies relating to the parent version (4-12 year olds) and conflicting findings as to reliability and construct validity. The parent reported SDQ appeared to perform worse than the teacher version, and internal consistency (Cronbach's alpha, α) was higher in teacher ratings when compared to parents' scores which were below the 0.70 benchmark for conduct (0.58) and peer problems (0.53), emotional symptoms (0.66) and prosocial behaviour (0.67) subscales.

A notable difficulty with interpreting these combined (weighted mean) internal consistency scores is that 21 of the 48 studies in the Stone review contained children above the cut-off age range (i.e. above 12 years) and it was not clear to what degree these were omitted. The most recent systematic review of the properties of the SDQ (Kersten, Czuba, McPherson, Dudley, Elder, ... et al., 2016) identified even poorer alpha subscale reliabilities and concluded that this was "an indication of inadequate internal consistency of those subscales" (p. 70).

Regarding reliability, alpha is one of the most pervasive reliability indices (Dunn, Baguley, & Brunnsden, 2013) but researchers tend not to consider its significant limitations (Sijtsma, 2009; Graham, 2009). Namely, it relies on a tau- equivalence model, whose assumptions are rarely met in practice, and it can under (or over) estimate reliability (Sijtsma, 2009; Dunn, Baguley, & Brunnsden, 2013). Practical alternatives (such as McDonald's Omega) have therefore been strongly recommended (Dunn, Baguley, & Brunnsden, 2013; Revelle & Zinbarg, 2008; Zinbarg & Alden, 2015).

Two recent studies (Stone, Jansens, Vermulst, Van der Maten, Engels, & Otten 2015; Stone, Otten, Engels, Vermulst, & Jansens 2010) found that this McDonald's Omega demonstrated consistently higher reliability than alpha when testing the SDQ. Indeed there is increasing credence to the argument presented by the authors that when data is non-normally distributed or there are few response categories (as in the SDQ) that alternative measures such as Omega are more appropriate.

Construct validity

Only four (of eight) studies in Stone et al.'s 2010 review conducted Confirmatory (CFA) rather than exploratory Factor Analysis (EFA) or provided factor loadings for interpretation (Dickey & Blumberg, 2004; Van Leeuwen, Meerschaert, Bosmans, De Medts, & Braet, 2006; Van Roy, Veenstra, Clench-Aas, 2008; Dickey & Blumberg, 2001; Hill & Hughes, 2007; Mellor & Stokes, 2007) and these found varying degrees of support for different factorial models. This represents a significant methodological difficulty in the area, that construct validity findings presented in many studies reflect EFA rather than CFA, which is inappropriate when analysing the pre- specified theory driven factor model.

Since the 2010 review several additional studies continue to question the five-factor model (Boman et al., 2016; Mc Crory & Layte, 2015; Kersten et al., 2016). In Sweden, Boman et al.

(2016) found that the parent SDQ 5-factor model was not supported in the total sample or in subgroup analyses by gender, ethnicity, or education level. Notably, the methods in the latter study did not produce important standard fit statistics (CFI, RMSEA etc) and used Exploratory Factor Analysis (EFA). Mc Crory and Layte's (2015) analysis also could not confirm the five-factor structure in a sample of 8568 nine year olds in Ireland. Goodman et al., (2010) suggest that using a three-subscale division (internalising, externalising, prosocial) may be more appropriate in general population samples. Similarly, some studies suggest using the total difficulties score rather than relying on the five subscales scores is preferable.

Measurement invariance

It is feasible that individuals with different demographic and social characteristics may interpret items on a measure differently and these factors have an important influence on health and well-being (Longevity Science Advisory Panel, 2012). A key research question is therefore how the properties of the SDQ measure might vary across these sub groups and consequently impact (or invalidate) subsequent interpretation.

A key benefit of psychometrically sound and validated clinical and research measures is that they can allow comparison across diverse groups and settings. In order for such comparisons to be valid, a measure needs to demonstrate factorial invariance (that its items and constructs hold the same meaning for individuals across these groups or subpopulations). This stability of measurement is a cornerstone of psychometrically sound measurement instruments. In contrast to this accepted principle, measures are often developed and tested based on homogenous populations or overall samples without consideration of variance across subgroups (Ortuno-Sierra, Fonseca-Pedrero, Aritio-Solana, Velasco...et al., 2015; Limbers, Daniel, Newman & Varni, 2009).

Recently, studies examining invariance in the SDQ have found significant differences between cultures (Ortuno-Sierra et al., *ibid.*; Mieloo, Raat, van Oort, Bevaart, Vogel, & Donker, 2012; Stevanovic, Urban, Atilola, Vostanis, Balhara... & Petrov, 2015). One cross cultural study found factor structures were not equivalent across seven countries (Stevanovic et al., 2015). Another study across five European countries involving 3012 adolescents suggests that the five-factor model should be modified (Ortuno-Sierra et al., 2015).

Generally, there has been limited research evaluating invariance of the SDQ parent version (4-12) and rarely are potentially important demographic factors such as socio-economic status, or gender investigated. Further, to the author's knowledge, no study has tested invariance according to responding parents gender. This is surprising given that invariance here is necessary to conclude that both mother and father ratings provide reliable and similar information (Chiorri, Hall, Casely-Hayford, & Malmberg, 2016). One study has assessed measurement invariance according to parent status in dyadic pairs (comparing responses of both parents of the same child [Chiorri et al. *ibid.*]). Comparing mothers and fathers generally is very different from comparing parents of the same child. In fact it removes within family factors that may influence responding. Our study does not aim to compare parents in such a way. Because only one parent completed data about their child in this study, it assesses the experience of being a parent generally, rather than what being a parent of the same child might mean. Thus, it will not be possible to isolate whether a gender or parent effect is being assessed.

Aims

The aims of the current study are to test psychometric properties of the SDQ in a large nationally representative sample of school aged children. With uncertainty over the construct validity and reliability of the parent version (4-12- year olds), the current study seeks to:

1. Examine competing factorial models and reliability under different data treatment scenarios. We expect that (Mc Donald's Omega) will produce higher reliability findings than Cronbach's alpha in SDQ subscales. On balance, we also expect a 5-factor structure to reveal itself and predict differential results based on whether the SDQ is treated as continuous or categorical in nature.
2. Following these analyses and if appropriate we aim to test measurement invariance for gender, socio-economic status, and parent status.
3. Establish up to date normative data.

Method

Participants and procedure

Data were accessed via the Welsh Health Survey (WHS) databases. The WHS is an annual health, lifestyle and well-being survey that samples approximately 15,000 households throughout Wales, one of the composite nations of the United Kingdom, stratified by local authority, through the Post Office's Postcode Address File. The National Centre for Social Research (NatCen) are commissioned by the Welsh Government to carry out the survey with primary aims of providing national estimates of health, facilitating sub-group analysis and providing evidence to inform national and local policy and strategy. It takes account of selection probabilities with the aim of matching the Welsh population. Full methodological detail is presented elsewhere in easily accessible technical reports (Doyle, Brown, & Alvarez, 2015).

Following database checking we were able to combine data from the 3 years (2013; 2014; 2015). A sample of 4828 children aged 4-12 was achieved (2516 boys and 2312 girls). Families were located in South East (n = 2264, 47%), Mid and West (n= 1362, 28%), and North Wales (1202, 25%). We chose not to combine further editions of the survey because the sampling framework excludes participants for 2 subsequent years. This ensured that people who had participated in previous years would not be re-sampled.

Of the total sample, 681 (14%) of parents were fathers and 3941 mothers (82%), 30 were 'step parents' (<1%), and 106 (2%) 'someone else'. 1811 (37%) held Managerial/professional occupations, 845 (17%) held intermediate occupations, and 1822 (38%) held routine and manual occupations, according to the National Statistics Socio-economic Classification (NS-SEC, see Rose & Pevalin, 2003). The remaining 207 (4%) were unemployed or had never worked. Data was completed on 2516 boys (52%) and 2312 girls (48%).

Statistical analysis

Analyses were conducted and results structured as follows.

1. First, we adopt a similar analysis procedure to Croft, Stride, Maughan, & Rowe (2015) whereby the factor structure of the SDQ is first examined using confirmatory factor analysis (CFA) and then the recognised five-factor model compared to Goodman et al.'s (2010) three-factor (internalising; externalising; and prosocial factors) and 1-factor model (Harman's 1967 single factor test).
2. Second, alpha and omega reliability estimates were calculated for each subscale and additional validity indicators considered.
3. Third, we examined measurement invariance, that is, the stability of the factor structure across groups, using Multiple group CFA with covariates (mimic) with categorical factor indicators and a threshold structure. This was assessed across gender (males vs. females), socio-economic status (NSEC categories), and parental relationship to child (mother vs father).
4. Fourth we present normative and subgroup data.

For comparative purposes, we conducted and provide the findings of CFAs using two different approaches which treated the SDQ data as either continuous or categorical in nature. Differential results could then be attributed to these methods. When the SDQ was treated as categorical the Weighted Least Squares estimator (WLMSV) was chosen. This method uses estimations based on polychoric correlations and appears robust to non-normality (Flora & Curran, 2004). Wirth & Edwards (2007) point out that using standard estimation methods designed for continuous data (such as Maximum Likelihood estimation [Lawley & Maxwell, 1963]) which rely on Pearson correlations “ignores the categorical nature of the data and implicitly introduces misspecification into the series of equations” (p. 19). Research

demonstrates that WLMSV is more accurate in identifying the true fit of a model than standard estimators when data is non-normal or has fewer (<5) response categories (Newsom, 2015; Cai, Maydeu-Olivares, Coffman, & Thissen, 2006; Wirth & Edwards, 2007). While this is arguably the most appropriate method due to the nature of the SDQ data (there are 3 ordinal response options and data is typically skewed in distribution) it is not always implemented in practice.

Because of the large size of our sample and the sensitivity of the χ^2 test (for example, to large samples and normality violations [Muthen & Muthen, 1998; Chen, 2007]) we relied on alternative fit indices, including the root mean square error of approximation (RMSEA, Steiger & Lind, 1980), Comparative Fit Index (CFI, Bentler, 1990), and Tucker-Lewis index (TLI; Tucker & Lewis, 1973) to evaluate model fit. While there has been variation in thresholds applied to interpret these model fit statistics we adopted commonly used recommendations (Hu & Bentler, 1999; Hooper, Coughlan, & Mullen, 2008; He, Burnstein, Shmitz, & Merikangas, 2013) whereby CFI and TLI values above 0.90 or “close to” 0.95 (Hu & Bentler, 1999, p. 27) alongside RMSEA values below 0.06 are indicative of reasonable model fit.

Changes in CFI (Δ CFI) in combination with global fit indices were examined to assess change in model fit for measurement invariance analyses. If the increasingly restrictive models lead to changes in fit indices (and specifically a decrease in Δ CFI greater than or equal to .01) they should be rejected (Cheung & Rensvold, 2002).

Missing data was excluded from the analyses (listwise method) when calculating SDQ subscale reliability scores, as opposed to imputing missing data, to reduce the risk of introducing bias (Graham, 2009; Mazza, Enders, & Ruehlman, 2015). Missing data was negligible with less than 2.6% on SDQ items and subscale scores were computed for >98.2% of respondents.

Analyses were run with and without missing data (using the ‘Listwise = on’ function) and revealed negligible differences.

Data were analysed using SPSS (version 22 [IBM Corp, 2013]) for descriptive, relational and alpha scale-reliability analyses. The R software package (R Core team, 2013) was used to calculate Mc Donald’s Omega using the MBESS package (Kelley, 2016) and Mplus (version 7.4 [Muthen & Muthen, 2015]) was used for subsequent confirmatory factor analyses (CFA), and multigroup CFA analyses. Norms were calculated using weighted quantiles, according to household and child level, using the R software ‘Hmisc’ package (Harrell, 2017).

Results

Confirmatory Factor Analysis

Table 1 presents the fit indices for the one, three, and five-factor models for analyses treating the SDQ as either continuous or categorical in nature. Neither the 1-factor model (Harman’s 1-factor test [Harman, 1967]) nor the 3-factor model indicated reasonable fit. The 5-factor model did not indicate reasonable fit when analysed as a continuous variable using the Maximum Likelihood (ML) estimation method ($\chi^2 = 4651.873$, $df = 265$, $P < .005$, $RMSEA = 0.059$, $CFI = 0.856$). When analyses treated the data as categorical using the Weighted Least Squares (WLSMV) estimation method, acceptable fit indices were evident ($\chi^2 = 1546.709$, degrees of freedom [df] = 265, $P < .0005$, $RMSEA = 0.058$, $CFI = 0.914$). This difference may be due to the ML estimator attenuating results when applied to categorical data (Brown, 2006). The five-factor model was the most acceptable, and explained 52% of the variance. See table 1 for an overview of fit indices.

Table 1 Comparison of fit indices for alternative factor models of the Strengths and Difficulties

Data treated as continuous in nature

Model	χ^2	df	Ch χ^2	Ch df	Change χ^2	CFI	TLI	RMSEA	SRMR
1-factor	10645.03	275	----	----	----	0.66	0.629	0.089	0.075
3-factor	6297.385	272	4347.641	3	0.0000	0.803	0.782	0.068	0.061
5-factor	4651.873	265	1645.512	7	0.0000	0.856	0.837	0.059	0.049

Data treated as categorical in nature

Model	χ^2	df	Ch χ^2	Ch df	Change χ^{2**}	CFI	TLI	RMSEA	SRMR
1-factor	11447.8	275	----	----	----	0.774	0.754	0.092	---
3-factor	6261.16	272	5186.63	3	0.0000	0.879	0.867	0.068	---
5-factor	4507.23	265	1753.93	7	0.0000	0.914	0.903	0.058	---

Questionnaire (continuous vs categorical) in the combined (2013/14/15) WHS dataset.

Note. * Indicates statistically significant difference ($p < .001$). ***Note.* The χ^2 (chi square goodness of fit statistic) cannot be used in the regular way for ordinal variables. A chi-square difference test for the WLSMV estimator was therefore conducted (covariances among the factors are fixed at zero) using the DIFFTEST option (Muthen & Muthen, 1998). Ch= change; df = degrees of freedom; RMSEA = Root-Mean-Square Error of Approximation; CFI = Comparative Fit Index; TLI = Tucker Lewis Index; WRMSR = Weighted Root Mean Square Residual.

Standardised factor loadings ranged from 0.70 to 0.80 for hyperactivity, 0.48 to 0.80 for emotional problems, 0.66 to 0.80 for conduct problems, 0.51 to 0.81 for peer problems, and 0.66 to 0.81 for pro-sociality. Guadagnoli & Velicer (1998) suggest that a factor can be interpreted as reliable if it has a minimum of four loadings above 0.6. All except two item-factor loadings were above 0.6 (see table 2) indicating that this criterion was met. Two item-

factor loadings (Q3 and Q23) can be interpreted as ‘fair’ to ‘good’, with the remainder ranging from ‘good’ to ‘excellent’ according to Tabachnick and Fidell’s (2007) cut-off points (0.32 = poor, 0.45 = fair, 0.55 = good, >0.63 = very good or excellent). See table 2 for item-factor loadings.

Table 2 Standardised item factor loadings for the 5-factor confirmatory factor analysis in the combined (2013/14/15) WHS dataset.

Factor	Items	Estimate	S.E.	Est./S.E.
Hyperactive	Q2 Restless	0.787	0.009	83.442
	Q10 Fidgets	0.804	0.010	78.619
	Q1 Distractible	0.796	0.009	90.840
	Q21 Reflective	0.707	0.012	60.128
	Q25 Persistent	0.775	0.010	78.573
Emotional symptoms	Q24 Fears	0.765	0.014	54.739
	Q16 Clingy	0.709	0.015	47.924
	Q8 Worries	0.683	0.016	42.482
	Q3 Somatic complaints	0.481	0.021	22.611
	Q13 Unhappy	0.809	0.018	45.754
Conduct problems	Q18 Lies	0.664	0.016	41.946
	Q12 Fights	0.756	0.020	38.592
	Q5 Temper	0.717	0.012	60.257
	Q22 Steals	0.722	0.029	24.940

	Q7 Obeys	0.739	0.013	56.837
Peer problems	Q6 Solitary	0.622	0.018	35.000
	Q11 Good friend	0.735	0.018	40.001
	Q19 Bullied	0.601	0.019	30.886
	Q23 Best with adults	0.513	0.018	27.757
	Q14 Popular	0.816	0.016	50.392
Prosocial	Q4 Shares	0.748	0.014	55.431
	Q9 Caring	0.771	0.013	58.287
	Q17 Kind	0.774	0.016	49.314
	Q20 Helps	0.667	0.014	47.880
	Q1 Considerate	0.816	0.011	71.130

Note. $p < .001$ for all item-factor loadings; Number of observations = 4768

Confirmatory analysis in an additional year (2012) also revealed acceptable fit, further confirming the 5-factor model ($\chi^2 = 1949.741$, $df = 265$, $P < .0005$, $RMSEA = 0.054$, $CFI = 0.919$).

Reliability and additional validity indicators

Alpha versus Omega

The Cronbach's alpha coefficient for the SDQ was acceptable for the whole measure ($\alpha = 0.845$) however the peer problems, conduct problems and emotional symptoms subscales did not meet the 0.70 benchmark (Nunnally & Bernstein, 1994).

To calculate the McDonald's omega coefficient the MBESS package (Kelley, 2016) for R was used (R Core Development Team, 2017). This indicated acceptable reliability for all subscales (table 4). It should be noted however, that the peer problems subscale remained slightly lower than 0.70. See table 3 for a visual comparison of reliability statistics.

Table 3 Internal consistency of the Strengths and Difficulty Questionnaire

Scale measured	Cronbach's alpha (α)	Omega (ω)
Total Scale (- Prosocial items)	.845	0.87
Hyperactive Scale	.79	0.85
Emotional Symptoms	.69	0.74
Conduct Problems	.63	0.71
Peer Problems	.618	0.68
Prosocial Scale	.755	0.77

In the literature, externalising behaviour is typically inversely correlated with pro-sociality. As would be expected of a good externalising measure, we found significant negative correlations ($p < .01$) between pro-sociality and hyperactivity (-.592), conduct disorder (-.589), peer problems (-.537), and the SDQ total score (-.525).

Criterion validity was indicated by significant differences between those currently being treated (versus not treated) for 'mental illness' ($u [23,958] = 62,503, p < .001$), and those with/ without longstanding limiting illness ($u [41,205] = 41,2585, p < .001$) on the SDQ total difficulties score.

Measurement Invariance

We conducted a series of logical and increasingly restrictive steps (Vandenberg & Lance, 2000; Millsap, 2011) testing configural, metric, and scalar invariance for the identified groupings

(gender; SES; parent status). In step 1 we tested whether the structure of the SDQ holds for groups (configural invariance), that is, do groups conceptualise the five SDQ constructs the same way? Here no equality constraints were imposed across groups. For identification purposes Mplus fixes the first factor loading for each latent variable is to one. In step two, we tested whether participants from groups respond similarly to the items, that is, are the factor loadings invariant? (metric invariance). Here the factor loadings were constrained equal across groups. Metric invariance is required to permit subsequent comparison across the factors (but not mean scores) of the SDQ. In step 3 we tested scalar invariance. This is the last step to allow subsequent comparisons in the latent mean scores to be made (Milfont & Fischer, 2010). In this step the thresholds of the items were constrained equal across groups. Table 4 presents the results of measurement invariance testing.

Table 4 Measurement Invariance testing of the SDQ with categorical factor indicators

Multiple group CFA (gender)						
Model (Gender)	χ^2	df	CFI	Δ CFI	TLI	RMSEA
1. Configural	4647.797	530	0.915	-	0.903	0.057
2. Metric invariance	4572.753	550	0.909	0.006	0.917	0.055
3. Scalar invariance	4398.929	570	0.921	0.012	0.916	0.053
Multiple group CFA (parent status)						
Model	χ^2	df	CFI	Δ CFI	TLI	RMSEA
1. Configural	4122.67	530	0.919	-	0.908	0.054
2. Metric invariance	3957.371	550	0.923	0.011	0.916	0.052
3. Scalar invariance	3747.626	570	0.929	0.006	0.925	0.049
Multiple group SES						
Model	χ^2	df	CFI	Δ CFI	TLI	RMSEA

1. Configural	49823.773	530	0.910	-	0.906	0.057
2. Metric invariance	38262.6	550	0.912	0.002	0.908	0.055
3. Scalar invariance	3910	570	0.914	0.004	0.910	0.054

Note. χ^2 = Chi square; df = degrees of freedom; CFI = Comparative Fit Index; TLI = Tucker-Lewis Index; RMSEA = Root Means Square Error of Approximation; WRMR = Weighted Root Mean Square Residual.

The analyses indicated measurement invariance of the 5-factor model at steps 1 (configural) and 2 (metric) across groups of boys and girls. Following recommendations in the MI literature, we therefore proceeded to the next step (Vandenberg & Lance, 2000). At step 3, the model fit was also supported, indicating scalar invariance. It is suggested in the literature that a change in CFI (Δ CFI) between models should not be substantial (a decrease of $\geq .01$), or it is unlikely that measurement invariance holds (Cheun & Rensfold, 2002). This guidance typically relates to interpretation of continuous data testing and is unlikely to apply to an increase or improvement in indices (Muthen, 2017; Kline, 2015). The fit of our model, as indicated by CFI and TLI, appeared to improve with increased constraints so we therefore accepted the model. While improvement is unexpected, there is limited understanding as to how fit indices behave with categorical data and other researchers have experienced the same phenomenon (Muthen, 2017). The same pattern was confirmed when we used a different software package (Jorgensen, Pornprasertmanit, Miller, Schoemann, Rosseel, 2016).

The analyses also showed that the SDQ was invariant according to parental status and SES for configural, metric and scalar analyses. In each of these there were acceptable levels of fit indices, which did not substantially decrease, suggesting the model was invariant across SES and parent status (i.e. mothers and fathers).

Normative data and bandings

As expected there were significant differences between males and females across all subscales) and for the total SDQ score (table 5). Because of this gender specific norms that follow should be interpreted unless there is specific justification to do otherwise.

Table 5 Analyses of SDQ scores by gender (4-12's)

	Gender	N	M Rank	Mean	SD	SE	U
SDQ Total (excl. Prosocial)	Male	2475	2511	8.118	6.1		
	Female	2278	2230	6.836	5.3	47,159	-7.056*
Prosocial	Male	2476	2168	8.227	1.96		
	Female	2278	2604	8.823	1.62	45149	11.475*
Hyperactivity	Male	2473	2600	3.644	2.59	46,812	
	Female	2277	2131	2.774	2.33		-11.871*
Emotional	Male	2477	2301	1.543	1.94	45,583	
	Female	2278	2460	1.698	1.90		4.137*
Conduct Disorder	Male	2477	2475	1.525	1.77	45,466	
	Female	2277	2270	1.123	1.47		-5.339*
Peer	Male	2473	2457	1.413	1.77	44,539	
	Female	2275	2284	1.123	1.47		-4.598*

Note. U = Mann Whitney U test (standardised test statistic); *p = <.001

Normative data is presented on the pages that follow, for the overall sample (table 6), and for boys and girls (table 7). A quick reference summary table for clinicians to easily identify cut-off points is provided in the Appendices.

We used frequency distributions to develop cut points to band scale scores according to percentile groups. The accuracy of the SDQ in community samples is dependent on the 90%

dichotomisation system (Vaz, Cordier, Boyes, Parsons, Joosten, Cicarelli, Falkmer, & Falkmer, 2016), that is, cut off points for the top 10% are used to identify those at increased risk (x 15) of having a mental health problem / psychiatric diagnosis (based on Goodman's criteria, 1997; 2001).

Previous UK data identifies the high-risk banding as a total score of 17 or over. However, only 7.8% (n = 375; 243 boys and 132 girls) of the current study fell within this range. Updated bands were therefore calculated (table 6 above) and revealed that the appropriate cut-off point was 16.

Data in the WHS is weighted according to household and child level. We applied weights using the R software 'Hmisc' package (Harrell, 2017). There were negligible differences between weighted data when calculated at household or child level, which likely indicates the representativeness of this study's sample. There were no differences between weighted and non-weighted data in terms of where the high-risk cut-off points fall. We provide an illustration of this in table 6, comparing weighted and unweighted total difficulty scores or TDS (shown as TDSH and TDSC respectively).

Table 6 Normative data for the SDQ: Percentile scores for all children

	5 th	10 th	15 th	20 th	25 th	30 th	35 th	40 th	45 th	50 th	55 th	60 th	65 th	70 th	75 th	80 th	85 th	90th	95th	100th
TDS	1	2	2	3	4	4	5	5	6	7	7	8	9	10	11	12	14	16	20	37
TDS (H)	1	2	2	3	4	4	5	5	6	7	7	8	9	10	11	12	14	16	20	37
TDS (C)	1	2	2	3	4	4	5	6	6	7	8	8	9	10	11	13	14	16	20	37
Prosocial	5	6	6	7	7	8	8	8	9	9	9	9	10	10	10	10	10	10	10	10
Hyper	0	0	1	1	1	2	2	2	3	3	3	4	4	4	5	5	6	7	8	10
Emot.	0	0	0	0	0	0	0	0	1	1	1	1	2	2	2	3	3	4	5	10
Conduct	0	0	0	0	0	0	1	1	1	1	1	2	2	2	2	3	3	4	5	10
Peer	0	0	0	0	0	0	0	0	1	1	1	1	2	2	2	3	3	4	5	10

Note: TDS=Total Difficulties Score; (H) = Household weights applied; (C) = Child level weights applied.

Table 7 Normative data for the SDQ. Percentile breakdown for boys and girls.

Percentile scores for boys																				
	5 th	10 th	15 th	20 th	25 th	30 th	35 th	40 th	45 th	50 th	55 th	60 th	65 th	70 th	75 th	80 th	85 th	90 th	95 th	100 th
TDS	1	2	3	3	4	4	5	6	6	7	8	9	9	11	12	13	14	17	20	36
Prosocial	4	5	6	6	7	7	8	8	8	9	9	9	9	10	10	10	10	10	10	10
Hyper	0	0	0	1	1	2	2	2	3	3	3	4	4	4	5	5	6	7	9	10
Emot.	0	0	0	0	0	0	0	1	1	1	1	1	2	2	2	3	4	4	6	10
Conduct	0	0	0	0	0	0	0	1	1	1	1	2	2	2	2	3	3	4	5	10
Peer	0	0	0	0	0	0	0	0	1	1	1	1	2	2	2	3	3	4	5	10
Percentile scores for girls																				
	5 th	10 th	15 th	20 th	25 th	30 th	35 th	40 th	45 th	50 th	55 th	60 th	65 th	70 th	75 th	80 th	85 th	90 th	95 th	100 th
TDS	0	1	2	3	3	4	5	5	6	6	7	8	9	9	11	12	13	15	19	37
Prosocial	5	6	7	7	8	8	8	9	9	9	9	10	10	10	10	10	10	10	10	10
Hyper	0	0	0	1	1	1	2	2	2	3	3	3	3	4	4	5	5	6	8	10
Emot.	0	0	0	0	0	1	1	1	1	2	2	2	2	3	3	4	4	5	6	10
Conduct	0	0	0	0	0	0	0	1	1	1	1	1	2	2	2	3	3	4	5	10
Peer	0	0	0	0	0	0	0	0	1	1	1	1	2	2	2	3	3	4	5	10

Note: TDS=Total Difficulties Score

Discussion

These findings make several important contributions to the related literature.

An important international priority is to improve the health and wellbeing of children and vulnerable subgroups. If key measures (and their normative interpretation) used for this purpose are inappropriate, they can impact this process and disenfranchise certain groups (particularly, those at increased risk).

Regarding the psychometric properties of the SDQ, findings identified the 5-factor model as most appropriate. This is contrary to suggestion that using a three-subscale division (internalising, externalising, prosocial) may be more appropriate in general population samples (Goodman et al, 2010).

While it has been recognised that the nature of the data plays a crucial role in the outcome of statistical analyses, only more recently is the extent of this phenomenon coming to light. Regarding methods of analyses, results demonstrate that had we incorrectly treated the SDQ as a continuous variable we would have rejected this model. This is in line with previous research showing superior model fit indices (CFI, TLI, and RMSEA) attributable to the estimator type (WLSMV rather than ML) because Pearson correlations underestimate the true relation, particularly for two and three category variables, i.e. the SDQ (Beauducel & Herzberg, 2006). This is important because this finding suggests that past studies and systematic reviews of SDQ studies (e.g. Kersten et al., 2016), may require re-interpretation as they have not taken account of the estimation method in their evaluation process. Currently, it may be that data continues to be treated this way because ubiquitous statistical programmes (such as SPSS) do not yet have in-built ability to conduct some of these techniques. This phenomenon has been noted by Borsboom (2005).

Related to this point findings highlight another neglected methodological consideration in the SDQ and wider literature, that of reliability estimates. Some studies suggest using the total difficulties score and that subscale scores should not be interpreted due to low alpha reliability (e.g. Vaz et al., 2016). Our findings provide evidence to the contrary, i.e. that it is appropriate to interpret the subscales. Omega revealed higher scores than alpha in our study. This is in line with findings from a recent study by Stone et al. (2015) on the preschool version and fits with increased recognition in the literature that alpha underestimates reliability. Despite the clear and generally accepted logic that alpha will rarely be an appropriate reliability statistic in practice (Revelle & Zinbarg, 2008; Sitsma, 2009; Zinbarg & Alden, 2015), authors continue to neglect the inadequacy of alpha. Thus, systematic reviews likely to influence future research in the area (Kersten et al., 2016) continue to draw conclusions based on questionable statistical approaches.

Some authors suggest that it is inappropriate to rely on a summary score in the absence of a one factor model. However, while the SDQ is not unidimensional (i.e. it consists of 5-factors), the reliability estimate (ω) of the overall variance in the SDQ data due to general and specific factors was high (0.87), suggesting it is appropriate for clinicians to also interpret the total difficulties score.

Regarding differences and relationships between groups, significant gender differences were identified. On average boys reported significantly more problems than girls overall (total SDQ score) and on all subscales, and this is comparable with similar findings elsewhere (Boman et al., 2016; Moriwaki et al., 2014; Achenbach et al., 2008). Clinicians are therefore reminded to use the gender specific norms provided when interpreting scores.

Regarding measurement invariance, our study found the 5-factor model to be invariant for gender, parent status and socio-economic status. To our knowledge this is the first study to examine this set of variables in the same sample. It is also the first study to examine parent status in this way. The results indicate that the SDQ measures the same constructs across these groups. Comparisons between these groups can therefore validly be made.

This study presents new and up to date normative data which has been lacking in this context. Findings identify a different proportion of children at risk than current UK bandings and different mean scores when compared to the older and less context specific data provided by SDQ publishers. While this is arguably not an unexpected finding, as mean scores have been found to vary by study, potentially due to culture or time, it is concerning for several reasons.

Firstly, because screening measures are often used to assess risk and contribute to decisions about who receives further assessment and treatment (Babor, Sciamanna, & Pronk, 2004), it is likely that high-risk children are being identified incorrectly (that is, under-represented) because of out of date norms.

Second, because the SDQ is routinely used to contribute to decision making (e.g. policy and funding) around intervention strategies to improve the health of children, accuracy of measurement and norms are pivotal in this process. The predictive validity of the SDQ will be reduced by inaccurate cut-off points, sensitivity and specificity analyses routinely cited relating to the SDQ (e.g. Bowman et al., 2016), will be inaccurate.

Up to date normative profiling may be appropriate in other regions, given our findings. What is clear is that cross-national comparisons are ill advised without population specific norms (Kersten et al., 2016; Goodman et al., 2012).

Our study has several limitations. Firstly, because ours was a population based sample, it may not generalise to clinical groups. Further evaluation of this point is warranted given recent studies conflictingly suggest invariance for clinical versus community settings (Smits, Theunissen, Reijneveld, Nauta, & Timmerman, 2016) but also have found structural differences according to the administration setting (Mieloo, Raat, van Oort, Bevaart, Vogel, & Donker, 2012).

A limitation of our analyses concerns SES and recent analysis that found that those from managerial and professional NSEC categories were more likely to participate in this research (p 32, Doyle, Brown, & Alvarez, 2015). It is therefore possible that there is an underestimate of other NSEC categories. Conclusions from this study are also limited to the parent version. Because the study uses a parental self-report method it reflects the parental perception of the child rather than the child's view or an independent professional view. There is no independent verification (for example, as to mental health diagnosis). Parents and teacher responses tend to differ (possibly due to the situational specificity of psychosocial problems [Achenbach, McConaughy, & Howell, 1987]), albeit with inconsistencies according to study. In some studies, parent estimation of risk has been found to be higher than teachers (e.g. Boman et al., 2016; Goodman et al., 2000) but this has not been replicated in other studies (Mellor & Stokes, 2007). Regardless, a combined parent and teacher approach remains as recommended practice.

Despite these limitations, our study makes an important contribution to the literature and has several strengths. It provides a large representative national sample, not often matched in this literature, answers important questions relating to mental health measurement, and identifies areas for action. In conclusion, this study supports the 5-factor SDQ for use in 4-12 year olds and highlights the importance of up to date and context specific norms. Findings suggest that

greater acknowledgement of the recurrent methodological difficulties relating to the analysis of the SDQ and other similar ordinal categorical measures is warranted. It is recommended that the literature, and relevant review processes, question studies that ignore or do not justify the methods used when analysing categorical measures. The new normative bandings and risk cut-off points should be implemented in Wales by researchers and clinicians.

Acknowledgements: The authors would like to thank all those who took time to complete the questionnaire; the holders of the WHS data for their advice, and in particular thanks is due to the following people: Cath Roberts of the Welsh Governments Health Statistics and Analysis Unit for her responsiveness; Dr. Kate Langley and other researchers of the WHS for time in discussing their own research and in identifying the existing literature; Dr Chris Stride from the University of Sheffield for statistical advice relating to preparing data for MPLUS.

References

- Achenbach, T. M., Becker, A., Döpfner, M., Heiervang, E., Roessner, V., Steinhausen, H. C., & Rothenberger, A. (2008). Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: research findings, applications, and future directions. *Journal of Child Psychology and Psychiatry*, 49(3), 251-275.
- Achenbach, T. M., Mc Conaughy, S. H., & Howell, C. T. (1987). Child/ adolescent behavioural and emotional problems: Implications of cross-informant correlations for situational specificity. *Psychology Bulletin*, 1987;101:213–32.
- Babor, T. F, Sciamanna, C. N., & Pronk, N. P. (2004). Assessing multiple risk behaviors [sic] in primary care: Screening issues and related concepts. *American journal of preventive medicine*, 27: 42–53.
- Bayer, J. K., Hiscock, H., Ukoumunne, C., Price A, Wake M. (2008). Early childhood aetiology of mental health problems: a longitudinal population- based study. *Journal of Child Psychology and Psychiatry*, 49(11),1166–1174
- Bayer, J., Hiscock, H., & Scalzo, K. (2009). Systematic review of preventive interventions for children's mental health: what would work in Australian contexts? *Australia and New Zealand Journal of Psychiatry*, 43(8), 695–710.
- Beauducel, A., & Herzberg, P.Y. (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, DOI: 10.1207/s15328007sem1302_2

- Boman, F., Stafström, M., Lundin, N., Moghadassi, M., Törnhage, C. J., & Östergren, P. O. (2016). Comparing parent and teacher assessments of mental health in elementary school children. *Scandinavian journal of public health, 44*(2), 168-176.
- Borsboom, D. (2005). *Measuring the mind. Conceptual issues in contemporary psychometrics*. Cambridge: Cambridge University Press.
- Brown, T. A. (2006). *Confirmatory factor analysis for applied research*. New York, NY: Guilford.
- Cai, L., Maydeu-Olivares, A., Coffman, D. L., & Thissen, D. (2006). Limited information goodness-of-fit testing of item response theory models for sparse 2P tables. *British Journal of Mathematical and Statistical Psychology, 59*, 173-194.
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling, 14*(3), 464-504. Doi:000248799400005.
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modelling, 9*(2), 233-255. doi: 10.1207/S15328007sem0902_5 WOS:000180828700005.
- Chiorri, C., Hall, J., Casely-Hayford, J., & Malmberg, L. (2016). Evaluating Measurement Invariance Between Parents Using the Strengths and Difficulties Questionnaire (SDQ). *Assessment, 23* (1), 63 – 74. doi: 10.1177/1073191114568301
- Croft, S., Stride, C., Maughan, B., & Rowe, R. (2015). Validity of the strengths and difficulties questionnaire in preschool-aged children. *Pediatrics, 135*(5), e1210-e1219.

- Dickey, W. C., & Blumberg, S. J. (2004). Revisiting the factor structure of the strengths and difficulties questionnaire: United States, 2001. *Journal of the American Academy of Child & Adolescent Psychiatry*, 43(9), 1159-1167.
- Doyle, M., Brown, L., & Alvarez, P. C. (2016). *Welsh Health Survey 2015: Technical Report*. Welsh Government. NatCen Social Research: London. Retrieved, 10-04-17 from, <http://gov.wales/docs/statistics/2016/160929-welsh-health-survey-2015-technical-report-en.pdf>
- Dunn, T. J., Baguley, T., & Brunsten, V. (2014). From alpha to omega: A practical solution to the pervasive problem of internal consistency estimation. *British Journal of Psychology*, 105(3), 399-412.
- Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological methods*, 9(4), 466. Doi: 10.1037/1082-989X.9.4.466.
- Goodman, R., & Scott, S. (1999). Comparing the Strengths and Difficulties Questionnaire and the Child Behavior Checklist: is small beautiful?. *Journal of abnormal child psychology*, 27(1), 17-24.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.
- Goodman, R. (2001). Psychometric properties of the Strengths and Difficulties Questionnaire. *Journal of American Academy of Child Adolescent Psychiatry*, 40, 1337-1345. doi:10.1097/00004583-200111000-00015

- Goodman, R., Ford, T., Simmons, H., Gatward, R., & Meltzer, H. (2000). Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *The British Journal of Psychiatry*, *177*(6), 534-539.
- Goodman, A., Lamping, D. L., & Ploubidis, G. B. (2010). When to use broader internalising and externalising subscales instead of the hypothesised five subscales on the Strengths and Difficulties Questionnaire (SDQ): data from British parents, teachers and children. *Journal of abnormal child psychology*, *38*(8), 1179-1191.
- Graham, J. W. (2009). Missing data analysis: Making it work in the real world. *Annual Review of Psychology*, *60*, 549–576.
- Green, H., Mc Ginnity, A., Meltzer, H., Ford, T., & Goodman, R. (2005). *Mental health of children and young people in Great Britain, 2004. Summary report*. The office for National Statistics: Crown copyright.
- Guadagnoli, E., & Velicer, W. (1998). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, *103*, 265-275.
- Harrell, F. E. (2017). ‘Hmisc’ package for R. Retrieved, 05-02-17 from, <https://github.com/harrelfe/Hmisc>
- Harman, H. H. (1967). *Modern factor analysis*. Chicago: University of Chicago Press.
- Hill, C. R., & Hughes, J. N. (2007). An examination of the convergent and discriminant validity of the Strengths and Difficulties Questionnaire. *School Psychology Quarterly*, *22*(3), 380.
- He, J. P., Burstein, M., Schmitz, A., & Merikangas, K. R. (2013). The Strengths and Difficulties Questionnaire (SDQ): the factor structure and scale validation in US adolescents. *Journal of abnormal child psychology*, *41*(4), 583-595.

- Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, 6(1), 53-60.
- Hu, L. T., & Bentler, P. M. (1999). Cu-toff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modelling: a multidisciplinary journal*, 6(1), 1-55. doi: 10.1080/10705519909540118
- IBM Corp. (2013). *IBM SPSS Statistics for Windows*. Version 22.0. Armonk, NY: IBM Corp.
- Jorgensen, T. D., Pornprasertmanit, S., Miller, P., Schoemann, A., Rosseel, Y. et al. (2016). 'SemTools'. Useful Tools for Structural Equation Modelling. The R project. Retrieved 10-01-17 from, <https://cran.r-project.org/web/packages/semTools/semTools.pdf>
- Kersten, P., Czuba, K., McPherson, K., Dudley, M., Elder, H., Tauroa, R., & Vandal, A. (2016). A systematic review of evidence for the psychometric properties of the Strengths and Difficulties Questionnaire. *International Journal of Behavioral Development*, 40, 64-75. doi:10.1177/0165025415570647
- Kelley, K. 2016. The MBESS R package. The R core development team. Retrieved, 10-01-17 from, <https://cran.r-project.org/web/packages/MBESS/MBESS.pdf>
- Kline, R. B. (2015). *Principles and practice of structural equation modelling*. Guilford publications: New York.
- Lawley, D. N. & Maxwell, A. E. *Factor analysis as a statistical method*. London: Butterworth, 1963.
- Leeuwen, K. V., Meerschaert, T., Bosmans, G., De Medts, L., & Braet, C. (2006). The strengths and difficulties questionnaire in a community sample of young children in flanders. *European Journal of Psychological Assessment*, 22(3), 189.
- Limbers, C. A., Newman, D. A., & Varni, J. W. (2009). Factorial invariance of child self-report across race/ethnicity groups: a multigroup confirmatory factor analysis approach

- utilizing the PedsQL™ 4.0 Generic Core Scales. *Annals of Epidemiology*, 19(8), 575-581.
- Longevity Science Advisory Panel. (2012). *Life expectancy: Past and Future variations by gender in England and Wales. LSAP paper 2*. Retrieved 10-01-16 from, http://www.longevitypanel.co.uk/_files/life-expectancy-by-gender.pdf
- Mazza, G. L., Enders, C. K., & Ruehlman, L. S. (2015). Addressing item-level missing data: A comparison of proration and full information maximum likelihood estimation. *Multivariate behavioral research*, 50(5), 504-519. doi: 10.1080/00273171.2015.1068157
- McCrorry, C., & Layte, R. (2012). Testing competing models of the Strengths and Difficulties Questionnaire's (SDQ's) factor structure for the parent-informant instrument. *Personality and Individual Differences*, 52, 882-887. doi:10.1016/j.
- Mellor, D., & Stokes, M. (2007). The factor structure of the Strengths and Difficulties Questionnaire. *European Journal of Psychological Assessment*, 23(2), 105.
- Meltzer, H., Gatward, R., Goodman, R., & Ford, F. (2000). *Mental health of children and adolescents in Great Britain*. London: The Stationery Office.
- Milfont, T. L., & Fischer, R. (2015). Testing measurement invariance across groups: Applications in cross-cultural research. *International Journal of psychological research*, 3(1), 111-130.
- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. New York: Taylor and Francis Group.
- Mieloo, C., Raat, H., van Oort, F., Bevaart, F., Vogel, I., Donker, M., & Jansen, W. (2012). Validity and reliability of the Strengths and Difficulties Questionnaire in 5–6 year olds: differences by gender or by parental education?. *PloS one*, 7(5), e36805.

- Moriwaki, A., & Kamio, Y. (2014). Normative data and psychometric properties of the strengths and difficulties questionnaire among Japanese school-aged children. *Child and adolescent psychiatry and mental health*, 8(1), 1.
- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49, 115–132.
- Muthén, L.K. and Muthén, B.O. (1998-2015). *Mplus User's Guide. Seventh Edition*.
- Muthen, B. O. (2017). Forum discussion, March 01, 2017 - 7:56 am. Available from, <http://www.statmodel.com/discussion/messages/23/25.html>
- Newsom, J. T. (2015). *Longitudinal structural equation modelling: A comprehensive introduction*. Routledge.
- Nunnally, J. C., & Bernstein, I. R. (1994). *Psychometric theory*. New York, USA: McGraw Hill Book Company.
- Ortuño-Sierra, J., Fonseca-Pedrero, E., Aritio-Solana, R., Velasco, A. M., De Luis, E. C., Schumann, G., ... & Bokde, A. (2015). New evidence of factor structure and measurement invariance of the SDQ across five European nations. *European child & adolescent psychiatry*, 24(12), 1523-1534.
- R Core Team. (2013). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria. Retrieved 10-01-17 from, <http://www.R-project.org/>.
- Revelle, W., & Zinbarg, R. E. (2008). Coefficients alpha, beta, omega and the glb: comments on Sijtsma. *Psychometrika*, 74, 1, 145-154.
- Revelle, W. (2017). 'Psych': Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, <https://CRAN.R-project.org/package=psych> Version = 1.7.5.

- Schweizer, K. (2011). On the changing role of Cronbach's α in the evaluation of the quality of a measure. *European Journal of Psychological Assessment*, 27, 143–144.
- Sijtsma, K. (2009). On the Use, the Misuse, and Very Limited Usefulness of Cronbach's Alpha. *Psychometrika*, 74, 107–120.
- Smits, I. A., Theunissen, M. H., Reijneveld, S. A., Nauta, M. H., & Timmerman, M. E. (2016). Measurement Invariance of the Parent Version of the Strengths and Difficulties Questionnaire (SDQ) Across Community and Clinical Populations. *European Journal of Psychological Assessment*. doi: 10.1027/1015-5759/a000339.
- Steiger, J. H., & Lind, J. C. (1980). Statistically based tests for the number of common factors. In *annual meeting of the Psychometric Society, Iowa City, IA* (Vol. 758, pp. 424-453).
- Stevanovic, D., Urbán, R., Atilola, O., Vostanis, P., Balhara, Y. S., Avicenna, M., & Petrov, P. (2015). Does the Strengths and Difficulties Questionnaire–self report yield invariant measurements across different nations? Data from the International Child Mental Health Study Group. *Epidemiology and psychiatric sciences*, 24(04), 323-334.
- Statistics Canada. (2000). National Longitudinal Survey of Children and Youth (NLSCY) Cycle 3 Survey Instruments: Parent Questionnaire. Ottawa, Canada: Statistics Canada.
- Stone, L. L., Otten, R., Engels, R. C., Vermulst, A. A., & Janssens, J. M. (2010). Psychometric properties of the parent and teacher versions of the strengths and difficulties questionnaire for 4-to 12-year-olds: a review. *Clinical child and family psychology review*, 13(3), 254-274.
- Stone, L. L., Janssens, J. M., Vermulst, A., Van Der Maten, M., Engels, R. C., & Otten, R. (2015). The Strengths and Difficulties Questionnaire: psychometric properties of the parent and teacher version in children aged 4–7. *BMC Psychology* (2015) 3 (4), 1-12. Doi: 10.1186/s40359-015-0061-8

- Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics*. Fifth Edition. Pearson Education Inc.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, *38*, 1-10. Doi:10.1007/BF02291170
- Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational research methods*, *3*(1), 4-70.
- Vaz, S., Cordier, R., Boyes, M., Parsons, R., Joosten, A., Ciccarelli, M., ... & Falkmer, T. (2016). Is using the strengths and difficulties questionnaire in a community sample the optimal way to assess mental health functioning?. *PloS one*, *11*(1), e0144039.
- Van Roy, B., Veenstra, M., & Clench- Aas, J. (2008). Construct validity of the five- factor Strengths and Difficulties Questionnaire (SDQ) in pre- , early, and late adolescence. *Journal of Child Psychology and Psychiatry*, *49*(12), 1304-1312.
- Watkins, M. W. (2013). *Omega* [Computer software]. Phoenix, AZ: Ed & Psych Associates. Retrieved 10-02-17 from, <http://edpsychassociates.com/Watkins3.html>
- Wirth, R. J., & Edwards, M. C. (2007). Item factor analysis: current approaches and future directions. *Psychological methods*, *12*(1), 58.
- Youngstrom, E., Loeber, R., & Stouthamer-Loeber, M. (2000). Patterns and correlates of agreement between parent, teacher, and male adolescent ratings of externalizing and internalizing problems. *Journal of consulting and clinical psychology*, *68*(6), 1038.
- Zinbarg, R. E. & Alden, A. R. 2015. Coefficient Alpha and Coefficient Omega hierarchical. *The Encyclopedia of Clinical Psychology*, 1–6.

Chapter 3.

Contributions to theory and clinical practice

Overview

This thesis concerned itself with approaches to intervention and measurement in child and adolescent mental health. First, it systematically reviewed and synthesised the evidence relating to the efficacy of Positive Psychology Interventions. Second, it evaluated the psychometric properties of one of the most commonly used measures of child and adolescent mental health, the SDQ, and updated outdated normative data. The implications for practice, further research and policy are considered in this paper. This is structured in two sections, first in relation to the findings relating to Positive Psychology Intervention, and second, in relation to the SDQ findings.

Section 1. Implications from the systematic review

Clinical implications

Despite an inherent lack of logic, clinical and positive psychology intervention (PPI) approaches have often been considered separately within the literature and there are calls for greater integration and testing of positive psychology interventions (PPI's) within clinical contexts (Freire et al., 2014; Lindley, Joseph, Harrington, & Wood, 2006; Pawelsky, 2016a; 2016b; Drvaric, Gerritsen, Rashid, Bagby, & Mizrahi, 2015).

While it is recognized that reducing distress or psychopathology is not the same as improving well-being and happiness (Bech, Olsen, Kjoller, & Rasmussen, 2003; Bergsma, Have, Veenhoven, De Graaf, 2011) some authors have criticised Positive Psychology for creating an

unhelpful “dichotomy out of positive and negative rather than integrating them” (Lazarus, 2003, p. 173). We concur, and suggest that integration is common sense.

At the same time, those implementing psychological interventions are increasingly recommended to rely on evidence based practice which has been lacking. Reviews of mental health interventions in children and adolescents have typically neglected PPI’s (for example, Hetrick, Cos, Witt, Bir, & Merry, 2016; Stevens, Roberts, & Shiell, 2010). Positive findings identified in this review suggest this position be rectified.

Previously, the evidence base for PPI’s was limited and referred to 3 interventions in the RCT literature. Following this review, a more detailed picture of the evidence is available, accompanied by summarised findings for 14 RCT’s and related recommendations. This information is important because it identifies which interventions are effective and to what degree. Two PPI’s in clinical or sub-clinical populations that led to significant improvements in mental health and are likely applicable within clinical settings were identified. Clinical Health Psychologists should also note the beneficial effect of well-being therapy on physical well-being and somaticisation. Complex interconnected pathways likely exist and remain uncovered but evidence suggests that interventions aimed at improving happiness may have an impact on a range of clinical and non-clinical groups. For example, a recent study of PPI’s in an adolescent clinical setting (inpatient eating disorder) found meaningful improvement in happiness and life satisfaction (Harrison, Al-khairulla, & Kikoler, 2016).

Based on findings clinicians should note that the efficacy of interventions can depend on baseline levels of certain traits. For example, Tobach et al. (2016) found that those unaware of their character strengths prior to the intervention showed greater improvement than those that were aware. It is notable that in Toback et al. (2016) no differences were identified for

those implementing the coping skills versus those that did not. This implies that just identifying character strengths using an inventory might be sufficient and that the coping skills element is not the key factor. Similarly, those low in Positive Affect initially see greater improvements over time (Froh et al., 2009). These findings suggests that targeting PPI's accordingly could lead to greater benefits.

The review identified that PPI's affected outcome differently. For example, improvement on subjective well-being measures such as life satisfaction did not necessarily correspond to improvement in psychopathology (e.g. Suldo, Savage & Mercer, 2014). This provides further evidence that these are distinguishable constructs (Suldo & Shaffer, 2008). Clinicians should therefore be mindful that targeting one domain may not lead to improvements in the other. It seems clear that there may be no direct causal relationship at work. However, the intriguing question in this regard is whether there may be mediating mechanisms which could possibly connect the same two constructs, and if so whether a program to improve well-being could be developed or manipulated in order to increase or produce an improvement in psychopathology, or the reverse.

Implications for future research and theory

A variety of factors that facilitate or inhibit implementation, and can contribute to future research and theory were identified by the review.

The mechanisms by which PPI's might work

An important question that remains to be answered is, what are the mechanisms of change underlying PPI's? There are likely complex pathways in operation, for example, physical activity can lead to changes in hope via increased self-worth (McDavid et al., 2015).

A key implication of this review is that studies should strongly consider more complex statistical analyses to account for mediating moderating pathways. This is important because then we can titrate interventions according to these mechanisms. It is also likely that mechanisms will be different based on the type of PPI (for example, gratitude may well work differently than a character strengths intervention) and therefore require comparative testing. Studies identified suggest that changes in gratitude (but not hope) mediate the PPI effect on depression and satisfaction and hope mediates effects on life satisfaction (Kwok et al., 2016). Additionally, levels of Positive Affect (Froh et al., 2009) and awareness of character strengths are important to intervention impact. Findings are in line with a previous review suggesting that depressed participants have greater gains (Sin & Lyubomirsky, 2009).

Froh et al (2008) suggests that ‘happy people tend to be more sensitive to rewards’ (p. 230). We suggest, equally, that a neglected consideration in the research is the role of attentional processes in mediating PPI’s. There is considerable evidence to suggest that attention can play a mediating role in a variety of physical and mental symptoms (for a review see Connolly, 2016). Rather than being more sensitive to rewards, happy people may direct (and be predisposed to direct) their attention more towards rewards.

An interesting direction then would be to test whether general attentional differences in how one attends to stimuli (e.g. through the ‘Rubber Hand Illusion’ experiment pioneered by Botvinich & Cohen [1998]), and specific attentional differences (such as how one notices or is grateful for positive events) correlate with PPI outcomes. Innovative studies from adult populations, for example, such as identifying neurological correlates of gratitude (Kini et al., 2016), can contribute to this process. Fecteau et al. (2007) used transcranial stimulation to modulate prefrontal cortex activity, finding it affected emotion regulation and reduced risk-

taking behaviours in adults. How neurological interventions such as this could influence emotion regulation, and might apply to adolescent populations to modify risk-taking behaviour, raises interesting possibilities (and research and ethical considerations).

From a theoretical perspective, it might be relevant to note that synonyms for risk-aversion include the descriptors cautious, vigilant, discreet, and attentive (Cambridge Advanced Learner's Dictionary & Thesaurus, 2017), all of which suggest very opposite qualities to the SDQ factor items for hyperactivity and conduct disorder.

Positive Affect is one of the most common outcomes assessed in PPI literature but there has been no study of personality or coping typologies. There is a convincing evidence base showing that people with certain personality characteristics (e.g. those high in neuroticism and low in positive affect) pay more attention to negative than positive information and use less productive coping strategies (e.g. emotion and avoidance focused rather than problem focussed strategies; Connolly, 2013). This may be a fruitful avenue for future research, particularly because it could lead to more targeted intervention and screening.

Consider important confounding variables

The range of account for confounding variables was described in this review, and indicates areas for action as well as the current quality of the RCT evidence. The role of many of these variables can be controlled and / or adjusted for, but this occurs rarely. It is commendable that studies compare baseline characteristics of some variables (primarily age and gender).

However, the argument put forward, that if groups are equal at baseline they do not require adjusting for (e.g. Froh et al., 2009) in subsequent analyses, may be untrue. The more problematic issue is likely that the small sample sizes identified in many studies mean that studies are underpowered to detect difference, and for multivariate analyses. Groups are

rarely identical, and demographic factors may become relevant under different conditions during the intervention phase. Future research should consider these variables and allow for them when considering the statistical power required for multivariate analyses. Previous experience of treatment, contamination between groups, current activities and medications are relevant factors that are consistently neglected. Additionally, the role of age and gender to content and format acceptability (as implicated in Manicavasagar et al., 2014) requires greater elucidation.

Future research should also consider the nature of the active control group. Findings may be very different when comparing a 'negative' control group (for example, listing everyday hassles, Froh et al., 2008) versus a neutral group (e.g. treatment as usual). Some studies that label their control group neutral may in fact be dealing with an active one, for example, who are writing about feelings relating to activities (Froh et al.'s, 2009). In this case a no treatment control would be beneficial in facilitating robust conclusions.

Age and developmental stage is a salient factor that has been neglected. Gratitude develops in middle childhood (Emmons & Shelton, 2002) and it seems logical that it may alter or express differently alongside changes in cognitive or socio-emotional capability that accompany developmental stage such as early, middle, and late adolescence. Interestingly, however, there was no difference in gratitude according to age or grade in Froh et al. (2009).

Consider the design impact of the context on outcomes

The potential role of motivation in PPI findings requires further consideration. We concur with Froh et al. (2009) regarding the importance of volition in the classroom setting. When children are expected to complete the intervention, it removes an element of volition. When designing interventions to improve well-being, happiness or positive emotion, it is important

that the participants want to change and engage voluntarily with the intervention. Indeed, Sin & Lyubovitsky (2009) found a greater intervention effect for self-selected participants. At a minimum, one way of exploring this aspect is to have a non-active control as well as an active control.

Assess the long-term effects of PPI's and use appropriate measures

Many studies involved two time points (pre-and post). Increased follow up is therefore required to assess maintenance over time. Some studies used adult measures (e.g. the HADS, PANAS) despite existing child versions (e.g. PANAS-C) and this is ill-advised.

Consider the context

It has been suggested that a tendency to focus on the positive often neglects the contributing role of context (Ciarrochi et al., 2016). Because affect and positive affect are impacted by multiple contextual factors it is important to consider them in future research. For example, children of depressed mothers require greater intensity in the emotional stimulus than controls (Joorman, Gilbert, & Gotlib, 2010).

Qualitative research might be particularly important when evaluating the positive outcomes that children often report following major life challenges such as illness (Tong, Gow, Wong, Henning, & Carroll, 2015). Indeed there is particular interest in applying PPI's to populations with chronic and acute illness and distressing situations, for example, palliative care (e.g. Chavez et al., 2016). Research could benefit from evaluating this evidence further.

Consider and adapt the format and delivery accordingly

Innovative approaches to delivering PPI's in different delivery formats, for example, culturally informed storytelling or reading materials (Wood, Mayaba, & Theron, 2012), music

video / lyrics (Burns, Robb & Haase, 2009) or gaming may be a productive avenue in the future. Because funding is an issue when designing such innovative and resource-consuming evaluations, this review identifies the funding sources of the RCT literature (see the process and implementation table in the Appendices).

The intensity of intervention is most probably important, but remains to be fully evaluated, particularly regarding acceptability and cost. For example, daily gratitude intervention and repeated measurement may irritate children (e.g. Froh et al., 2008) and weekly gratitude exercises have been found more beneficial elsewhere (Sin & Lyobomirsky, 2004). The Bite back PPI (Burckhardt et al., 2015) highlighted difficulty in implementation. It was poorly accepted when delivered in a structured online format as part of the school program, in part due to its intensive resource requirements. How these PPI's perform under less controlled and resource constrained scenarios is unclear. RCT evaluations of interventions can be implemented in real world situations and in a cost-effective manner (Malloy, 2013). These implementation considerations are pivotal (although often neglected) in child and adolescent intervention evaluation literature.

This review focussed on interventions aimed at the individual child. Future reviews could consider indirect effects, and focus on interventions aimed at broader contexts (for example, community, family, parents). PPI's might be well suited to address recent findings of the moderating role of parental resiliency in child outcome (Rosenberg, Wolfe, Bradford, Shaffer, Yi & Frazier, Curtis, & Baker, 2014). How positive psychology can be used to benefit parents, particularly those of ill children, shows promise and could be the focus of specific reviews in the future.

Increasing conceptual clarity in the area

Some authors outline a logical theoretical and clinical framework underlying PPI's (for example, Ruini & Fava, 2012) while others do not, and this position could and should be addressed to understand the mechanisms by which an intervention might work.

There continue to be major issues in the literature. During the course of the review we identified other reviews that relate to Positive Psychology (Chodkiewicz & Boyle, 2017; Green, S., Anthony, T., & Rynsaardt, 2007). These tend to be unsystematic and discursive, and the plethora of interventions included under the PPI umbrella is bewildering, ranging from life coaching to CBT.

The challenge of adequately defining the parameters of positive psychology interventions remains within the Positive Psychology discipline. Recently authors have begun the process of deconstructing this quite complex issue (Pawelsky, 2016a; 2016b). This review contributes to this process, building on previous systematic reviews in the area. Its search protocol bypassed the problem caused by use of the term 'positive psychology'. It is an approach rather than an intervention term and is not a necessary limiting search term for future reviews. Over-dependence on the term in search strategies will lead to the omission of studies, as has been the case in previous reviews. Alternatively, it may lead to the return of an unmanageable number of search results, and a requirement to add additional limiters which may also lead to the omission of important evidence.

Process and implementation

Regarding the paper by Jaser, Russell, Rothman, Choi, & Whitemore (2014), an understanding of the non-response pattern would be an important part of any future replication, as it is unclear why 23 participants refused to participate at initial approach.

Despite this there was a high retention throughout the study, potentially be due to the nature

of the setting (i.e. participants may feel obliged to participate as they are dependent on the clinics and must return). While participants reported satisfaction with intervention (Jaser et al., 2014b) half of this group also reported not using the exercises, suggesting the intervention may not be acceptable. Reported satisfaction, despite non-engagement, may be due to the accompanying qualitative information which highlighted a significant motivation being due to financial / material reward (Jaser et al., *ibid.*). In Froh et al. (2009) contamination can be considered highly likely given that the intervention group discussed their experiences in the same classroom as the control. Data were collected and checked immediately after the author facilitated reflections in the classroom and participants were then instructed to complete missing data. This design decreases volition and increases the likelihood of a social desirability response. It is positive that there were implementation fidelity checks but these were not completed by two of the intervention arms (the 8th and 12th grades).

In Kwok et al.'s (2016) study there was no attrition following randomisation and the intervention was therapist-implemented under protocol. Social services assisted in recruitment and participants were selected based on having mild depression according to the HADS.

It was noted that in Froh et al. (2008) introductions to class and teachers were made independently prior to start by author; there were random integrity checks from author and principle; data were collected daily in class; on day eight one class failed to complete measures (the teacher forgot); and three 6th grade classes completed pre-test measures on a different day than other classes due to a field trip. Understandably, some children were observed to be irritated with repeatedly having to complete questionnaires.

In Suldo et al. (2014) randomisation was ineffective in balancing group differences.

Positively, however, fidelity checks led to 100% completion of planned activities. Attrition was 16.4%, mainly due to children leaving the area. Of those completing the intervention all completed every session of the 10 sessions.

In Roth et al. (2017) there was considerable implementation checks. A high attendance during all sessions was indicated with at least twenty of 21 participants attending. Homework completion was checked in session, and verbally through conversations with child and parent, and tracked through a points system. Approximately 67% of parents (n= 14) attended the information session despite repeated engagement attempts for the remaining seven parents and the offer of four sessions. Qualitative feedback from children was positive; for example, that they enjoyed everything about the sessions (19%) or would not change anything about them (50%), or commented on growth during intervention (57%).

Manicavasagar et al. (2014) was a self-directed intervention and a substantial proportion of under 16's did not go on to participate because they needed parent consent. Significant changes were only found for highly engaged participants (at least 30 minutes and 3 times per week). Feedback suggested content acceptability may vary according to age, with older adolescents (>16) reporting less acceptability. Treatment dose was not maintained but was deemed particularly important. In fact, it determined whether intervention affected outcome.

In Burkhardt et al. (2015) the school based online program ('Bite back') was delivered as part of the curriculum during school time, and the considerable resources required are highly relevant here. These involved supporting and training the school; teacher time in delivery; developing content; website moderator to pre-screen user posts; and clinical psychologist

advisor. This directly influenced PPI implementation as shown, for example, where reliance on teacher time in delivery led to implementation failure in one school.

In Ruini et al (2009) one class (n = 25) assigned to intervention was lost to follow up because they graduated. We note that effect sizes in the unexpected negative direction on environmental mastery were found for both groups (in fact these were the largest effect sizes found, at follow up; $d = .42$ in intervention; $-.40$ for control). We also noted the duration of sessions (two hours) in this study, and the extra session, compared to the 2006 study. Further, while the authors suggest ease of application underlined measure choice, the 92 item symptoms measure imposed a significant response burden in practice.

Section 2 Implications for child and adolescent mental health measurement and the SDQ

Positive Psychological Interventions, however scattered, uncertain, and ill-defined, are potentially of significant benefit to children with mental health dysfunction. However, even using the limited number of measures which are available, it is essential that interventions be developed from a measured and balanced baseline, and followed up with a post-intervention assessment. Without this being available, clinical input from a range of disciplines cannot be properly assessed for effectiveness and direction, or may be inappropriate to the situation.

Importantly, it will not allow what should be an integral contribution to the development of public policy and mental health and well-being.

Apart from developing towards clinical and assessment excellence, the informing of public policy is essential for the development of better services. This will be discussed later in this section.

The instrument of focus in our investigation, the Strengths and Disabilities Questionnaire, scans for child mental health and pathology. It is not the first measure developed to assess maladjustment among children in an objective manner. The history of assessment identifies several measures with related underlying concepts, that have attempted to address the perceived limitations of its predecessors. Most have been incorporated into research relating to public health monitoring and clinical assessment, similarly to the SDQ

For example, one of the first instruments that became available was the Bristol Social Adjustment Guides (BSAG; Stott, 1963;1969;1987), which assessed 12 ‘syndromes,’ including for example, Unforthcomingness, Withdrawal, Restlessness/Impulsiveness, peer relationships, Hostility towards adults, and Anxiety. Factor analyses (Shepherd, 2013; Mc Dermot, 1980) suggest underlying domains (Under reaction and Over reaction; Externalised or Internalised behaviour) that appear to bear, on a simple level at least, some indication of comparison with the three-factor groupings often used in the SDQ.

Rutter (1967) criticised the BSAG and other instruments for reasons relating to diagnostic distinction, lack of recent validity data, that it was too long, and that it was not suitable for pre-adolescents. He expressed the need for a reliable, valid and short questionnaire which could be easily (and quickly) administered. From this position, he developed his Children’s Behaviour Questionnaire (Children’s Behaviour Checklist or CBCL), a 118-item checklist focused entirely on psychopathology or difficulties.

The SDQ, published in 1997, and as underlined in chapter 2, has become perhaps the most widely used screening instrument. Goodman (1997) pointed to four primary differences between the CBCL and SDQ. The latter was brief (25 items), it could be administered by a parent or teacher, it evaluated strengths as opposed to difficulties alone, and was accepted by

parents of non-maladjusted children. This comparison study underlined many similarities between two scales, but underlined the significant difference in time needed to administer them, and the presence of the strengths or Prosocial factor.

Subsequently, as described in chapter 2, it became evident that there were conflicting psychometric findings in relation to the SDQ, that the literature had been affected by several methodological issues. A major conclusion of our study was the fact that inappropriately analysing the SDQ using a continuous rather than categorical measure had seriously affected model choice and reliability findings. This continues to occur (e.g. Bowman et al., 2016) despite available techniques to treat data as ordered. Our study concluded that it was more appropriate to use and interpret the 5-factor rather than the 3-factor structure, and confirmed it was acceptable to interpret subscale as well as total scores. Important findings were the invariance of gender, SES status and parental status, and the updating of norms and revision of the total difficulties score cut-off point.

Given the prevalence rates for mental health at diagnoses in the population group is estimated at between 10-20%, the 10% cut off point is likely a conservative one, although now likely to include a larger group of maladjusted as the total difficulties score has reduced. The study identified several expected relationships which support the validity of the SDQ and poorer scores were correlated with the presence of currently treated mental health condition, and with chronic limiting physical conditions.

The question of why some children have emotional difficulty and others do not is of interest to those introducing strategies to manage, mitigate, and improve the health of children.

Accurate cut-off points are vital. To highlight this, we consider an example where we might wish to identify relationships between known risk variables (SES) and mental health

problems, perhaps with a view to funding services. When we compare the top two versus bottom two SES indicator groups (NSEC 5) using the old -offs the percentage at risk of mental health problems is over twice as high (8.04% [n= 226/2809]) in the bottom 2 NSEC indicator groups cut-offs (excluding never worked/unemployed) compared to the top SES groups (1.4% [n=339/2329]). When we use the new cut-off points this rises significantly to 14.05% and 6.7% respectively. Thus 11% more cases are identified in the new cut-off system.

It is notable that according to the old cut-offs, of the 1315 children with the long-standing illness, 21.6% are classified as high risk of having a mental health problem, which contrasts with 6.39% (n=310/4849) children without long-standing illness. This point should be highlighted because it has considerable ramifications for public health policy and service provision within this population. More research is warranted here to monitor and support this group and to indicate intervention. National databases such as the WHS allow vulnerable and at risk groups, such as those with chronic illness, to be identified and supported. Related to this group, interventions that improve physical and mental well-being might be particularly important. The influence of Well-being therapy (Ruini et al., 2006; 2009; Tomba et al., 2010) on these aspects warrants further exploration.

One point arguing persistently towards the use of the SDQ is that it looks at strengths as well as difficulties. However, the four difficulty factors use 20 items, the strengths factor only five, and the reason for this imbalance is unclear. Normative data in table 7 is both revealing and puzzling and raises a question about the usefulness of the prosocial strengths factor. It appears that higher scores tell us nothing useful as most children score at the maximum. Because of this limited spread of scores the prosocial scale lacks discriminative ability and is of limited use. In fact, it identifies difficulties on the scale, in contrast to stated purpose. All

cases upwards of the 70th percentile present maximum points for boys, and this also happens from the 60th percentile for girls. It would seem a sensible cut-off point would have to be at least as low as the 35th percentile for boys, and the 25th percentile for girls.

Once this has been pointed out it should draw attention in future studies, or indeed in clinical use. Significant directed work needs to be done to establish why the problem exists and how to solve it. The prosocial scale does not demonstrate strengths, based on these findings. How these limited items can be useful to clinicians remains to be seen. It indicates difficulties in few items that arguably relate to social communication, and have no meaningful spread of strengths that can inform intervention. This raises the question as to measure appropriateness in clinical contexts. How useful is the SDQ clinically in identifying strengths and is a longer measure more helpful in informing intervention? The comprehensive Via-Youth measure identified in chapter 1 seems better suited to this task.

On a further critical note, item-factor loadings on Q3 and Q 23 can only be interpreted as ‘fair’ to ‘good,’ with all remaining questions ranging from ‘good’ to ‘excellent.’ Question 3 contributes to the Emotional factor, and question 23 to the Peer problem factor. It is not clear why this should occur, or indeed whether both items are expressed in too general fashion to be useful, and might even be altered or substituted by more effective questions in a subsequent revision.

Returning to the issue of public health policy. The WHS used the SDQ since its inception in 2003. Now, however, the Welsh Government has ceased the Welsh Health Survey. The Mental Health Foundation (Mental Health in Wales: Fundamental Facts, 2016) states that: “The Welsh Government are in the process of developing a new survey of adults which will

include health related questions and will replace the Welsh Health Survey Welsh Government's 10-year strategy for mental health and wellbeing..." (p.2).

The Mental Health (Wales) Measure (2010) and the Wellbeing of Future Generations (Wales) Act (2015) place a legal responsibility on local authorities to monitor and support the well-being of citizens. They provide a useful framework that can inform design of future population research aimed at improving well-being and mental health. They identify a 'healthier Wales' as paramount and key national 'indicators are being produced to monitor progress towards these goals' (p. 3). We hope that these national indicators take account of strengths to a further degree than did the WHS (and its SDQ indicator) and that there is a role for children in creating and contributing to their mental health, well-being, and related treatment from services. How do children and adolescents understand well-being and happiness, and what do they view as important?

It is essential that nothing is lost in the transition to these new indicators. However, the fact that the WHS was terminated in 2016, before a follow-on program was thought out and tested is worrying. From the little information available, it appears that reduction in cost is one of the prime movers of the change in policy. It is now unclear how the authorities plan to monitor and improve the provision of well-being services and support for children with mental health problems. It is unclear how new developing and older but improved PPI approaches will be supported.

This study has focused on the literature around PPI, and pointed out a direction to follow, and the steps and corrections it needs to take. It has also re-normed, and clearly validated a ubiquitous measure of child mental health.

It is pivotal that the work started before Stott, and by him, and continued by Rutter and then Goodman to produce this very workable version of the SDQ, be continued. Because the SDQ was developed to relate to diagnostic criteria which have since been updated, development to match current understanding of diagnosis may be appropriate. Importantly, a measure that more comprehensively indicates strengths that can inform intervention is required.

The existing WHS database only makes practical sense if it is continued. Informed research needs to be carried out continuously to provide policymakers with the information and direction they need: simple statistical information on the presence and extent of maladjustment and general ill-health, but in addition informed advice on new interventions, new test measures, new notions such as wellness and well-being, and advice on how these might best be developed within a supportive health service. It is hoped that the Government understand this, and will provide a productive, constructive set within which families and care providers can flourish.

References

- Botvinick, M., & Cohen, J. (1998). Rubber hands' feel'touch that eyes see. *Nature*, *391*(6669), 756.
- Bech, P., Olsen, L. R., Kjoller, M., & Rasmussen, N. K. (2003). Measuring well- being rather than the absence of distress symptoms: a comparison of the SF- 36 Mental Health subscale and the WHO- Five well- being scale. *International journal of methods in psychiatric research*, *12*(2), 85-91.
- Bergsma, A., Have, M. T., Veenhoven, R., & Graaf, R. D. (2011). Most people with mental disorders are happy: A 3-year follow-up in the Dutch general population. *The Journal of Positive Psychology*, *6*(4), 253-259.
- Boman, F., Stafström, M., Lundin, N., Moghadassi, M., Törnåge, C. J., & Östergren, P. O. (2016). Comparing parent and teacher assessments of mental health in elementary school children. *Scandinavian journal of public health*, *44*(2), 168-176.
- Burns, D. S., Robb, S. L., & Haase, J. E. (2009). Exploring the feasibility of a therapeutic music video intervention in adolescents and young adults during stem-cell transplantation. *Cancer nursing*, *32*(5), E8-E16.
- Chodkiewicz, A. R., & Boyle, C. (2017). Positive psychology school- based interventions: A reflection on current success and future directions. *Review of Education*, *5*(1), 60-86.
- Ciarrochi, J., Atkins, P. W., Hayes, L. L., Sahlra, B. K., & Parker, P. (2016). Contextual positive psychology: policy recommendations for implementing positive psychology into schools. *Frontiers in psychology*, *7*.
- Connolly, J. F. (2013). *Stress and coping in university employees: a longitudinal evaluation of stress, personality, coping and psychological distress* (Doctoral dissertation, Queen Margaret University, Edinburgh).

- Connolly, J. (2016). The Role of Attentional Processes in Medically Unexplained Symptoms. *Irish Psychologist*, 43(02), 48-55.
- Drvaric, L., Gerritsen, C., Rashid, T., Bagby, R. M., & Mizrahi, R. (2015). High stress, low resilience in people at clinical high risk for psychosis: Should we consider a strengths-based approach?. *Canadian Psychology/Psychologie canadienne*, 56(3), 332.
- Emmons, R. A., & McCullough, M. E. (2004). *The psychology of gratitude*. Oxford University Press.
- Fecteau, S., Pascual-Leone, A., Zald, D. H., Liguori, P., Théoret, H., Boggio, P. S., & Fregni, F. (2007). Activation of prefrontal cortex by transcranial direct current stimulation reduces appetite for risk during ambiguous decision making. *Journal of Neuroscience*, 27(23), 6212-6218.
- Freire, T., Teixeira, A., Silva, E., & Matias, G. P. (2015). Interventions for youth depression: From symptom reduction to well-being and optimal functioning. *Journal of Behavior, Health & Social Issues*, 6(2), 9-19.
- Froh, J. J., Kashdan, T. B., Ozimkowski, K. M., & Miller, N. (2009). Who benefits the most from a gratitude intervention in children and adolescents? Examining positive affect as a moderator. *The Journal of Positive Psychology*, 4(5), 408-422.
- Froh, J. J., Sefick, W. J., & Emmons, R. A. (2008). Counting blessings in early adolescents: An experimental study of gratitude and subjective well-being. *Journal of school psychology*, 46(2), 213-233.
- Ghodsian, M. (1977). Children's behaviour and the BSAG: some theoretical and statistical considerations. *British Journal of Social and Clinical Psychology*, 16(1), 23-28.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A Research Note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.

- Green, S., Grant, A., & Rynsaardt, J. (2007). Evidence-based life coaching for senior high school students: Building hardiness and hope. *International Coaching Psychology Review*, 2(1), 24-32.
- Harrison, A., Al-Khairulla, H., & Kikoler, M. (2016). The feasibility, acceptability and possible benefit of a positive psychology intervention group in an adolescent inpatient eating disorder service. *The Journal of Positive Psychology*, 11(5), 449-459.
- Hetrick, S. E., Cox, G. R., Witt, K. G., Bir, J. J., & Merry, S. N. (2016). Cognitive behavioural therapy (CBT), third- wave CBT and interpersonal therapy (IPT) based interventions for preventing depression in children and adolescents. *The Cochrane Library*.
- Jaser, S. S., Patel, N., Rothman, R. L., Choi, L., & Whittemore, R. (2014a). Check it! A randomized pilot of a positive psychology intervention to improve adherence in adolescents with type 1 diabetes. *The Diabetes Educator*, 40(5), 659-667.
- Joormann, J., Gilbert, K., & Gotlib, I. H. (2010). Emotion identification in girls at high risk for depression. *Journal of Child Psychology and Psychiatry*, 51(5), 575-582.
- Kini, P., Wong, J., McInnis, S., Gabana, N., & Brown, J. W. (2016). The effects of gratitude expression on neural activity. *NeuroImage*, 128, 1-10.
- Kwok, S. Y., Gu, M., & Kit, K. T. K. (2016). Positive psychology intervention to alleviate child depression and increase life satisfaction: A randomized clinical trial. *Research on social work practice*, 26(4), 350-361.
- Lazarus, R. S. (2003). Does the positive psychology movement have legs?. *Psychological inquiry*, 14(2), 93-109.
- Linley, P. A., Joseph, S., Maltby, J., Harrington, S., & Wood, A. M. (2009). Positive psychology applications. Oxford Handbooks Online.

- Manicavasagar, V., Horswood, D., Burckhardt, R., Lum, A., Hadzi-Pavlovic, D., & Parker, G. (2014). Feasibility and effectiveness of a web-based positive psychology program for youth mental health: randomized controlled trial. *Journal of medical Internet research, 16*(6), e140.
- McDavid, L., McDonough, M. H., & Smith, A. L. (2015). An empirical evaluation of two theoretically-based hypotheses on the directional association between self-worth and hope. *Journal of adolescence, 41*, 25-30.
- McDermott, P. A. (1980). Principal components analysis of the revised Bristol Social Adjustment Guides. *British Journal of Educational Psychology, 50*(3), 223-228.
- Molloy, L. E., Moore, J. E., Trail, J., Van Epps, J. J., & Hopfer, S. (2013). Understanding real-world implementation quality and “active ingredients” of PBIS. *Prevention science, 14*(6), 593-605.
- Pawelski, J. O. (2016a). Defining the ‘positive’ in positive psychology: Part I. A descriptive analysis. *The Journal of Positive Psychology, 11*(4), 339-356.
- Pawelski, J. O. (2016b). Defining the ‘positive’ in positive psychology: Part II. A normative analysis. *The Journal of Positive Psychology, 11*(4), 357-365.
- Rosenberg, A. R., Wolfe, J., Bradford, M. C., Shaffer, M. L., Yi-Frazier, J. P., Curtis, J. R., & Baker, K. S. (2014). Resilience and psychosocial outcomes in parents of children with cancer. *Pediatric blood & cancer, 61*(3), 552-557.
- Ruini, C., Belaise, C., Brombin, C., Caffo, E., & Fava, G. A. (2006). Well-being therapy in school settings: a pilot study. *Psychotherapy and Psychosomatics, 75*(6), 331-336.
- Ruini, C., & Fava, G. A. (2009). Well-being therapy for generalized anxiety disorder. *Journal of Clinical Psychology, 65*(5), 510-519.

- Ruini, C., & Fava, G. A. (2012). Role of Well-Being Therapy in Achieving a Balanced and Individualized Path to Optimal Functioning. *Clinical Psychology & Psychotherapy*, *19*(4), 291–304.
- Rutter, M. (1967). A children's behaviour questionnaire for completion by teachers: preliminary findings. *Journal of child psychology and psychiatry*, *8*(1), 1-11.
- Shepherd, P. (2013). Bristol Social Adjustment Guides at 7 and 11 years. *Centre for Longitudinal Studies*.
- Sin, N. L., & Lyubomirsky, S. (2009). Enhancing well- being and alleviating depressive symptoms with positive psychology interventions: A practice- friendly meta-analysis. *Journal of clinical psychology*, *65*(5), 467-487.
- Stevens, M., Roberts, H., & Shiell, A. (2010). Research review: Economic evidence for interventions in children's social care: revisiting the what works for children project. *Child & Family Social Work*, *15*(2), 145-154.
- Stott, D. H., & Sykes, E. G. (1963). *Bristol Social-adjustment Guides*. University of London Press. Retrieved, 10-04-17 from, <http://www.cls.ioe.ac.uk>.
- Suldo, S. M., & Shaffer, E. J. (2008). Looking beyond psychopathology: The dual-factor model of mental health in youth. *School Psychology Review*, *37*(1), 52.
- Suldo, S. M., Savage, J. A., & Mercer, S. H. (2014). Increasing middle school students' life satisfaction: Efficacy of a positive psychology group intervention. *Journal of happiness studies*, *15*(1), 19-42.
- Toback, R. L., Graham-Bermann, S. A., & Patel, P. D. (2016). Outcomes of a Character Strengths–Based Intervention on Self-Esteem and Self-Efficacy of Psychiatrically Hospitalized Youths. *Psychiatric Services*, *67*(5), 574-577.

Tomba, E., Belaise, C., Ottolini, F., Ruini, C., Bravi, A., Albieri, E., ... & Fava, G. A. (2010).

Differential effects of well-being promoting and anxiety-management strategies in a non-clinical school setting. *Journal of Anxiety Disorders*, 24(3), 326-333.

Wood, L., Theron, L., & Mayaba, N. (2012). 'Read me to resilience': Exploring the use of

cultural stories to boost the positive adjustment of children orphaned by AIDS.

African Journal of AIDS Research, 11(3), 225-239.

Appendices

Appendix 1a Jadad (1996) rating scale

Item	Jadad rating scale	Score
1	Was the study designed as randomised?	
2	Was the study designed as double blind?	
3	Was there a description of withdrawals and drop outs?	
4	Were the objectives of the study defined?	
5	Were the outcome measures defined clearly?	
6	Was there a clear description of the inclusion and exclusion criteria?	
7	Was the sample size justified (for example, power calculation)?	
8	Was there a clear description of the interventions?	
9	Was there at least one control (comparison) group?	
10	Was the method used to assess adverse effects described?	
11	Were the methods of statistical analysis described?	
Jadad total:		/13

Note: (1 = Yes; 0 = No); additional points can added or deducted (creating a total of 13)

depending on the appropriateness of randomisation, blinding etc.

Appendix 1b Method Appraisal Scale

Item	Method appraisal scale	Score
1	Are the results of the study valid? (See CASP criteria)	
2	Representative sample?	
3	Appropriate comparison group?	
4	Baseline response >60%?	
5	Follow-up (80% in cohort/ 60% in cross section)?	
6	Adjustment for non-response and drop out?	
7	Conclusions substantiated by the data?	
8	Adjustment for confounders?	
9	All intervention group exposed; non-contaminated comparison group?	
10	Appropriate statistical tests?	

Method appraisal scale total:

/ 10

Note: Adapted from Connolly et al. (2013); (1 = Yes; 0 = No); In calculating baseline response, if a study states that there was an initial sample (e.g. various schools were invited) but only provides statistic on those that participated, they receive a 0 mark.

Appendix 2 Adjustment for confounding variables

	Burckhardt (2015)	Manicavasagar et al. (2014)	Jaser et al., (2014)	Kwok et al., (2016)	Froh et al. (2009)	Froh et al., (2008)	Toback et al., (2016)	Roth et al., 2017	Suldo et al., 2013	Ruini et al., 2006	Ruine et al., 2009	Tomba et al., 2010
Age	X	X ^B	X ^B	X ^B	X ^B	X	X	X	X ^B	X	X ^B	X
Gender	X	X ^B	X ^B	X ^B	X ^B	√	X ^B	X	X ^B	X	X ^B	√
Ethnicity	X	X	X ^B	X ^B	X ^B	X	X ^B	X	X ^B	X	X	X
Adherence	X ^B	√	√	-		X	-	X	X	X	X	X
Religiosity (e. g. importance of god)	X	X	X	X	X ^B	X	X	X	X	X	X	X
Parental marital status or guardianship	X	X	X	X	X	X	X	X ^B	X ^B	X	X	X
Unit of randomisation (e.g. if block randomised by school class)	Block	Ind	Ind	Block	Ind	Block	Ind	Stratified individual	Ind (& propensity)	Ind	Block	Block
Drop out analysis	√	X ^{SNR}	X	-	-	X		-	X	X	X	X
Group Interaction	X	-	X	-	X	-	X	X	X	X	X	X
Previous treatment effects	X	X	X	-	X	X	X	X	X	X	X	X
Medical diagnoses / medication	X	X	√	X	X	X	X	X	X	X	X	X
Current activities	X	X	X	X	X	√	X	X	X	X	X	X
Analysis of group differences on outcome measure at baseline	√	√	√	√	√	X	√	√	√	X ^{SNR}	X ^{SNR}	√

Adjustment for confounder (x = no; √ = yes) = refers to statistical adjustment / analysis of variable on outcome post intervention. B = Baseline statistical analyses of group differences according to this variable. RCT = Randomised controlled trial; SNR = Stated but methods, analyses, or test values not fully reported; - = not applicable; 'Religiosity' was included because religious people report more gratitude (see Froh et al., 2009 for a discussion) and may therefore confound results; Kwok et al (2016) explored if children were still living with their parents (all children were); Group interaction refers to attempts to control or manage potential interaction between groups, deemed relevant in intervention setting where participants may have contact with each other i.e. schools; Dropout adjustment = refers to statistical analysis of dropouts; Suldo et al., 2014 used a propensity score matching process because randomisation did not lead to equivalence between groups on outcome; Table adapted from Connolly et al., 2013.

Appendix 3 Study appraisal scores and Inter-rater reliability

Study	Author	Method appraisal score: Reviewer 1-JC	Method appraisal score: Reviewer 2-JM	Jadad score Reviewer 1-JC	Jadad score Reviewer 2-JM
1	Manicavasagar et al., (2014)	4	5	9	9
2	Burckhardt et al., 2015	1	4	8	8
3	Jaser et al., 2016	6	7	9	8
4	Kwok et al., (2016)	6	7	10	10
5	Froh et al., (2009)	7	8	8	8
6	Froh et al., 2008	4	4	5	7
7	Toback et al., 2016	4	4	8	7
8	Suldo et al., 2014	4	4	9	9
9	Roth et al., 2017	6	6	8	7
10	Ruini et al., 2006	5	4	7	7
11	Ruini et al., 2009	3	4	7	7
12	Tomba et al., 2010	8	8	8	8

Note: The agreement between raters was high (Intra-class correlation coefficients = .916) on the method appraisal scores.

Appendix 4 Normative data for the SDQ parent version (4-12 years) – a quick reference of cut-off points for clinicians.

		High-risk (90-95 th percentile)	Very High-risk (top 5%, 95-100 th percentile)
Total difficulties	Boys	17-20	21-40
	Girls	15-18	19-40
	Total	16-18	20-40
Prosocial behaviour	Boys	4-5	<4
	Girls	5-6	<5
	Total	5-10	0-4
Hyperactivity	Boys	7-8	9-10
	Girls	6-7	8-10
	Total	7	8-10
Emotional symptoms	Boys	4-5	6-10
	Girls	5	6-10
	Total	4	5-10
Conduct problems	Boys	4	5-10
	Girls	4	5-10
	Total	4	5-10
Peer problems	Boys	4	5-10
	Girls	4	5-10
	Total	4	5-10

We subdivide the top 10% of scorers as this approach has been observed in the literature and allows both a top 5 and top 10% analysis.

Appendix 5 The Strengths and Difficulties Questionnaire

Strengths and Difficulties Questionnaire

Shift+Ctrl+S

For each item, please mark the box for Not True, Somewhat True or Certainly True. It would help us if you answered all items as best you can even if you are not absolutely certain or the item seems daft! Please give your answers on the basis of the child's behaviour over the last six months or this school year.

Child's Name

Male/Female

Date of Birth

	Not True	Somewhat True	Certainly True
Considerate of other people's feelings	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Restless, overactive, cannot stay still for long	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often complains of headaches, stomach-aches or sickness	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Shares readily with other children (treats, toys, pencils etc.)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often has temper tantrums or hot tempers	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Rather solitary, tends to play alone	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Generally obedient, usually does what adults request	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Many worries, often seems worried	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Helpful if someone is hurt, upset or feeling ill	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Constantly fidgeting or squirming	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Has at least one good friend	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often fights with other children or bullies them	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often unhappy, down-hearted or tearful	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Generally liked by other children	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Easily distracted, concentration wanders	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Nervous or clingy in new situations, easily loses confidence	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Kind to younger children	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often lies or cheats	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Picked on or bullied by other children	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Often volunteers to help others (parents, teachers, other children)	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Thinks things out before acting	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Steals from home, school or elsewhere	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Gets on better with adults than with other children	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Many fears, easily scared	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sees tasks through to the end, good attention span	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Appendix 6 Confirmation of ethical approval from Bangor University Research Ethics Committee

ethics@bangor.ac.uk

Mon 25/07/2016, 09:19 John Francis Connolly

Inbox

Dear John,

2016-15761 Measurement and predictors of child emotional health in a representative national sample: A Welsh Health Survey Analysis.

Your research proposal number 2016-15761 has been reviewed by the Psychology Ethics and Research Committee and the committee are now able to confirm ethical and governance approval for the above research on the basis described in the application form, protocol and supporting documentation. This approval lasts for a maximum of three years from this date.

Ethical approval is granted for the study as it was explicitly described in the application

If you wish to make any non-trivial modifications to the research project, please submit an amendment form to the committee, and copies of any of the original documents reviewed which have been altered as a result of the amendment. Please also inform the committee immediately if participants experience any unanticipated harm as a result of taking part in your research, or if any adverse reactions are reported in subsequent literature using the same technique elsewhere.

Appendix 7 Additional process and implementation data pertaining to PPI's

Study	Response / Implementation / process
Jaser et al., 2014	<p><i>Response/ Implementation / process:</i> 24 of the 73 families approached refused to participate and 9 were ineligible. No drop outs in either condition; 95% completed measures at 3 & 6 months; 75% participation in calls for intervention and 90% in control; Research assistant made follow up phone calls. Biggest challenge reported by participants was completing phone calls.</p> <p><i>On-going Program?</i> No</p> <p><i>Funded?</i> Grant funded (NIH & Yale School of Nursing)</p>
Kwok et al., 2016	<p><i>Response/ Implementation / process:</i> 9 suicidal students excluded and referred elsewhere; school social workers met all participants. Attrition = 0%; Therapist ran groups; followed program protocol. Participants and therapists asked to not to share information with others outside groups; Groups equal at baseline.</p> <p><i>On-going Program?</i> No</p> <p><i>Funded?</i> -</p>
Froh et al., 2009	<p><i>Response/ Implementation / process:</i> Both groups conducted in the same classroom Data collected in the classroom immediately after the facilitated reflection with the author (contamination). Donuts were provided regardless of participation. No parent forms were returned for 8th & 12th grader conditions.</p> <p><i>On-going Program?</i> No</p> <p><i>Funded?</i> In part by a National Institute of Mental Health grant</p>
Toback et al., 2016	<p><i>Response/ Implementation / process:</i> 15 (37%) failed to implement intervention; Completion at 2 week follow up was 61 (75%); at 3 months was 49 (60%); VIA-Youth (Publicly available at www.viame.org) completed online with support to clarify questions and monitor online use; 88% (36) described the intervention as positive and agreed with identified strengths. Control group activity significantly shorter than intervention;</p> <p><i>On-going Program?</i> no</p> <p><i>Funded?</i> -</p>

Study	Response / Implementation / process
Froh et al., 2008	<p><i>Response/ Implementation / process:</i> Introductions to class and teachers independently prior to start by author; Random integrity checks from author and principle; Data collected daily in class; On day 8, one class failed to complete measures (teacher forgot); 3x 6th grade classes completed pre-test measures on a different day than other classes due to a field trip; 2/3 teachers were blinded to hypotheses and other conditions; however 1 ‘blinded’ teacher implemented two conditions; Part of mandatory curriculum; donuts offered.</p> <p><i>On-going Program?</i> No</p> <p><i>Funded?</i> -</p>
Roth et al., 2017	<p><i>Response/ Implementation / process:</i> All 274 students completed a Life satisfaction scale (BMSLSS); low scorers (<6) invited to participate (n = 111). Subsequent to parental consent, candy bars and entry to raffle was received (\$25 gift card); 42 (38%) consented. Booster session timing adapted to fit school timetable; Groups facilitated by a professor/psychologist and a coleader (doctoral candidate); 10 hours of intervention received on average. Considerable implementation & fidelity checks; Positive qualitative feedback from participants (enjoyment and satisfaction with program).</p> <p><i>On-going Program?</i> No</p> <p><i>Funded?</i> -</p>
Suldo et al., 2014	<p><i>Response/ Implementation / process:</i> Opt out letter sent to all parents; 1 refused; 333 then completed a Life satisfaction scale (BMSLSS); low scorers (<6) invited to participate (n = 132).; 67 (58%) consented; Attrition: 11 dropped out (16%) during intervention mainly due to moving from the area (n=10). Final analyses based on n = 40 /67 (60%); Leaders (n =6) were school psychologists and doctoral candidates; collaboratively planned the next session together and co facilitator checked off activities and paced sessions. 100% fidelity with planned activities; Sessions chosen by child to take the place of an elective class; small treats received for homework completion;</p> <p><i>On-going Program?</i> No</p> <p><i>Funded?</i> : University of South Florida Collaborative for Children, Families and Communities.</p>
Ruini et al., 2006	<p><i>Response/ Implementation / process:</i> 2 clinical psychologists implemented both interventions.</p> <p><i>On-going Program?</i> No</p> <p><i>Funded by:</i> Part funded by Fondazione Carisbo (Bologna, Italy) and Ministero dell’Università e della Ricerca Scientifica e Tecnologica (Rome).</p>

Study	Response / Implementation / process
Ruini et al., 2009	<p><i>Response/ Implementation / process:</i> ‘various’ schools invited, one participated; 1 entire class (n = 25) assigned to intervention was lost to follow up (graduated); Twice as many missing in the intervention group at baseline (7 v 3) and post-test (13 v 6)</p> <p><i>On-going Program?</i> No</p> <p><i>Funded?</i> Fondazione Cassa di Risparmio di Cesena and SOS Telefono Azzurro Bologna.</p>
Tomba et al., 2010	<p><i>Response/ Implementation / process:</i> At post-test 3 intervention and 4 control participants were missing; at six months 5 intervention and 8 control were missing; Implemented by 2 clinical Psychologists; Sessions audio taped, randomly checked by blinded evaluator as to which group they belonged to.</p> <p><i>On-going Program?</i> No</p> <p><i>Funded?</i> SOS Telefono Azzurro Bologna.</p>
Burckhardt et al., 2015	<p><i>Response/ Implementation / process:</i> High attrition (59%); 2 schools dropped out, one citing negative feedback, lack of child understanding of purpose, and boredom; the other due to staffing; Of 572, 338 completed measures and enrolled (110 could not complete measure due to technical difficulties); treatment dose varied by school.</p> <p><i>On-going Program?</i> Yes; website content (www.biteback.org.au).</p> <p><i>Funded?</i> Grant from the Department of Health and Ageing</p>
Manicavasagar et al., 2014	<p><i>Response/ Implementation / process:</i> High attrition. Of 695 interested, 235 were enrolled. Of these, 71% (167) remained at follow up. Attrition in intervention (58 / 52%) more than twice that of control (23 / 20%); Treatment dose was not achieved by most and varied by condition; Many younger participants (<16) did not proceed due to requiring parental consent; Not all intervention group exposed (n = 13 went to wrong website and were excluded).</p> <p><i>On-going Program?</i> Yes; website content available (www.biteback.org.au).</p> <p><i>Funded?</i> Grant from the Department of Health and Ageing</p>