

**Measuring Facilitator Competent Adherence and Examining its Role in the Outcomes of Parenting Programme Beneficiaries: An Investigation of the Broader Literature and the Delivery of Parenting for Lifelong Health for Parents and Adolescents (PLH-Teens) at Scale in Tanzania**

Doctoral dissertation, in partial fulfilment of the requirements of the Doctor of Philosophy  
in Social Intervention and Policy Evaluation

Mackenzie Martin

Hertford College

Department of Social Policy and Intervention

University of Oxford

June 2023

**Supervisors:** Professor Frances Gardner and Dr Jamie Lachman

**Word count:** 79,551

## Table of Contents

<b>TABLE OF CONTENTS .....</b>	<b>2</b>
<b>LIST OF TABLES AND FIGURES .....</b>	<b>7</b>
<b>ACKNOWLEDGEMENTS .....</b>	<b>10</b>
<b>ABSTRACT .....</b>	<b>14</b>
<b>1. INTRODUCTION .....</b>	<b>20</b>
<b>2. BACKGROUND LITERATURE AND THEORETICAL FOUNDATION ..</b>	<b>22</b>
<b>2.1 Parenting Programmes .....</b>	<b>22</b>
2.1.1 Defining Parenting Programmes .....	23
2.1.2 Parenting Programmes to Reduce Violence Against Children and Child Behavioural and Emotional Problems.....	24
2.1.3 Parenting Programmes at Scale.....	28
<b>2.2 Implementation Science .....</b>	<b>31</b>
2.2.1 RE-AIM Framework .....	32
2.2.2 Defining Implementation .....	33
2.2.3 Implementation Outcomes .....	33
2.2.4 Implementation Fidelity .....	34
2.2.5 Facilitator Competent Adherence .....	38
2.2.6 Conceptualisation of the Relationship between Implementation and Outcomes.....	40
2.2.7 Approach to the Study of Implementation .....	40
<b>2.3 Parenting for Lifelong Health .....</b>	<b>41</b>
2.3.1 Evidence Base for PLH-Teens .....	42
2.3.2 PLH-Teens in Tanzania.....	44
2.3.3 Furaha Adolescent Implementation Research (FAIR) Study.....	49
2.3.4 Note on Candidate's Role .....	50
<b>3. REVIEW OF LITERATURE UNDERPINNING THE METHODS.....</b>	<b>51</b>
<b>3.1 Measures of Facilitator Competent Adherence.....</b>	<b>52</b>
3.1.1 Domains of Competent Adherence .....	53
3.1.2 Mode of Data Collection and Assessors .....	54
3.1.3 Response Types.....	57
3.1.4 Dominant Measure Types .....	58

3.1.5	Challenges of Measuring Competent Adherence using Observational Methods.....	60
<b>3.2</b>	<b>Reliability and Validity of Measures of Competent Adherence.....</b>	<b>62</b>
3.2.1	Psychometric Analyses .....	62
3.2.2	Measure Reliability .....	63
3.2.3	Measure Validity .....	68
3.2.4	Parenting for Lifelong Health for Young Children-Facilitator Assessment Tool .....	71
<b>3.3</b>	<b>Competent Adherence and Relationship to Outcomes .....</b>	<b>75</b>
3.3.1	Evidence from the Broader Implementation Science Literature .....	75
3.3.2	Evidence from the Parenting Programme Literature.....	76
3.3.3	Potential Reasons for Mixed Evidence .....	77
<b>4.</b>	<b>DISSERTATION OVERVIEW .....</b>	<b>78</b>
<b>4.1</b>	<b>Rationale for the Dissertation .....</b>	<b>78</b>
<b>4.2</b>	<b>Research Questions .....</b>	<b>80</b>
<b>4.3</b>	<b>Overview of Papers .....</b>	<b>81</b>
<b>4.4</b>	<b>Hypotheses .....</b>	<b>82</b>
<b>4.5</b>	<b>Ethics .....</b>	<b>83</b>
<b>5.</b>	<b>PAPER 1 - THE ASSOCIATION BETWEEN FACILITATOR COMPETENT ADHERENCE AND OUTCOMES IN PARENTING PROGRAMS: A SYSTEMATIC REVIEW AND SWIM ANALYSIS.....</b>	<b>84</b>
<b>5.1</b>	<b>Abstract .....</b>	<b>84</b>
<b>5.2</b>	<b>Background.....</b>	<b>86</b>
<b>5.3</b>	<b>Methods .....</b>	<b>89</b>
5.3.1	Systematic Review .....	89
5.3.2	Synthesis Without Meta-Analysis.....	92
5.3.3	Reporting.....	93
<b>5.4</b>	<b>Results.....</b>	<b>93</b>
5.4.1	Programs and Outcomes .....	98
5.4.2	Designs and Analysis Methods .....	99
5.4.3	Control Variables .....	100
5.4.4	Clustering and Multiple Comparison .....	100
5.4.5	Associations of Competent Adherence with Outcomes.....	101
<b>5.5</b>	<b>Discussion of Clinical Implications .....</b>	<b>103</b>
5.5.1	Overall Findings .....	103
5.5.2	Conceptualizing the Role of Facilitator Competent Adherence.....	105
5.5.3	Methodological Issues and Study Quality.....	106
5.5.4	Strengths and Limitations .....	107
5.5.5	Suggestions for Future Research.....	108

<b>5.6</b>	<b>Conclusion.....</b>	<b>109</b>
<b>6.</b>	<b>PAPER 2 - ASSESSING FACILITATOR DELIVERY OF A PARENTING PROGRAM AT SCALE IN TANZANIA: A PSYCHOMETRIC EVALUATION OF AN OBSERVATIONAL MEASURE OF FACILITATOR COMPETENT ADHERENCE.....</b>	<b>111</b>
<b>6.1</b>	<b>Abstract .....</b>	<b>111</b>
<b>6.2</b>	<b>Background.....</b>	<b>113</b>
<b>6.3</b>	<b>Parenting for Lifelong Health for Parents and Adolescents .....</b>	<b>114</b>
<b>6.4</b>	<b>Current Study .....</b>	<b>115</b>
<b>6.5</b>	<b>Methods .....</b>	<b>116</b>
6.5.1	Facilitator Assessment Tool.....	116
6.5.2	Observational Assessors and Assessment Procedure.....	117
6.5.3	Contextual Challenges .....	118
6.5.4	Data Collection Procedure .....	119
6.5.5	Facilitators and Coaches .....	119
6.5.6	Psychometric Evaluation.....	120
6.5.7	Descriptive Analysis of Level of Facilitator Competent Adherence .....	127
<b>6.6</b>	<b>Results.....</b>	<b>127</b>
6.6.1	Content Validity .....	127
6.6.2	Intra- and Inter-Rater Reliability.....	128
6.6.3	Construct Validity and Internal Consistency .....	129
6.6.4	Level of Competent Adherence using PLH-FAT-T.....	137
6.6.5	Analyses with the PLH-FAT-T Short Form.....	137
<b>6.7</b>	<b>Discussion .....</b>	<b>142</b>
6.7.1	Reliability and Validity of the PLH-FAT-T.....	143
6.7.2	Facilitator Competent Adherence at Scale .....	145
6.7.3	Limitations .....	146
6.7.4	Clinical Implications and Future Research .....	147
<b>6.8</b>	<b>Conclusion.....</b>	<b>148</b>
<b>7.</b>	<b>PAPER 3 - THE PREDICTIVE VALIDITY OF A MEASURE OF COMPETENT ADHERENCE: EVALUATING THE ROLE OF FIDELITY IN ADOLESCENT AND PARENT OUTCOMES USING ROUTINE DATA COLLECTED DURING PARENTING PROGRAM DELIVERY AT SCALE .....</b>	<b>149</b>
<b>7.1</b>	<b>Abstract .....</b>	<b>149</b>
<b>7.2</b>	<b>Background.....</b>	<b>150</b>
7.2.1	Current Study .....	152
<b>7.3</b>	<b>Methods.....</b>	<b>153</b>
7.3.1	Intervention .....	153

7.3.2	Study Setting and Sample .....	154
7.3.3	Data Collection Procedures .....	155
7.3.4	Measures .....	157
7.3.5	Data Analyses.....	161
<b>7.4</b>	<b>Results.....</b>	<b>163</b>
7.4.1	Participant and Facilitator Characteristics .....	163
7.4.2	Pre-Post Outcomes .....	163
7.4.3	Associations between Competent Adherence and Outcomes .....	165
<b>7.5</b>	<b>Discussion .....</b>	<b>168</b>
7.5.1	Overall Findings.....	168
7.5.2	Findings in the Context of the Broader Literature .....	170
7.5.3	Potential Explanations for Findings .....	170
7.5.4	Limitations .....	174
<b>7.6</b>	<b>Conclusion.....</b>	<b>175</b>
<b>8.</b>	<b>OVERALL DISCUSSION.....</b>	<b>177</b>
<b>8.1</b>	<b>Overview of Dissertation.....</b>	<b>177</b>
<b>8.2</b>	<b>Summary of the Findings of Each Paper .....</b>	<b>178</b>
8.2.1	Paper 1: The Relationship between Competent Adherence and Outcomes.....	178
8.2.2	Paper 2: Psychometric Properties of the PLH-FAT-T and Level of Competent Adherence Achieved by PLH-Teens Facilitators .....	179
8.2.3	Paper 3: Predictive Validity of the PLH-FAT-T .....	181
<b>8.3</b>	<b>Discussion of Findings.....</b>	<b>182</b>
8.3.1	The Role of Facilitator Competent Adherence is Inconsistent .....	182
8.3.2	The PLH-FAT-T Short Form Requires Improvement .....	185
8.3.3	Community Facilitators Delivered PLH-Teens to a High Level of Competent Adherence .....	186
<b>8.4</b>	<b>Limitations and Challenges .....</b>	<b>187</b>
8.4.1	Focus on Observational Measures of Facilitator Competent Adherence .....	188
8.4.2	Reliance on Secondary Data Collected in a Challenging Context .....	188
8.4.3	Issues of Internal Validity .....	194
8.4.4	Issues of External Validity .....	197
<b>8.5</b>	<b>Strengths.....</b>	<b>198</b>
8.5.1	Use of ‘FUPS’ Data .....	199
8.5.2	Examination of Gold Standard Measurement Approach .....	200
8.5.3	Timely Evidence for PLH Programmes .....	200
8.5.4	Contribution to Implementation Science and Parenting Programme Literature .....	201
<b>8.6</b>	<b>Implications for Future Research and Practice.....</b>	<b>202</b>
8.6.1	Conceptualisation of the Role of Facilitator Competent Adherence .....	203
8.6.2	Who Delivers Parenting Programmes .....	203

8.6.3	Competent Adherence in Real-World Settings .....	204
8.6.4	Training for Assessors.....	205
8.6.5	Future of Competent Adherence Measurement in PLH.....	205
<b>8.7</b>	<b>Recommendations for Future Research and Practice .....</b>	<b>206</b>
8.7.1	Pursue Research on Measure Reliability and Validity .....	207
8.7.2	Enhance Study Reporting.....	208
8.7.3	Improve Study Rigour.....	208
8.7.4	Investigate Mechanisms Achieving Participant Outcomes .....	209
8.7.5	Explore How to Balance Rigor and Practicality .....	210
8.7.6	Reflect on Observational Measurement .....	212
<b>8.8</b>	<b>Dissemination Strategies.....</b>	<b>214</b>
<b>9.</b>	<b>CONCLUSION.....</b>	<b>216</b>
	<b>APPENDICES.....</b>	<b>218</b>
	<b>Appendix 1: Study Protocols .....</b>	<b>218</b>
	<b>Appendix 2: Glossary .....</b>	<b>219</b>
	<b>Appendix 3: List of Acronyms.....</b>	<b>222</b>
	<b>Appendix 4: PLH-FAT-YC.....</b>	<b>223</b>
	<b>Appendix 5: Summary of Changes to PLH-FAT-YC .....</b>	<b>232</b>
	<b>Appendix 6: Ethics Forms and Approvals .....</b>	<b>234</b>
	<b>Appendix 7: Paper 1 Supplementary Files.....</b>	<b>240</b>
	<b>Appendix 8: Paper 2 Supplementary Files.....</b>	<b>274</b>
	<b>Appendix 9: Paper 3 Supplementary Files.....</b>	<b>299</b>
	<b>REFERENCES .....</b>	<b>330</b>

## **List of Tables and Figures**

### **BACKGROUND LITERATURE AND THEORETICAL FOUNDATION**

<b>Figure 1</b> <i>Relationship between Facilitator Delivery and Participant Outcomes</i> .....	40
---	----

### **PAPER ONE AND SUPPLEMENTARY FILES**

<b>Table 1</b> <i>Inclusion and Exclusion Criteria</i> .....	91
<b>Figure 1</b> <i>PRISMA Flowchart of Study Screening and Selection</i> .....	92
<b>Table 2</b> <i>Summary of Study Results</i> .....	94
<b>Table 3</b> <i>Search Terms</i> .....	241
<b>Table 4</b> <i>Location, Gender, Ethnic Make-Up of Study Participants</i> .....	251
<b>Table 5</b> <i>Summary of Study Data</i> .....	253
<b>Table 6</b> <i>Psychometric Evidence of Measures in Included Studies</i> .....	272
<b>Table 7</b> <i>Study Risk of Bias, Quality, and Measure Practicality</i> .....	273

### **PAPER TWO AND SUPPLEMENTARY FILES**

<b>Table 1</b> <i>Lead Assessor Intra-Rater Reliability</i> .....	132
<b>Table 2</b> <i>Lead Assessor-Coach Inter-Rater Reliability</i> .....	133
<b>Table 3</b> <i>Exploratory Factor Analysis of Home Activity Items – Round 2</i> .....	134
<b>Table 4</b> <i>Exploratory Factor Analysis of Role-play Items – Round 2</i> .....	135
<b>Table 5</b> <i>Exploratory Factor Analysis of Skills Items – Round 2</i> .....	136
<b>Table 6</b> <i>Summary of the Level of Competent Adherence Achieved by PLH-Teens Facilitators Using the PLH-FAT-T</i> .....	137
<b>Table 7</b> <i>PLH-FAT-T Short Form</i> .....	138
<b>Table 8</b> <i>Lead Assessor Intra-Rater Reliability with PLH-FAT-T Short Form</i> .....	141
<b>Table 9</b> <i>Lead Assessor-Coach Inter-Rater Reliability with PLH-FAT-T Short Form</i> ....	141

<b>Table 10</b> <i>Summary of Level of Competent Adherence Achieved by PLH-Teens Facilitators using the PLH-FAT-T Short Form .....</i>	<i>142</i>
<b>Table 11</b> <i>Summary of Recommendations and Changes to PLH-FAT-T .....</i>	<i>274</i>
<b>Table 12</b> <i>Descriptive Statistics for the PLH-FAT-T Home Activity Items .....</i>	<i>285</i>
<b>Table 13</b> <i>Exploratory Factor Analysis of Home Activity Items – Round 1 .....</i>	<i>289</i>
<b>Table 14</b> <i>Exploratory Factor Analysis of Home Activity Items – Round 2 .....</i>	<i>290</i>
<b>Table 15</b> <i>Exploratory Factor Analysis of Role-play Items – Round 1 .....</i>	<i>291</i>
<b>Table 16</b> <i>Exploratory Factor Analysis of Role-play Items – Round 2 .....</i>	<i>292</i>
<b>Table 17</b> <i>Exploratory Factor Analysis of Skills Items – Round 1 .....</i>	<i>293</i>
<b>Table 18</b> <i>Exploratory Factor Analysis of Skills Items – Round 2 .....</i>	<i>295</i>
<b>Table 19</b> <i>Exploratory Factor Analysis of Skills Items – Round 3 .....</i>	<i>297</i>

### **PAPER THREE AND SUPPLEMENTARY FILES**

<b>Figure 1</b> <i>Loss of Facilitator Data During Merging with Pre-Post Surveys .....</i>	<i>156</i>
<b>Table 1</b> <i>Primary and Secondary Outcomes Administered Using Pre-Post Surveys.....</i>	<i>158</i>
<b>Table 2</b> <i>Primary and Secondary Outcomes at Pre- and Post-Test of Subset of 3,057 Families .....</i>	<i>164</i>
<b>Table 3</b> <i>Associations between Parent- and Adolescent-Reported Outcomes and Facilitator Competent Adherence .....</i>	<i>167</i>
<b>Table 4</b> <i>Intra-class Correlations for Fixed and Random Effects Included in Multi-Level Poisson Models .....</i>	<i>299</i>
<b>Table 5</b> <i>Association between Facilitator Competent Adherence and Parent-Reported Outcomes .....</i>	<i>302</i>
<b>Table 6</b> <i>Associations between Facilitator Competent Adherence and Adolescent-Reported Outcomes .....</i>	<i>305</i>
<b>Table 7</b> <i>Associations between Parent- and Adolescent-Reported Outcomes and Facilitator Skills .....</i>	<i>307</i>



<b>Table 8</b> <i>Associations between Parent-Reported Outcomes and Facilitator Adherence to Home Activity</i> .....	312
<b>Table 9</b> <i>Associations between Parent-Reported Outcomes and Facilitator Adherence to Role-play</i> .....	315
<b>Table 10</b> <i>Association between Adolescent-Reported Outcomes and Facilitator Adherence to Home Activity</i> .....	318
<b>Table 11</b> <i>Association between Adolescent-Reported Outcomes and Facilitator Adherence to Role-play</i> .....	321
<b>Table 12</b> <i>Summary of Pre-Post Surveys</i> .....	324

## **Acknowledgements**

I am privileged to have an incredible community of family, friends, mentors, colleagues, and funders who helped me pursue my doctorate. Although I have learned a great deal from my studies, it is from this community that I have derived the most inspiration, learning, and growth. In thanking this community for their time, effort, and commitment to my development as a person, as a thinker, and as a researcher, I share a short list of the main learnings I have gained from each of them over the years and through the DPhil process.

### **Family**

From my Mom, I have learned how to work hard; to be a logical thinker and communicator (or at least really try); and be curious about how to make the world a better place. Mama, the wonderful life you have created for us is the source of my interest in advancing the health and well-being of children and families through interventions. It is my hope that research can find a way to ensure that every child has a champion like you have been for me – someone who loves them fiercely, creates opportunities for them to learn and grow, supports them in developing strength and confidence, and ensures reaching their potential is possible. Thank you, thank you, thank you.

From my husband and partner of ten years, I have learned to put family and relationships first; trust the good intentions of others; not sweat the small stuff; keep calm when the going gets tough; and “just Google it”. Fred, you have been my biggest cheerleader. Your unwavering commitment to me and confidence in me is the stuff of dreams.

## **Friends**

From Bridget Steele, I have learned what it is to have a true friend and sounding board. Thank you for being there through the “dark days”. The DPhil journey sometimes felt like being out to sea and you have been both a sail and an anchor.

From the Department of Social Policy and Intervention DPhils, I have learned what a difference being surrounded by positive, generous, and uplifting friends and colleagues makes in life and work. Your dedication to others is inspiring. Thank you for creating a wonderful doctoral experience.

## **Mentors**

From Jamie Lachman, I have learned to bring fun into learning; be passionate about a cause; believe in the power of families; and be unafraid to think out of the box. Thank you for affording me wonderful opportunities to learn both in the office and in the field.

From Frances Gardner, I have learned to consider the big picture; really ‘have a think’ about my scientific choices; be critical and transparent; and strive to become a rigorous scientist. Thank you for being a steady and committed source of guidance over the years.

## **Colleagues**

From Joyce Wamoyi, I have learned of the richness that gathering multiple perspectives via qualitative research and analysis brings to intervention science. Thank you for dedicating your time and expertise to our team and to my learning.

From Yulia Shenderovich, I have learned to grapple with the complexity of studying the implementation of programmes in the real-world. I am so appreciative of your

collaboration and mentorship. Thank you for always being there to talk through the big challenges.

From Francisco Calderon and Qing Han, I learned a great deal about statistics and R. Thank you for answering my many questions and helping to debug my code.

From colleagues at Clowns Without Borders South Africa (Nyasha, Stivin, Levita, and Sibongile) and Pact Tanzania (Esther, Amal, Amon, and Asheri), I have learned more about what is possible and what works on the ground. Thank you for sharing your experience and expertise.

## **Funders**

From the Rhodes Trust, I have learned to expand my scope and network globally. Your support has been a generous, life-changing, and empowering gift. I am grateful to the Social Sciences and Humanities Research Council of Canada and Hertford College for their substantial financial assistance and to the Evaluation Fund for funding the study in Tanzania (EF-2019-1107). I am also grateful to the European Research Council under the European Union's Horizon 2020 research and innovation program (Grant No. 737476 and No. 771468); Research England; UK Research and Innovation Global Challenges Research Fund (Grant No. ES/S008101/1); National Research Foundation of South Africa (Grant No. 118571); the Complexity and Relationships in Health Improvement Programmes of the Medical Research Council (MRC-UK) and Chief Scientist Office (Grant No. MC\_UU\_00022/1, CSO SPHSU16, MC\_UU\_00022/3, and CSO SPHSU18); Centre for the Development and Evaluation of Complex Interventions for Public Health (DECIPHer) with funding from the Welsh Government through Health and Care Research Wales, and the

Wolfson Centre for Young People's Mental Health (with support from the Wolfson Foundation) as these funders provided support to the FAIR study and its team members.

## **Abstract**

**Background:** Implementation fidelity is a critical component of intervention science research, which aims to understand how interventions unfold in practice to improve their outcomes. A key element of fidelity is facilitator competent adherence - the extent to which a programme is delivered as prescribed with the specified level of quality. The dissertation endeavoured to better understand how to measure facilitator competent adherence and the role facilitator competent adherence plays in achieving intended parent/caregiver (parent) and child outcomes in the parenting programme literature and, specifically, within Parenting for Lifelong Health for Parents and Adolescents (PLH-Teens). PLH-Teens is a parenting programme designed to reduce violence against children and child behavioural and emotional problems in low- and middle-income countries (LMICs). The dissertation is composed of three studies – one which synthesised data from the parenting programme literature and two which analysed data from the 2020-2021 scale-up of PLH-Teens in Tanzania to 75,061 participants by community facilitators (school teachers and community health workers;  $N=444$ ).

**Objectives:** The dissertation had three objectives with each corresponding to an individual paper. The first objective was to synthesise the evidence on the relationship between observational measures of facilitator competent adherence and parent and child outcomes in the parenting programme literature. The second objective was to examine whether the observational measure of facilitator competent adherence used in the large-scale implementation of PLH-Teens in Tanzania is reliable and valid for use in research and practice and to determine the level of competent adherence with which community

facilitators delivered PLH-Teens in Tanzania. The third objective was to determine the predictive validity of the observational measure of competent adherence used in PLH-Teens by examining whether competence adherence is associated with parent and adolescent outcomes.

**Methods:** Paper 1 synthesised the results of a systematic review of studies on parenting programmes aiming to reduce violence against children and child behavioural and emotional problems to examine the associations between observational measures of facilitator competent adherence and parent and child outcomes. Due to study heterogeneity and poor reporting, Synthesis Without Meta-Analysis (SWiM) guidelines were followed.

Paper 2 used 95 facilitator assessments collected by implementing partners during the 2020-2021 delivery of PLH-Teens in Tanzania. The paper evaluated the reliability and validity of the measure used to assess facilitator competent adherence in PLH-Teens - the Facilitator Assessment Tool (PLH-FAT-T). Reliability was assessed by conducting intra-rater reliability, inter-rater reliability, and internal consistency analyses using percentage agreements, intra-class correlations, Cronbach's alphas, and omegas. Validity was assessed via consultations with stakeholders (content validity) and exploratory factor analyses (construct validity). This paper also estimated the level of competent adherence with which community facilitators delivered PLH-Teens by calculating the average PLH-FAT-T score achieved by facilitators.

Paper 3 investigated the relationship between facilitator competent adherence and the pre-post outcomes of PLH-Teens participants. Analyses used 24 PLH-FAT-T

assessments that could be linked to the pre-post surveys of 3,057 families. This analysis was conducted using multi-level Poisson regressions with fixed and random effects.

**Results:** Paper 1 found 18 studies reporting on the relationship between observational measures of facilitator competent adherence and parent and child outcomes. The review found that most studies ( $n=13$ ) reported a statistically significant positive relationship with at least one of the parent or child outcomes reported. However, eight studies reported inconsistent findings across outcomes. Four studies found no significant association with outcomes.

Paper 2 found that the PLH-FAT-T showed strong content validity, poor to moderate intra- and inter-rater reliability, strong internal consistency, and moderate construct validity. Iterative exploratory factor analyses produced a shortened PLH-FAT-T, the PLH-FAT-T Short Form, comprised of 19 fewer items which had stronger psychometric properties. Analyses of the PLH-FAT-T Short Form found that community facilitators delivered PLH-Teens at scale in Tanzania to a high-level of competent adherence (82.3% average).

Using the PLH-FAT-T Short Form, Paper 3 found that the relationship between facilitator competent adherence and outcomes was mixed with some positive, some insignificant, and some negative associations. A positive association was found between competent adherence and the primary outcome of interest, child maltreatment, as reported by adolescents. The analysis found that increased competent adherence had a positive association with two of the 12 parent-reported outcomes and seven of the 10 adolescent-reported outcomes (including child maltreatment). Yet, increased competent adherence also



had a negative association with five parent-reported outcomes, as well as insignificant associations with five parent-reported outcomes and three adolescent-reported outcomes.

**Discussion:** Paper 1 suggests that better facilitator competent adherence is generally associated with positive parent and child outcomes. However, this finding is weakened by the methodological heterogeneity of included studies and due to the wide variety of ways in which studies conceptualised competent adherence-outcome relationships. As a result, the paper reveals that there is substantial methodological work to be done in the broader parenting programme community to improve the rigour of and reporting on investigations regarding this relationship. As the amount of literature on the measurement and role of facilitator competent adherence grows in the behavioural intervention literature, the recommendations made in Paper 1 have relevance for other implementation scientists conducting and sharing studies on competent adherence.

Paper 2 reports on the first psychometric evaluation of the PLH-FAT-T and is the first study of its kind to report on the fidelity achieved by facilitators during routine parenting programme delivery at scale in a low-income country. Findings suggest that the PLH-FAT-T had poor to moderate reliability and sufficient validity and that the PLH-FAT-T Short Form had stronger psychometric properties. Although the tool was stronger following iterative exploratory factor analyses, the findings indicate that further work is needed to strengthen the reliability and validity of the PLH-FAT-T Short Form. Findings also suggest that community facilitators with minimal background in and training on parenting programmes delivered PLH-Teens to a high level of quality at scale in a low-income community setting despite significant barriers. Thus, the findings of Paper 2 suggest that it may be possible for community facilitators to deliver behavioural

interventions to a high level of competent adherence in low-income routine delivery settings at scale.

The findings of Paper 3 are similar to the findings of Paper 1 in that Paper 3 does not provide a clear answer as to whether, and to what extent, facilitator competent adherence impacts participant outcomes. Potential explanations of the findings include the PLH-FAT-T Short Form has poor predictive validity; the PLH-FAT-T Short Form assessments were not reliable; a variety of methodological challenges may have prevented an examination of the true relationship between competent adherence and outcomes; competent adherence does not relate to outcomes in the manner theorised; competent adherence plays a less important role in the achievement of outcomes than anticipated or, at some point, plays a negative role; and only certain programme components are achieving outcomes so the PLH-FAT-T Short Form is not capturing the important aspects of programme delivery. The alignment of the findings of Papers 1 and 3 with some other systematic reviews and meta-analyses in the broader implementation science literature suggests that the role facilitator competent adherence plays in participant outcomes is not fully understood. Thus, there is reason to further investigate the theorised relationship between facilitator competent adherence and outcomes outlined in seminal implementation science theories and models to fully illuminate the inner workings of the ‘black box’ of interventions. A fuller understanding of the role that facilitator competent adherence plays in participant outcomes is essential to maximise the benefits to be reaped from evidence-based behavioural interventions.

**Conclusion:** The dissertation provides important evidence regarding the measurement and role of facilitator competent adherence in the parenting programme literature and in

Parenting for Lifelong Health. As a result, the dissertation provides a series of recommendations for the future of competent adherence monitoring in research and practice that are relevant to both the parenting programme literature and the broader implementation science literature. As parenting programmes continue to be delivered and scaled worldwide, it is intended that the findings and recommendations herein will be used to benefit both Parenting for Lifelong Health and the broader parenting programme community in the quest to maximise opportunity for vulnerable children and families globally to benefit from evidence-based parenting programmes.

## 1. Introduction

### Summary of Dissertation

The dissertation investigates the competent adherence with which facilitators deliver parenting programmes aiming to reduce violence against children and child behavioural and emotional problems to better understand how to measure competent adherence as well as to determine whether, and to what extent, facilitators play a role in achieving intended parent/caregiver (parent) and child outcomes. The dissertation builds on my master's thesis, which consisted of a systematic review exploring the measures of competent adherence reported in the parenting programme literature, now published in *Clinical Child and Family Psychology Review* (Martin et al., 2021b) and a psychometric evaluation of the observational measure of competent adherence used to assess Parenting for Lifelong Health for Young Children (PLH-YC) programme facilitators, now published in *Child: Care, Health and Development* (Martin et al., 2022a). The first doctoral paper draws on the aforementioned systematic review to synthesise the relationship between facilitator competent adherence and outcomes. The second and third papers utilise data collected as part of a larger study on the scale-up of the Parenting for Lifelong Health for Parents and Adolescents (PLH-Teens) programme in Tanzania – the Furaha Adolescent Implementation Research (FAIR) study – to evaluate the psychometric properties of the observational tool used to assess facilitator competent adherence and investigate the relationship between facilitator competent adherence and outcomes. The plans for the FAIR study are summarised in a protocol published in *Implementation Science Communications* (Martin et al., 2021a) (see [Appendix 1](#)).

## **Focus of Dissertation**

Implementation fidelity, as a part of intervention science research, seeks to examine the degree to which programmes are implemented as intended (Fixsen et al., 2005). Two key elements of implementation fidelity are the extent to which facilitators deliver a programme as prescribed and with the specified level of quality – or ‘competent adherence’ (Breitenstein et al., 2010a; Forgatch et al., 2005; Proctor et al., 2011). The dissertation reports on three interlinked papers examining these elements of programme delivery in the wider parenting programme literature and in the 2020-2021 scale-up of PLH-Teens in Tanzania. This research is one of few studies on the fidelity of parenting interventions at scale; it is also larger than many studies in high-income countries and is the largest known study of its kind in low- and middle-income countries (LMICs).

## **Organisation of Dissertation**

This document is organised into eight chapters: introduction (Chapter 1); background literature and theoretical foundation (Chapter 2); review of literature underpinning the methods (Chapter 3); dissertation overview (Chapter 4); Paper 1, which is published in *Prevention Science* (Chapter 5); Paper 2, which is in submission (Chapter 6); Paper 3, which is close to submission (Chapter 7); and discussion of the overall findings, implications of the dissertation, and strategies used to disseminate findings (Chapter 8). A glossary of key terms and a list of acronyms used throughout the dissertation are in Appendix 2 and Appendix 3.

## **2. Background Literature and Theoretical Foundation**

This chapter describes the background literature and theoretical foundation on which the dissertation was developed. It is composed of three sections. The first section (2.1) outlines the rationale for focusing on parenting programmes aiming to reduce violence against children and child behavioural and emotional problems. The section also provides an overview of the relevant evidence, including the challenges and gaps in knowledge, with respect to the implementation of parenting programmes in LMICs and at scale. The second section (2.2) describes the implementation science frameworks and models that informed the dissertation. It also defines and provides a rationale for focusing on two dimensions of implementation fidelity related to programme facilitators (competence and adherence). Drawing on this background, the third section (2.3) summarises the existing evidence on Parenting for Lifelong Health (PLH) programmes and situates Papers 2 and 3 as part of the FAIR study conducted in Tanzania.

### **2.1 Parenting Programmes**

Parenting programmes were chosen as the focus of the dissertation as there is a sizeable body of evidence on the effectiveness of parenting programmes aiming to reduce violence against children and child behavioural and emotional problems yet there are gaps in knowledge regarding the measurement and role of facilitator competent adherence. Little is known regarding how to effectively measure facilitator delivery as well as whether, and to what extent, facilitator delivery plays a role in intended beneficiary outcomes – particularly in low-income community settings at scale. Investigating competent adherence in the parenting programme literature may have relevance in other behavioural intervention fields with limited evidence on the reliability and validity of implementation fidelity

measures, the extent to which programmes can be delivered with fidelity in practice with limited resources, and the role that implementation fidelity plays in participant outcomes.

### **2.1.1 Defining Parenting Programmes**

Parenting programmes are interventions that seek to support parents – those responsible for the care and/or upbringing of a child between the ages of 0 and 18 irrespective of biological relationship – to acquire the knowledge and skills that will enable them to maintain and improve the health and well-being of their children (Barlow & Coren, 2018). Parenting programmes vary in their theoretical underpinning, core components, programme facilitator types, session formats, and intended outcomes (Barlow & Coren, 2018). Parenting programmes are typically rooted in either social learning theory (Bandura & Walters, 1977) or attachment theory (Bowlby, 1974). Parenting programme components found to be effective in enhancing parent and child outcomes address topics including child development, parental self-management, parental warmth and sensitivity, positive parent-child interaction and relationships, child behaviour management, and effective disciplinary approaches (Kaminski et al., 2008; Leijten et al., 2022; Melendez-Torres et al., 2019; WHO, 2023). These programme topics are often delivered using participatory approaches such as modelling, role-plays, discussions, home practice activities, and opportunities to provide feedback and receive support for challenges experienced when applying parenting skills at home (Kazdin, 1997). Some programmes are offered to groups of parents whereas others are provided to individual parents; some programmes include children or teens as participants whereas others only engage parents; some have few sessions whereas others have many sessions; some are delivered in-person whereas others are delivered online;

some are delivered in community settings whereas others are delivered in family homes; and some are delivered by community facilitators whereas others are delivered by highly trained professionals.

### **2.1.2 Parenting Programmes to Reduce Violence Against Children and Child**

#### **Behavioural and Emotional Problems**

Parenting programmes commonly aim to address two interrelated global public health and human rights issues experienced by and impacting children worldwide - violence against children and/or child behavioural and emotional problems (referred to throughout as child behaviour problems). Violence against children and child behaviour problems are often interrelated in that each exacerbates the other in a coercive cycle (i.e., child behaviour problems intensify parental use of harsh discipline and vice versa) (Patterson, 1989; Patterson et al., 1992). Prior to describing the gaps in knowledge as it relates to the implementation of parenting programmes, it is worth defining violence against children and child behaviour problems and providing a rationale for focusing on interventions that address these interconnected issues.

According to the World Health Organization (WHO) (2022), “violence against children includes all forms of violence against young people under 18 years old, whether perpetrated by parents or other caregivers, peers, romantic partners, or strangers” (p.1). Of the many types of violence against children, this dissertation focuses on programmes that target physical and emotional abuse perpetrated by parents. It is estimated that each year, over one billion children experience violence with disproportionate numbers of children impacted in LMICs (Hillis et al., 2016; Stoltenborgh et al., 2013; UNICEF, 2010). A



substantial body of research indicates that such violence has serious short- and long-term negative consequences for children, including for mental health, substance use, peer violence, delinquency, and intergenerational transfer of violence via intimate partner violence and child maltreatment (Cowell et al., 2015; Cuartas, 2022; Font & Berger, 2015; Gershoff, 2010; Gershoff et al., 2018; Gershoff & Grogan-Kaylor, 2016; Heilmann et al., 2021; Mills et al., 2011; Moylan et al., 2010; Walker et al., 2011; Widom et al., 2012).

For the purposes of this dissertation, child behaviour problems involve conduct or emotional problems characterised by externalising (e.g., aggression, antisocial behaviour) or internalising (e.g., anxiety, depression) behaviours by children (Campbell et al., 2000; Eisenberg et al., 2001; Liu et al., 2011; McMahon & Frick, 2019). A global meta-analysis of literature from 27 countries estimated that 5.7% of children have disruptive disorder (3.6% have oppositional defiant disorder and 2.1% have conduct disorder) (Polanczyk et al., 2015). These estimates indicate that behaviour-related problems are prevalent mental health issues affecting children globally. While most children in the world live in LMICs, there are fewer mental health supports in these contexts (Mayo-Wilson et al., 2014; Rose et al., 2022; United Nations Department of Economic Affairs, 2019). Child behaviour problems have a series of negative consequences for children across the life course (e.g., Colman et al., 2009; Erskine et al., 2016; Reef et al., 2011). To illustrate, a 25-year study of a birth cohort in New Zealand ( $N=1,265$ ) by Fergusson et al. (2005) found that conduct problems at age 7 were associated with a range of negative outcomes in adulthood, including poor relationships, involvement in crime, use of substances, and mental health problems. Additionally, the societal costs of coping with and treating child behaviour

problems are substantial. A study by Foster et al. (2005) using family-reported data from a subset of 664 families who participated in the Fast Track project in the United States found that the public sector costs associated with child conduct disorder was approximately \$70,000 USD per child over seven years.

The interconnection between violence against children and child behaviour problems is illustrated by evidence that numerous parenting behaviours serve as either risk or protective factors for child behaviour problems. Studies over many decades have found that higher levels of harsh discipline, physical aggression, and parental inconsistency as well as lower levels of parental warmth, monitoring, and involvement are associated with higher levels of child behaviour problems across the lifespan (Carroll et al., 2013; Hinshaw & Lee, 2003; Pasalich et al., 2016; Racz & McMahon, 2011; Stormshak et al., 2000; Widom, 2017). This relationship is found across multiple contexts globally. A study using data from the International Parenting Study from Asia, Europe, and North America by Rebellon and Straus (2017) found that higher levels of parental use of violent discipline at age 10 was associated with higher rates of antisocial behaviours among the same children when they were adults ( $N=8,901$ ).

There is now considerable evidence that parenting programmes that aim to increase the use of positive parenting strategies, reduce harsh discipline, and improve parent-child relationships also reduce both violence against children and child behaviour problems (Barlow & Coren, 2018; Buchanan-Pascall et al., 2018; Burkey et al., 2018; Jeong et al., 2021; Knerr et al., 2013; McCoy et al., 2020). Due to their effectiveness, parenting interventions have received international attention and calls to action from

organisations such as the WHO. For instance, the multi-agency *INSPIRE: Seven Strategies to End Violence Against Children* and the recently published *WHO Guidelines on Parenting Interventions to Prevent Maltreatment and Enhance Parent-Child Relationships with Children 0-17 Years* recommend parenting programmes as a key strategy to prevent abuse (WHO, 2016, 2023).

### *Parenting Programmes in LMICs*

While the majority of evidence on the effectiveness of parenting programmes is from high-income countries, there is growing evidence of their effectiveness on reducing harsh punishment and increasing positive parenting in LMICs as well (Gardner et al., 2016; Knerr et al., 2013; McCoy et al., 2020; WHO, 2023). Parenting programmes are needed in LMICs due to higher rates of violence against children than in high-income countries (Stoltenborgh et al., 2013; Ward et al., 2016a). A UNICEF survey of 35 LMICs found that 75% of children reported experiencing violent punishment by members of their household (2010). Further, a study by Cuartas et al. (2019) which analysed data from 107,063 children in 49 LMICs collected as part of the Multiple Indicator Cluster Surveys estimated that approximately two-thirds of children in LMICs under the age of 5 experience physical discipline from their parents. The study also estimated higher rates of physical discipline among families in sub-Saharan Africa and South Asia (Cuartas et al., 2019). Additionally, the delivery of parenting programmes is particularly pertinent in LMICs due to the presence of multiple risk factors for violence against children in these contexts (e.g., Pelton, 2015; Stith et al., 2009; Walker et al., 2011). One risk factor for violence against children in many LMICs is the belief by many parents that corporal punishment is a normative child

disciplinary practice (Lansford & Deater-Deckard, 2012; Stith et al., 2009). Another risk factor for violence against children is family poverty (e.g., Pelton, 2015).

### **2.1.3 Parenting Programmes at Scale**

Given the encouraging evidence regarding the effectiveness of parenting programmes aiming to reduce violence against children and child behaviour problems in LMICs, there has been increasing research attention on the scale-up of parenting programmes in these countries (Gardner et al., 2016; Mikton et al., 2013; Shenderovich, 2021; Ward et al., 2015). There have also been numerous calls to build the capacity of governments and agencies to implement such programmes at scale (WHO, 2016). Scale-up may be defined as “deliberate efforts to increase the impact of health innovations successfully tested...so as to benefit more people and foster the development of sustainable policies and programs” (Cash, 2011, p. 3). The scale-up of evidence-based interventions is often the long-term goal of the development, implementation, and evaluation of interventions to ensure that as many beneficiaries as possible benefit from improved outcomes.

Despite attention, research on the scale-up of parenting programmes is relatively limited, particularly in LMICs (Shapiro et al., 2010; Shenderovich, 2021). A cluster randomised trial of the large-scale implementation of the Triple P programme in South Carolina, USA (estimated reach of  $N=8,883-13,560$  families) reported benefits in reducing child maltreatment using a tiered approach which aimed to bolster an entire community’s knowledge and awareness of child development and evidence-based parenting practices through media campaigns (Prinz et al., 2009). While the study found large reductions in

child maltreatment, injuries, and child welfare placements (Prinz et al., 2009), commentators have pointed to a range of methodological challenges with this study, including selective reporting (Eisner, 2014). An evaluation of Triple P in Scotland by Marryat et al. (2017) examined the pre-post outcomes of the six-year delivery of Triple P at the population level in Glasgow using routine data (estimated reach of at least  $N=30,748$  families in addition to the families reached via mass media) and found that the programme was not effective in improving intended outcomes (Marryat et al., 2017). Limitations of this study included a lack of a comparison group and difficulty determining each family's level of exposure to the programme. Further, a study by Gray and colleagues (2018) in England examined the uncontrolled effects of various evidence-based parenting programmes delivered at a large scale, including Triple P and Incredible Years, and found community delivery ( $N=3,706$  parents) produced similar effect sizes as researcher delivery produced ( $N=1,390$  parents). Although Gray et al.'s use of a non-randomised study design may have led to an overestimation of effect sizes, the finding suggests that large-scale delivery is possible and can be effective for children and families (2018). Further still, there is some convincing evidence of the effectiveness of the Parent Management Training-Oregon Model (PMTO) programme via large-scale and routine service delivery in numerous countries (Forgatch & Kjøbli, 2016; Kjøbli et al., 2013). For instance, a study of the scale-up of PMTO in Norway by Tommeraas and Ogden (2017) found that programme effects were just as substantial at scale as they were in randomised trials.

### *Implementation and Evaluation Challenges at Scale*

There may be limited evidence on parenting programmes implemented at scale due to the numerous critical questions and challenges associated with scale-up (Gottfredson et al., 2015). Foremost among these questions is whether programmes are still effective when delivered beyond the scope of their original testing. Consider the Triple P programme, for instance. Although there is significant evidence from randomised controlled trials that the programme is effective (Sanders et al., 2014), population-level studies have not been able to consistently replicate these findings – as was the case in Glasgow (Marryat et al., 2017). Another key question concerns whether parenting programmes transported to new groups and contexts are culturally acceptable and appropriate for the intended beneficiaries and stakeholders (Lau, 2006; Palinkas et al., 2009). To date, some research has shown parenting programmes to be effective when implemented with new groups and in new contexts (Gardner et al., 2016). As cultural appropriateness and acceptability are important programme factors, implementers and researchers must study and reflect upon participant perspectives regarding the suitability of the programme for families in their context and the extent to which programmes delivered at scale should be regulated or more flexible in response to participant needs (Kemp, 2016). Yet another question about scale-up asks whether the relevant stakeholders believe in the value of moving the programme to scale and are willing to commit the required time and other resources to the effort (Cash, 2011). In response to this potential challenge, stakeholder buy-in can be assessed by asking for their perspectives on the delivery of the programme at scale. A final question about scale-up asks whether it is administratively

feasible and sustainable to deliver programmes to large numbers of people (Cash, 2011). Embedded in this question is whether programmes can be delivered to large numbers of beneficiaries with fidelity – a challenge this dissertation explores. Programme drift is a major consideration at scale as deviations from the programme model often occur when the original developers are not directly involved in implementation (Bond et al., 2000; Botvin, 2004). Programme drift can be studied by using implementation science approaches to examine whether core intervention components remain intact and to identify how flawed implementation might be rectified (Mowbray et al., 2003).

## **2.2 Implementation Science**

Implementation science is a field that examines how interventions unfold in practice and uses this information to improve intervention implementation and outcomes (Bhattacharyya et al., 2009; Mihalic, 2004; Peters et al., 2013). As the previously mentioned studies of the Triple P programme illustrate, evidence of programme effectiveness from randomised trials is not enough to ensure that parenting programmes will achieve their intended outcomes in practice or when scaled. Implementation science provides useful frameworks and models for evaluating programme delivery beyond randomised trials and at scale, which have informed the theoretical foundation of the dissertation.

This section summarises the six implementation science frameworks and models used, outlines the aspects of these frameworks and models that are focused upon in the dissertation (implementation fidelity), and discusses the importance of implementation fidelity in the context of parenting programmes. The RE-AIM framework informed what

aspect of parenting programme delivery to study – implementation (2.2.1). Fixsen’s framework, which defines implementation, was utilised to narrow the focus of the dissertation to the role of facilitators delivering interventions (2.2.2). Proctor’s Taxonomy of Implementation Outcomes supported the selection of the specific components of implementation fidelity to examine (competence and adherence) (2.2.3). Berkel’s Integrated Model of Implementation and Carroll’s Conceptual Framework of Implementation Fidelity informed the conceptualisation of the potential relationship between facilitator delivery and participant outcomes (2.2.6). Durlak’s model of the steps to follow to study implementation was employed in Papers 2 and 3 (2.2.7).

### **2.2.1 RE-AIM Framework**

Many factors influence the success of public health interventions in practice. The RE-AIM framework proposed by Glasgow and colleagues positions the successful scale-up of public health programmes as the result of five key factors – reach, efficacy, adoption, implementation, and maintenance (1999). “Reach” describes the people and organisations affected by an intervention. “Efficacy” refers to the evidence regarding the impact – both the positive and negative outcomes – of an intervention. “Adoption” refers to the extent of programme uptake. “Implementation” refers to the degree to which an intervention is delivered as outlined according to the programme theory and logic model. “Maintenance” refers to the extent to which a programme is feasible to deliver in the long-term with the ability to support positive outcomes. Of these five, the dissertation focuses on the fourth factor – implementation and more specifically, the implementation of parenting programmes. The RE-AIM framework also conceptualises of efficacy and implementation as interacting to produce the degree of programme effectiveness (i.e., efficacy x



implementation = effectiveness). The framework suggests that research on programme implementation is fundamental to determining whether an intervention is practical and effective, particularly when delivered in community settings.

### **2.2.2 Defining Implementation**

To conceptualise implementation, the dissertation utilised a framework outlined by Fixsen and colleagues (2005) which defines implementation as comprised of five aspects – the programme, implementing organisation, facilitators, feedback mechanisms (i.e., information that signals how the programme is working), and sphere of influence (i.e., the larger socio-political and historical contexts affecting the programme). Of these components, the dissertation focuses on the facilitators implementing parenting programmes. Facilitators are thought to be a vital aspect of programme delivery to study as they are the vehicle through which participants receive an intervention; programmes do not implement themselves (Petersilia, 1990). According to Fixsen and colleagues, “in human services, the practitioner is the intervention” (2009, p. 532). Although facilitators are considered a critical aspect of interventions, there is little research on how to select, train, assess, and supervise facilitators to ensure high-quality delivery and whether, and to what extent, facilitator delivery matters for outcomes - particularly in LMICs and in routine delivery at scale (Ward et al., 2016b).

### **2.2.3 Implementation Outcomes**

Proctor’s Taxonomy of Implementation Outcomes (2011) specifies eight outcomes to examine when studying programme implementation - adoption (the extent of programme uptake), acceptability (participant satisfaction), appropriateness (programme fit), implementation fidelity (adherence to the programme theory and model), cost (time and

other resources required), feasibility (the extent to which the programme can be delivered successfully, including consideration of its benefits and challenges), penetration (the extent to which programme delivery is embedded within existing services and systems), and sustainability (the practicality of long-term delivery). Of these eight implementation outcomes, the dissertation focuses on fidelity.

#### **2.2.4 Implementation Fidelity**

Implementation fidelity refers to the extent to which an intervention is implemented as intended by programme developers and as outlined in its logic model or programme manual (Bumbarger & Perkins, 2008; Dane & Schneider, 1998; Dusenbury et al., 2003). Implementation fidelity is an aspect of implementation science that examines how and why an intervention works as well as how it can be improved (Carroll et al., 2007). At times, implementation fidelity is also referred to in other ways including as ‘intervention fidelity’ (e.g., Santacroce et al., 2004) or ‘integrity of implementation’.

##### *Importance of Implementation Fidelity*

Understanding the degree to which a programme is implemented as intended is important for several reasons. First, data on implementation fidelity provides information about the extent to which the programme theory was implemented in practice (Breitenstein et al., 2010b). Even though a programme may have a strong theoretical foundation or is efficacious in randomised trials, it does not necessarily mean that the programme will be delivered as planned in practice or that it will be used by stakeholders as expected (Petersilia, 1990). Without collecting and analysing implementation fidelity data, it is difficult to understand if, and to what degree, intervention components were fulfilled, in

which case the magnitude of its Type III error is unknown (Carroll et al., 2007; Dobson & Cook, 1980). Type III error is the uncertainty arising from not knowing whether poor intervention effects are due to poor programme delivery or the ineffectiveness of a programme's theory of change (Dobson & Cook, 1980). Ultimately, there are a variety of reasons why an intervention may be unsuccessful and ascertaining what was implemented and how can help uncover the root of the problem (Hawe et al., 2004). Conversely, even if an intervention is found to be effective, some degree of implementation failure may mean intervention effects are misattributed or reduced thereby giving an incomplete picture of the viability of the programme theory (Breitenstein et al., 2010b; Carroll et al., 2007).

Second, information about whether a programme was implemented as intended provides insight into the potential mechanisms through which an intervention effects its outcomes (Astbury & Leeuw, 2010; Fixsen et al., 2005; Scriven, 1999). Uncovering such mechanisms may illuminate what programme elements are contributing to the achievement of outcomes so that programme outcomes can be maximised, and programme costs can be minimised, thereby creating a more sustainable intervention.

Third, research on implementation fidelity data provides information about how programmes can be improved (Breitenstein et al., 2010b). For instance, fidelity data can help establish what dimensions a programme facilitator struggles to deliver, which can inform future support (e.g., training and coaching).

Fourth, implementation fidelity may be associated with programme outcomes. The general theory is that better programme delivery will translate into better outcomes for intervention beneficiaries (Durlak & DuPre, 2008; Dusenbury et al., 2003). As a part of

programme delivery, implementation factors such as high participant attendance, high levels of participant engagement in sessions, and high-quality programme delivery by facilitators are theorised to be associated with enhanced participant outcomes (Carroll et al., 2007). A systematic review conducted by Durlak and colleagues (2008) consisting of over 500 studies on health promotion interventions found positive associations between implementation (e.g., attendance, engagement, facilitator delivery) and outcomes. Similar findings have emerged in the parenting programme literature regarding the role of facilitator delivery, but the literature is inconsistent (e.g., Breitenstein et al., 2010a; Cantu et al., 2010; Forgatch & DeGarmo, 2011).

Fifth, research on implementation fidelity can provide information about what is needed to dependably deliver a high-quality intervention in various contexts during programme dissemination and scale-up (Glasgow et al., 2003). This is especially the case as programme effectiveness is often weakened at scale due to poor implementation or programme drift, which often occurs once a programme becomes widely used (Araujo et al., 2021; Bond et al., 2000; Botvin, 2004). This phenomenon has been coined the “scale-up penalty” (Institute of Medicine & National Research Council, 2014).

#### *Components of Implementation Fidelity*

There are five commonly acknowledged components of implementation fidelity – adherence (extent to which the programme is delivered according to a manual or protocol), quality of delivery (style or skill with which the programme is delivered), dosage (amount of the intervention delivered), participant responsiveness (degree of participant engagement), and programme differentiation (extent to which the intervention is different

from other interventions) (Dane & Schneider, 1998; Dusenbury et al., 2003; Mihalic, 2004; Proctor et al., 2011). Berkel et al.'s (2011) Integrated Model of Program Implementation suggests that there are two lines of inquiry concerning implementation fidelity – programme facilitators and programme participants. Although there are a variety of aspects of facilitator implementation that could be studied, the dissertation focuses on adherence and quality of delivery – otherwise known as facilitator competent adherence.

### *Fidelity and Adaptation*

Before going on to define and discuss facilitator competent adherence, it is worth pausing to consider the tension between fidelity and adaptation. Although the five components of implementation fidelity are thought to be important, there is debate regarding whether, and to what extent, they should be prioritised over unplanned or planned adaptations, particularly in community settings and at scale (Kumpfer et al., 2017; Mazzucchelli & Sanders, 2010; O'Connor et al., 2007). As explained by Conduct Problems Prevention Research Group (2002), the challenge is finding the right balance between “maintaining intervention fidelity while responding appropriately to the local norms and needs of communities that vary widely in their demographic and cultural/ethnic composition” (p.1). To address the tension between fidelity and adaptation, some researchers and practitioners recommend striving for “flexibility with fidelity” (Kendall & Beidas, 2007, p. 13), ‘fidelity-consistent’ adaptations (Stirman et al., 2015), and function over form (Skivington et al., 2021).

While it is acknowledged that adaptation is potentially an impactful ingredient in achieving participant outcomes, the dissertation focused on implementation fidelity for two

key reasons. First, although adaptations may be valuable, it is worthwhile to investigate the extent to which intervention models found to be effective via randomised trials are and can be implemented as intended in practice. Second, there is concern and evidence that - in comparison to high fidelity programme delivery via randomised trials - interventions delivered in community settings and at scale will suffer from diminished effects due to poor implementation fidelity (e.g., Araujo et al., 2021; Institute of Medicine & National Research Council, 2014). However, to know whether there is poor implementation in these contexts, there is need to investigate how to measure fidelity and what role it plays in observed participant outcomes. In addition, focusing on fidelity is a worthwhile line of investigation in PLH as these programmes are being disseminated on a large-scale with little control over programme drift.

### **2.2.5 Facilitator Competent Adherence**

Adherence and quality of delivery were chosen for study because they are directly related to the individuals (facilitators) who implement programmes. Adherence, also referred to as ‘fidelity’ or ‘integrity’, is the strictness with which a facilitator implements prescribed programme content (e.g., explaining concepts), activities (e.g., group discussion), and strategies (e.g., praising parents) (Dane & Schneider, 1998; Dusenbury et al., 2003; Fixsen et al., 2005). Quality of delivery, also referred to as ‘competence’, refers to the skill and style with which a facilitator delivers programme components in practice (Dane & Schneider, 1998; Dusenbury et al., 2003; Fixsen et al., 2005). These components may be outlined by programme developers or dictated by professional standards (e.g., social workers, teachers, nurses, psychologists). Some researchers argue that a facilitator

cannot be competent without being adherent (Muse & McManus, 2013; Waltz et al., 1993). Although this argument has some merit and may be accurate in some instances, facilitators may be skilled at delivering certain programme components without necessarily delivering all components. Further, as it is useful to analyse both what was delivered and how, the view that adherence is subsumed within competence has not been adopted. However, delivering a programme competently is not possible if adherence to the intervention is too low (Stirman, 2020).

The combination of facilitator adherence and competence is also referred to as ‘competent adherence’ (Breitenstein et al., 2010a; Forgatch et al., 2005). According to Forgatch et al., “current zeitgeist calls for assessing two dimensions of method fidelity: adherence to core program criteria as specified in manuals and competent delivery of the program” (2011, p. 236). As a composite of two aspects of implementation fidelity, competent adherence refers to both the skill with which a facilitator delivers intervention components and the strictness with which a facilitator follows the activities outlined in the programme manual (Fixsen et al., 2005). Many parenting programmes have developed a measure to assess the competent adherence of their facilitators.

For the purposes of this dissertation, a facilitator is defined as a practitioner or lay person who delivers a programme to participants (Fixsen et al., 2005). In the literature, a facilitator may also be referred to using terms such as ‘purveyor’ (Fixsen et al., 2005, p. 537), ‘clinician’, ‘therapist’, or ‘practitioner’, or ‘group leader’.

### 2.2.6 Conceptualisation of the Relationship between Implementation and Outcomes

Conceptualisation of the potential relationship between facilitator delivery and outcomes was informed by Berkel et al.'s (2011) Integrated Model of Program Implementation and Carroll et al.'s (2007) Conceptual Framework for Implementation Fidelity. This model and framework were selected because they hypothesise potential pathways through which implementation components influence participant outcomes. Both Berkel and Carroll hypothesise that participant outcomes are a function of the degree to which the intervention is delivered with implementation fidelity. As such, Berkel's model and Carroll's framework informed Paper 3 of the dissertation, which investigates the relationship between facilitator delivery and participant outcomes using data from PLH-Teens in Tanzania. A visual representation of the relationship between facilitator delivery and outcomes is shown in Figure 1.

**Figure 1**

*Hypothesised Relationship between Facilitator Delivery and Participant Outcomes*



### 2.2.7 Approach to the Study of Implementation

The dissertation utilised the four steps proposed by Durlak (1998) to study the competent adherence of facilitators delivering PLH-Teens in Tanzania: delineating the



programme's active ingredients; creating a reliable and valid method for assessing implementation; measuring implementation during programme delivery; and examining whether programme implementation is associated with participant outcomes. These steps were followed in two of the three papers herein by testing the reliability and validity of a measure (steps 1 and 2, Paper 2), employing the measure to assess the delivery of facilitators (step 3, Paper 2), and using the assessment results to examine the relationship between facilitator delivery and outcomes (step 4, Paper 3).

### **2.3 Parenting for Lifelong Health**

Building on the parenting programme literature and theoretical models and frameworks in the implementation science literature, Papers 2 and 3 examine the measurement and role of facilitator competent adherence within the context of the PLH-Teens programme. This section provides background information about PLH programmes, the delivery of PLH-Teens in Tanzania, and situates the research as part of the FAIR study.

PLH has developed a suite of group-based parenting programmes, initially to be delivered in-person but more recently investigated for digital and hybrid delivery, aiming to reduce child behaviour problems and familial violence against children (Lachman et al., 2016). PLH programmes are targeted at parents with children across the development spectrum: PLH for Babies (prenatal-6 months), PLH for Toddlers (10-60 months), PLH for Young Children (PLH-YC) (2-9 years), and PLH for Parents and Adolescents (PLH-Teens) (10-17 years). PLH programmes were developed and evaluated by researchers and practitioners from several universities and non-profit organisations, including the University of Oxford, the University of Cape Town, the University of Stellenbosch, Clowns

Without Borders South Africa (CWBSA), UNICEF, and the WHO. PLH programmes are being implemented in over 35 LMICs in sub-Saharan Africa, Southeast Asia, Latin America, and Eastern Europe, with PLH-Teens being delivered in 19 LMICs. To date, PLH programmes have reached hundreds of thousands of beneficiaries (Shenderovich et al., 2020). The implementation of PLH programmes in a variety of contexts provides an opportunity to understand the translation of a promising intervention from research into practice.

### **2.3.1 Evidence Base for PLH-Teens**

PLH-Teens is among the few low-cost parenting interventions for families with adolescents that has been rigorously tested in LMICs (Cluver et al., 2018). Originally developed and tested in South Africa, PLH-Teens aims to reduce adolescent exposure to violence in the home and community by improving positive parenting and parent-child communication while reducing familial conflict, harsh discipline, parenting stress, child behaviour problems, risky behaviour, and mental ill-health (Cluver et al., 2016). Trained community facilitators engage both parents and adolescents in 14 weekly sessions. Ten of the 14 sessions are delivered to groups of parent-adolescent dyads and four of the 14 sessions are delivered to groups of adolescents and parents separately. Facilitators use non-didactic, participatory methods including group discussions, role-plays, problem-solving, and experiential activities (Wessels & Ward, 2015). Facilitators also assist families in developing child safety plans, responding to abuse, budgeting, and accessing medical and social services. Thus, PLH-Teens tackles a multitude of upstream and downstream

contextual factors that lead to increased risk of violence against children (e.g., Fraser, 1997; Garbarino, 1977; Merrill et al., 1996).

A recent cluster randomised trial of PLH-Teens in South Africa by Cluver et al. (2018) ( $N=40$  clusters, 552 parent-adolescent dyads) found longer-term intervention effects, as reported by parents, for reduced abuse and corporal punishment and improved positive parenting, involvement, and monitoring at five- to nine-months post-intervention. The trial also found improved positive parenting and less poor supervision as reported by adolescents but did not find decreases in adolescent-reported abuse or corporal punishment. The trial did not find decreases in parent- or adolescent-reported neglect. Positive intervention effects on secondary outcomes, as reported by parents, included reductions in parenting stress, depression, endorsement of corporal punishment, financial stress, and substance use (Cluver et al., 2018). Adolescents did not report substantial differences in secondary outcomes except for use of substances, which was found to decrease. No harmful effects were detected. A recently published mixed methods evaluation of the delivery of PLH-Teens in the Philippines ( $N=30$  families) conducted by Jocson et al. (2023) found positive intervention effects (e.g., reductions in child maltreatment, child behaviour problems, neglect, acceptability of corporal punishment, parenting stress, parent and adolescent depression) as well as positive participant views regarding the programme's feasibility and acceptability. A cost-effectiveness analysis of PLH-Teens estimated that the intervention cost \$972 USD per case of abuse prevented and had a total estimated cost savings of \$2,724 USD per case of abuse (Redfern et al., 2019).

Although there is strong, emerging evidence of the effectiveness of PLH programmes, there are gaps in understanding about its implementation. Specifically, there are gaps in understanding about the competent adherence of facilitators delivering the programme at scale and about the role competent adherence plays in the achievement of parent and adolescent outcomes. A study on the implementation of PLH-Teens in South Africa, which did not use the PLH-FAT-T, investigated the role of facilitator delivery, participant attendance, and participant engagement on family outcomes (Shenderovich et al., 2019). Aside from better fidelity predicting greater reductions in violence against children based on adolescent-reports, implementation was not found to predict participant outcomes. To build on this PLH-Teens research, the dissertation focuses on the competent adherence of facilitators delivering the PLH-Teens programme in Tanzania.

### **2.3.2 PLH-Teens in Tanzania**

The encouraging results from the cluster randomised trial (Cluver et al., 2018) has contributed to the rapid dissemination of PLH-Teens. In Tanzania, the large-scale implementation of PLH-Teens is one of a series of interventions being implemented as part of the Kizazi Kipya Project by an organisation called Pact Tanzania (Pact). Kizazi Kipya is funded by the USAID-PEPFAR (United States Agency for International Development-President's Emergency Plan for AIDS Relief) DREAMS (Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe) Initiative, which aims to reduce the incidence of HIV among adolescent girls, support family economic strengthening, and enhance parenting skills. PLH-Teens is a sanctioned intervention in the USAID-PEPFAR DREAMS Initiative in sub-Saharan African and the Caribbean. The delivery of the programme in

Tanzania is important as over 72% of youth aged 13 to 24 years in the country report experiencing physical violence before age 18 (UNICEF, 2011). Caregivers, other adult relatives, and teachers are the most commonly reported perpetrators of physical and emotional violence against children in Tanzania, with corporal punishment considered normative (UNICEF, 2011).

Prior to programme delivery in Tanzania, PLH-Teens went through a cultural adaptation process, involving both surface and deep adaptations, to create an HIV-enhanced version of the programme (the Furaha Caring Families Programme for Parents and Teens, locally known as Furaha Teens or “Happy Families” in Kiswahili). The adaptation process resulted in revisions to the images and characters in the programme; translation of the programme materials into Kiswahili; the addition of content on HIV prevention and risk reduction (e.g., role-play on HIV medication and HIV disclosure); and modification of the family economic strengthening content to suit the Tanzanian context (CWBSA, 2017). Following adaptation, the revised programme was tested with several hundred families and revised further before the programme was implemented on a larger scale.

In 2020-2021, Pact scaled up PLH-Teens with 444 trained facilitators and 69 coaches to reach 75,061 beneficiaries ( $n=36,259$  parents and  $n=38,802$  adolescent girls). The 2020-2021 delivery of PLH-Teens in Tanzania offered an unprecedented opportunity to examine the intervention and its implementation at scale. Papers 2 and 3 use data collected on the delivery of PLH-Teens in Tanzania as part of the FAIR study.

#### *Delivery, Locations, and Timeline*

PLH-Teens was delivered in eight districts of rural and semi-urban Tanzania: Kyela District Council (DC), Mbeya DC, Muleba DC, Shinyanga DC, Shinyanga Municipal Council (MC), Kahama Town Council (TC), Msala DC, and Ushetu DC. The programme was delivered in-person to groups comprised of approximately 20 dyads or 40 people (approximately 20 adolescent girls and 20 parents, including mothers, fathers, and other biological and non-biological caregivers). Pact planned for PLH-Teens to be delivered three times with each wave of delivery involving 14 weekly sessions. Wave 1 was to take place from approximately December 2019 to February 2020, Wave 2 was to take place from approximately February to April 2020, and Wave 3 was to take place from approximately April to July 2020. However, programme delivery experienced many delays due to the COVID-19 pandemic and the practicalities of delivery at scale. Wave 1 was delivered from January to March 2020 and then completed in July 2020 as there was a three month pause due to school closures and lockdowns. Wave 2 was delivered from July to October 2020. Wave 3 was partially delivered from December 2020 to March 2021. Only part of Wave 3 was delivered because higher than anticipated costs meant that there were insufficient funds remaining to continue programme delivery and the latter half of Wave 3 was not able to proceed.

#### *PLH-Teens Content*

The 14 weekly sessions in the programme are intended to be approximately three hours each. Using an experiential and strengths-based approach, facilitators guide groups of parents and adolescents to discuss and practice skills on a range of topics including establishing goals, quality time together, praising each other, talking about feelings, managing anger, problem solving, saving and making a budget, dealing with problems

without conflict, establishing rules and routines, keeping safe in the community, responding to crisis, and widening social supports (Parenting for Lifelong Health, 2018). As mentioned, the content of PLH-Teens was enhanced via a programme adaptation process. According to the programme manual, each weekly session involves a discussion on the week's topic as well as a role-play activity to support parents and adolescents to practice what they have learned. The second and each subsequent weekly session includes a conversation on the home activities from the previous week. Programme sessions also include physical exercises as well as exercises to relax (i.e., "taking a pause") and recognise emotions (i.e., "emotional check-in").

#### *PLH-Teens Implementers*

PLH-Teens was delivered in schools and community centres by community facilitators who were either school teachers or community health workers ( $N=444$ ) with programme coaches ( $N=69$ ) providing facilitators with ongoing supportive supervision (i.e., mentorship and feedback). Community facilitators delivered the programme via the coordination of five local implementing organisations - Humuliza, Tadepa, Integrated Rural Development Organisation, Caritas, and Red Cross - under the leadership of Pact. Facilitators and coaches were recruited by the local implementing organisations.

#### *PLH-Teens Facilitator and Coach Training*

The 444 facilitators were provided with five consecutive days of training on how to deliver the programme. Facilitators were trained by instructors from CWBSA, which was contracted to provide Pact with support throughout the delivery of PLH-Teens in Tanzania. The facilitator training was developed by CWBSA and PLH to provide facilitators with an

understanding of the purpose of the programme, a session by session overview of the curriculum, and an understanding of PLH's key facilitation approaches. CWBSA delivered the facilitator training using experiential learning techniques including presentations, group discussions, skill demonstrations, and practice activities. The training was completed using a facilitator manual to support programme delivery. The 69 coaches were provided with three consecutive days of training on how to provide supportive supervision (e.g., support, feedback, mentorship) to the facilitators and two days of training on how to conduct facilitator assessments. Coach training was also delivered by instructors from CWBSA. The coach training used similar materials and delivery approaches to the facilitator training, but included additional topics and practice activities related to the role of coaches and how they would provide facilitators with supportive supervision.

#### *Facilitator Assessments*

As part of the delivery of PLH-Teens, Pact planned for each facilitator to receive at least one assessment of their delivery using the Parenting for Lifelong Health for Parents and Adolescents-Facilitator Assessment Tool (the PLH-FAT-T). PLH-FAT-T assessments of facilitator programme delivery were to be conducted by coaches. However, only 95 facilitator assessments were collected during programme delivery due to a variety of factors including the COVID-19 pandemic and the practicalities of programme delivery at scale.

#### *Informed Consent and Eligibility*

Pact asked facilitators and coaches for their consent to participate in the research after their recruitment as programme staff.



### **2.3.3 Furaha Adolescent Implementation Research (FAIR) Study**

The FAIR study aims to provide vital information on how to establish, implement, improve, and sustain high-quality delivery of PLH-Teens. FAIR was led by the National Institute for Medical Research in Mwanza, Tanzania in collaboration with the University of Oxford, Cardiff University, CWBSA, and Pact. The FAIR study is part of an even larger study called the Scale-Up of Parenting Evaluation Research, which is investigating the implementation of PLH programmes in multiple LMICs (Shenderovich et al., 2020). The FAIR study is examining the quality of delivery and impact of PLH-Teens on preventing and reducing violence against children at scale in Tanzania and further, how implementation quality can be improved to optimise intervention impact (Martin et al., 2021a).

As a mixed-methods study, FAIR involved the integration of qualitative (primary) data and quantitative (secondary) data. Qualitative (including focus group discussions and in-depth semi-structured interviews collected by the FAIR team) and quantitative (merged secondary data collected via routine monitoring and evaluation by Pact, local implementing partners, and CWBSA) methods were used to answer the FAIR research questions. As randomisation to intervention and control groups was not possible, the study made the most of the routine service delivery data available. Analysing this data allowed for a unique inquiry into the real-world implementation of a parenting programme at scale. Altogether, the FAIR study collected primary data from 56 programme coaches, 102 programme facilitators, 16 Pact and local implementing partner staff members, seven school principals, three CWBSA staff members, 100 parents, and 60 adolescent girls. The FAIR study also

analysed anonymised secondary data from parents, adolescent girls, programme facilitators, programme coaches, and implementing partner staff. Some of the secondary data collected from parents, adolescents, and facilitators were analysed in the dissertation.

#### **2.3.4 Note on Candidate's Role**

I played a significant role in the FAIR study as the Oxford team's Research Manager wherein I supported the FAIR team by co-writing and revising the grant application, drafting and refining ethics applications, designing primary data collection tools (e.g., semi-structured interview guides), creating and conducting staff training, developing PLH-FAT-T assessment training materials, drafting definitions for PLH-FAT-T items, revising the PLH-FAT-T, updating the facilitator assessment manual, developing a research uptake strategy, analysing results (e.g., coding qualitative data), crafting dissemination materials (e.g., published evidence brief) (Martin et al., 2022b), and presenting findings (see [8.8](#)).

### **3. Review of Literature Underpinning the Methods**

This chapter outlines the literature underpinning the methodological aspects of the dissertation. The chapter is divided into three sections, the first two of which are based on studies from my master's thesis. My master's thesis was composed of two papers – a systematic review of the measures of competent adherence used in the parenting programme literature and their psychometric properties as well as a psychometric evaluation of the observational measure of competent adherence used in the PLH for Young Children programme (PLH-YC) using data from Southeastern Europe (North Macedonia, Romania, and Moldova). Drawing on the findings of the systematic review conducted for my master's (Martin et al., 2021b), the first section (3.1) summarises the measures of competent adherence used by parenting programmes aiming to reduce violence against children and child behaviour problems. The second section (3.2) describes the psychometric properties of the observational measures of facilitator competent adherence used in the parenting programme literature (Martin et al., 2021b). This section also summarises the existing psychometric evidence on the tool used to assess the competent adherence of PLH-YC facilitators using the findings from an evaluation of the tool conducted during my master's (Martin et al., 2022a). The third section (3.3) describes the evidence on the relationship between facilitator competent adherence and parent and child outcomes in the parenting programme literature and argues in favour of the need to investigate this relationship both in the broader parenting programme literature and in the context of PLH.

### **3.1 Measures of Facilitator Competent Adherence**

There is a wide range of measures of facilitator competent adherence documented in the parenting programme literature. The measures are varied with respect to the parenting programmes for which they were designed; facilitators assessed; domains of competence and/or adherence captured; modes through which assessments are collected; individuals who conduct assessments; and response types used to rate facilitators. Overall, most measures assess facilitator adherence, use observational methods, are completed by researchers, and employ Likert scales to rate highly educated and experienced facilitators delivering well-known parenting programmes in high-income countries (Martin et al., 2021b).

There is a paucity of research on the competent adherence of facilitators delivering parenting programmes in LMICs. A systematic review conducted as part of my master's identified 65 observational and non-observational measures of facilitator competent adherence in 63 different parenting programmes (Martin et al., 2021b). The programmes that reported on measures of facilitator competent adherence most frequently were Multi-Systemic Therapy, Incredible Years, Parent-Child Interaction Therapy, Parent Management Training – Oregon Model, and Triple P (Martin et al., 2021b). Among the 65 measures, the most reported measures were Therapy Adherence Measure; Fidelity of Implementation Rating Scale; Therapy Adherence Measure-Revised; COACH Rating System; and the Leadership Observation Tool. PLH and other parenting programmes designed for or delivered in LMICs are underrepresented in this literature, with only eight known studies

reporting on measures of facilitator competent adherence used in LMICs (Martin et al., 2021b).

As it relates to the facilitators assessed in the measures documented in the literature, studies typically assessed small samples of highly educated and experienced Caucasian female facilitators of various ages, with a few US studies reporting African American facilitators (Martin et al., 2021b).

### **3.1.1 Domains of Competent Adherence**

Measures are designed to capture varying combinations of competence and adherence, including competence-only, adherence-only, competent adherence, or competence and adherence (Martin et al., 2021b). Competence-only measures are those that seek to assess the quality with which facilitators deliver a programme (e.g., use of body language, ability to engage participants, tone of voice, confidence speaking in front of groups, making connections between participant experiences and programme content). For example, in a study by Roggman and colleagues (2001), participating parents were asked to rate the quality with which facilitators delivered the Early Head Start home visitation programme. A limitation of competence-only measures is they do not capture the extent to which programme components are implemented (i.e., adherence).

Adherence-only measures aim to assess the extent to which facilitators implement a programme as designed. For example, in a study by Furlong and colleagues (2015) the authors describe the self-report measure completed by facilitators each week to document what aspects of the Incredible Years programme they delivered. Previous findings in the literature indicate that adherence-only measures are most common, possibly because they

are quick and simple to design and use (Goense et al., 2015). These types of tools are limited as they do not allow for the measurement of the skill and style with which facilitators deliver programmes (i.e., competence).

Studies that assess competence and adherence are those that measure both concepts using separate scales. For example, Webster-Stratton and colleagues (2014) report on the use of multiple measures completed by different programme stakeholders. After every session, facilitators completed a checklist to assess their adherence during programme delivery and parents were asked to complete a different measure to assess facilitator competence (using four questions rated on a 1-4 Likert scale). Independent observers also completed a competent adherence measure which assessed facilitators on eight domains using a five-point Likert scale.

Measures of competent adherence are those that assess both competence and adherence using one tool. These measures detail whether, and how well, facilitators deliver a programme as intended. For example, in a study by Smith and colleagues (2019), the authors report on the COACH Rating System which captures the integrity and quality of facilitator delivery on five aspects of the Family Check-Up programme. Measures that assess both adherence and competence are particularly insightful as they attempt to establish a complete picture of facilitator delivery.

### **3.1.2 Mode of Data Collection and Assessors**

Measures of competent adherence can be broadly classified as observational or non-observational and are completed by various assessor types. Observational methods require assessors to conduct facilitator assessments by watching live or video-recorded

programme sessions (Eames et al., 2008). Observational assessments are typically conducted by researchers or supervisors (those responsible for facilitator quality assurance). Generally, this mode of data collection results in assessments that are more accurate than those conducted using non-observational methods (Girard & Cohn, 2016). Observational methods are the gold standard form of assessment as they garner more accurate results due to heightened objectivity as well as their ability to capture the body language of facilitators and participants (Eames et al., 2008; Gardner, 2000). Observational assessments also have drawbacks as they are time and resource intensive and assessments may be impacted by reactivity bias whereby those being assessed deliver the programme differently due to being observed (Breitenstein et al., 2010b; Girard & Cohn, 2016).

For the dissertation, non-observational methods are defined as those in which assessors use techniques such as listening to audio recordings of programme sessions (typically completed by researchers or supervisors) or in which facilitators complete self-reports of their delivery. Other non-observational approaches include participant recollection of delivery or the use of automated/machine learning approaches. Non-observational methods, especially assessments based on facilitator self-report, may not be as reliable as observational methods due to social desirability bias and because memory can be faulty (Mowbray et al., 2003; Stone et al., 1999). Also, assessments based on audio recordings or automated/machine learning approaches may not be as reliable as observational methods because assessors may not have the opportunity to assess certain elements of programme delivery (e.g., body language, participation reactions, participant engagement) (Breitenstein et al., 2010b).

Video-based observation (alone or in combination with other assessment approaches) was the most common assessment method documented in the literature, which demonstrates that programmes frequently use the more rigorous method of data collection (Martin et al., 2021b). Among non-observational methods, memory-based assessments were most common with the majority being facilitator self-reports. In some cases, assessments were based on parents or supervisors recounting facilitator delivery. Facilitators and independent assessors (e.g., researchers, trainers) were most often reported as conducting assessments, with a significant number of studies using multiple types of assessors (Martin et al., 2021b).

Video-based assessments are an ideal form of observational assessment for multiple reasons. Conducting facilitator assessments based on video recordings may reduce reactivity bias compared to assessments conducted in-person. If an assessor is viewing programme delivery in-person, facilitator awareness of being observed may be heightened (Gardner, 2000; Girard & Cohn, 2016). Additionally, if video recordings are taken during every programme session, facilitators may become accustomed to the presence of the camera thereby reducing the extent to which they perform differently (Feely et al., 2018). Using video recordings also reduces the time assessors spend travelling to and from programme sessions to observe delivery (Feely et al., 2018). However, facilitators may perform differently with the presence of a video camera (Breitenstein et al., 2010b). In comparison to in-person assessments, video-based assessments save time and other resources. Using video recordings also allows assessors to complete assessments on their own timeline (Breitenstein et al., 2010b). Finally, the use of video recordings to conduct assessments makes the establishment of intra- and inter-rater reliability simpler. A video



recording of programme delivery allows assessors to assess the same programme multiple times (for intra-rater reliability) and allows multiple assessors to assess the same sessions (for inter-rater reliability) (Breitenstein et al., 2010b). However, video-based assessments may reduce the confidentiality of parenting programme participants or miss programme activities that happen outside of the view of the camera (Breitenstein et al., 2010b). Videos can pose additional challenges for implementers in low-income contexts as they need the funds to purchase video equipment and the capacity to store and transmit video recordings.

### **3.1.3 Response Types**

Measurement tools designed to assess facilitator competent adherence typically use response types with dichotomous or Likert scale response options, or response types which count the frequency of certain facilitator activities or behaviours (Martin et al., 2021b). With dichotomous items, assessors are given two choices, such as ‘yes’ and ‘no’, in responding to an item. These response options are clear, making it easier to establish inter-rater reliability between assessments. While dichotomous items are likely well-suited to capturing facilitator adherence (e.g., ‘completed’ or ‘not completed’), having only two response options may mean that nuances in programme delivery are missed, particularly intricacies associated with measuring facilitator skills. With Likert scale items, assessors have more options which means that they can capture gradations in facilitator delivery. As a result, these response options require finer judgements to be made by assessors, making it harder to establish inter-rater reliability. Likert scale items are well-suited to questions capturing facilitator competence as skills are hard to assess with simple ‘yes’ and ‘no’ responses. The fidelity checklist utilised to capture facilitator competent adherence in the Chicago Parent Program is an example of a measure that uses both Likert scale and

dichotomous response types (Breitenstein et al., 2010a). This session-specific checklist is comprised of up to 31 items (depending on the session) wherein questions relating to adherence are rated as either ‘yes’ or ‘no’ and those relating to competence are rated on a scale of one to three (Breitenstein et al., 2010a). Another response type is frequency counts, which asks assessors to count the number of instances of a facilitator delivering a particular intervention component or using a particular skill. Frequency counts are beneficial as they allow assessors to watch only a fixed period of programme delivery (e.g., 20 minutes). However, in sampling short periods of programme delivery, the validity of these assessments may not be known. In other words, it is difficult to establish whether overall facilitator delivery will be captured within the interval of time selected for assessment. It is also difficult to determine what frequency of a specific facilitator behaviour represents high-quality delivery. An example of such a measure is the Leadership Observation Tool used in the Incredible Years BASIC programme, which has assessors code the frequency of 18 facilitator behaviours in intervals of ten minutes over a two-hour period (Eames et al., 2008).

### **3.1.4 Dominant Measure Types**

Bringing together the domains of competent adherence, mode of data collection, assessors, and response types commonly used in the literature, the dominant types of measures identified in the literature can be illustrated using four groupings (Martin et al., 2021b). The first type measures adherence by asking facilitators to self-report on the implementation of specific programme activities using dichotomous items. An example of such a non-observational tool is used by Lester (2015) in the Positive Parenting Skills Training Program wherein facilitators complete session-specific forms with 11 to 12 ‘yes’

or ‘no’ questions. A strength of this type of measure is that it is quick, cheap, and simple, however, reliability may be questionable due to factors such as social desirability (Mowbray et al., 2003; Stone et al., 1999).

The second type is a non-observational measure that asks recipients of the intervention to assess one or more aspects of facilitator competent adherence. For instance, Chapman and Schoenwald (2011) asked families in MultiSystemic Therapy to rate facilitator adherence using the Revised Treatment Adherence Measure. This measure captures nine aspects of delivery using 28 items rated on a five-point scale from ‘not at all’ to ‘very much’. Using participant assessors is valuable in that these assessments capture an important perspective and are cheaper than using trained assessors. However, reliability may be limited due to factors such as a lack of training on how to conduct assessments and, as with facilitator self-reports, reliance on memory.

The third type involves researchers, supervisors, or other independent observers completing video-based or live assessments of facilitator competent adherence. A study by Bywater et al. (2019) on the Incredible Years, for instance, reports on the Parent Program Implementation Checklist used to capture competent adherence. This checklist is composed of 18 items rated from ‘not at all’ (1 point) to ‘excellent’ (5 points). A well-cited measure of competent adherence found in the literature is the Fidelity of Implementation Measure used in the Parent Management Training-Oregon Model (PMTO), which assesses facilitators on five components (“PMTO knowledge, structure, teaching skills, clinical skills, and overall quality”, p.7) using a nine-point scale ranging from ‘needs work’ to ‘good work’ (Forgatch et al., 2005). Although this approach provides rich and objective

information about facilitator delivery, it is time consuming and resource intensive.

Additionally, the independence of some assessors may still be limited using this approach, such as if assessments are conducted by supervisors who have a prior relationship with facilitators.

The fourth type is a measure that uses multiple assessors to capture one or more aspects of competence and adherence using observational and/or non-observational approaches. An example is the Alternatives for Families: A Cognitive Behavioral Therapy Program Treatment Adherence Form described by Herschell et al. (2019). In this study, three different assessor-types were used – caregivers, facilitators, and experts. Caregivers and facilitators recorded their recollection of delivery whereas experts conducted their assessments based on audio recordings. All assessors were asked to complete nine dichotomous items (‘occurrence’ or ‘non-occurrence’) to indicate facilitator adherence. The use of multiple assessors is recommended by some researchers (Ginsburg et al., 2021) as it may enhance the reliability of the assessments, but it also requires considerably more time and effort to analyse the results, especially if there is disagreement among assessors. As a result, using multiple assessors increases the resources required.

### **3.1.5 Challenges of Measuring Competent Adherence using Observational Methods**

There are numerous challenges associated with measuring competent adherence using observational methods. First, competence is a subjective concept making the development of a tool to measure it a complex task as it is difficult to pinpoint, and therefore assess, the precise impactful outward manifestations of high-quality delivery (Mowbray et al., 2003). Further, using such a tool is challenging as interpretations of what

is being observed may differ by assessor, resulting in variations in their ratings on measure items. Second, the presence of an external observer who is there to assess a facilitator or the presence of a video camera may alter their natural delivery, and thus may introduce reactivity bias (Gardner, 2000). Third, other implementation factors such as participant attendance, participant engagement, and parent-facilitator therapeutic alliance may influence facilitator delivery making accurate measurement difficult (Feeley et al., 1999; Webb et al., 2012). Fourth, it is unclear when and how often facilitators should be assessed to accurately establish their competent adherence (Mowbray et al., 2003). It is debateable, for instance, if the assessment results from a single programme session can be an accurate indication of a facilitator's overall delivery (Waltz et al., 1993). Yet, capturing facilitator delivery at multiple timepoints is typically not feasible in routine service delivery settings (Stirman, 2020). Fifth, the role of co-facilitation may need to be considered during assessments when programmes are delivered by two or more facilitators working together with each delivering different components. Sixth, the process for measuring facilitator competent adherence is complicated, time consuming, and resource intensive (Horvath et al., 2011) as it necessitates that researchers and practitioners implement multiple steps such as designing a measure; testing the validity and reliability of the measure; improving the measure after testing; creating an assessor training programme and training materials on how to use the measure; recruiting assessors; conducting assessor training; having assessors conduct the assessments (either by observing live sessions or by watching videotaped sessions); compensating assessors; providing support and feedback to assessors throughout the assessment process; collecting the assessments; entering and tabulating the assessment data; analysing the data; and then using the data, for example, to make changes to improve

programme implementation via revisions to facilitator training, programme manuals, and supervision processes (e.g., Fixsen et al., 2005; Ginsburg et al., 2021; Sanders et al., 2020; Stirman, 2020). Seventh, the challenges of measuring facilitator competent adherence are only heightened during the delivery of interventions in community settings and at scale (Ginsburg et al., 2021). In these contexts, additional challenges, such as missingness in the data collected, are often experienced (Ginsburg et al., 2021; Wolpert & Rutter, 2018).

### **3.2 Reliability and Validity of Measures of Competent Adherence**

This section provides an overview of how to define and evaluate measure reliability and validity; summarises the findings of a systematic review conducted during my master's which reported on the psychometric properties of 30 observational measures of facilitator competent adherence found in the parenting programme literature (Martin et al., 2021); and describes the psychometric properties of the Facilitator Assessment Tool used in PLH-YC (Martin et al., 2022a).

#### **3.2.1 Psychometric Analyses**

Even though many measures of facilitator competent adherence exist, they may not be reliable and/or valid. It is fundamental that fidelity measures are reliable and valid as these properties speak to whether a tool can be administered consistently and capture what it intends to measure (Mowbray et al., 2003). The guidance and criteria provided by the Consensus-Based Standards for the Selection of Health Measurement Instruments (COSMIN) are typically used to evaluate measure reliability and validity (Mokkink et al., 2016). Evaluations of measure reliability and validity provide practitioners and researchers with a sense of the quality of implementation measures currently available; support

researchers in determining which measures need further testing and analysis; and allow practitioners and researchers to quickly assess the desirability of using a certain measure in future programme implementation and evaluation. It is commonly acknowledged that further evidence on the reliability and validity of measures of facilitator competent adherence is needed (Ginsburg et al., 2021; Ruud et al., 2020; Stirman, 2020).

### **3.2.2 Measure Reliability**

Establishing the reliability of a measure is critical to determining whether assessments of facilitator competent adherence can be trusted and completed consistently over time. Reliability is the degree to which a measure produces consistent outputs across multiple measurements under various conditions (Fan & Randall, 2018; Mokkink et al., 2010a). It refers to the extent to which assessments conducted using a measure document competent adherence as opposed to variations by an assessor (intra-rater reliability), between assessors (inter-rater reliability), in items (internal consistency), and in facilitator delivery (test-retest reliability). Reliability does not provide any information about the accuracy with which a measure captures the construct(s) of interest, but instead captures the extent to which assessment results can be reproduced over time and is therefore a necessary precursor to establishing validity (Gushta & Rupp, 2010).

Three types of reliability – internal consistency, inter-rater reliability, and intra-rater reliability – are evaluated in Paper 2. Reliability is best assessed when it is possible to examine all three types, or as many aspects as possible, together (Gushta & Rupp, 2010). Each type of reliability has advantages, disadvantages, and underlying assumptions to consider when interpreting study results (Stemler & Tsai, 2008). The following section

describes the definitions of and methods for evaluating each of these types of reliability and describes the evidence on the reliability of facilitator competent adherence measures in the parenting programme literature.

### *Defining Reliability*

Internal consistency measures the degree to which there is variability among measure questions (items) and is determined by examining how responses are statistically interrelated (Heinl et al., 2016; Terwee et al., 2007). Internal consistency is an important aspect of reliability as it establishes whether items that address similar concepts result in similar responses. Thus, the level of consistency between or among items intended to measure similar constructs is an indicator of reliability. However, the measurement of internal consistency has limitations. For one, while high internal consistency may suggest that the items are intercorrelated, it may also be that items are too similar to make useful distinctions between them (Barchard, 2010). For another, analysing internal consistency is not useful if the developers of a measure hypothesised that certain items would not be associated with each other (Barchard, 2010). Thus, it may not make sense to analyse internal consistency if some of the items are not intended to cluster in a conceptual grouping. Finally, the internal consistency of a measure is impacted by the length of the test and who completes the test, with longer measures and measures completed by more heterogeneous groups of individuals tending to be more reliable (Furr, 2017). As a result, these factors and influences should be considered carefully when assessing the internal consistency of a measure.



Inter-rater reliability gauges the consistency of different assessors rating the programme delivery of the same facilitator (Chen & Krauss, 2004; Cho, 1981; Hallgren, 2012; Heidl et al., 2016). Variability between assessor ratings signals inconsistent application of the tool by assessors. When assessing inter-rater reliability, researchers must consider whether to prioritise assessor consistency or consensus; distinguishing between these types of agreement has implications for the time and resources necessary to achieve high inter-rater reliability (Stemler & Tsai, 2008). Consensus indices – represented by percentage agreements and Cohen’s Kappas – are useful for determining how often assessors rate measure items in exactly the same way, whereas consistency indices – represented by ICCs, correlations, and Cronbach alphas – are useful for determining the extent to which assessors generally rate items in the same way (Stemler & Tsai, 2008). Typically, more time and resources are necessary to ensure consensus than consistency.

Intra-rater reliability quantifies the degree of assessor stability in rating the same programme delivery at different times (Gwet, 2014; Heidl et al., 2016) and is less frequently calculated in the competent adherence literature (Martin et al., 2021b). For facilitator competent adherence measures, establishing intra-rater reliability is done by having assessors watch and assess a video recording of the same programme delivery twice – with time between each of the two observations and assessments. The time and resources required to do the observation and assessment twice may explain why this property is rarely studied. However, intra-rater reliability is an important component of reliability as it captures the reproducibility of assessments by the same assessor on different occasions (Gwet, 2014). A limitation of this type of reliability is that intra-rater reliability may reflect

the skill of the assessor rather than the reliability of the measure. As with inter-rater reliability, when assessing intra-rater reliability, decisions must be made about whether to employ consistency or consensus indicators.

### *Evaluating Reliability*

Percentage agreements, ICCs, Cohen's Kappa, and Pearson and Spearman correlations are often used to assess intra- and inter-rater reliability (Heinl et al., 2016). It is ideal if studies do not report and reach conclusions only based on percentage agreements, even though they are simple to calculate and easy to interpret, as this type of agreement is more susceptible to chance (McHugh, 2012; Stemler & Tsai, 2008). Statistical analyses such as ICCs address this issue by taking chance agreement into account and by considering how close item responses are to each other (Hallgren, 2012). Strong reliability is usually indicated by ICCs greater than 0.70, Kappas greater than 0.70, or Pearson's correlations greater than 0.80 (Heinl et al., 2016). When interpreting the results, it is important to acknowledge that Kappa statistics can become skewed when certain results occur infrequently, and the strength of ICCs can be reduced if the data are not normally distributed (Snow et al., 2005). Cronbach alphas and person separation indices are commonly used to assess internal consistency (Heinl et al., 2016). Additional methods for calculating internal consistency include producing raw Cronbach's alphas, standardised coefficient alphas, raw alphas for binary items ( $KR_{20}$ ), and omega coefficients (Furr, 2017). These methods all make different assumptions and are applicable to different types of measures (Furr, 2017). While it is commonly agreed that higher Cronbach alphas mean higher consistency, there is debate over what level of consistency is acceptable (O'Leary-

Kelly & Vokurka, 1998). Although alphas and person separation indices of 0.70 or above are generally seen as indicative of good reliability, some researchers argue that alphas of 0.50 or above are sufficient for new measures (Heinl et al., 2016; Nunnally & Bernstein, 1967; O'Leary-Kelly & Vokurka, 1998; Terwee et al., 2007). It is advisable for researchers to take these various considerations and limitations into account when interpreting studies of measure reliability (Stemler & Tsai, 2008).

#### *Reliability of Measures of Facilitator Competent Adherence*

The systematic review conducted for my master's found two studies that reported on measure intra-reliability, 41 that reported on inter-rater reliability, and 20 that reported on internal consistency (Martin et al., 2021b). Of the two studies reporting on intra-rater reliability, one study met COSMIN criteria and the other did not. Poor intra-rater reliability may indicate that assessor understanding of items changed or that assessors weighed different considerations due to poor item clarity and/or training deficiencies (Multon & Colemon, 2018). However, there is insufficient literature available to draw meaningful conclusions about this psychometric property. Of the 41 studies reporting on inter-rater reliability, 21 met COSMIN criteria. While this suggests that assessors are somewhat inconsistent in their understanding and application of measure items, a considerable number of results were close to the quality cut-offs or produced findings just below the quality threshold. Further, several studies were not rated as high-quality as they only reported percentage agreements. Yet, all studies reported agreements above 70.0% - a level many researchers consider sufficient (Aspland & Gardner, 2003). The inter-rater reliability results suggest that more work is necessary to improve inter-reliability, especially as there is heavy

reliance on percentage agreements. Of the 41 studies, 20 reported on internal consistency with 12 rated as high-quality. Yet, an additional five were close to the quality threshold suggesting that, in total, internal consistency was strong.

### **3.2.3 Measure Validity**

Validity is the degree to which a measure accurately reflects its intended constructs (Mokkink et al., 2010a; Thorkildsen, 2010). Put another way, validity captures the extent to which a measure actually assesses competent adherence as defined by those who developed the measure. Five types of measure validity are often examined – content, construct, convergent/divergent, predictive, and criterion. The guidance and criteria provided by COSMIN are typically used to evaluate validity (Mokkink et al., 2016). As Paper 2 of the dissertation analysed the content and construct validity of the PLH-FAT-T, the COSMIN definitions and methods are discussed in detail hereafter.

#### *Defining and Evaluating Validity*

Assessing content validity is often the informative first step in establishing the psychometric properties of a measure. Content validity refers to the degree to which a measure appears to adequately capture its underlying construct(s) according to the perspectives of the tool's key stakeholders, such as those who develop and use the measure (Markus & Smith, 2010). It is sometimes positioned as the most important psychometric property (Terwee et al., 2018) as a measure should be meaningful in relation to its underlying constructs to ensure it can be utilised in practice (Haynes et al., 1995). Establishing content validity is difficult because there is limited information available regarding how to conduct such analyses. Key steps to establishing content validity include

defining the domain of interest, articulating the purpose of the measure, crafting items, and investigating the extent to which the measure captures the domain(s) of interest (McKenzie et al., 1999; Thorkildsen, 2010). The latter is typically established by examining how subject matter experts and other stakeholders perceive the comprehensiveness, comprehensibility, and relevance of the measure and its items (Bornstein, 2004; Mokkink et al., 2010a; Terwee et al., 2018; Thorkildsen, 2010) and then using their input to modify and thereby move the measure closer to being able to accurately evaluate competent adherence. This process is qualitative and often iterative; even the manner through which stakeholders are asked to provide their input is consequential (Sireci, 1998). For instance, in some cases, subject matter experts may be asked to rate each item based on pre-specified criteria. However, as such constraints can limit the feedback they provide or encourage socially desirable responses (Sireci, 1998), care must be taken in the process of asking for input.

Construct validity, or structural validity, is the extent to which a measure statistically reflects its underlying constructs as hypothesised by its developers (Peng & Mueller, 2004; Terwee et al., 2007). This type of validity statistically determines whether the measure actually captures competent adherence. An analysis of a measure's construct validity attempts to determine whether certain items 'group' together to reflect one or more underlying constructs. Factor analyses are commonly used to evaluate construct validity as these analyses test how many underlying constructs underpin a measure and how its items correlate to these constructs (Kline, 2014). There are issues that need to be considered with the two main approaches to conducting and analysing construct validity – exploratory and confirmatory factor analyses (O'Leary-Kelly & Vokurka, 1998). Exploratory factor

analyses are more subjective whereas confirmatory factor analyses are more objective (O'Leary-Kelly & Vokurka, 1998). Exploratory factor analysis does not require an a priori hypothesis as to how items should relate. By not pre-specifying a theory, studies utilising exploratory factor analysis are at risk of “data snooping” or “data dredging” (O'Leary-Kelly & Vokurka, 1998; Salkind, 2010) – a practice wherein researchers look at their data and then make a decision regarding which analyses to perform. In contrast, confirmatory factor analysis is hypothesis testing in that researchers must outline how they expect items to map onto constructs in advance. With both types of analyses, sample size and the number of items being tested per construct need to be considered because too few facilitators or too few items will not result in a robust analysis (Hair et al., 1998). When an exploratory or confirmatory factor analysis is conducted, information on the number of factors that emerged from the analyses, factor loadings, percentage of variance explained by the factors (which should be at least 50%), eigenvalues, and model fit (e.g., model chi square, confirmatory factor indices above 0.95, root mean square of approximation less than 0.06, standardised root mean square of approximation less than 0.08) are all indices that shed light on the measure's construct validity (Terwee et al., 2007).

#### *Validity of Measures of Facilitator Competent Adherence*

There is limited evidence on the validity of measures of facilitator competent adherence in the parenting programme literature (Martin et al., 2021b). The systematic review conducted during my master's identified three studies reporting on content validity and nine on construct validity. While the quality of the evidence was mediocre based on COSMIN standards, the validity of studies deemed low-quality was arguably better than the

COSMIN checklist ratings indicated, with many measures having borderline or mixed evidence. Content validity requires substantial future attention since the systematic review only identified three studies that provided information about this property, two of which did not provide sufficient information. Overall, the lack of reporting means that there is a lack of understanding about the foundation on which measures have been built. Of the nine studies identified as reporting construct validity, four were high-quality (e.g., factors explaining at least 50.0% of the variance) and five were low-quality (Mokkink et al., 2010b). Among the latter, factor loadings were close to the threshold in many of the studies with some just above and below the cut-off and other indicators (e.g., model fit) close to those considered high-quality. Despite over half of the studies not meeting COSMIN criteria, the measures appear to be mostly capturing their intended constructs.

#### **3.2.4 Parenting for Lifelong Health for Young Children-Facilitator Assessment Tool**

The PLH Facilitator Assessment Tool (FAT) is an observational measure of facilitator competent adherence with various versions of the tool being used in approximately 25 countries to accomplish numerous goals including to assess the programme delivery of PLH facilitators. The FAT is used to assess facilitator competent adherence based on either live or video recorded delivery during experimental trials and routine service delivery. Despite extensive use of various versions of the tool, only one study has been conducted to establish the content validity, intra-rater reliability, and inter-rater reliability of the version of the FAT used during the implementation of PLH-YC (the PLH-FAT-YC). This psychometric evaluation took place in late 2019 and early 2020 in

Southeastern Europe (Martin et al., 2022a). Several insights stemming from this study informed Paper 2.

### *Version of the PLH-FAT-YC in Southeastern Europe*

The original PLH-FAT-YC was designed to assess the extent to which programme facilitators adhered to programme activities and skills using two subscales. The first subscale measured facilitator adherence by examining how effectively a facilitator led core activities (Activity Subscale; e.g., “Explore at least one specific challenge experienced by a parent regarding the main home activity”). The second subscale measured facilitator competence by examining how well a facilitator utilised process skills (Skills Subscale; e.g., “Model behaviours with co-facilitator” and “Facilitator is situated within the group, is at the level of the parents, and in a different place than the co-facilitators”). Each item on these two subscales was rated using a four-point Likert scale ranging from zero to three, where zero was ‘inadequate’, one was ‘needs improvement’, two was ‘good’, and three was ‘excellent’. The points allotted to each item from the two subscales were summed to reach an overall impression score represented as a percentage.

### *Content Validity*

A revised version of the PLH-FAT-YC used to assess PLH-YC facilitators in Southeastern Europe was developed by way of a content validity exercise with three stakeholder groups (see [Appendix 4](#)) – PLH trainers from CWBSA; three parenting and PLH programme experts; and 11 assessors in Southeastern Europe (five in Moldova with backgrounds in teaching and family therapy; three in North Macedonia who were all later-career psychologists; and three in Romania who were all early-career psychologists). This



stakeholder consultation process resulted in substantial modifications to the original PLH-FAT-YC based on feedback from each of the three groups. Stakeholders recommended splitting complicated items into two simpler items; eliminating redundant items; changing item wording; adding additional items to capture components and skills missing from the tool; creating new items to assess the frequency of certain facilitator behaviours (reflexive statements, specific praise, and unspecific praise); and creating definitions for each item and point on the Likert scale (see [Appendix 5](#) for examples). The changes resulted in the revised PLH-FAT-YC having nine more items and an additional subscale (Frequency) with three items. According to the stakeholders, the revised PLH-FAT-YC was a more understandable, specific, and practical tool. The intra- and inter-rater reliability of the revised PLH-FAT-YC was then examined.

#### *Intra-Rater Reliability*

Each assessor assessed a video recording of a facilitator delivering the programme. Then, a few weeks later, each assessor watched the same video recording again and conducted a second assessment. The overall percentage agreements across assessments by assessors in the three countries ranged from 57.6-91.5% with ICCs of 0.52-0.94. Intra-rater reliability was strong in North Macedonia, moderately strong in Romania, and moderate in Moldova. The results indicated that the assessors in all three countries were largely able to consistently complete PLH-FAT-YC assessments. The ICCs were, with few exceptions, larger than the percentage agreements with most exceeding 0.50. This result indicated that when assessors did not select the same score on the items during their two assessments of the same video(s), they were more likely than not to select scores closer together rather than

further apart. Thus, even though assessors did not always exactly replicate how they rated an item, they were more consistent than not in their application of the revised PLH-FAT-YC (Stemler & Tsai, 2008). The results also indicated that it was challenging for assessors to reach consensus on frequency items.

### *Inter-Rater Reliability*

To examine inter-rater reliability, each assessor also completed PLH-FAT-YC assessments of three facilitators using video recordings of programme delivery. In order to capture a facilitator leading all activities on the Activities Subscale, as per the recommendation of one of the parenting programme experts, videos of two specific programme sessions were observed. The overall percentage agreements between assessors ranged from 18.1-74.0% with ICCs of 0.49-0.91. Inter-rater reliability was moderate in Romania, moderate in Moldova, and strong in North Macedonia based on ICC 95% confidence intervals. There may have been varying levels of reliability in the three countries due to assessors having different amounts of experience with conducting facilitator assessments. However, the moderate to strong results suggested that the assessors in all three countries were largely consistent with each other in their assessments. Although the percentage agreements were on the lower end, the ICCs indicated the inter-rater reliability as stronger demonstrating that assessors were more similar to each other than the percentage agreement results suggested. This finding revealed that assessors were more consistent than not in their application of the measure and its items, yet it was still difficult for assessors to achieve consensus on many occasions.

## *Insights*

This psychometric evaluation generated two key insights for future measurement of facilitator competent adherence in PLH that were applicable to the study of the PLH-FAT-T. The first insight was that the tool should have a three-point scale instead of a four-point scale because the latter resulted in discrepancies in assessor use of the middle two points ('1' and '2') on the scale. The second insight was that facilitators should only be assessed on one programme activity instead of two. PLH facilitators deliver the programme in pairs and during each session they alternate delivery of programme activities. This delivery method meant that to assess two activities (and therefore to complete one assessment of a facilitator), an assessor had to attend two sessions. This assessment method was found to be impractical, time consuming, and costly. These two insights informed the content validation process conducted as part of the psychometric evaluation of the PLH-FAT-T used to assess PLH-Teens facilitators in Tanzania (see Paper 2).

### **3.3 Competent Adherence and Relationship to Outcomes**

This section describes the existing evidence on the relationship between facilitator competent adherence and outcomes in the broader implementation science literature and the parenting programme literature. This section also discusses potential reasons for inconsistent evidence on the relationship as found in the parenting programme literature.

#### **3.3.1 Evidence from the Broader Implementation Science Literature**

In recent years, there has been increasing interest in the fidelity with which evidence-informed interventions are implemented as it is theorised that better implementation fidelity is associated with better participant outcomes. Systematic review

and meta-analytic evidence from a variety of fields suggests that implementation fidelity is a mechanism through which to enhance participant outcomes (Carroll et al., 2007). In the field of health promotion, a well-cited systematic review of over 500 studies reported evidence of a relationship between implementation fidelity and participant health and well-being at the study level (Durlak & DuPre, 2008). In the field of educational interventions, several systematic reviews found a positive relationship between fidelity and outcomes. Of these, a review of randomised trial results found higher levels of teacher fidelity were associated with improved student achievement outcomes (Hill & Erickson, 2019). Further, a review of the implementation of 29 school-based physical activity programs found programme delivery by more highly competent teachers consistently predicted better student outcomes (Naylor et al., 2015) and a meta-analysis of 221 school-based child behaviour programs found implementation fidelity was a key contributor to positive changes in student behaviour (Wilson et al., 2003).

### **3.3.2 Evidence from the Parenting Programme Literature**

There is substantial evidence which demonstrates that parenting programmes are an effective way to support parents acquire the knowledge and/or skills to enhance their children's health and well-being and thereby improve child outcomes (Barlow & Coren, 2018; Barlow et al., 2006b; Chen & Chan, 2016; Furlong et al., 2013; Knerr et al., 2013). However, the role of implementation fidelity and its relationship to intended outcomes in these programmes is unclear. Few studies report on parenting intervention fidelity (e.g., Gardner et al., 2016), and among these there is limited evidence on the relationship between implementation fidelity and outcomes (Rojas-Andrade & Bahamondes, 2019). Among the growing number of primary studies on the relationship between fidelity and

outcomes, the evidence is inconsistent with some studies finding associations and others finding no association (Breitenstein et al., 2010a; Cantu et al., 2010; Durlak, 1998; Fixsen et al., 2005; Forgatch et al., 2005; Olofsson et al., 2016). This literature will be discussed in greater detail in Chapters 5 and 8.

### **3.3.3 Potential Reasons for Mixed Evidence**

The results in the literature may be inconsistent due the challenges of studying the relationship between fidelity and outcomes, including publication bias wherein studies that find insignificant or negative associations may be less likely to be published; selective reporting wherein authors are less likely to report insignificant or negative associations; the potential influence of confounding variables (Breitenstein et al., 2010b); interaction effects with other aspects of implementation (Berkel et al., 2011); inaccurate measurement due to the use of tools which have not had their reliability and validity established (Breitenstein et al., 2010b); insufficient power to examine the relationships due to small sample sizes; and little variation in fidelity or intervention outcomes, making analyses lack sensitivity to discern associations (Durlak & DuPre, 2008). These issues will be discussed further in Chapters 7 and 8.

## **4. Dissertation Overview**

This chapter outlines the dissertation's rationale (4.1), sets out the research questions (4.2), provides an overview of the three papers (4.3), sets out the hypotheses (4.4), and details the acquired ethics approvals (4.5). The three subsequent chapters provide the background, methods, results, and discussion regarding each paper.

### **4.1 Rationale for the Dissertation**

In view of the literature outlined in the previous chapter, this dissertation aims to advance knowledge within the implementation science field in general, and within the parenting programme field in particular, concerning the measurement of facilitator competent adherence, the psychometric properties of facilitator competent adherence measures, and the relationship between facilitator competent adherence and parent and child outcomes.

Although numerous parenting programmes have assessed facilitator competent adherence, in 2021, no studies had systematically reviewed the measures used by parenting programmes aiming to reduce violence against children and child behaviour problems. As a result, for my master's, I conducted a systematic review of the available measures which is now published in *Clinical Child and Family Psychology Review* (Martin et al., 2021b). A subset of the studies found via this systematic review were analysed in Paper 1 to investigate the relationship between competent adherence and participant outcomes. This paper aims to clarify the inconsistent evidence on this relationship in the parenting programme literature. Given the widespread dissemination of parenting programmes, knowledge about the role that facilitator competent adherence plays in outcomes would enhance understanding of whether it is a key mechanism by which parenting programmes

achieve their positive results for children and families. Such understanding could then be used to inform future programme delivery, particularly as programmes are translated into community settings and taken to scale (Fixsen et al., 2005; Mowbray et al., 2003). As facilitator competent adherence is potentially an important factor in enhancing parenting programme outcomes, Paper 1 synthesises the existing literature on the relationship.

Paper 2 contributes to the scant literature on observational measures of facilitator competent adherence used during routine service delivery, particularly in LMIC settings and at scale. As few studies report on such measures, little is known about facilitator competent adherence in these contexts. The dissertation provides the first analysis of the level of competent adherence achieved by community facilitators during the routine delivery of a parenting programme at scale in a LMIC. There is also limited psychometric evidence on observational measures of competent adherence used in community settings and in LMICs. Findings as to whether measures of facilitator competent adherence are reliable and valid will help establish if facilitator assessment results can be relied on and used to enhance facilitator delivery of programmes (Fixsen et al., 2005). That fewer parenting programmes are available for families in LMICs heightens the need in these contexts for solid evidence regarding how to improve programme delivery and outcomes (Knerr et al., 2013; Mercy et al., 2008; Mikton & Butchart, 2009). Although there is a study on the psychometric properties of the PLH-FAT-YC (Martin et al., 2022a), there are no studies that have assessed the psychometric properties of the version of the FAT used in PLH-Teens. This dissertation advances the literature by examining the reliability and validity of the PLH-FAT-T as part of its use in the scale-up of PLH-Teens in Tanzania.

Paper 3 contributes to the sparse and inconsistent parenting program literature on the relationship between facilitator competent adherence and parent and child outcomes, particularly in LMICs and at scale. Further knowledge about the role that competent adherence plays in parent and child outcomes would lead to an enhanced understanding of whether implementation is a key mechanism through which parenting programmes achieve positive results for children and families. Such understanding could inform future programme delivery of parenting programmes globally, particularly as they are taken to scale in additional contexts. Paper 3 also aims to contribute to the global dissemination of PLH programmes by providing useful information for the continued scale-up of PLH programmes in LMICs.

#### **4.2 Research Questions**

The dissertation aims to answer the following four research questions related to the measurement and role of facilitator competent adherence:

1. What is the evidence on the relationship between observational measures of facilitator competent adherence and parenting programme outcomes? (Paper 1)
2. What are the psychometric properties of the observational measure used to assess the competent adherence of PLH-Teens facilitators during routine service delivery at scale in Tanzania? (Paper 2)
3. What is the level of competent adherence with which facilitators deliver PLH-Teens during routine service delivery at scale in Tanzania? (Paper 2)
4. What is the predictive validity of the observational measure used to assess the facilitator competent adherence of PLH-Teens facilitators during routine service delivery at scale in Tanzania? (Paper 3)



### 4.3 Overview of Papers

To address the above research questions, three papers were conducted using studies from the broader parenting programme literature and data from the FAIR study in Tanzania.

#### *Paper 1 – Synthesis Without Meta-Analysis*

Building on the results of a systematic review of the parenting programme literature (Martin et al., 2021b), this paper conducted a synthesis without meta-analysis to compile the evidence on the relationship between facilitator competent adherence and outcomes. Synthesis without meta-analysis was used as a meta-analysis was deemed methodologically unfeasible. This paper answers research question 1.

#### *Paper 2– Psychometric Analysis and Level of Competent Adherence*

Using data from the large-scale delivery of PLH-Teens in Tanzania, this paper evaluated the PLH-FAT-T's content validity, intra-rater reliability, inter-rater reliability, internal consistency, and construct validity and determined the level of competent adherence achieved by facilitators during programme delivery. To evaluate intra-rater reliability, inter-rater reliability, internal consistency, and construct validity, the paper analysed the assessments of 95 facilitators. This paper answers research questions 2 and 3.

#### *Paper 3 – Multi-Level Regression Analyses*

Using data from PLH-Teens in Tanzania, this paper linked assessments of 24 facilitators to the pre-post surveys from 3,057 families who participated in the scale-up of PLH-Teens in Tanzania. Multi-level Poisson regressions with an interaction term as well as fixed and random effects were used to investigate whether higher competent adherence

scores are associated with greater improvements in parent- and adolescent-reported primary (child maltreatment) and secondary (child conduct problems, child emotional problems, positive parental involvement, parental support of education, poor supervision, acceptability of corporal punishment, sexual health communication, depression, parenting stress, financial insecurity, intimate partner violence (IPV) perpetration, IPV victimisation, and school violence) outcomes. This paper answers research question 4.

#### **4.4 Hypotheses**

The following hypotheses were developed in response to the research questions. For Paper 1, it was hypothesised that higher levels of competent adherence would be associated with better beneficiary outcomes. For Paper 2, it was hypothesised that changes would be made to the PLH-FAT-T following the content validity process (e.g., addition, deletion, and revision of items) and that the revised tool would then demonstrate sufficient reliability (intra-rater, inter-rater, and internal consistency). It was also hypothesised that the construct validity analysis would indicate that at least two factors (adherence and competence) underpin the tool and that some items should be revised or deleted from the tool. Lastly, it was hypothesised that on average facilitators would, at minimum, reach the PLH-designated certification standard of 60%. For Paper 3, it was hypothesised that the PLH-FAT-T would demonstrate predictive validity wherein higher PLH-FAT-T scores relate to greater improvements in parent- and adolescent-reported outcomes, particularly for the primary outcome of interest (child maltreatment). In other words, it was hypothesised that there would be a positive linear relationship between competent adherences and parent and child outcomes.

## 4.5 Ethics

The analyses conducted for the dissertation were approved by several ethics committees - in Tanzania (NIMR/HQ/R.8a/Vol.IX/3459 and NIMR/HQ/R.8a/Vol.IX/2902) and at Oxford (RE002 HEY BABY; R64777/RE001; SPR\_DREC\_20-21\_026) (see [Appendix 6](#)). The dissertation involved an analysis of papers found via a systematic review (Paper 1) as well as analyses of anonymised quantitative secondary data from Pact and CWBSA (Papers 2 and 3).

## 5. Paper 1 - The association between facilitator competent adherence and outcomes in parenting programs: A systematic review and SWiM analysis

The following chapter is published in *Prevention Science*. This paper has been written with US spelling in alignment with the journal's submission guidelines. Please find the tables throughout the chapter and the supplementary materials in [Appendix 7](#).

**Authors:** Martin, M.<sup>1</sup>, Steele, B.<sup>1</sup>, Spreckelsen, T. F.<sup>2</sup>, Lachman, J. M.<sup>1</sup>, Gardner, F.<sup>1</sup>, & Shenderovich, Y.<sup>3,4</sup>

### **Affiliations:**

<sup>1</sup>Department of Social Policy and Intervention, University of Oxford, United Kingdom

<sup>2</sup>School of Social and Political Sciences, University of Glasgow, United Kingdom

<sup>3</sup>Wolfson Centre for Young People's Mental Health, Cardiff University, United Kingdom

<sup>4</sup>Centre for the Development, Evaluation, Complexity and Implementation in Public Health Improvement (DECIPHer), School of Social Sciences, Cardiff University, United Kingdom

**Citation:** Martin, M., Steele, B., Spreckelsen, T.F., Lachman, J. M., Gardner, F., & Shenderovich, Y. (2023). The association between facilitator competent adherence and outcomes in parenting programs: A systematic review and SWiM analysis. *Prevention Science*. <https://doi.org/10.1007/s11121-023-01515-3>

**My role:** I conceptualised the paper, screened studies, extracted data, conducted the analyses, drafted the manuscript, revised the manuscript, and led the response to reviewers.

BS did double coding for the screening and extraction of studies. BS, TS, YS, JML, and FG contributed sections and revisions to the manuscript. All authors read and approved the manuscript before submission to *Prevention Science*.

### 5.1 Abstract

**Objective:** There is increasing interest about the fidelity with which interventions are implemented because it is theorized that better implementation fidelity by facilitators is associated with better participant outcomes. However, in the parenting program literature, there is mixed evidence on the relationship between implementation fidelity and outcomes. This paper provides a synthesis of the evidence on the relationship between facilitator delivery and outcomes in the parenting program literature. **Method:** Following PRISMA

guidelines, this paper synthesizes the results of a systematic review of studies on parenting programs aiming to reduce violence against children and child behavior problems. Specifically, it examines associations between observational measures of facilitator competent adherence and parent and child outcomes. A meta-analysis was not feasible due to study heterogeneity. As a result, Synthesis Without Meta-Analysis guidelines were followed. **Results:** Searches in electronic databases, reference searching, forward citation tracking, and expert input identified 9,653 articles. After screening using pre-specified criteria, 18 articles were included. The review found that most studies (n=13) reported a statistically significant positive relationship with at least one parent or child outcome. However, eight studies reported inconsistent findings across outcomes and four studies found no association with outcomes. **Conclusions:** The results suggest that better facilitator competent adherence is generally associated with positive parent and child outcomes. However, this finding is weakened by the methodological heterogeneity of included studies and due to the wide variety of ways in which studies conceptualized competent adherence-outcome relationships.

*Key Words:* parenting, fidelity, behavioral interventions, systematic review, violence prevention

## 5.2 Background

In recent years, there has been increasing interest in the fidelity with which evidence-informed interventions are implemented as it is theorized that better implementation fidelity - the extent to which an intervention is implemented as intended (Dane & Schneider, 1998) - is associated with better participant outcomes. Systematic review and meta-analytic evidence from a variety of fields now empirically supports that improving implementation fidelity is an important mechanism through which to enhance participant outcomes (Carroll et al., 2007). In the field of health promotion, a well-cited systematic review of over 500 studies reported evidence of a relationship between higher implementation fidelity and improved participant health and well-being at the study level (Durlak & DuPre, 2008). In the field of educational interventions, several systematic reviews found a positive relationship between fidelity and outcomes. A review of randomized trials found that higher levels of teacher fidelity were associated with improved student achievement outcomes (Hill & Erickson, 2019). Further, a review of the implementation of 29 school-based physical activity programs found program delivery by more highly competent teachers consistently predicted better student outcomes (Naylor et al., 2015) and a meta-analysis of 221 school-based child behavior programs found that implementation fidelity was a key contributor to positive changes in student behaviour (Wilson et al., 2003).

Substantial evidence demonstrates that parenting programs are an effective way to support parents to acquire the knowledge and skills to enhance their children's health and well-being and thereby improve child outcomes (e.g., Barlow & Coren, 2018; Barlow et al., 2006a; Chen & Chan, 2016; Furlong et al., 2013; Knerr et al., 2013). However, the role of

implementation fidelity and its relationship to intended outcomes in these programs is unclear. Few studies report on parenting intervention fidelity (Gardner et al., 2016), and among these there is limited evidence on the relationship between implementation fidelity and outcomes (Rojas-Andrade & Bahamondes, 2019). One exception is a recently published systematic review of 24 studies on programs aiming to reduce child externalizing behaviors (Leitão et al., 2020). It reported on the role of several facilitator factors, including facilitator adherence, on program outcomes (Leitão et al., 2020). While this review concluded that facilitator delivery mattered for outcomes, it only included interventions specifically addressing child behavior problems and did not summarize the results of each study in detail. Another example is a meta-analysis of 156 studies on nine home visiting programs aiming to reduce child maltreatment which found that several implementation fidelity components, including facilitator adherence, were positively related to reductions in child maltreatment (Casillas et al., 2016).

Among the growing number of primary studies on the relationship between fidelity and outcomes, the evidence is inconsistent with some studies finding an association and others finding no association (Breitenstein et al., 2010a; Cantu et al., 2010; Durlak, 1998; Fixsen et al., 2005; Forgatch et al., 2005; Olofsson et al., 2016). Some studies have found the relationship between fidelity and outcomes to be curvilinear wherein the highest levels of fidelity were detrimental to participant outcomes (e.g., Hogue et al., 2008). The literature may be inconsistent due to a variety of challenges connected with studying the relationship between fidelity and outcomes, including publication bias; the potential influence of confounding variables (Breitenstein et al., 2010b); interaction effects with other aspects of implementation (Berkel et al., 2011); inaccurate measurement due to the

use of tools which have not had their reliability and validity established (Breitenstein et al., 2010b); insufficient power to examine the relationships due to small sample sizes; and little variation in fidelity or intervention outcomes, making analyses lack sensitivity to discern associations.

Given the widespread dissemination of parenting programs, the relationship between fidelity and outcomes should be clarified. Knowledge about the role that fidelity plays in outcomes would lead to an enhanced understanding of whether implementation is one of the key mechanisms via which parenting programs achieve their positive results for children and families. Such an understanding could then be used to inform future program delivery, particularly as programs are translated into community settings and taken to scale (Fixsen et al., 2005; Mowbray et al., 2003). As implementation fidelity is potentially an important factor in enhancing parenting program outcomes, there is a need to take stock of the existing literature on the relationship between fidelity and outcomes.

This paper synthesizes the research on the relationship between implementation fidelity and outcomes found via a systematic review of the existing literature on parenting programs aiming to (a) reduce child maltreatment; harsh or dysfunctional parenting; and/or child conduct problems and/or (b) improve positive child behavior management strategies; parent-child bonding/attachment and relationships; and/or early childhood development outcomes. The review focused on two aspects of implementation fidelity outlined in Proctor's taxonomy (2011) – adherence (strictness with which a facilitator implements the prescribed content) and competence (skill and style with which a facilitator delivers program components). Although distinct, these aspects were selected for this review as it is commonly thought that they should be assessed simultaneously (Breitenstein et al., 2010b;



Forgatch et al., 2005; Martin et al., 2021b). Together, competence and adherence (or ‘competent adherence’) refers to the quality and strictness with which facilitators deliver an intervention as intended (Carroll et al., 2007; Forgatch et al., 2005). This review included studies that reported on competence and/or adherence.

This study is the first synthesis of the evidence on the relationship between observational measures of facilitator competent adherence and parenting program outcomes and summarizes the methods used to study the relationship. It specifically focuses on observational assessments of facilitator competent adherence – completed on facilitator program delivery based on their live or video-taped delivery – as these are considered most rigorous and provide a more detailed account of program delivery (Dusenbury et al., 2003; Eames et al., 2008). Further, the study provides critical insight into whether better fidelity is associated with improved family outcomes.

### **5.3 Methods**

This study builds on a systematic review conducted by Martin et al. (2021b) that compiled the observational and non-observational measures of facilitator competent adherence found in the parenting program literature and synthesized the psychometric properties of the measures found. Using articles reporting on observational measures of facilitator competent adherence from the review, this paper synthesizes the evidence on the association, if any, between facilitator competent adherence and parent and/or child outcomes.

#### **5.3.1 Systematic Review**

The full details of the methods used for the systematic review are described by Martin et al. (2021b). In sum, a tested search strategy was implemented in 12 electronic

bibliographic databases (see [Supplementary File 1](#)). To find additional studies, the database searches were supplemented by: reviewing articles included in a systematic review of parenting programs in low- and middle-income countries to ensure representation from these contexts (Gardner et al., forthcoming); conducting backward citation tracking; conducting forward citation tracking using Google Scholar; and seeking input from parenting program experts. The inclusion and exclusion criteria applied are summarized in Table 1. The review tested inter-rater reliability between two coders at the title/abstract, full-text, and data extraction stages. Percentage agreements ranged from 92.8-94.4% and were thus sufficiently high.

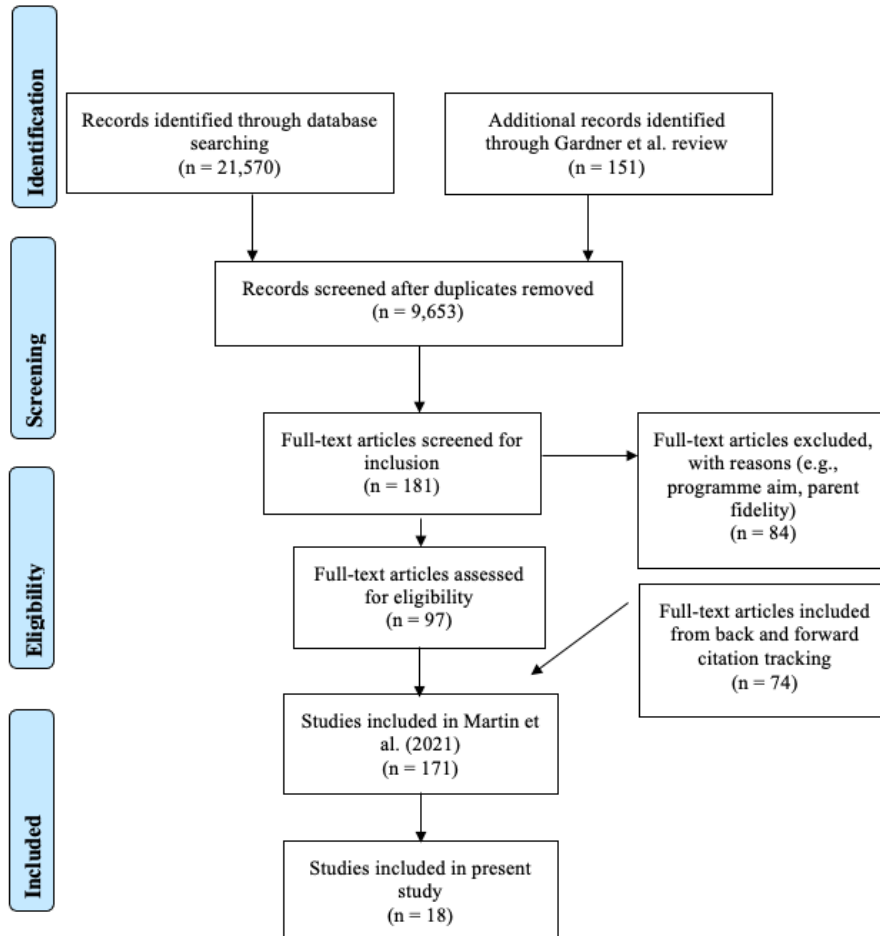
The systematic review by Martin et al. found 9,653 articles as of August 2021 (see Figure 1). To be included in the review, articles must have been written in English; reported on observational measures of facilitator competent adherence; and used an observational, quasi-experimental, or experimental approach to analyze the association between facilitator competent adherence and family outcomes. Of the original articles, 18 articles reported on the relationship between observational measures of facilitator competent adherence and parent and child outcomes and were thereby included in the review.

**Table 1***Inclusion and exclusion criteria*

<b>Inclusion Criteria</b>	<b>Exclusion Criteria</b>
Report on parenting programs aiming to (a) reduce child maltreatment; harsh or dysfunctional parenting; and/or child conduct problems and/or (b) improve positive child behavior management strategies; parent-child bonding/attachment and relationships; and/or early childhood development outcomes.	Parenting programs with other aims or programs which (1) narrowly focus on specific child risks such as poisoning or accidents or on skills training for children's specific medical conditions or physical disabilities (e.g., developmental disability) or (2) primarily deliver financial support (e.g., conditional case transfer programs) or other support to parents, but which do not aim to change parents' knowledge or behavior concerning their child(ren)
Report on parenting programs wherein at least 50% of the program content is delivery to parents/caregivers	Parenting programs wherein children or others (not parents/caregivers) are the main focus of the intervention
Report on observational measures of facilitator competence and/or adherence	Reports solely on treatment alliance or working relationship Reports on facilitator competence and/or adherence without some reference to how it was measured or reports on a non-observational measure of facilitator competent adherence
Report on parenting programs wherein parents are 18 years or older and children are 17 years or younger	Reports on parenting programs for teenage parents (17 years and younger) and their children
Data surfacing from academic or grey publications including peer-reviewed articles, unpublished manuscripts, ongoing studies, and theses/dissertations	Data surfacing from books, newspapers, and magazines

**Figure 1**

*PRISMA Flowchart of Study Screening and Selection*



### 5.3.2 Synthesis Without Meta-Analysis

Upon reviewing the 18 included studies, a meta-analysis was deemed methodologically unfeasible. As a result, a synthesis without meta-analysis was performed based on the Synthesis Without Meta-Analysis (SWiM) guidelines outlined by Campbell and colleagues (2020) (see [Supplementary File 2](#)). These guidelines specify nine key categories of information that should be provided when a quantitative synthesis is not

possible. The revised Meta-Analysis Reporting Standards was followed where possible (Appelbaum et al., 2018).

### **5.3.3 Reporting**

Findings are reported following the PRISMA guidelines (Liberati et al., 2009) (see Supplementary File 3). For studies which differentiated competent adherence-outcome relationships by subscale (e.g., adherence subscale score, competent subscale score, overall competent adherence score), overall competent adherence scores were extracted.

## **5.4 Results**

The 18 included studies are disparate in terms of the programs studied, parent and child outcomes considered, and methods used. This heterogeneity meant that a meta-analysis could not be conducted. Further, the small number of studies limited the feasibility of meta-analyzing subgroups. The results of the 18 studies are summarized in Table 2 and Supplementary File 4. Using a modified version of the Thomson and Thomas (2013) effect direction plot visual display system, results are visualized using: a sideways arrow ( $\Leftrightarrow$ ) indicating no statistically significant association between competent adherence and parent/child outcomes; an upwards arrow ( $\Uparrow$ ) indicating a positive, statistically significant association between stronger competent adherence and better parent/child outcomes; or a downwards arrow ( $\Downarrow$ ) indicating a negative, statistically significant association between stronger competent adherence and poorer parent/child outcomes.

**Table 2**  
*Summary of Study Results*

<b>Paper</b>	<b>Program and Country</b>	<b>Prevention Classification</b>	<b>Fidelity Measure</b>	<b>Mean Fidelity (SD)</b>	<b>Measure Timepoint</b>	<b>Findings</b>	<b>Summary</b>
(Cantu et al., 2010) or “S1”	Strengthening Families Program, US	Universal	Adherence only; assessed by researchers	81.00% (7.00%)	Immediately post-intervention	Adherence was not related to program outcomes	<b>Parenting:</b> ⇔
(Chiapa et al., 2015) or “S2” *data Smith, 2013	Family Check-Up, US	Treatment	Competent adherence; assessed by a third party	61.10% (14.20%)	5.5-6.5 years post-registration	Decline in competent adherence associated with less behavior change (drift)	<b>Child behavior:</b> ↑
(Eames et al., 2010) or “S3”	Incredible Years BASIC program, UK	Treatment	Competent adherence; assessed by a third party	Not reported	Immediately post-intervention	Facilitator competent adherence associated with better parenting skills	<b>Parenting:</b> ↑
(Forgatch et al., 2005) or “S4”	Parent Management Training Oregon Model, US	Selected	Competent adherence; no assessor information	Not reported	12 months post-intervention	Higher competent adherence predicted improved parenting	<b>Parenting:</b> ↑
(Forgatch & DeGarmo, 2011) or “S5”	Parent Management Training Oregon Model, Norway	Indicated	Competent adherence; assessed by trained non-participant observer	67.78% (15.67%)	9 months post-intervention	Higher competent adherence predicted improved parenting	<b>Parenting (path analysis):</b> ↑ <b>Parenting (correlations):</b> ⇔
(Giannotta et al., 2019) or “S6”	Incredible Years, Sweden	Treatment	Competent adherence; assessed by independent raters	76.50% (14.40%)	Immediately post-intervention	Competent adherence was not associated with parent and child outcomes	<b>Parenting:</b> ⇔ <b>Child behavior:</b> ⇔
(Hukkelberg & Ogden, 2013) or “S7”	Parent Management Training Oregon	Indicated	Competent adherence; assessed by researchers	73.12% (11.80%)	Timepoint 4 (not fully described)	Competent adherence predicted reductions in behavior issues (parent-report only); found alliance and competent adherence to be independent	<b>Child behavior (parent-report):</b> ↑

	Model, Norway					from each other; data only presented for timepoint three	<b>Child behavior (teacher-report):</b> ↔
(Hogue et al., 2008) or “S8”	Multi-dimensional Family Therapy, US	Treatment	Adherence and competence measured separately; assessed by program supervisors	<u>Parent Interventions:</u> Adherence: 67.71% Competence: 80.29%  <u>Family Intervention:</u> Adherence: 46.87% Competence: 74.57%	At 6 month follow-up	Better adherence predicted greater reductions in parent-reported externalizing but not youth-reported externalizing behavior Competence did not predict either internalizing or externalizing behavior as reported by parents and youth Some evidence of a curvilinear relationship for adherence-internalizing relationship, but not reported for externalizing despite this being a main outcome Curvilinear relationships between competence and outcomes are not explored Only report aggregate results not broken down by timepoint	<b>Adherence:</b> <b>Parent-reported externalizing (linear):</b> ↑ <b>Youth-reported externalizing (linear):</b> ↔ <b>Parent-reported internalizing (linear):</b> ↔ <b>Parent-internalizing (curvilinear):</b> ↑↑ <b>Competence:</b> <b>Parent-reported internalizing (linear):</b> ↔ <b>Parent- and youth-reported externalizing (linear):</b> ↔ <b>Externalizing curvilinear not reported</b>
(Maaskant et al., 2016) or “S9”	Parent Management Training Oregon Model, Netherlands	Universal	Competent adherence; assessed by program supervisors	79.80% (5.30%)	Both post-intervention and 4 month follow-up	Higher competent adherence associated with better improvements in some parenting dimensions but not others	<b>Post-Test:</b> <b>Stress:</b> ↑↑ <b>Warmth:</b> ↔ <b>Responsiveness:</b> ↔ <b>Explaining:</b> ↑↑ <b>Autonomy:</b> ↑↑ <b>Strictness:</b> ↔ <b>Discipline:</b> ↔ <b>Child behavior:</b> ↔ <b>Follow-Up:</b> <b>Stress:</b> ↑↑

							<b>Warmth:</b> ⇌ <b>Responsiveness:</b> ↑ <b>Explaining:</b> ↑ <b>Autonomy:</b> ↑ <b>Strictness:</b> ⇌ <b>Discipline:</b> ⇌ <b>Child behavior:</b> ⇌
(Rendu, 2004) or “S10”	BASIC Parent-Training Program, UK	Treatment	Competent adherence; assessed by researchers	76.25%	5-7 months post-intervention	Some facilitator competent adherence dimensions associated	<b>Group facilitation and child behavior:</b> ⇌ and ↑ <b>Practicalities and child behavior:</b> ⇌
(Robbins et al., 2011) or “S11”	Brief Strategic Family Therapy, US	Treatment	Competent adherence; assessed by researchers	74.00%	12 months post-randomization	Competent adherence associated with better parent and child outcomes	<b>Family functioning:</b> ↑ <b>Drug use:</b> ↑ <b>Some outcomes not reported</b>
(Roggman et al., 2016) or “S12”	Early Head Start, US	Selected	Competent adherence; assessed by a third party	57.28% (8.00%)	Child reaches age 36 months (registration at 5 months)	Competent adherence associated with better parent and child outcomes	<b>Parenting:</b> ↑ <b>Child academics:</b> ↑
(Satterfield, 2013) or “S13”	Functional Family Therapy, Ireland	Treatment	Competent adherence; assessed by program supervisors	66.00% (16.20%)	Final treatment	Competent adherence predicted behavior reductions from parent perspective but not teen perspective	<b>Youth reported behavior:</b> ⇌ <b>Parenting reported behavior:</b> ↑
(Scott et al., 2008) or “S14”	Incredible Years, UK	Treatment	Competent adherence; assessed by researchers	68.00% (8.00%)	5-7 months post-intervention	Greater competent adherence predicted better behavior	<b>Child behavior:</b> ↑
(Smith et al., 2013) or “S15”	Family Check-Up, US	Treatment	Competent adherence; assessed by researchers	62.11% (14.22%)	One and two years post-intervention	Greater competent adherence not directly associated with better improvements in parenting and behavior	<b>Parenting:</b> ⇌ <b>Child behavior:</b> ⇌



(Snider, 2019) or “S16”	Parent-Child Interaction Therapy, US	Universal and Treatment	Competent adherence; assessed by researchers	Competence: 56.00% (32.00%) Adherence: 70.00% (25.00%)	Baseline, 3 months follow-up, 6 months follow-up, and 12 months follow up (model includes all points as growth)	Competent adherence not associated with child behavior or parenting; results for all timepoints not reported	<b>Child behavior:</b> ⇔ <b>Parenting:</b> ⇔
(St. George et al., 2016) or “S17”	Familias Unidas, US	Selected	Competence and adherence measured separately; assessed by independent raters	51.43% (8.00%)	Baseline and 6 months follow-up	Higher competence associated with reductions in substance abuse but no relationship between adherence or competence and family functioning and no relationship between adherence and substance use	<b>Family functioning:</b> ⇔ (but data not provided) <b>Adherence and substance use:</b> ⇔ <b>Competence and substance use:</b> ↑
(Thijssen et al., 2017) or “S18”	Parent Management Training Oregon Model, Netherlands	Treatment	Competent adherence; assessed by independent raters	80.0%	Difference between scores at T1(6 months), T2 (12 months), and T3 (18 months) with baseline	Associations were not significant, but sub-constructs of facilitator competent adherence were	<b>Child behavior:</b> ⇔ <b>Parenting stress T1 and T2:</b> ⇔ <b>Parenting stress and T3:</b> ↑
<b>Note:</b> Using a modified version of the Thomson and Thomas (2013) effect direction plot visual display system, results are visualised using: a sideways arrow (⇔) indicating no statistically significant association between competent adherence and parent/child outcomes; an upwards arrow (↑) indicating a positive, statistically significant association between stronger competent adherence and better parent/child outcomes; or a downwards arrow (↓) indicating a negative, statistically significant association between stronger competent adherence and poorer parent/child outcomes.							

### 5.4.1 Programs and Outcomes

The 18 studies examined 11 different parenting programs delivered in high-income countries in Europe and the United States: Strengthening Families Program (n=1; Study 1 - S1), Multi-Dimensional Family Therapy (n=1; S8), Basic Parent Training Program (n=1; S10), Brief Strategic Family Therapy (n=1; S11), Early Head Start (n=1; S12), Functional Family Therapy (n=1; S13), Parent Child Interaction Therapy (n=1; S16), Familias Unidas (n=1; S17), Family Check-Up (n=2; S2, S15), Incredible Years (n=3; S3, S6; S14), and Parent Management Training Oregon Model (n=5; S4, S5, S7, S9, S18). The studies reported an average of 38 facilitators and 159 families. The average level of fidelity reported in the studies was 69.46%. According to the Institute of Medicine (2009) classification system, two programs were universal prevention (S1, S9), two were indicated prevention (S5, S7), three were selected prevention (S4, S12, S17), 10 were treatments (S2, S3, S6, S8, S10, S11, S13, S14, S15, S18), and one was a combination of universal and treatment approaches (S16). Ten of the 18 studies included programs that targeted caregivers only, with the remaining eight targeting both caregivers and children (S1, S2, S8, S11, S13, S15-S17).

Studies assessed the association of competent adherence with five outcomes – child development (n=1; S12), parenting stress (n=2; S9, S18), family functioning (n=2; S11, S17), parenting behaviors and skills (n=9; S1, S3-S6, S9, S12, S15, S16), and child behavior (n=13; S2, S6-S11, S13-S18). These outcomes were measured using numerous scales. For instance, in the 10 studies reporting on parenting behaviors and skills, the outcomes were measured using 10 different scales - the Intervention Targeted Parenting Attitude and Behavior Scale, the Dyadic Parent-Child Interaction Coding System, the

Family and Peer Process Code, the Parenting Sense of Competence Scale, Parenting Behavior Questionnaire, the Home Observation Measure of the Environment, Relationships Process Code, Coders Impression Inventory, Alabama Parenting Questionnaire, and Caregiver Wish List. In the 13 studies reporting on child behavior, the outcomes were measured using eight different scales – the Child Behavior Checklist (which was used most often; CBCL), Eyberg Child Behavior Inventory (ECBI), Swanson Nolan and Pelham-IV Questionnaire (SNAP-IV), Strengths and Difficulties Questionnaire, Parent Account of Child Symptoms Interview, Parent Daily Report, and two different researcher-created self-reported drug use measures. The variability in the scales used to measure parent and child outcomes contributed to the methodological heterogeneity of the studies.

#### **5.4.2 Designs and Analysis Methods**

The studies employed a wide range of methods. All studies examined facilitator delivery using data from the intervention arms of randomized trials. As expected, fidelity to outcome associations were observed as they occurred rather than experimentally manipulated. Associations between competent adherence and outcomes were analyzed using correlations (n=1; S18), regression and one-way ANOVA (n=1; S3), SEM/path analysis and correlation (n=2; S5, S15), latent growth curve modelling (n=3; S2, S8, S11), SEM/path analysis (n=3; S7, S17, S4), and regression (n=8; S1, S6, S9, S10, S12, S13, S14, S16). Competent adherence variables were modelled as categorical (n=1, three categories: “no exposure”, “low exposure”, and “high exposure”; S3) or continuous (n=17; all but S3) (Durlak & DuPre, 2008). Five studies conducted associations between fidelity captured at more than one timepoint, yet not all results were reported (S7, S8, S9, S16, S17).

### **5.4.3 Control Variables**

Of the 18 studies, 17 reported controlling for potential confounding variables to estimate the relationship between competent adherence and outcomes (all but S18). The controls varied considerably with only two studies conducted by the same researcher reporting the same combination of variables. The control variables included therapeutic alliance, program site, baseline participant outcomes, facilitator characteristics, organizational characteristics, and participant characteristics (e.g., child age, child gender). For one study, control variables were only partially used as facilitator, organizational, family, and attendance variables were included in other parts of the structural equation model (S17). The studies did not discuss the rationale for selection of control variables.

### **5.4.4 Clustering and Multiple Comparison**

Six of the 18 studies reported that they accounted for clustering of observations due to the nested design of delivering group-based parenting programs, often by facilitators working in pairs (S1, S4, S6, S8, S11, S14). Two of the five studies accounted for the same combination of clustering variables. The clustering variables accounted for included program level (n=1; S1), within-couple dependence (n=1; S4), family level (n=1; S11), parenting group (n=1; S6), and unspecified (n=2, S8, S14). In one of the 18 studies, it was unclear whether clustering was used because multi-level modelling accounted for repeated measures and multiple respondents per family, but not for multiple families per facilitator (S9). Further, when associations with several outcomes are investigated, it is best practice to account for multiple comparisons. None of the studies reported accounting for multiple comparisons.

#### **5.4.5 Associations of Competent Adherence with Outcomes**

Most studies found that facilitator competent adherence was positively associated with one or more parent and/or child outcomes. Of the 18 studies, six found statistically significant positive associations ( $\Uparrow$ ) between competent adherence and all parent and child outcomes examined (S2-S4, S11, S12, S14). A further eight studies found mixed evidence (S5, S7-10, S13, S17, S18) wherein at least one outcome was positively associated with competent adherence and one outcome was not ( $\Leftrightarrow$ ). Of these eight, all found that while some outcomes had a statistically significant positive association with competent adherence others had no significant association. None of the studies had a negative association with competent adherence. Of the 18 studies, four found no significant association between facilitator competent adherence and any of the outcomes studied (S1, S6, S15, S16). In conducting these analyses, most studies reported on the average level of fidelity achieved (all but S3 and S5).

#### ***Parenting Behaviors and Skills***

Of the nine studies examining competent adherence and parenting behaviors and skills, three found a positive association (S3, S4, S12), two found mixed associations based on the two types of analyses performed (S5) or in the types of parenting behaviors measured (S9), and four found no associations (S1, S6, S15, S16).

#### ***Parenting Stress and Family Functioning***

Of the two studies examining the relationship between competent adherence and parenting stress, one found a positive association (S9) and the other found mixed associations (S18). Similar findings were observed in the two studies examining competent

adherence and family functioning wherein one found a positive association (S11) and one found no association (S17).

### ***Child Behavior and Development***

Of the 13 studies examining the relationship between competent adherence and child behavior outcomes, three found a positive association (S2, S11, S14); five found a mix of positive and no association based on who reported outcomes (e.g., parent- versus teacher-report) or based on varying dimensions of competent adherence examined (S8, S9, S10, S13, S17); and five found no associations (S6, S7, S15, S16, S18). Finally, in the one study examining child developmental outcomes, a positive association was observed between competent adherence and child academic attainment (S12).

### ***Dimensions of Competent Adherence***

Two studies examined competence and adherence separately and found differences in their association with outcomes (S8, S17). Utilizing linear models, Hogue et al. (2008; S8) found that adherence was related to greater reductions in child externalizing behavior while competence was not. Furthermore, neither adherence nor competence were associated with child internalizing behavior. In exploring curvilinear relationships, this study found some evidence of a curvilinear relationship between adherence and internalizing behavior where medium levels of adherence were positively associated with outcomes. However, the study did not report the results of an analysis of curvilinear relationships between adherence and externalizing behavior issues and did not report on competence and either internalizing or externalizing behavior issues. St George et al. (St. George et al., 2016; S17) found that competence was related to decreased substance use, but not related to

improvements in family functioning and that adherence was not related to either reduced substance use or improved family functioning.

Four studies examined whether specific dimensions of competent adherence were associated with outcomes (S5, S10, S11, S18). For example, in one of two models tested, Rendu (2004; S10) found that one dimension of competent adherence – a group facilitation technique – was related to reductions in child behavior issues but another facilitation technique was not so related in either model tested. As another example, Robbins et al. (2011; S11) found that a facilitation approach called “joining” was associated with improved family functioning and reduced adolescent drug use, but three other facilitator approaches were not.

Several studies found differences in relationships between facilitator competent adherence and participant outcomes when reported using different measures (S7, S8, S13). To illustrate, in the paper by Hukkelberg et al. (2013; S7) competent adherence was associated with reductions in child behavior problems based on parent-reports, but not based on teacher-reports. As another example, the paper by Satterfield (2013; S13), competent adherence was associated with reductions in child behavior issues based on parent-reports, but no association was found based on youth-reports.

## **5.5 Discussion of Clinical Implications**

### **5.5.1 Overall Findings**

The synthesis considered 18 studies reporting on the relationship between observational measures of facilitator competent adherence and parent/child outcomes, with most interventions having a treatment focus. Studies focused variously on selective or indicated prevention, with most evaluating treatment programs. Treatment studies

nevertheless have considerable implications for prevention, as treatment for child behavior problems is intended also to serve as prevention of their long-term adverse outcomes, including offending and poor mental health, education, and employment prospects. Of the 18 included studies, studies reported on child behavior, nine on parenting skills and behaviors, two on parenting stress, two on family functioning, and one on child development. The studies were highly heterogeneous in their design and analysis methods. Five studies conducted analyses on associations between fidelity captured at more than one timepoint.

Most studies found that facilitator delivery is associated with at least one parent or child outcome - eight of the 13 studies examining child behavior, five of the nine studies examining parenting skills and behaviors, both studies examining parenting stress, one of the two studies examining family functioning, and the one study examining child development. These findings generally suggest better competent adherence is associated with better parent and child outcomes. There was no discernible difference in associations between competent adherence and outcomes based on the aspects of competent adherence measured. Still, there was a sizeable number of studies with mixed findings where some outcomes are associated with competent adherence and others are not. Lack of detected associations between competent adherence and outcomes has several potential explanations, such as that fidelity does not matter for outcomes (low fidelity has no negative impact), that our efforts regarding fidelity are not worthwhile (high fidelity has no positive impact), issues due to poor measurement, and/or lack of statistical power to detect associations due to small sample sizes. Fidelity may also be indirectly associated with outcomes. Indirect associations were explored in several studies - such as Smith et al.



(2013; S15) who found that although fidelity was not directly associated with parenting or child behavior, it was indirectly associated with some outcomes. Finally, other implementation variables, such as participant responsiveness or engagement during program sessions, participant attendance, and facilitator-participant working alliance, may mediate or moderate the relationship between fidelity and outcomes (Berkel et al., 2011; Carroll et al., 2007). For example, if the effect of fidelity on child outcomes is mediated by participant engagement, statistical models that adjust for this variable may reduce the observed relationship between fidelity and child outcomes. Further research on competent adherence-outcome relationships would benefit from a more systematic theoretical understanding of the variable relationships.

### **5.5.2 Conceptualizing the Role of Facilitator Competent Adherence**

The finding that better facilitator competent adherence is generally associated with better outcomes is limited by the diverse conceptualizations of the relationship between competent adherence and outcomes in the studies. This diversity is exemplified by the range of controls used in the models tested, including facilitator (e.g., therapeutic alliance), organizational (e.g., amount of coaching support provided to facilitators), and participant characteristics (e.g., child age, baseline outcomes), with only two studies using the same combination of controls. These differences reveal considerable variation in how researchers theorize about the potential mechanisms impacting, and dissensus on how they hypothesize, the relationship between facilitator competent adherence and outcomes. As few papers articulated a clear rationale or delineated a conceptual framework for their choice of controls, such as through directed acyclic graphs (DAGs) (Pearl et al., 2016), research in the field may be at risk of including unnecessary controls and overcontrol bias

(incorporating inappropriate variables leading to spurious results or underpowered models) (Achen, 2005; Rohrer, 2018) . As a result, future research would benefit from documentation about the theory underpinning the research - the mechanism(s) linking facilitator, organizational, and participant characteristics with competent adherence and its association with outcomes.

### **5.5.3 Methodological Issues and Study Quality**

There are limitations in the studies reviewed. In particular, issues concerning the robustness of the analyses, reliability and validity, and quality of reporting are discussed using ROBINS-I (Risk of Bias In Non-randomized Studies of Interventions) (Sterne et al., 2016).

#### ***Robustness of Analyses***

The sample size in the studies was generally small, with data collected from an average of 38 facilitators and 159 families. Further, none of the studies performed power calculations to determine the number of observations necessary to examine the relationship between competent adherence and outcomes, or accounted for multiple comparisons. Studies differed considerably as to whether and which type of clustering was accounted for in the analyses. Clustering may occur at the parent group level (parents often receive an intervention in a group format) and the facilitator level (programs are often delivered by more than one facilitator and facilitators typically deliver a program to more than one parent) leading to non-independent observations. However, only six studies accounted for clustering. A wide range of statistical approaches (such as latent growth curve modelling, regression, and SEM/path analysis) were used, demonstrating variation in researcher thinking about how an analysis of competent adherence-outcome relationships should be

examined. This variation was one factor which prevented a meta-analysis. If future studies report bivariate correlations and unstandardized regression coefficients, these studies will produce results that are easier to standardize and incorporate in meta-analyses.

### ***Reliability and Validity***

Little is known about the accuracy and reliability of observational measures of competent adherence meaning conclusions drawn from this synthesis should be made with caution (Breitenstein et al., 2010b). Eleven of the 18 studies synthesized herein provide some information on the reliability and/or validity of the measures of competent adherence used – with ten of these reporting on inter-rater reliability, nine reporting on internal consistency, and five reporting on construct validity (Martin et al., 2021b) (see [Online Resource 4](#)).

### ***Reporting***

In future research, study reporting could be more detailed. Five of 18 papers did not provide information on facilitator sample size. Additionally, several studies made claims about competent adherence-outcome relationships yet did not provide the numerical results. Other studies did not provide information about key outcomes mentioned in their methods.

### **5.5.4 Strengths and Limitations**

Although this paper makes an important contribution to implementation science as it relates to parenting programs, it has limitations. This review did not include studies reporting on non-observational measures of facilitator competent adherence (e.g., self-report measures). A synthesis of such measures would be worthwhile to conduct when greater study homogeneity permits meta-analyses. In addition, this review focused on parenting programs aiming to (a) reduce child maltreatment; harsh or dysfunctional

parenting; and/or child conduct problems and/or (b) improve positive child behavior management strategies; parent-child bonding/attachment and relationships; and/or early childhood development outcomes. However, as studies reported on associations between facilitator competent adherence and several secondary outcomes (e.g., parenting stress), we synthesized information on these analyses as well. Thus, the associations reported herein between facilitator competent adherence and these secondary outcomes are likely not inclusive of all literature reporting on these outcomes. Further, this review was unable to explore selective reporting bias or publication bias. Aggregate associations and publication bias could not be explored as the substantial methodological heterogeneity prevented a meta-analysis from being conducted. While there are limitations, the review is the first to synthesize the data from studies examining the relationship between observationally measured facilitator competent adherence and parenting program participant outcomes to clarify the mixed evidence found in the literature.

### **5.5.5 Suggestions for Future Research**

Future studies, and the parenting intervention field at large, would benefit from investigations of competent adherence-outcome relationships that clearly conceptualize the mechanisms hypothesized to influence competent adherence-outcome relationships; utilize larger sample sizes; account for clustering variables at the parent- and facilitator-level; incorporate carefully chosen control variables; follow best practices in open science, including pre-registration to minimize the risk of selective reporting; and report bivariate correlations and unstandardized regressions. As a result, this literature may benefit from reporting guidelines.

The field would also benefit from an examination of understudied and novel aspects of the relationship between competent adherence and family outcomes. For instance, there was some evidence from one study that the relationship between fidelity and outcomes was not linear and at high levels fidelity could be associated with worse program outcomes (Hogue et al., 2008). Exploration of curvilinear relationships would provide information on whether there is a tipping point at which further attention to fidelity is unnecessary or even unhelpful. Other topics to examine in future studies analyzing competent adherence-outcome relationships include examining the test-retest reliability of measures to see how competent adherence fluctuates over time as this review only identified two studies that examined associations between competent adherence and outcomes at more than one timepoint; determining how to weigh fidelity with adaptation; exploring whether competent adherence-outcome relationships are significant in the long-term; and testing whether fidelity is indirectly associated with parent/child outcomes or whether other implementation variables such as engagement or working alliance mediate the relationship between fidelity and outcomes. Further, it may be valuable to study competent adherence in different contexts, since all of studies found were nested within randomized trials, which may not always be fully representative of routine delivery contexts.

## **5.6 Conclusion**

This review aimed to provide clarity on the evidence regarding the role facilitator competent adherence plays in achieving parent and child outcomes. While this paper finds that the evidence is inconsistent, the synthesis also finds a general trend indicating that higher levels of facilitator competent adherence are related to improved parent and child outcomes. The latter finding is limited by a high number of the studies having found mixed

evidence and no associations with outcomes as well as the diverse methodological approaches employed by the studies. The review highlights the need for further research on whether there is an association between facilitator competent adherence and outcomes and recommends how researchers and practitioners can advance the field.

## **6. Paper 2 - Assessing facilitator delivery of a parenting program at scale in Tanzania: A psychometric evaluation of an observational measure of facilitator competent adherence**

This paper is in submission to an academic journal. This paper has been written with US spelling in alignment with the journal's submission guidelines. Please find the tables throughout the chapter and the supplementary materials in [Appendix 8](#). Note that several of the methodological choices made in conducting this study are expanded upon herein but are not included in the submitted paper.

**Authors:** Martin, M.<sup>1</sup>, Lachman, J. M.<sup>1,2</sup>, Calderon, F.<sup>1</sup>, Han, Q.<sup>1</sup>, Shenderovich, Y.<sup>3,4</sup>, Ndyetabura, E.<sup>5</sup>, Mwakakeke, S.<sup>6</sup>, Chikola, L.<sup>6</sup>, Mgunga, S.<sup>7</sup>, Manjengenja, N.<sup>7</sup>, Exavery, A.<sup>5</sup>, Ally, A.<sup>5</sup>, Gardner, F.<sup>1</sup>, and Wamoyi, J.<sup>7</sup>

### **Affiliations:**

<sup>1</sup>Department of Social Policy and Intervention, University of Oxford, United Kingdom

<sup>2</sup>MRC Social and Public Health Sciences Unit, University of Glasgow, United Kingdom

<sup>3</sup>Wolfson Centre for Young People's Mental Health, Cardiff University, United Kingdom

<sup>4</sup>Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement, School of Social Sciences, Cardiff University, United Kingdom

<sup>5</sup>Pact Tanzania, United Republic of Tanzania

<sup>6</sup>Clowns Without Borders South Africa, South Africa

<sup>7</sup>Department of Sexual and Reproductive Health, National Institute for Medical Research, United Republic of Tanzania

**My role:** I conceptualised of the study, liaised with field staff from the implementing partner organizations responsible for collecting facilitator assessments, prepared the data, planned and conducted the analyses, drafted the manuscript, and revised the manuscript. Other authors contributed support for the analyses and by reading and approving the manuscript.

### **6.1 Abstract**

There is limited evidence on the extent to which facilitators deliver parenting programs with competent adherence at scale in low-income community settings. The literature is also sparse with respect to the reliability and validity of measures to assess facilitator competent adherence, particularly in low-income community settings. Yet, strong

assessment tools are vital to address the challenges of maintaining fidelity at scale. The present study contributes to the scant literature by examining the psychometric properties of the observational measure used to assess facilitators who delivered Parenting for Lifelong Health for Parents and Adolescents (PLH-Teens) to 36,259 parents/caregivers and 38,802 adolescents in Tanzania and reporting on the level of competent adherence facilitators achieved. COSMIN guidelines were followed to assess the content validity, intra-rater reliability, inter-rater reliability, construct validity, and internal consistency of the PLH-Teens Facilitator Assessment Tool (PLH-FAT-T) used to assess facilitator competent adherence. Ninety-five PLH-FAT-T assessments were collected in practice. The PLH-FAT-T showed strong content validity, poor to moderate intra- and inter-rater reliability, strong internal consistency, and moderate construct validity. Iterative exploratory factor analyses produced a shortened PLH-FAT-T, the PLH-FAT-T Short Form, comprised of 19 fewer items. Using the PLH-FAT-T Short Form, total and subscale competent adherence scores were calculated as an indication of facilitator overall competent adherence. Analyses of the PLH-FAT-T Short Form found it had stronger psychometric properties. Using the tool, facilitators were found to have high levels of competent adherence (82.3% on average). This paper is the first psychometric evaluation of the PLH-FAT-T and the first to report on the competent adherence of a parenting program delivered at scale in a low-income country. Findings suggest that, while the PLH-FAT-T Short Form needs further work, it is a promising tool for measuring the delivery of PLH-Teens facilitators. Findings also suggest that facilitators implemented PLH-Teens to a high level of quality at scale in a low-income community setting despite significant barriers.

*Key Words:* implementation, fidelity, parenting, scale-up, psychometric properties



## 6.2 Background

There is extensive evidence that parenting programs are an effective way to reduce violence against children and child behavioural and emotional problems (e.g., Chen & Chan, 2016; Furlong et al., 2013; Leijten et al., 2019). Although hundreds of randomized trials indicate that parenting programs have a positive impact across multiple child and family outcomes, few are implemented at scale with mixed evidence of their effectiveness (e.g., Gray et al., 2018; Marryat et al., 2017; Shapiro et al., 2010). One explanation for reduced effects at scale is program drift, which can be studied by collecting implementation fidelity data to examine whether core intervention components remain intact and to identify how flawed implementation might be rectified (Mowbray et al., 2003). However, relatively little attention has been paid to the implementation fidelity with which parenting interventions are delivered at scale (e.g., Shenderovich, 2021; Tomlinson et al., 2018). One of many potential reasons for this gap is that monitoring and evaluating the delivery of evidence-based programs requires the development and use of tools and processes to measure implementation components, such as adherence, competence, dosage, and participant engagement (Dane & Schneider, 1998; Dusenbury et al., 2003; Mihalic, 2004; Proctor et al., 2011). Further reasons include insufficient knowledge about the usefulness of implementation tools by organizations responsible for program delivery, the extent to which implementation tools capture the constructs they intend to measure (validity), and whether implementation tools can be applied consistently by implementing staff over time (reliability) (Glasgow & Riley, 2013). Lack of knowledge about the reliability and validity of implementation measures is further evidenced by few psychometric evaluations (Schoenwald & Garland, 2013; Stirman, 2020). A systematic review of 41 studies reporting

on observational measures of facilitator competent adherence in the parenting program literature found 20 studies examining inter-rater reliability, two examining intra-rater reliability, 20 examining internal consistency, three examining content validity, nine examining construct validity, and one examining convergent validity (Martin et al., 2021b). The gaps in knowledge regarding the psychometric properties of implementation fidelity measures are important to address as the “widespread use of fidelity scales requires reliability and validity” (Ruud et al., 2020, p. 871).

### **6.3 Parenting for Lifelong Health for Parents and Adolescents**

The recent large-scale delivery of Parenting for Lifelong Health for Parents and Adolescents (PLH-Teens, locally known as Furaha Teens) in Tanzania was a unique opportunity to assess the competent adherence – the skill and strictness with which a facilitator delivers intervention components (Breitenstein et al., 2010a; Fixsen et al., 2005) – of PLH-Teens facilitators. The scale-up of PLH-Teens was also a unique opportunity to examine the psychometric properties of the measure used to assess facilitator competent adherence to intervention activities and skills. From January 2020 to March 2021, PLH-Teens was delivered by multiple non-governmental organizations to 75,061 participants - 38,802 adolescent girls and 36,259 parents/caregivers (parents) - in eight rural and semi-urban districts of Tanzania as part of a large-scale HIV-prevention program (Martin et al., 2021a). Participating families were affected by several disadvantages and vulnerabilities, including HIV/AIDS and financial insecurity (Lachman et al., forthcoming). PLH-Teens was one of several interventions delivered as part of a package of services provided via Kizazi Kipya, which supported programming for over a million adolescent girls in sub-Saharan Africa with funding from USAID.

Delivered in over 19 low- and middle-income countries (LMICs), PLH-Teens is a low-cost, evidence-based parenting program aiming to reduce violence against children and child behavioral and emotional problems as well as improve mental health and overall functioning among vulnerable families (Cluver et al., 2016; WHO, 2020). According to the program manual, parents and adolescents participate in 14 weekly group sessions with each session delivered over three hours by pairs of facilitators trained in the program's participatory, strengths-based, and non-didactic techniques (Cluver et al., 2018). The large-scale delivery of PLH-Teens in Tanzania represents the largest implementation of PLH worldwide and one of the largest implementations of a parenting program in a LMIC (Shenderovich, 2021).

#### **6.4 Current Study**

This study contributes to the literature on the psychometric properties of implementation fidelity measures used in parenting programs and on the competent adherence with which parenting programs are delivered in low-income community settings at scale. The paper reports on an analysis of the reliability and validity of the Parenting for Lifelong Health for Parents and Adolescents-Facilitator Assessment Tool (PLH-FAT-T), an observational tool developed and used to assess the competent adherence of facilitators delivering PLH-Teens. The tool is used for a variety of purposes including by implementers to provide facilitators with feedback on their delivery. The psychometric evaluation was conducted to assess whether the tool was understandable to stakeholders; investigate whether the tool could be implemented consistently by assessors; examine the variability among measure items; evaluate the measure's underlying constructs; and determine whether any items could be removed from the tool. The paper also reports on the competent

adherence with which community facilitators delivered PLH-Teens. To this end, the paper addresses the following research questions: (1) What are the psychometric properties of the observational measure used to assess the competent adherence of PLH-Teens facilitators during routine service delivery at scale in Tanzania? and (2) What is the level of competent adherence with which facilitators deliver PLH-Teens during routine service delivery at scale in Tanzania?

## **6.5 Methods**

### **6.5.1 Facilitator Assessment Tool**

Worldwide, PLH programs use two different tools to assess facilitator competent adherence in research and practice – one for PLH-Teens (PLH-FAT-T) and one for PLH for Young Children (PLH-FAT-YC). Various versions of these tools have been used for several years. Although the tools have different items, they have similarities due to shared interventions components. There has been one study on the PLH-FAT-YC, which found that the tool had sufficient content validity, intra-rater reliability, and inter-rater reliability to recommend its continued use (Martin et al., 2022a) but there has not been a study on the PLH-FAT-T. The present study focuses on the PLH-FAT-T.

The initial version of the PLH-FAT-T examined in this study consists of two subscales (i.e., Activities and Skills). The Activities Subscale measures facilitator adherence based on two activities – the home activity discussion (11 items) and the role-play (7 items). The home activity discussion is a facilitator-guided conversation held with participants to support them with challenges experienced when implementing learned skills at home (e.g., “The facilitator identifies at least one specific challenge experienced by a participant regarding the main home activity”). The role-play involves facilitators

supporting parents and adolescents in practicing new skills (e.g., “The facilitator discusses with participants about how the role-play relates to their lives”).

The Skills Subscale measures facilitator competence using items that examine how well facilitators utilize core facilitation skills (19 items). These items assess the degree to which facilitators model parenting behaviors (e.g., “The facilitator gives positive reinforcement and specific praise to participants”) (5 items), demonstrate a core PLH facilitation technique called ‘Accept-Explore-Connect-Practice’ (e.g., “The facilitator explores participant experiences and opinions using open-ended questions”) (4 items), and utilize collaborative leadership skills (e.g., “The facilitator assures equal and active participation among participants”) (10 items).

Each of the 37 PLH-FAT-T items are rated using a four-point Likert scale ranging from zero to three (*0=inadequate, 1=needs improvement, 2=good, 3=excellent*). Final PLH-FAT-T scores, represented as a percentage, are calculated out of a total possible score of 111 (37 items x 3 points per item).

Assessments using the PLH-FAT-T are conducted by program coaches – individuals selected by local implementing partners and trained by Clowns Without Borders South Africa (CWBSA) – to provide supportive supervision including ongoing mentorship and feedback to facilitators. CWBSA is a non-profit organization which provides implementation support to organizations delivering PLH globally.

### **6.5.2 Observational Assessors and Assessment Procedure**

Several steps were taken to train coaches (also referred to as assessors) to conduct PLH-FAT-T assessments in Tanzania. To start, two Kiswahili-speaking lead assessors from CWBSA were trained on how to deliver PLH-FAT-T training to coaches. Then, the lead

assessors provided two days of training (approximately 14 hours) to program coaches on how to use the PLH-FAT-T. The training workshop provided an overview of the measure items, outlined how to conduct assessments, reviewed potential challenges assessors might face and potential solutions to those challenges, required assessors to conduct one practice assessment, and required assessors to participate in an evaluation of that assessment (wherein assessors' practice assessment results were compared to the assessment results reached by the lead assessors of the same observed program delivery). Training was supported by a PLH-FAT-T manual, which outlined the coding procedure and criteria for each PLH-FAT-T item. After training, observational assessments of facilitators were then conducted by the coaches who attended live program sessions.

### **6.5.3 Contextual Challenges**

Although video assessments are simpler and less resource-intensive (Breitenstein et al., 2010b; Feely et al., 2018), assessments were conducted in-person instead of via video due to budget and connectivity challenges. Budget limitations meant that there were insufficient resources to purchase enough video cameras. Even if there had been sufficient funds, as program delivery was conducted in rural areas, implementers often experienced poor internet connectivity and access to electricity, which made it impossible for video files to be shared with coaches online. However, as coaches often lived a substantial distance away from program sites, transportation was also a barrier to in-person assessments. These barriers were compounded by the already considerable challenges of monitoring implementation in routine service delivery settings with limited budget, staff, and data collection procedures. As PLH-Teens was delivered during COVID-19, the pandemic also

affected the ability of coaches to conduct assessments due to lockdowns and school closures as well as decisions made by local implementing partners.

#### **6.5.4 Data Collection Procedure**

Although the initial plan was for all 444 facilitators to receive an assessment, 95 PLH-FAT-T assessments (representing 21.4% of the facilitator population) were collected due to the challenges mentioned. The assessments collected represent a convenience sample as researchers had no control over how facilitators were selected for assessments and which of the 14 sessions were observed. Coaches chose which facilitators and sessions to assess. Facilitators and coaches were asked for their consent to participate in the research by implementing partners after their recruitment to deliver the program. All data was collected and anonymized by implementers before sharing with the research team.

#### **6.5.5 Facilitators and Coaches**

PLH-Teens was delivered by 444 community facilitators who were recruited by local implementing partners and who received five days of training from CWBSA to become PLH facilitators. Of the 95 facilitators who received assessments, demographic data was only collected from a subset of 34 facilitators (35.8%) due to practical challenges faced by implementing partners. Among this subset, the average age was 33.4 years, 76.5% were male, 91.8% were parents, and 100.0% identified as being part of the same community as the PLH-Teens participants. The highest level of education achieved by facilitators ranged from primary school to professional designations (2.9% certificates; 5.9% primary school; 8.8% diplomas; 8.8% professional degrees; 32.4% secondary school; and 41.2% high secondary school).

Facilitators received ongoing supportive supervision from 69 coaches. Although all coaches were expected to complete PLH-FAT-T assessments, only 31 conducted assessments due to practical challenges faced by implementing partners. Of the 31 coaches, demographic data was collected from a subset of eight coaches. Among this subset, the average age was 35.9 years, 50.0% were female, and 100.0% were parents. The highest level of education achieved by coaches ranged from primary school to undergraduate degrees (12.5% primary school; 12.5% secondary school; 25.0% high secondary; 25.0% professional degree; and 25.0% bachelor's degree). Coaches were employed in jobs outside of their role with PLH-Teens including in teaching, social work, public health, child protection, and childcare.

#### **6.5.6 Psychometric Evaluation**

Consistent with the approaches used in the wider literature (Martin et al., 2021b), this study examined the psychometric properties of the PLH-FAT-T by analyzing its content validity, intra-rater reliability, inter-rater reliability, internal consistency, and construct validity. Content validity, the degree to which a measure appears to adequately capture its underlying construct(s) according to key stakeholders (Markus & Smith, 2010), was examined by reviewing insights from the study on the PLH-FAT-YC (Martin et al., 2022a) and by collecting feedback from lead assessors and other personnel at CWBSA with international expertise in conducting and coordinating the training and assessment of PLH facilitators. The insights and feedback were then used to update the PLH-FAT-T as well as its manual, training, and assessment procedures. Utilizing the insights from the study on the PLH-FAT-YC as part of the content validity process was warranted since PLH-Teens and



PLH-YC use similar facilitation approaches grounded in social learning theory with participatory, non-didactic, and strengths-based methods as well as similar items.

Intra-rater reliability was examined by having the lead assessors conduct two PLH-FAT-T assessments of the same program delivery on two separate occasions. Using video recordings, the lead assessors each assessed the program delivery of two facilitators in August 2020 and then repeated the assessments in November 2020. A three-month gap between assessments was selected to reduce the likelihood that assessors remembered their previous assessments.

Inter-rater reliability between assessors was examined by having the two lead assessors conduct independent assessments at the same time as 22 randomly selected coaches. As facilitators deliver PLH-Teens in pairs, each lead assessor and 11 of their coaches attended 11 program sessions to assess 22 facilitators. Altogether, the inter-rater reliability was based on 44 PLH-FAT-T assessments. To reduce reporting bias, lead assessors and coaches were instructed not to discuss how they rated each facilitator (Walton et al., 2017). Random selection of coaches across geographic regions was used for logistical reasons as it was not possible to conduct inter-rater reliability analyses with all 69 coaches (Walton et al., 2017).

Intra- and inter-rater reliability were examined by calculating percentage agreements and intra-class correlations (ICCs) using R statistical software (v4.1.1) and the 'irr' package (Bruton et al., 2000; Gamer et al., 2017; Koo & Li, 2016; R Core Team, 2021b). Percentage agreements and ICCs were also used in the two studies on intra-rater reliability as well as in many of the 20 studies on inter-rater reliability found in a systematic

review of the parenting program literature (Martin et al., 2021b). Percentage agreements were calculated at the item level with agreements above 70% considered acceptable (Aspland & Gardner, 2003). When calculating ICCs, decisions regarding the ‘model’, ‘definition’, and ‘type’ were made (McGraw & Wong, 1996). For intra-rater reliability, a two-way mixed-effects model with an absolute agreement definition was used for ICCs (Shrout & Fleiss, 1979). Absolute agreement is the appropriate definition as agreement between multiple measurements is of interest (Koo & Li, 2016). For ‘type’, single-rater was selected as the mean values of the assessors is not of interest (Koo & Li, 2016). For inter-rater reliability, a one-way random-effects model was used for ICCs as although there is a set pool of possible assessors, facilitators were not evaluated by the same assessors (Koo & Li, 2016). Absolute agreement is also appropriate because agreement between multiple assessors is what this study sought to examine (Koo & Li, 2016). For ‘type’, single-rater was selected as the mean values of the assessors is not of interest (Koo & Li, 2016). For both intra- and inter-rater reliability, ICCs within the 95% confidence interval under 0.50 were considered poor, between 0.50 and 0.75 moderate, between 0.75 and 0.90 good, and above 0.90 excellent (Koo & Li, 2016). Koo and colleagues indicate that interpreting ICCs based on a range of possible scores within the 95% confidence interval is critical as ICCs are only estimates of the true ICC score (2016).

Internal consistency was determined by examining how responses to the PLH-FAT-T were statistically interrelated (Heinl et al., 2016; Terwee et al., 2007). Cronbach’s alphas and omegas were calculated using the ‘psych’ and ‘ltm’ packages in R (v4.1.1) (Furr, 2017; Mair, 2018; R Core Team, 2021b; Revelle, 2017; Revelle & Zinbarg, 2009; Rizopoulos,

2007). This item-level approach was selected because it allows for the examination of associations between items (Furr, 2017). In the parenting program literature, Cronbach's alphas are the most common way of representing internal consistency allowing for an easy comparison with other studies (Martin et al., 2021b). However, since it is unknown if errors are random or if item scores are correlated in a linear fashion, the reliability of the tool may be underestimated with alphas (Furr, 2017). To address this limitation, an omega coefficient was also calculated as this approach does not require the same assumptions and is thus more broadly applicable (Furr, 2017; Revelle & Zinbarg, 2009). Values above 0.70 were sufficient (Terwee et al., 2007).

Construct validity, the internal structure or dimensionality of the PLH-FAT-T was analysed using a common factor model exploratory factor analysis (EFA). EFA was appropriate for two reasons. First, the relationship between adherence and competence for this tool was unknown. As a result, it was difficult to pre-specify how the two dimensions were expected to relate to one another and how the items would map onto these constructs. Second, as the PLH-FAT-T was created and revised as part of a content validity process to simplify coding during assessments, there was uncertainty regarding whether all items in the version of the tool being tested should be retained. EFA helps determine what items are not serving their intended purpose and can be considered for deletion or revision (Furr, 2017). A common factor model was selected over other approaches (e.g., principal components model) as the model is appropriate "when the goal of research is to identify latent constructs for theory building or to create a measurement instrument in which the researcher plans to make the case that the resulting measurement instrument reflects a

meaningful underlying construct” (Fabrigar & Wegener, 2011, p. 33). Further, a principal component model assumes that there is no measurement error, which is often an inappropriate assumption when using real-world data (Fabrigar & Wegener, 2011).

Three separate EFAs, one for each subscale, were conducted because there were no overlapping responses between the home activity and role-play items. Each EFA was conducted using five steps. First, the suitability of the data was tested by determining the amount of missing data, checking whether normality and linearity assumptions were met, and using Bartlett’s Test of Sphericity and the Kaiser-Meyer-Olkin statistic to test for sufficiently large correlations (Bartlett, 1954; Kaiser, 1974; Watkins, 2018). The amount of missing data was handled using maximum likelihood as this approach is recommended when larger amounts of data are missing (Baraldi & Enders, 2010; Watkins, 2018). Further, each item was inspected for normality and linearity. These characteristics of the data are important to determine as factor analyses typically rely on normally distributed and linear data. If these assumptions are not met, other approaches for calculating correlations among items are recommended (e.g., polychoric correlations) (Watkins, 2018). The normality of each item was calculated using the Shapiro-Wilk Normality Test, wherein p-values less than 0.05 indicate that the data is not normally distributed (Shapiro & Wilk, 1965). Finally, the data were checked to ensure correlations among items were sufficiently high to proceed with the factor analysis. Sufficient correlations were checked using Bartlett’s Test of Sphericity and the Kaiser-Meyer-Olkin (KMO) statistic (Watkins, 2018). Bartlett’s Test of Sphericity was used to determine whether there was a statistically significant difference between items using a chi-square test wherein the data were considered sufficient for the

factor analysis if a statistically significant chi-square was produced (Bartlett, 1954; Watkins, 2018). In addition, the KMO statistic was calculated to determine the shared variance between variables (Watkins, 2018). Ideally, the amount of variance between items is 0.70 or greater with values in the 0.60s considered mediocre, values in the 0.50s considered poor, and values below 0.50 considered unacceptable (Kaiser, 1974; Watkins, 2018). A KMO was calculated as Bartlett's approach is often critiqued for being highly sensitive to randomness (Watkins, 2018).

Second, an iterated principal axis approach was selected as this method is best suited to small sample sizes and when multivariate normality is not met (Fabrigar & Wegener, 2011; Watkins, 2018).

Third, the number of factors for the EFA were selected based on the results of a scree plot, parallel analysis, and maximum average partials test (Fabrigar & Wegener, 2011; Watkins, 2018). Although often used by researchers, the Kaiser criterion (wherein eigenvalues greater than or equal to one indicate the presence of a factor) was not used as this approach is now considered inappropriate due to its inaccuracy (Costello & Osborne, 2005; Fabrigar & Wegener, 2011; Fabrigar et al., 1999; Furr, 2017; Watkins, 2018). Scree plots illustrated the eigenvalues and factor numbers, which together revealed how many factors were present (Furr, 2017); the number of factors can be visually determined from the plot by identifying the point at which the line levels off (Furr, 2017). This visual determination was conducted by producing a plot of the total variance explained wherein the eigenvalues for each item are portrayed (Furr, 2017). The plot shows where values become small thereby illustrating the levelling-off point (Furr, 2017). Parallel analysis is an

approach that compares the real data to simulated random data to extract the number of eigenvalues produced from the real data that is greater than the number of eigenvalues produced from the simulated data (Horn, 1965; Watkins, 2018). By inspecting a plot of the parallel analysis results, the number of factors is determined by finding the point before the lines intersect (Humphreys & Ilgen, 1969; Humphreys & Montanelli Jr, 1975). The maximum average partials test criterion determines the number of factors that emerge when the common variance in the data is removed leaving the unique variance (Velicer, 1976; Watkins, 2018). However, at times, the maximum average partials test has been found to overestimate the number of factors to extract (Warne & Larsen, 2014), so should be weighed with the results of the scree plot and parallel analysis. From these analyses, the number of factors was determined and then used in the factor analysis.

Fourth, if there was an indication of multi-dimensionality, the factors were rotated using an oblique rotation as this approach allows for factors to be correlated or uncorrelated whereas a orthogonal rotation treats factors as uncorrelated (Furr, 2017).

Fifth, associations between PLH-FAT-T items and the tool's underlying factors in the EFA were analyzed wherein factor loadings equal to or above 0.50 were considered sufficient (Field, 2013). The factor analyses were performed using the “psych” package in R (v4.1.1) (R Core Team, 2021b; Revelle, 2017). To determine whether a shortened version of the PLH-FAT-T would perform better, this study conducted several iterations of the EFA for each of the subscales to determine if any items could be removed. Iterative EFAs were performed until none of the items were below the 0.50 loading threshold, variance was as close to 50% as possible, and indices of internal consistency were at or above 0.70. As a

shortened version of the PLH-FAT-T (the PLH-FAT-T Short Form) was produced from the iterative EFAs, additional calculations of assessor intra- and inter-rater reliability were conducted.

#### **6.5.7 Descriptive Analysis of Level of Facilitator Competent Adherence**

The level of competent adherence was computed (using both the revised PLH-FAT-T produced from the content validity exercise and the PLH-FAT-T Short Form produced from the iterative EFAs) by calculating a total (activities and skills) score for each facilitator and by calculating the average facilitator score. Data are provided for facilitators who received assessments with no missing data.

### **6.6 Results**

#### **6.6.1 Content Validity**

Four major modifications were made to the original PLH-FAT-T via the content validity process, which integrated feedback from experts as well as the results of the psychometric evaluation of the PLH-FAT-YC (see [Supplementary File 1](#)). First, drawing upon insights generated from the study of the PLH-FAT-YC, the PLH-FAT-T's assessment procedure was modified so that both facilitators in a pair could be assessed at the same time instead of one at a time as the tool originally required. This change meant that each facilitator was assessed on only one activity – either the home activity discussion or role-play – and on all of the skills items. The change was made because assessors in the PLH-FAT-YC study indicated that the original coding procedure was impractical, time-consuming, and too costly to be conducted in a community setting. Since each facilitator in a pair takes turns leading program activities in each session, the original version of the

PLH-FAT-T required coaches to observe two sessions in order to complete an assessment of each facilitator. Second, drawing on insights from the Martin et al. (2022a) study, the PLH-FAT-T measurement scale was altered from a four-point to a three-point Likert scale to simplify the assessment procedure and reduce discrepancies between assessors. This change was made as it was hypothesized that the low inter-rater reliability found in the study on the PLH-FAT-YC was due to low agreement on how to apply a ‘1’ or ‘2’ and that discrepancies would be minimized following the removal of one point on the scale (Martin et al., 2022a). Following this change, each PLH-FAT-T item was assessed on a three-point Likert scale ranging from zero to two (*0=inadequate, 1=good, 2=excellent*). Third, specific definitions for each PLH-FAT-T item and point on the Likert scale were created. These definitions were generated based on insights from the study on the PLH-FAT-YC (Martin et al., 2022a). Following all these modifications, a copy of the revised PLH-FAT-T was shared with lead assessors in Tanzania for their review and input. Fourth, several items were added, several items were removed, and minor adaptations were made to the wording of PLH-FAT-T items to suit the Tanzanian context and based on expert input. Once the changes were finalized, the revised tool (the revised PLH-FAT-T) was translated into Kiswahili for use by coaches to conduct assessments in the field (see [Supplementary File 2](#)). The revised PLH-FAT-T was comprised of 39 items (11 on the Activities Subscale and 28 on the Skills Subscale). Final PLH-FAT-T scores, represented as a percentage, were calculated out of a total possible score of 78 (39 items x 2 points per item).

#### **6.6.2 Intra- and Inter-Rater Reliability**

Intra- and inter-rater reliability results are shown in Tables 1 and 2 respectively. Using the revised tool, lead assessors had overall intra-rater reliability of 83.3% (ICC: 0.65,



95% CI: 0.50-0.76) and 76.9% (ICC: 0.65, 95% CI: 0.19-0.57) – both above the 70% threshold yet in the poor to moderate ICC range (Aspland & Gardner, 2003; Koo & Li, 2016). To determine whether assessors differed drastically in their ratings, intra-rater reliability was also examined where agreement was defined as the same score or one point away; in this case, the lead assessors reached 100.0% intra-rater agreement meaning that they did not, for instance, score a “0” on their first assessment of an item and then a “2” on their second. Analyses of the overall inter-rater reliability of each lead assessor with their respective group of coaches was 63.4% (ICC: 0.46, 95% CI: 0.40-0.51, range: 28.2-79.5%) and 69.3% (ICC: 0.41, 95% CI: 0.35-0.46, range: 51.3-97.4%) – both close to the 70% threshold yet in the poor range for ICCs (Aspland & Gardner, 2003; Koo & Li, 2016). When agreement was defined as the same score or one point away, the lead assessors reached 99.9% and 100.0% agreement indicating that it was rare that assessors dramatically differed in their assessments. Findings also indicated that intra- and inter-rater reliability were higher on the Skills Subscale than the Activities Subscale.

### **6.6.3 Construct Validity and Internal Consistency**

**Description of PLH-FAT-T data.** Descriptive statistics for the PLH-FAT-T items are included in [Supplementary File 3](#). The tables therein present item-level data on the missing data (with missing values for each item ranging from 0-17); mean response to each item including the standard deviation; and number of responses for each point on the scale (with some items never having been scored a ‘0’ and others being scored a ‘2’ in 79 out of 89 instances). Due to the small sample size of facilitator assessments and the large

percentage of missing data, EFAs were conducted with all available data. In other words, non-complete cases were utilized.

**Exploratory factor analyses (EFA).** The EFAs revealed one factor for the Home Activity Subscale, two factors for the Role-play Subscale, and one factor for the Skills Subscale. Shapiro-Wilk normality tests indicated that the PLH-FAT-T data was not normally distributed, which is to be expected as the PLH-FAT-T utilizes a three-point scale. Further, visual inspections of scatterplot matrices indicated non-linearity. As a result, EFAs were run using polychoric correlations rather than Pearson correlations (Watkins, 2018).

**Home activity subscale.** The Home Activity Subscale had a total of 43 observations across 11 items with 2.1% missing data. Two rounds of EFAs were conducted (see [Supplementary File 3](#)). Bartlett's test found sufficiently large correlations between items ( $\chi^2=346.65$ ,  $df=55$ ,  $p\text{-value}<0.001$ ) while the Kaiser-Meyer Olkin ( $KMO=0.61$ ) was above the minimum of 0.50 (Kaiser, 1974; Watkins, 2018). A parallel analysis, Velicer MAP test, and visual inspection of the scree plot revealed one factor. The two rounds of EFAs suggested that four items be removed (HA\_1, HA\_3, HA\_4, HA\_5) (see Table 3). The final proportion of variance explained was over the 50% threshold at 0.53. Internal consistency among items was strong ( $\alpha=0.81$ ,  $\omega=0.82$ ).

**Role-play subscale.** Assessments of the Role-play Subscale had a total of 43 observations across 11 items with 10.8% missing data. Two rounds of EFAs were conducted (see [Supplementary File 3](#)). Bartlett's test found sufficiently large correlations between items ( $\chi^2=272.42$ ,  $df=55$ ,  $p\text{-value}<0.001$ ) while the Kaiser-Meyer Olkin ( $KMO=0.53$ ) was just above the minimum of 0.50 (Kaiser, 1974; Watkins, 2018). A

parallel analysis, Velicer MAP test, and visual inspection of the scree plot revealed two factors. An oblique promax rotation was used (Watkins, 2018). Factor loadings suggest the removal of three items (RP\_6, RP\_7, RP\_11) (see Table 4). The proportion of variance explained was at 32.8% for factor 1 and 25.1% for factor 2 with a total cumulative variance above the 50% threshold at 0.58. Internal consistency among items was moderate to strong ( $\alpha=0.68$ ,  $\omega=0.76$ ).

**Skills subscale.** The Skills Subscale had a total of 95 observations across 28 items with 12.0% missing data. Three rounds of EFAs were conducted (see [Supplementary File 3](#)). Bartlett's test found sufficiently large correlations between items ( $\chi^2=1300.03$ ,  $df=378$ ,  $p\text{-value}<0.001$ ) while the Kaiser-Meyer Olkin ( $KMO=0.61$ ) was above the minimum of 0.50 (Kaiser, 1974; Watkins, 2018). A Velicer MAP test and visual inspection of the scree plot revealed one factor whereas a parallel analysis suggested two factors with one component therefore a one factor solution was employed. Factor loadings suggest the removal of twelve items (MB\_1, MB\_4, AECP\_1, AECP\_5, AECP\_6, CL\_6, CL\_7, CL\_11, CL\_12, CL\_13, CL\_14, CL\_15) (see Table 5). The proportion of variance explained was near the 50% threshold at 0.45. Internal consistency among items was strong ( $\alpha=0.87$ ,  $\omega=0.87$ ).

Table 1																
Lead Assessor Intra-Rater Reliability																
Overall Results				Assessment 1			Assessment 2			Skills Items			Activities Items			
Lead	%	ICC	95% CI	%	ICC	95% CI	%	ICC	95% CI	%	ICC	95% CI	%	ICC	95% CI	
Assessor																
1	83.3	0.66	0.50-0.76	79.5	0.55	0.28-0.73	87.2	0.75	0.57-0.86	87.5	0.73	0.59-0.84	72.7	0.47	0.07-0.75	
2	76.9	0.40	0.19-0.57	74.4	0.15	-0.17-0.44	79.5	0.55	0.29-0.44	82.1	0.50	0.28-0.68	63.6	0.17	-0.29-0.55	
Note: % refers to percentage agreement. ICC refers to intra-class correlation. The 95% CI is the 95% confidence interval for the ICC.																

<b>Table 2</b>										
<i>Lead Assessor-Coach Inter-Rater Reliability</i>										
Overall Results					Skills Items			Activities Items		
<u>Lead Assessor</u>	<u>%</u>	<u>ICC</u>	<u>95% CI</u>	<u>% Range</u>	<u>%</u>	<u>ICC</u>	<u>95% CI</u>	<u>%</u>	<u>ICC</u>	<u>95% CI</u>
1	63.4	0.46	0.40-0.51	28.2-79.5	63.8	0.44	0.35-0.52	62.4	0.53	0.38-0.64
2	69.3	0.41	0.35-0.46	51.3-97.4	72.1	0.43	0.36-0.49	62.4	0.37	0.25-0.47
<i>Note:</i> % refers to the average percentage agreement. ICC refers to intra-class correlation. The 95% CI is the 95% confidence interval for the ICC. % Range refers to the range of percentage agreements. Each row summarizes the results of 22 PLH-FAT-T assessments conducted by each lead assessor and 11 coaches.										

Table 3		
Exploratory Factor Analysis of Home Activity Items – Round Two		
Home Activity Item	Factor Loading	Item Wording The facilitator...
HA_2	0.772	Reviews the core building blocks from previous session with participants at the beginning of the discussion
HA_6	0.559	Helps participants connect their experiences to the core building blocks
HA_7	0.775	Identifies at least one specific challenge experienced by a participant regarding the main home activity
HA_8	0.872	Explores solutions to challenges shared and help participants choose a specific solution
HA_9	0.857	Practices the chosen solution with parents and teens
HA_10	0.655	Debriefs with the participants after practicing and encourages them to try the solution at home
HA_11	0.804	Thanks and praises participants for sharing experiences (at the end of the home activity discussion)
SS Loadings		3.69
Proportion Variance		0.53
Cronbach alpha		0.81
Omega		0.82
RMSEA Index		0.19
KMO		0.69
VSS	The Velicer MAP achieves a minimum of 0.09 with 1 factor	
Note: Bolded items indicate factor loadings above  0.5 . HA stands for home activity discussion.		

**Table 4***Exploratory Factor Analysis of Role-play Items – Round Two*

<u>Role-play Item</u>	<u>Factor 1 Loading</u>	<u>Factor 2 Loading</u>	<u>Item Wording</u> <i>The facilitator...</i>
RP_1	<b>0.752</b>	-	Provides introductory context to the role-play
RP_2	<b>0.934</b>	-	Acts out the role-play following the steps for leading the role-play (fidelity)
RP_3	<b>0.809</b>	0.171	Acts out the role-play following tips for leading the role-play (quality)
RP_4	<b>0.594</b>	-0.116	Discusses the role-play with participants after facilitators acted it out
RP_5	0.278	<b>0.599</b>	Explores possible solutions for negative role-plays
RP_8	0.181	<b>0.752</b>	Debriefs with participants about experiences and feelings after acting out the positive scenario
RP_9	-0.114	<b>0.607</b>	Discusses with participants about how the role-play relates to their lives
RP_10	-0.230	<b>0.819</b>	Connects the role-play to the building blocks of the session
SS Loadings	2.622	2.010	
Proportion Variance	0.328	0.251	
Cumulative Variance	0.325	0.579	
Cronbach alpha			0.68
Omega			0.76
RMSEA Index			0.19
KMO			0.62
VSS	The Velicer MAP achieves a minimum of 0.07 with 2 factors		
<i>Note: Bolded items indicate factor loadings above  0.5 . RP stands for role-play.</i>			

**Table 5***Exploratory Factor Analysis of Skills Items – Round Three*

<u>Skills Items</u>	<u>Factor Loading</u>	<u>Item Wording</u>
MB_2	<b>0.759</b>	Gives positive, specific, and realistic instructions
MB_3	<b>0.588</b>	Maintains commitments to time management principles
MB_5	<b>0.703</b>	Demonstrates respectful behaviour towards participants
AECP_2	<b>0.655</b>	Accepts parent responses verbally
AECP_3	<b>0.648</b>	Explores participant experiences and opinions using open-ended questions
AECP_4	<b>0.579</b>	Explores thoughts and feelings
AECP_7	<b>0.555</b>	Connects experiences to the building blocks the from session
AECP_8	<b>0.575</b>	Identifies opportunities to practice skills (in addition to the structured group practice)
CL_1	<b>0.555</b>	Arranges the room in a way that encourages equal and active participation
CL_2	<b>0.893</b>	Facilitator is situated within the group, is at the level of the participants, and in a different place than the co-facilitator
CL_3	<b>0.735</b>	Participants appear comfortable and engaged in the session
CL_4	<b>0.762</b>	Participant-facilitator speaking ratio
CL_5	<b>0.752</b>	Assures equal and active participation among participants
CL_8	<b>0.551</b>	Keeps participants focused on the topic of discussion
CL_9	<b>0.639</b>	Demonstrates knowledge of session content
CL_10	<b>0.707</b>	Delivers the session with confidence
SS Loadings		7.24
Proportion Variance		0.45
Cronbach alpha		0.87
Omega		0.87
RMSEA index		0.11
KMO		0.78
VSS	The Velicer MAP achieves a minimum of 0.03 with 1 factor	
<i>Note: Bolded items indicate factor loadings above  0.5 . MB stands for modelling behaviour, AECP stands for Accept-Explore-Connect-Practice, and CL stands for collaborative leadership.</i>		



#### 6.6.4 Level of Competent Adherence using PLH-FAT-T

Using PLH-FAT-T, facilitators achieved a mean total competent adherence score of 79.9% (see Table 6). Facilitators achieved higher scores on the Skills Subscale (items assessing their competence in delivering program components) than on the Activity Subscale (items assessing their adherence to the program manual).

<b>Table 6</b> <i>Summary of Level of Competent Adherence Achieved by PLH-Teens Facilitators using the PLH-FAT-T</i>			
	<u>Total</u> <u>(N=66)</u>	<u>Individual %</u> <u>Activities</u> <u>(N=74)</u>	<u>Skills</u> <u>(N=71)</u>
Mean (SD)	79.9 (11.5)	74.7 (17.6)	81.8 (12.8)
Median	81.0	77.0	82.0
Scores	50.0-99.0	14.0-100.0	43.0-100.0
<i>Note:</i> This table presents the assessment results for facilitators for whom complete PLH-FAT-T assessments (no missing data) were received. Thus, a complete case analysis was conducted for which there were no missing values on any of the items involved in the computation.			

#### 6.6.5 Analyses with the PLH-FAT-T Short Form

The EFA results were used to produce the PLH-FAT-T Short Form comprised of 19 fewer items with a total of seven home activity items, eight role-play items, and 16 skills items (see Table 7).

**Table 7**  
*PLH-FAT-T Short Form*

<u>Item</u>	<u>Subscale</u>	<u>Construct</u>	<u>Item Wording</u>
HA_2	Home Activities	Home Activity Discussion and Practice	Reviews the core building blocks from previous session with participants at the beginning of the discussion
HA_6	Home Activities	Home Activity Discussion and Practice	Helps participants connect their experiences to the core building blocks
HA_7	Home Activities	Home Activity Discussion and Practice	Identifies at least one specific challenge experienced by a participant regarding the main home activity
HA_8	Home Activities	Home Activity Discussion and Practice	Explores solutions to challenges shared and help participants choose a specific solution
HA_9	Home Activities	Home Activity Discussion and Practice	Practices the chosen solution with parents and teens
HA_10	Home Activities	Home Activity Discussion and Practice	Debriefs with the participants after practicing and encourages them to try the solution at home
HA_11	Home Activities	Home Activity Discussion and Practice	Thanks and praises participants for sharing experiences (at the end of the home activity discussion)
RP_1	Role-play Activity	Role-play Steps	Provides introductory context to the role-play
RP_2	Role-play Activity	Role-play Steps	Acts out the role-play following the steps for leading the role-play (fidelity)
RP_3	Role-Play Activity	Role-play Steps	Acts out the role-play following tips for leading the role-play (quality)
RP_4	Role-play Activity	Role-play Steps	Discusses the role-play with participants after facilitators acted it out
RP_5	Role-play Activity	Role-play Support	Explores possible solutions for negative role-plays
RP_8	Role-play Activity	Role-play Support	Debriefs with participants about experiences and feelings after acting out the positive scenario
RP_9	Role-play Activity	Role-play Support	Discusses with participants about how the role-play relates to their lives
RP_10	Role-play Activity	Role-play Support	Connects the role-play to the building blocks of the session
MB_2	Skills	Empowering and Experiential Facilitation	Gives positive, specific, and realistic instructions

MB_3	Skills	Empowering and Experiential Facilitation	Maintains commitments to time management principles
MB_5	Skills	Empowering and Experiential Facilitation	Demonstrates respectful behaviour towards participants
AECP_2	Skills	Empowering and Experiential Facilitation	Accepts parent responses verbally
AECP_3	Skills	Empowering and Experiential Facilitation	Explores participant experiences and opinions using open-ended questions
AECP_4	Skills	Empowering and Experiential Facilitation	Explores thoughts and feelings
AECP_7	Skills	Empowering and Experiential Facilitation	Connects experiences to the building blocks from the session
AECP_8	Skills	Empowering and Experiential Facilitation	Identifies opportunities to practice skills (in addition to the structured group practice)
CL_1	Skills	Empowering and Experiential Facilitation	Arranges the room in a way that encourages equal and active participation
CL_2	Skills	Empowering and Experiential Facilitation	Facilitator is situated within the group, is at the level of the participants, and in a different place than the co-facilitator
CL_3	Skills	Empowering and Experiential Facilitation	Participants appear comfortable and engaged in the session
CL_4	Skills	Empowering and Experiential Facilitation	Participant-facilitator speaking ratio
CL_5	Skills	Empowering and Experiential Facilitation	Assures equal and active participation among participants
CL_8	Skills	Empowering and Experiential Facilitation	Keeps participants focused on the topic of discussion
CL_9	Skills	Empowering and Experiential Facilitation	Demonstrates knowledge of session content
CL_10	Skills	Empowering and Experiential Facilitation	Delivers the session with confidence

The PLH-FAT-T facilitator assessments were then re-analyzed using the PLH-FAT-T Short Form data to investigate whether intra- and inter-rater reliability or the level of competent adherence changed. Analyses of the overall intra-rater reliability of each lead assessor using the PLH-FAT-T Short Form data was 63.4% (ICC: 0.46, 95% CI: 0.40-0.51) and 85.1% (ICC: 0.45, 95% CI: 0.19-0.65) (see Table 8). Both assessors had higher percentage agreements and ICCs on the PLH-FAT-T Short Form, yet one assessor had a percentage agreement below the 70% threshold and both had ICCs in the poor to moderate range. Analyses of the overall inter-rater reliability of each lead assessor with their respective group of coaches was 65.1% (ICC: 0.47, 95% CI: 0.39-0.53) and 66.2% (ICC: 0.30, 95% CI: 0.22-0.38) (see Table 9). The inter-rater reliability was higher for one lead assessor using the PLH-FAT-T Short Form but was lower for the other lead assessor. Both assessors had percentage agreements below 70% and ICCs in the poor range.

Using the PLH-FAT-T Short Form data, facilitators achieved a mean total competent adherence score of 82.3% total (see Table 10). Thus, facilitators achieved higher scores on the PLH-FAT-T Short Form than on the revised PLH-FAT-T. As with the revised PLH-FAT-T, facilitators achieved higher scores on the Skills Subscale than on the Activity Subscale.

<b>Table 8</b>															
<i>Lead Assessor Intra-Rater Reliability with PLH-FAT-T Short Form</i>															
Overall Results				Assessment 1			Assessment 2			Skills Items			Activities Items		
<u>Lead</u>	<u>%</u>	<u>ICC</u>	<u>95% CI</u>	<u>%</u>	<u>ICC</u>	<u>95% CI</u>	<u>%</u>	<u>ICC</u>	<u>95% CI</u>	<u>%</u>	<u>ICC</u>	<u>95% CI</u>	<u>%</u>	<u>ICC</u>	<u>95% CI</u>
<u>Assessor</u>															
1	87.2	0.74	0.58-0.85	78.3	0.57	0.23-0.79	95.8	0.92	0.82-0.96	93.8	0.87	0.74-0.93	73.3	0.48	-0.50-0.79
2	85.1	0.45	0.19-0.65	78.3	-0.08	-0.45-0.33	91.7	0.76	0.52-0.89	90.6	0.63	0.36-0.80	73.3	0.20	-0.32-0.63
<i>Note: % refers to percentage agreement. ICC refers to intra-class correlation. The 95% CI is the 95% confidence interval for the ICC.</i>															

<b>Table 9</b>									
<i>Lead Assessor-Coach Inter-Rater Reliability with PLH-FAT-T Short Form</i>									
Overall Results				Skills Items			Activities Items		
<u>Lead Assessor</u>	<u>%</u>	<u>ICC</u>	<u>95% CI</u>	<u>%</u>	<u>ICC</u>	<u>95% CI</u>	<u>%</u>	<u>ICC</u>	<u>95% CI</u>
1	65.1	0.47	0.39-0.53	67.3	0.41	0.32-0.50	60.6	0.51	0.38-0.61
2	66.2	0.30	0.22-0.38	70.3	0.29	0.19-0.38	57.6	0.30	0.16-0.43
<i>Note: % refers to the average percentage agreement. ICC refers to intra-class correlation. The 95% CI is the 95% confidence interval for the ICC. Each row summarizes the results of 22 PLH-FAT-T Short Form assessments conducted by each lead assessor and 11 coaches.</i>									

**Table 10**

*Summary of Level of Competent Adherence Achieved by PLH-Teens Facilitators using the PLH-FAT-T Short Form*

	<u>Individual %</u>			
	<u>Total</u> (N=68)	<u>Home Activities</u> (N=41)	<u>Role-play</u> (N=37)	<u>Skills</u> (N=71)
Mean	82.3	72.5	77.5	85.9
(SD)	(13.2)	(22.79)	(15.5)	(13.0)
Median	84.1	78.6	75.0	87.5
Scores	32.6-100.0	14.29-100	43.8-100.0	28.1-100.0
<i>Note:</i> This table presents the assessment results for facilitators for whom complete PLH-FAT-T assessments (no missing data) were received. Thus, a complete case analysis was conducted for which there were no missing values on any of the items involved in the computation.				

## 6.7 Discussion

This study contributes to the improvement of PLH programmes as it is the first psychometric evaluation of the PLH-FAT-T, a tool used to assess the delivery of PLH-Teens facilitators in multiple LMICs. Given that the PLH-FAT-T is used for purposes including to inform decisions regarding the certification of PLH facilitators, to determine the quality of delivery in multiple contexts, and for coaches to provide facilitators with supportive supervision, it is critical that the reliability and validity of the tool be established. The paper also contributes to the sparse literature on the psychometric properties of competent adherence measures used in parenting programs; a small number of studies report on construct validity, content validity, and intra-rater reliability (Martin et al., 2021b). Further, this study contributes to both PLH and the parenting program literature as it is the first to report on the competent adherence with which community facilitators delivered a parenting program at scale in a community setting at scale in a LMIC.

### **6.7.1 Reliability and Validity of the PLH-FAT-T**

The content validity process used to modify the original PLH-FAT-T drew upon insights from a study on the PLH-FAT-YC (Martin et al., 2022a) as well as on feedback from PLH-Teens stakeholders. The content validity process was undertaken to ensure that PLH-Teens stakeholders viewed the revised tool as understandable, comprehensive, and relevant (Terwee et al., 2018). From the perspective of the PLH-Teens stakeholders who were consulted, the revised PLH-FAT-T was made more practical to use and suitable for the context of program delivery in Tanzania.

The intra- and inter-rater reliability results from both the revised PLH-FAT-T and PLH-FAT-T Short Form indicate that assessors achieved poor to moderate reliability, with assessors achieving higher reliability on the skills items than on the activity items. A potential explanation for this finding is that assessors found the skills items easier to evaluate. However, assessing competence is typically considered more challenging than assessing adherence (Mowbray et al., 2003). Although some reliability scores were below the expected thresholds, percentage agreements were above or close to 70%. Further, the findings from both versions of the PLH-FAT-T are well within the range of reliability results found in the broader literature; a systematic review of measures of competent adherence used in parenting programs found that intra-rater reliability ranged from 60.0-87.0% and that inter-rater reliability ranged from 50.0-100.0% (Martin et al., 2021b). Intra- and inter-rater reliability may have been somewhat low in the present study due to two factors that potentially limited the extent to which coaches understood the PLH-FAT-T and interpreted it in the same way as their colleagues. First, there were challenges conducting

coach training and using training materials translated from English into Kiswahili. Second, coaches were provided with minimal training including only one practice assessment.

Although intra- and inter-rater reliability were poor to moderate, the PLH-FAT-T Short Form had high internal consistency (Cronbach alphas and omegas greater than 0.70). This finding indicates that responses to the items in the tool were statistically interrelated. The skills items had particularly high internal consistency meaning that these items are likely to be asking about similar concepts. The strong internal consistency indicates that the PLH-FAT-T Short Form is performing well in the context of the wider literature as a systematic review found that only just over half of studies reporting on the internal consistency of competent adherence measures reported Cronbach alphas and omegas greater than the 0.70 threshold (Martin et al., 2021b).

The construct validity analyses suggest the revised PLH-FAT-T tested in the first EFA performed poorly. However, iterative EFAs found that the PLH-FAT-T Short Form had stronger psychometric properties except for the inter-rater reliability of one of the two lead assessors. The shortened version also has 19 fewer items, which may improve the feasibility of conducting assessments. With respect to the Home Activity Subscale, seven items remained which appear to capture one underlying construct. This construct can be described as “Home Activity Discussion and Practice” as the remaining questions all relate to facilitators supporting parents to discuss the home activities and practice as a group. Regarding the Role-play Subscale, eight items remained which appear to capture two underlying constructs of four items each. The first four items can be described as “Role-play Steps” as these items capture tasks to prepare parents to conduct the role-play whereas the second four items can be described as “Role-play Support” as these items capture tasks



to support parents to engage in and reflect on the role-play. Regarding the Skills Subscale, three rounds of factor analyses resulted in a final 16 items capturing one construct. This construct can be described as “Empowering and Experiential Facilitation” as these items relate to the soft skills necessary of facilitators to empower parents to learn and reflect as well as ensure that parents positively engage by practicing the key lessons learned.

### **6.7.2 Facilitator Competent Adherence at Scale**

The analyses of the revised PLH-FAT-T and PLH-FAT-T Short Form results indicate that facilitators appear to have delivered PLH-Teens to a high level of competent adherence at scale despite substantial implementation challenges (e.g., COVID-19 disruptions). This finding is promising for the continued scale-up of PLH-Teens as the program uses facilitators with limited background in and training on parenting programs – which is commonly the case for family interventions in LMICs where there are shortages of professional staff and limited funds to pay for highly-trained professionals (Tomlinson et al., 2017).

Although facilitator competent adherence was found to be high, findings using both the revised PLH-FAT-T and PLH-FAT-T Short Form demonstrate that facilitators had better competence than adherence. The latter finding contrasts with the perspective of some researchers that a facilitator cannot be competent without being adherent as the findings suggest that facilitators were able to implement the interventions components they were able to deliver with competence (Muse & McManus, 2013; Waltz et al., 1993). That adherence was lower than competence may mean that facilitators were not able to deliver all intervention components during a given program session (e.g., due to time constraints) and/or made responsive adaptations to the way activities were delivered to suit participant needs, while maintaining a high level of quality to the components that were delivered.

### 6.7.3 Limitations

This study has several limitations related to the quantity and quality of data collected and the potential impact of several sources of bias on the measurement of competent adherence. As this study used data from ‘real-world’ program delivery at scale during a global pandemic, practical challenges impacted the quantity and quality of data collected by implementing partners. Only a small number of coaches were involved in establishing reliability. As a result, the examination of intra- and inter-rater reliability was limited to assessments conducted by two lead assessors and 22 coaches. Ideally, intra- and inter-rater reliability analyses would have been conducted with the entire sample. Further, only a small convenience sample of 95 assessments was collected. The small sample means that the sample’s representativeness of the larger population of facilitators who delivered the program is unknown. The small sample of assessments also limited the amount of data points for analyses in the psychometric evaluation. The quality of the data was limited in that there was a noteworthy amount of missing item-level data on some of the facilitator assessments collected – again, likely due to challenges in coordinating assessments at scale during a multi-year global pandemic.

The results may have been affected by several sources of bias. As coaches selected which facilitators to assess, it is possible that the 95 assessments collected were from the best or worst facilitators thereby inflating or deflating the level of competent adherence achieved by facilitators. This potential source of bias is likely reduced by the fact that coaches had to travel significant distances to conduct facilitator assessments so may have chosen facilitators to assess based on proximity. Facilitators were assessed by their coaches, which could also present a source of bias due to their pre-existing relationship (Walton et

al., 2017). Facilitators may also have been affected by reactivity bias in that they may have delivered the program differently due to being observed (Girard & Cohn, 2016). Lastly, as facilitators were only assessed at one timepoint, the degree to which any given assessment is representative of a facilitator's overall performance is unknown. Future research would benefit from the collection of facilitator assessments at different timepoints to examine fluctuation over time.

#### **6.7.4 Clinical Implications and Future Research**

Results suggest three ways to improve the psychometric properties of the PLH-FAT-T Short Form and its application in future research and practice. First, steps need to be taken to achieve stronger levels of intra- and inter-rater reliability since the challenges experienced in achieving reliability in Tanzania are likely to be experienced in other contexts. These steps include increasing assessor training time, conducting more practice assessments during assessor training, further refining item wording in collaboration with assessors to enhance their understanding, and ensuring that item meaning is conveyed as intended in the language of the assessors (back translation). Second, the PLH-FAT-T Short Form should be considered for adoption wherever PLH-Teens is delivered in future as the PLH-FAT-T's functionality was improved using a shortened version. To do so, this research and the PLH-FAT-T Short Form should be reviewed by stakeholders responsible for the quality assurance of PLH-Teens and a decision made on whether to adopt the shortened tool. In the meantime, it is recommended that the PLH-FAT-T Short Form be used for future research on competent adherence of facilitators in Tanzania (e.g., when examining associations between competent adherence and outcomes). Third, the construct validity of the PLH-FAT-T Short Form should be tested using a confirmatory factor

analysis and with a larger sample to determine whether the EFA results are replicable. This analysis could be conducted in other countries where PLH-Teens is delivered to ascertain the suitability of the tool for other countries, cultures, and contexts.

## **6.8 Conclusion**

This paper reports on the first psychometric evaluation of the PLH-FAT-T and provides crucial information about the competent adherence with which facilitators delivered a parenting program at scale. The results indicate that while the PLH-FAT-T Short Form needs further work, it is a promising tool for measuring the competent adherence of PLH-Teens facilitators. The findings regarding the overall reliability and validity of the tool are encouraging considering the practical challenges experienced with routine delivery at scale during a pandemic. In addition, the findings suggest several ways that the tool can be improved to increase its reliability and validity.

The results also suggest that PLH-Teens was delivered to a high level of competent adherence in a resource-poor context by community facilitators who received five days of training and who did not have backgrounds in delivering parenting programs. This finding is supportive of future scale-up of parenting programs in LMICs using community facilitators.

By contributing to the scant literature on facilitator competent adherence in LMICs during routine service delivery at scale as well as on the psychometric properties of measures of competent adherence in low-income community-settings, this paper contributes to the overall aim of increasing access to high-quality delivery of evidence-based parenting programs to those in need.

## **7. Paper 3 - The predictive validity of a measure of competent adherence: Evaluating the role of fidelity in adolescent and parent outcomes using routine data collected during parenting program delivery at scale**

This paper is close to submission to an academic journal. This paper has been written with US spelling in alignment with the journal's submission guidelines. Please find the tables throughout the chapter and the supplementary materials in [Appendix 9](#). Note that several of the methodological choices made in conducting this study are expanded upon herein but will not be included in the submitted paper.

Authors: Martin, M.<sup>1</sup>, Lachman, J. M.<sup>1,2,3</sup>, Calderon, F.<sup>1</sup>, Han, Q.<sup>1</sup>, Shenderovich, Y.<sup>4,5</sup>, Gardner, F.<sup>1</sup>, Nydetabura, E.<sup>6</sup>, Manjengenga, N.<sup>7</sup>, & Wamoyi, J.<sup>8</sup>

<sup>1</sup>Department of Social Policy and Intervention, University of Oxford, United Kingdom

<sup>2</sup>MRC Social and Public Health Sciences Unit, University of Glasgow, United Kingdom

<sup>3</sup>Centre for Social Sciences Research, University of Cape Town, South Africa

<sup>4</sup>Wolfson Centre for Young People's Mental Health, Cardiff University, United Kingdom

<sup>5</sup>Centre for the Development and Evaluation of Complex Interventions for Public Health Improvement, School of Social Sciences, Cardiff University, United Kingdom

<sup>6</sup>Pact Tanzania, United Republic of Tanzania

<sup>7</sup>Clowns Without Borders South Africa, South Africa

<sup>8</sup>Department of Sexual and Reproductive Health, National Institute for Medical Research, United Republic of Tanzania

**My role:** I conceptualised of the study, prepared the data, planned and conducted the analyses, drafted the manuscript, and revised the manuscript. Other authors contributed to initial conceptualisation and support for the analyses.

### **7.1 Abstract**

There is limited and mixed evidence on the relationship between facilitator competent adherence and the outcomes of parents/caregivers and children participating in parenting programs aiming to reduce violence against children and child behavioral and emotional problems. The evidence is particularly limited from program delivery in low- and middle-countries, at scale, and in community settings. To contribute to the literature, this study examined the predictive validity of an observational measure used to assess the

competent adherence of facilitators delivering the Parenting for Lifelong Health for Parents and Adolescents (PLH-Teens) program in Tanzania. The study used data collected by implementing partners during the 2020-2021 scale-up of PLH-Teens to 75,061 beneficiaries in rural and semi-urban Tanzania. To examine the relationship between competent adherence and outcomes, multi-level Poisson regressions were conducted. A total of 24 facilitator assessments could be linked with pre-post surveys from 3,057 families. Analyses found mixed results, with competent adherence having positive, negative, and insignificant associations with participant outcomes. As a result, the observational measure used in PLH-Teens was not found to demonstrate consistent predictive validity across multiple outcomes. As competent adherence was positively associated with some participant outcomes, including the primary outcome of interest (child maltreatment) according to adolescent-reports, competent adherence appears beneficial for participants to some extent. There are a variety of potential explanations for the inconsistent results which suggest avenues for future research. Exploration of these avenues would benefit the ongoing delivery and dissemination of Parenting for Lifelong Health programmes as well as the broader parenting program literature as these communities strive to maximize the ability of vulnerable children and families globally to benefit from evidence-based parenting programs.

*Key Words:* fidelity, parenting programs, predictive validity, scale-up

## **7.2 Background**

Current thinking in the implementation science literature is that evidence-based interventions should be delivered with integrity to the models tested via randomized trials in order to maintain their effectiveness (Allen et al., 2012). The role of fidelity is theorized

in several frameworks and models including those developed by Berkel et al. (2011) and Carroll et al. (2007), which conceptualize of participant outcomes as a function of the degree to which an intervention is delivered with fidelity. Although there are many types of fidelity (Proctor et al., 2011), two types thought to play a key role are facilitator adherence and quality of delivery – or ‘competent adherence’ (Breitenstein et al., 2010b; Forgatch et al., 2005). Competent adherence – the accuracy and style with which a facilitator delivers program components – is assumed to be directly associated with participant outcomes as facilitators are the vehicle through which participants receive, or do not receive, planned intervention components (Petersilia, 1990). Several reviews of the relationship between facilitator delivery and outcomes in the broader intervention literature suggest that facilitators play a key role in the achievement of participant outcomes (e.g., Durlak & DuPre, 2008; Hill & Erickson, 2019; Naylor et al., 2015; Wilson et al., 2003).

Despite extensive randomized trial evidence that parenting programs are beneficial for the health and well-being of children and their parents/caregivers (parents) (Barlow & Coren, 2018; Buchanan-Pascall et al., 2018; Gardner et al., 2023; Jeong et al., 2021; Knerr et al., 2013; McCoy et al., 2020; WHO, 2023), there is limited and mixed evidence on the role facilitator competent adherence plays in achieving participant outcomes. Several primary studies have found a positive relationship. For instance, a study on Incredible Years parenting program found a significant and positive relationship between competent adherence and positive parenting, which was also associated with improvements in child behavior (Eames et al., 2009). Similarly, several papers by Forgatch et al. on Parenting Management Training-Oregon Model found better competent adherence to be associated with better parenting skills (Forgatch & DeGarmo, 2011; Forgatch et al., 2005). Further, a

systematic review by Furlong et al. (2012) found that better facilitator fidelity was associated with more positive intervention effects. A synthesis without meta-analysis by Martin et al. (2023) however, found that while most studies reported at least one positive associations between observational measures of competent adherence and outcomes, the literature was inconsistent with many studies finding no significant relationship and with some studies finding mixed results. A study on Parenting for Lifelong Health for Parents and Adolescents programme (PLH-Teens) in South Africa by Shenderovich et al. (2019) did not find significant associations between facilitator competent adherence and outcomes, with some models finding a negative relationship ( $N=270$  families and 25 facilitators). A study of the delivery of Parent-Child Interaction Therapy did not find that higher competent adherence was associated with improved parenting or child behaviour ( $N=32$  families and 17 facilitators) (Snider, 2019). A study of Familias Unidas by St. George et al. (2016) found mixed results with better competence associated with reductions in adolescent substance use, better adherence not significantly associated with adolescent reductions in substance use, and neither competence nor adherence significantly associated with improvements in family functioning ( $N=365$  families).

### **7.2.1 Current Study**

This study contributes to the inconsistent literature by examining the role of facilitator competent adherence on participant outcomes using data from the delivery of an evidence-based parenting program in community settings in a low-income country at scale. In investigating the scale-up of PLH-Teens by implementing organizations in Tanzania, this study is one of the first to examine the role of competent adherence in a low- and middle-income country (LMIC) at scale via routine delivery (Martin et al., 2021b; Martin et al.,



2023). Evaluating the impact of facilitator competent adherence in a low-income context makes an important contribution to the parenting program literature as children in LMICs suffer greatly from the burden of violence against children (Hillis et al., 2016; Stoltenborgh et al., 2013; UNICEF, 2010). To investigate the role of competent adherence in advancing child and family outcomes, the paper uses data collected by implementing partners to examine the predictive validity of the observational measure of competent adherence, the PLH-Teens Facilitator Assessment Tool or PLH-FAT-T, used to assess community facilitators delivering PLH-Teens. Versions of the PLH-FAT-T are used in multiple LMICs for purposes including certification, assessment of the quality of program delivery, and facilitator feedback on their delivery. Due to the PLH-FAT-T's use in practice, this first study on its predictive validity will contribute to an understanding of the role facilitator competent adherence plays in achieving PLH-Teens participant outcomes, which understanding can then be used to enhance the tool and programme outcomes for families.

## **7.3 Methods**

### **7.3.1 Intervention**

PLH-Teens is a low-cost, open-access parenting program developed by Parenting for Lifelong Health (PLH) to reduce violence against children and child behavioural and emotional problems among families in LMICs (Cluver et al., 2018). To date, PLH programs have been delivered to hundreds of thousands of beneficiaries in more than 35 LMICs (Shenderovich et al., 2020). Based on behavior change principles and social learning theory, PLH-Teens is delivered to groups of parents and adolescents over 14 weekly sessions by trained facilitators from the local community, with each session being approximately three hours in length. Ten of the 14 sessions are delivered to groups of

parent-adolescent dyads and four of the 14 sessions are delivered to groups of adolescents and parents separately. The program is delivered by facilitators using a participatory, non-didactic, and empowering approach. Program topics include spending time together, praising each other, communicating about emotions, dealing with conflict through problem-solving and positive discipline, family budgeting and saving, and responding to crisis situations such as sexual violence. A cluster-randomized trial ( $N=40$  clusters, 552 families, 25 facilitators) of PLH-Teens in South Africa found reductions in child maltreatment (e.g., corporal punishment, abuse) and improvements in positive parenting, involvement, and supervision (Cluver et al., 2018). Since its initial testing in South Africa, PLH-Teens has been disseminated to 19 LMICs, mainly in sub-Saharan Africa.

### **7.3.2 Study Setting and Sample**

This paper used data collected from the large-scale delivery of PLH-Teens by Pact Tanzania and five local implementing partner organizations in eight districts of rural and semi-urban Tanzania (Martin et al., 2021a). In 2020 and 2021 during the COVID-19 pandemic, PLH-Teens was delivered in-person to 75,061 participants ( $n=36,259$  parents and  $n=38,802$  adolescent girls) in schools and community centres over three waves of implementation. The program was implemented as part of USAID's Kizazi Kipya initiative aiming to support the health and well-being of a million adolescent girls affected by HIV in sub-Saharan Africa. Program sessions were delivered by 444 community facilitators (community health workers and school teachers) who received five days of facilitator training from Clowns Without Borders South Africa (CWBSA), a non-profit organization that provides capacity-building for the delivery of PLH programs in Africa.

### **7.3.3 Data Collection Procedures**

Two types of secondary data collected by implementing partner organizations as part of routine monitoring and evaluation were used in this study: 3,057 pre-post surveys from participant families and 24 facilitator assessments. All data was collected and anonymized by implementers before being shared with the research team. The pre-post surveys measured a wide range of outcomes – child maltreatment, child conduct problems, child emotional problems, positive parental involvement, poor supervision, parenting stress, acceptability of corporal punishment, depression, financial insecurity, parental support of education, intimate partner violence (IPV) perpetration, IPV victimization, sexual health communication, and school violence. For the purposes of this analysis, the primary outcome of interest was child maltreatment as this is the main objective of PLH-Teens (Martin et al., 2021a).

### **Sample Size**

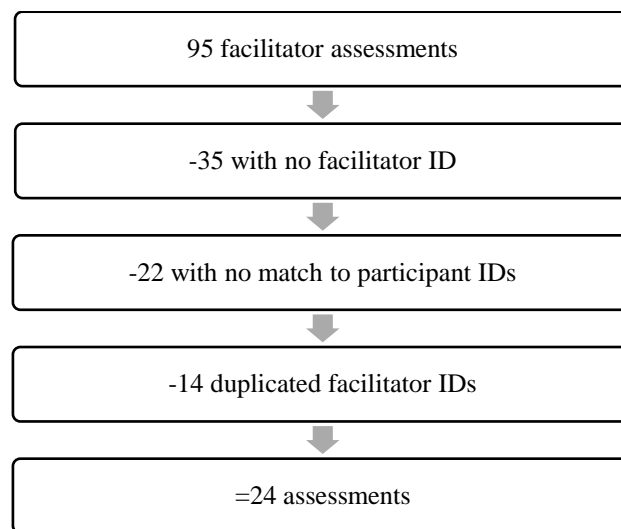
The secondary data used in this paper had limitations due to the difficulty of collecting and cleaning real-world data on such a large scale in a low-income setting during a global pandemic. As a result, this paper used the ‘flawed, uncertain, proximate and sparse’ data available (Wolpert & Rutter, 2018). A total of 67,456 parent pre-post surveys and 73,358 adolescent pre-post surveys were collected by implementing partner organizations. However, only 24,863 surveys were useable due to a variety of issues which surfaced during data cleaning. Reasons for excluding data included substantial amounts of missing data; participants with more than one pre-test or post-test survey; parent and adolescent survey data which did not match; implausible survey responses; and survey answers wherein participants answered “0” to every question, which was interpreted as an

error in data capturing. Many of the challenges with data collection were a function of the difficulty of conducting such a large number of surveys in a low-income setting at scale, including issues with paper-based data collection, inadequate allocation of identification numbers, lost forms, and data entry errors. Of 24,863 surveys which were useable, data from 3,057 families could be linked with the facilitator assessments collected by implementing partner organizations.

Regarding the facilitator assessments, a total of 95 facilitator assessments were collected by coaches during program delivery. However, only 24 of these assessments could be linked to pre-post surveys. Data from the 95 assessments were lost in three ways – 35 assessments did not have a facilitator identification number to link assessments to participant identification numbers; 27 did not have a facilitator identification number that matched participant identification numbers; and 14 facilitator identification numbers were duplicated. The loss of facilitator data is summarized in Figure 1.

**Figure 1**

*Loss of Facilitator Assessment Data During Merging with Pre-Post Surveys*



### 7.3.4 Measures

**Facilitator Competent Adherence.** PLH developed an observational assessment tool, the PLH-Teens-Facilitator Assessment Tool or PLH-FAT-T, for a number of reasons including to monitor facilitator competent adherence as part of program dissemination of PLH-Teens across multiple settings. In this paper, PLH-FAT-T scores were produced using the PLH-FAT-T Short Form which came to fruition as a result of a psychometric evaluation of the tool, which found that the shortened version had stronger psychometric properties (Martin et al., in submission-a). Using the tool, facilitators are assessed on their delivery of one of two program activities. The two program activities are the home activity discussion (conversation led by the facilitator to review and discuss the assigned home practice activities; 7 items) and the role-play activity (facilitator-supported exercise to support participants in practicing key skills; 8 items). In addition to being assessed on either the home activity discussion or the role-play activity, facilitators are assessed on an additional 16 skills items related to their ability to model key parenting skills (3 items), use PLH's 'Accept-Explore-Connect-Practice' technique (5 items), and demonstrate collaborative leadership skills (8 items). Each of the PLH-FAT-T Short Form items are rated using a three-point Likert scale ranging from zero to two (*0=inadequate, 1=good, 2=excellent*). Final PLH-FAT-T Short Form scores, represented as a percentage, are calculated out of a total possible score of 46 for facilitators assessed on the skills and home activity discussion (23 items x 2 points per item) or out of a total possible score of 48 for facilitators assessed on the skills and role-play items (24 items x 2 points per item).

**Pre-Post Surveys.** Implementing partner organizations administered pre-post surveys to parents who provided consent as well as to adolescents who provided assent and

received parental consent. The pre-test survey was administered during program registration, and the post-test survey during the final program session. A summary of the pre-post instruments used to assess the primary (child maltreatment) and secondary outcomes (child conduct problems, child emotional problems, positive parental involvement, poor supervision, parenting stress, acceptability of corporal punishment, depression, financial insecurity, parental support of education, sexual health communication, IPV perpetration, IPV victimization, and school violence) incorporated in the present analyses are included in Table 1. Implementing partners selected the instruments used to measure outcomes based on advice from CWBSA and PLH. The instruments were recommended as they are open-access and psychometrically validated. As the delivery of PLH-Teens was reaching so many participants, implementers chose to shorten the instruments. The choice to abbreviate the instruments may have limitations for the conclusions that can be drawn herein. More information on the measures can be found in [Supplementary File 3](#).

<b>Table 1</b> <i>Primary and Secondary Outcomes Administered Using Pre-Post Surveys</i>					
<u>Outcome</u>	<u>Type</u>	<u>Measure</u>	<u>Items</u>	<u>Unit Measure</u>	<u>Report</u>
Child Maltreatment	Primary	ISPCAN Child Abuse Screening Tools-Trial Version (Meinck et al., 2018)	4	0 to 7 and 8 for more than or equal to 8 in the past 4 weeks	Parent and adolescent
Child Conduct Problems	Secondary	Strengths and Difficulties Questionnaire (Goodman, 1997)	5	0 = Not true, 1 = Somewhat true, 2 = Very true	Parent and adolescent

Child Emotional Problems	Secondary	Strengths and Difficulties Questionnaire (Goodman, 1997)	4	0 = Not true, 1 = Somewhat true, 2 = Very true	Adolescent
Positive Parental Involvement	Secondary	Alabama Parenting Questionnaire (Frick, 1991)	3	0 = Never, 1 = Almost never, 2 = Sometimes, 3 = Often, 4 = Always	Parent and adolescent
Poor Supervision	Secondary	Alabama Parenting Questionnaire (Frick, 1991)	3	0 = Never, 1 = Almost never, 2 = Sometimes, 3 = Often, 4 = Always	Parent and adolescent
Parenting Stress	Secondary	Parental Stress Scale (Berry & Jones, 1995)	2	0 = Strongly disagree, 1 = Disagree, 2 = Neutral, 3 = Agree, 4 = Strongly agree	Parent
Acceptability of Corporal Punishment	Secondary	Multiple Indicator Cluster Survey (UNICEF, 2022)	1	0 = Strongly disagree, Disagree, and Not sure; 1 = Agree and Strongly agree	Parent and adolescent
Depression	Secondary	Centre for Epidemiologic Studies Depression Scale (CES-D 10) (Irwin et al., 1999)	3	0 = Rarely or none of the time, 1 = Some or a little of the time, 2 = Occasionally or a moderate amount of time, 3 = Most of the time	Parent and adolescent

Financial insecurity	Secondary	Family Financial Coping Scale (Shenderovich et al., 2020)	2	0 = Never, 1 = Rarely, 2 = Sometimes, 3 = Often	Parent and adolescent
Parental Support of Education	Secondary	Parental Support for School Scale (Ceballo et al., 2014)	2	1 = Never, 2 = Hardly ever, 3 = Sometimes, 4 = Most of the time, 5 = Almost every day	Parent and adolescent
Intimate Partner Violence Perpetration and Victimization	Secondary	Revised Conflict Tactics Scale Short Form (Straus et al., 1996)	4	0 to 7 and 8 for more than or equal to 8 in the past 4 weeks	Parent
Sexual Health Communication	Secondary	Risk Avoidance Planning Scale (Cluver et al., 2018)	3	0 = No, I find it too hard to talk about this, 1 = We have not made plans yet but I would like to talk about it, 2 = We have discussed this together	Parent and adolescent
School Violence	Secondary	Created based on ISPCAN Child Abuse Screening Tools-Trial Version	3	0 to 7 and 8 for more than or equal to 8 in the past 4 weeks	Adolescent



### 7.3.5 Data Analyses

To investigate the predictive validity of the PLH-FAT-T Short Form on the change in pre-post outcomes, analyses used multi-level Poisson regressions with an interaction term between pre-post outcomes and facilitator competent adherence. First, the data was inspected in preparation for the analyses. The distribution of each outcome was examined using Anderson-Darling Normality Tests. As outcomes were not normally distributed and measured frequencies, Poisson models were used except for one outcome (Fahrmeir et al., 2013). As acceptability of corporal punishment was binary, it was examined using logistic regression. Although some outcomes could have been run using negative binomial models, Poisson models were used to better address convergence and therefore model fit (Fahrmeir et al., 2013). Second, model variables were specified. The response variable was the adolescent- or parent-reported outcome; the fixed effects were pre- and post-test timepoint, wave of delivery, and district (region) of delivery; and the random effects were participants and facilitators with the former nested within the latter to address clustering. Clustering at the family level was addressed by running separate models for adolescent- and parent-reported outcomes. Wave of delivery was included as a fixed effect because each of the three rounds of implementation could have unique characteristics that influenced the ability of facilitators to deliver the program, especially during the COVID-19 pandemic. Wave of delivery could have also impacted the ability of participants to benefit from the program. The district of delivery was included as a fixed effect as the implementing partner organization coordinating program delivery in each district could have created an environment that was more or less conducive to facilitator delivery as well as participant outcomes. Third, as facilitators were assessed on either the home activity discussion (7

items) or role-play (8 items), PLH-FAT-T Short Form scores were adjusted to ensure the scores were comparable. To produce an equally weighted index measure of competent adherence, a standard deviation score was created by subtracting the mean of each subscale and dividing by the standard deviation of each subscale. This approach made the scores comparable by weighting scores by the distribution from which they were derived. Fourth, the multi-level Poisson regression models were run incorporating an interaction term to examine the relationship between facilitator competent adherence and time (pre to post). Models also included the “bobyqa” optimizer to address model convergence (Bates et al., 2015). Fifth, to account for the potential increase in Type I error resulting from running multiple comparisons, the Benjamini-Hochberg adjustment was applied to the resulting p-values. This adjustment was selected because this method does not reduce power as much as other methods and produces a less conservative estimate (Benjamini & Hochberg, 1995; Chen et al., 2017; Lee & Lee, 2018). Finally, the results were interpreted using an incidence rate ratio (IRR) and confidence interval if the interaction term was significant ( $p < 0.05$ ). The analyses were conducted in R v4.2.2 using the “iccCounts”, “lme4”, “dplyr”, “tidyverse”, and “nortest” packages (Bates et al., 2015; Carrasco, 2022; Gross & Ligges, 2015; R Core Team, 2021a; Wickham et al., 2019; Wickham et al., 2023).

In the present study, it was hypothesized that there would be a positive linear association between competent adherence and outcomes wherein higher PLH-FAT-T Short Form scores would predict greater improvements in adolescent- and parent-reported outcomes.

## **7.4 Results**

### **7.4.1 Participant and Facilitator Characteristics**

Of the 3,057 families at baseline, 59.7% of parents identified as women. Parents ranged from 18 to 95 years with a mean age of 45.4 years. Adolescent girls ranged from nine to 16 years with a mean age of 11.9 years. Families reported experiencing a range of vulnerabilities – 63.0% of parents were unemployed, 56.2% reported running out of money for food or essentials in the last month, 12.6% were affected by the drug abuse of a family member, 11.0% indicated that the family had an unwell caregiver, 7.2% shared that the family had at least one child who was sick, and 4.9% had experienced the death of a family member due to TB or HIV.

Among the facilitators who reported demographic data, the mean age was 33.6 years with a range from 25 to 54 years. Ten facilitators identified as male (41.7%), eight facilitators identified as female (33.3%), and six did not provide data on their gender (25.0%). Out of the 13 facilitators who reported their caregiving status, most facilitators were parents themselves (92.3%). Facilitators reported being assessed on sessions 1 through 9 with the median response being session 3. Only one facilitator reported being assessed on a prior occasion. On the three subscales, facilitators received a mean score of 8.91 out of 14 (63.6%, SD: 4.81, range: 0-14) on the home activities items, 11.86 out of 16 (74.1%, SD: 2.57, range: 8-16) on the role-play items, and 26.21 out of 32 (81.9%, SD: 4.65, range: 10-32) on the skills items.

### **7.4.2 Pre-Post Outcomes**

The pre-post outcomes for the participants whose data was linked with facilitator data is shown in Table 2 ( $N=3,057$  parent-adolescent dyads). Intra-class correlations

between each covariate and participant outcome is displayed in [Supplementary File 1](#). The pre-post outcomes of the 3,057 families aligns with the results of the larger sample of families, with all outcomes improving from pre- to post-test with the exception of parental support of education and positive parental involvement (Lachman et al., forthcoming).

**Table 2**  
*Primary and Secondary Outcomes at Pre- and Post-Test of Subset of 3,057 Families*

<u>Outcome</u>	<u>Parent-Reported Pre-Test</u>	<u>Parent- Reported Post-Test</u>	<u>Adolescent- Reported Pre- Test</u>	<u>Adolescent- Reported Post- Test</u>
Child Maltreatment	2.74 (3.19)	1.23 (1.73)	2.92 (3.21)	1.14 (1.73)
Conduct Problems	1.44 (1.75)	1.25 (1.81)	1.61 (1.80)	1.14 (1.63)
Positive Involvement	6.75 (4.02)	1.45 (2.14)	6.23 (3.90)	2.49 (2.05)
Poor Supervision	1.12 (1.77)	0.84 (1.38)	1.17 (1.76)	0.99 (1.58)
Parenting Stress	3.45 (2.62)	2.48 (2.38)	Not reported	Not reported
Acceptability of Corporal Punishment*	860 (or 28.13% of parents)*	242 (or 7.92% of parents)*	498 (or 16.29% of adolescents)*	101 (or 3.30% of adolescents)*
Depression	3.91 (1.63)	3.20 (1.64)	3.04 (2.10)	2.34 (2.03)
Financial Insecurity	3.03 (2.16)	2.46 (2.10)	Not reported	Not reported
Sexual Health Communication	2.77 (2.05)	4.16 (2.27)	2.71 (2.15)	4.11 (2.25)
Parental Support for Education	6.47 (2.80)	3.61 (1.41)	3.94 (2.63)	0.62 (1.24)

Intimate Partner Violence Perpetration	1.16 (2.21)	1.11 (1.49)	Not reported	Not reported
Intimate Partner Violence Victimization	1.28 (2.36)	1.19 (1.58)	Not reported	Not reported
Emotional Problems	Not reported	Not reported	1.80 (2.12)	2.06 (1.91)
School Violence	Not reported	Not reported	3.52 (4.64)	1.47 (2.09)
<i>Notes:</i> Standard deviations are in brackets after means. Depending on the outcome, increases or decreases from pre- to post-test are representative of either a positive or negative outcome for families. *As acceptability of corporal punishment is a binary variable, the data presented is the number of participants who agreed with the statement supporting the use of corporal punishment with the percentage of participants in brackets.				

### 7.4.3 Associations between Competent Adherence and Outcomes

Table 3 shows the results of the models investigating associations between facilitator competent adherence scores and family outcomes. Results are displayed using adjusted p-values. The results of associations between each subscale (home activities, role-play, and skills) and outcomes are included in [Supplementary File 2](#).

#### ***Main PLH-Teens Outcomes***

Increased competent adherence was not significantly associated with parent-reported child maltreatment or child conduct problems. As it relates to adolescent-reported outcomes, increased competent adherence was associated with a 14% decrease in child maltreatment (IRR=0.86 [95% CI=0.82-0.90,  $p<0.001$ ]), a 27% decrease in child conduct problems (IRR=0.73 [95% CI=0.70-0.77,  $p<0.001$ ]), and a 23% decrease in child emotional problems (IRR=0.77 [95% CI=0.69-0.86,  $p<0.001$ ]).

### ***Secondary Outcomes - Parenting***

Increased competent adherence had mixed associations with poor supervision as better delivery was associated with an 8% increase in parent-reported poor supervision (IRR=1.08 [95% CI= 1.03-1.15,  $p<0.01$ ]) yet a 7% decrease in adolescent-reported poor supervision (IRR=0.93 [95% CI=0.89-0.97,  $p<0.01$ ]). Increased competent adherence was associated with a 52% increase in parent-reported positive parental involvement (IRR=1.52 [95% CI=1.37-1.69,  $p<0.001$ ]) but was not significantly associated with adolescent-reported positive parental involvement. Increased competent adherence was not significantly associated with parent-reported parental support of education but was associated with a 122% increase in adolescent-reported parental support of education (IRR=2.22 [95% CI=1.90-2.58,  $p<0.001$ ]). Increased competent adherence was associated with a 28% increase in parent-reported acceptability of corporal punishment (IRR=1.28 [95% CI=1.11-1.49,  $p<0.01$ ]) but was not significantly associated with adolescent-reported acceptability of corporal punishment. Increased competent adherence was not significantly associated with parent-reported sexual health communication but was associated with a 10% increase in adolescent-reported sexual health communication (IRR=1.10 [95% CI=1.07-1.12,  $p<0.001$ ]).

### ***Secondary Outcomes – Family Well-Being and Other Violence***

Increased competent adherence was associated with a 4% decrease in parent-reported depression (IRR=0.96 [95% CI=0.94-0.99,  $p<0.01$ ]) but was not significantly associated with adolescent-reported depression. Increased competent adherence was associated with a 5% increase in parent-reported financial insecurity (IRR=1.05 [95% CI=1.03-1.08,  $p<0.001$ ]). Increased competent adherence was associated with 11% increase

in parenting stress (IRR=1.11 [95% CI=1.07-1.15,  $p<0.001$ ]). Increased competent adherence was also associated with a 31% decrease in adolescent-reported school violence (IRR=0.69 [95% CI=1.49-1.92,  $p<0.001$ ]). Increased competent adherence was not significantly associated with parent-reported IPV victimization but was associated with a 81% increase in parent-reported IPV perpetration (IRR=1.81 [95% CI=1.52-2.16,  $p<0.001$ ]).

**Table 3**

*Associations between Parent- and Adolescent-Reported Outcomes and Facilitator Competent Adherence*

<u><b>Outcome</b></u>	<u><b>Parent-Report</b></u>	<u><b>Adolescent-Report</b></u>
Child Maltreatment N=24 facilitators N=3,057 families	IRR=1.05, SE=0.02, 95% CI=1.00-1.10, p=0.10	IRR=0.86, SE=0.02, 95% CI=0.82-0.90, $p<0.001^{***}$
Child Conduct Problems N=24 facilitators N=3,057 families	IRR=1.01, SE=0.02, 95% CI=1.07-1.15, p=0.73	IRR=0.73, SE=0.02, 95% CI=0.70-0.77, $p<0.001^{***}$
Child Emotional Problems ^ N=24 facilitators N=3,057 adolescents	Not reported	IRR=0.77, SE=0.06, 95% CI=0.69-0.86, $p<0.001^{***}$
Poor Supervision N=24 facilitators N=3,057 families	IRR=1.08, SE=0.03, 95% CI= 1.03-1.15, p<0.01**	IRR=0.93, SE=0.02, 95% CI=0.89-0.97, $p<0.01^{**}$
Positive Parental Involvement ^ N=22 facilitators N=1,654 families	IRR=1.52, SE=0.05, 95% CI=1.37-1.69, p<0.001***	IRR=1.00, SE=0.04, 95% CI=0.92-1.08, p=0.90
Parental Support of Education ^ N=22 facilitators N=1,654 families	IRR=1.04, SE=0.03, 95% CI=0.97-1.11, p=0.50	IRR=2.22, SE=0.08, 95% CI=1.90-2.58, $p<0.001^{***}$
Parenting Stress N=24 facilitators N=3,057 parents	IRR=1.11, SE=0.02, 95% CI=1.07-1.15, p<0.001***	Not reported

Acceptability of Corporal Punishment $\diamond$ N=24 facilitators N=3,057 families	IRR=1.28, SE=0.07, 95% CI=1.11-1.49, p<0.01**	IRR=0.94, SE=0.08, 95% CI=0.80-1.11, p=0.57
Sexual Health Communication N=24 facilitators N=3,057 families	IRR=1.01, SE=0.01, 95% CI=0.99-1.04, p=0.59	IRR=1.10, SE=0.01, 95% CI=1.07-1.12, p<0.001***
Depression N=24 facilitators N=3,057 families	IRR=0.96, SE=0.01, 95% CI=0.94-0.99, p<0.01**	IRR=0.99, SE=0.01, 95% CI=1.07-1.12, p=0.33
Financial Insecurity N=24 facilitators N=3,057 parents	IRR=1.05, SE=0.01, 95% CI=1.03-1.08, p<0.001***	Not reported
School Violence $\wedge$ N=22 facilitators N=1,684 adolescents	Not reported	IRR=0.69, SE=0.07, 95% CI=1.49-1.92, p<0.001***
IPV Victimization $\wedge$ N=22 facilitators N=1,654 parents	IRR=1.61, SE=0.09, 95% CI=1.36-1.90, p=0.83	Not reported
IPV Perpetration $\wedge$ N=22 facilitators N=1,654 parents	IRR=1.81, SE=0.09, 95% CI=1.52-2.16, p<0.001***	Not reported
Note: Significance codes: ***p<0.001, **p<0.01, *p<0.05. The outcomes with the “ $\wedge$ ” symbol were run without wave as a fixed effect due to rank deficiency. The outcome with the “ $\diamond$ ” was run as a logistic regression as it is binary. IRR is the incidence rate ratio. LL is the lower bound and UL is the upper bound of the 95% confidence interval.		

## 7.5 Discussion

### 7.5.1 Overall Findings

This paper examined the predictive validity of an observational measure of facilitator competent adherence using data collected during the large-scale routine delivery of a parenting program in a low-income country. Results show that across the 14 outcomes



examined, the relationship between the PLH-FAT-T Short Form assessment results and outcomes was mixed with competent adherence having some positive associations (higher PLH-FAT-T Short Form scores associated with improved outcomes), some not statistically significant associations, and a few negative associations (higher PLH-FAT-T Short Form scores associated with worsened outcomes). For the primary outcome of interest, competent adherence was not significantly associated with parent-reported child maltreatment but was positively associated with adolescent-reported child maltreatment.

Of the 12 parent-reported outcomes, associations with the PLH-FAT-T Short Form were positive for two outcomes (positive parental involvement, parental depression), insignificant for five outcomes (child maltreatment, child conduct problems, parental support of education, sexual health communication, IPV victimization), and negative for five outcomes (poor supervision, acceptability of corporal punishment, parenting stress, financial insecurity, IPV perpetration).

Amongst the ten adolescent-reported outcomes, associations with the PLH-FAT-T Short Form were positive for seven outcomes (child maltreatment, child conduct problems, child emotional problems, poor supervision, parental support of education, sexual health communication, school violence) and insignificant for three outcomes (positive parental involvement, acceptability of corporal punishment, teen depression).

Taken together, the results are not as hypothesized as higher PLH-FAT-T Short Form scores do not consistently predict better participant outcomes. In the context of routine program delivery at scale, the PLH-FAT-T Short Form has not been found to demonstrate predictive validity across multiple outcomes. Although higher levels of competent adherence were positively associated with the primary outcome based on

adolescent-reports, the PLH-FAT-T Short Form's overall predictive validity is not consistent. As a result, competent adherence may be beneficial for participants to some extent.

### **7.5.2 Findings in the Context of the Broader Literature**

Although the insignificant and negative associations found are in contrast to the commonly theorized relationship between competent adherence and participant outcomes (Berkel et al., 2011; Carroll et al., 2007), the findings regarding insignificant and mixed associations are not unlike those found in the broader parenting program literature. A systematic review of papers reporting on the relationship between competent adherence and outcomes in the parenting literature also found insignificant, positive, and mixed associations with outcomes (Martin et al., 2023). Yet, unlike the broader literature, this paper found negative associations between competent adherence and some outcomes. Insignificant and negative associations were also found in the study of PLH-Teens in South Africa by Shenderovich et al. (2019), which did not use the PLH-FAT-T Short Form. In that study, while most outcomes had an insignificant relationship, higher competent adherence was associated with higher levels of adolescent-reported child maltreatment.

### **7.5.3 Potential Explanations for Findings**

While there may be more, seven potential explanations for the inconsistent findings are explored herein. First, the results may indicate that the PLH-FAT-T Short Form has poor predictive validity as the tool is not associated with outcomes in the manner hypothesized. While positive linear associations between competent adherence and family outcomes were found in some cases, negative and insignificant associations were also found.

Second, the degree to which PLH-FAT-T Short Form assessments can be relied upon may be limited due to poor to moderate assessor reliability found in a psychometric evaluation of the tool by Martin et al. (in submission-a). The poor to moderate reliability may mean that the assessments did not consistently capture facilitator delivery. If the competent adherence was not consistently captured, the results of the present study may be over or underestimating facilitator delivery.

Third, the true nature of the relationship between competent adherence and outcomes may be difficult to detect due to a range of methodological challenges when studying implementation fidelity during routine service delivery. Detecting the true nature of the relationship between competent adherence and outcomes may have been limited because of the sample size. As a systematic review found that the average sample size in randomized trials to be 38 facilitators and 159 families (Martin et al., 2023), the present study (with data from 24 facilitators and 3,057 families) was larger than many studies.

Fourth, it is possible that fidelity does not relate to outcomes as hypothesized. For instance, there may not be a linear relationship between competent adherence and family outcomes. Some studies have found that fidelity plays an indirect role in achieving participant outcomes. Smith et al. (2013) found that participant engagement had an intervening effect on the relationship between facilitator competent adherence and outcomes. Further studies may need to examine a variety of implementation fidelity components (e.g., participant engagement and responsiveness, therapeutic alliance) or examine the role of multiple implementation fidelity components simultaneously to capture the complex relationships that could be at play in achieving participant outcomes (Berkel et al., 2011).

Fifth, the interaction between competent adherence and real-time program adaptations may explain why competent adherence does not always relate to outcomes as expected. It may be that ‘fidelity-consistent’ adaptations that maintain intended intervention ‘functions’ produce similar program effects (Skivington et al., 2021; Stirman et al., 2015). For example, facilitators may make planned or responsive adaptations which achieve positive outcomes without implementing the model exactly as intended (Miller et al., 2021). Qualitative data collected as part of a larger study of the implementation of PLH-Teens in Tanzania (Martin et al., 2021a) indicated that facilitators made responsive adaptations to the program by translating it from Kiswahili into local languages since some participants did not speak the national language (Shenderovich et al., forthcoming). While this responsive adaptation may have prevented facilitators from delivering all program components, it allowed many more families to engage and therefore benefit from the program. While such adaptations may be extremely beneficial in achieving outcomes, there is no consensus on how best to capture adaptations and then quantitatively examine their impact on participant outcomes (McHugh et al., 2009). Future research may benefit from testing novel approaches for taking adaptations into account to determine whether, and how, the interaction of fidelity and adaptations play a role in participant outcomes. As an example, a paper by Owen and Hilsenroth (2014) examined within-case variabilities in therapist delivery of adult psychotherapy, which was associated with better patient outcomes and explained 10% of the variance. The dynamic relationship between responsive adaptation and competent adherence may find support in a paper by Hogue et al. (2008). This paper found a curvilinear relationship wherein higher levels of facilitator adherence were associated with less improvement in participant outcomes and mediocre levels of

facilitator adherence were associated with the best outcomes. This finding may indicate that facilitators who less strictly adhered to the intervention made meaningful adjustments that helped participants to achieve better results. Future studies on the PLH-FAT-T Short Form should examine curvilinear relationships by testing whether ‘medium’ levels of competent adherence relate to better participant outcomes.

Sixth, competent adherence may play a less significant role in achieving desired outcomes than anticipated or, at some point, may even play a negative role. Although negative associations are infrequently observed in the literature, it is difficult to know the extent to which publication bias and selective reporting have prevented negative associations from being reported and published (Breitenstein et al., 2010b; Martin et al., 2023). It is possible that competent adherence could play a negative role in that too closely adhering to intervention components does not allow for facilitators to make responsive adaptations.

A seventh potential explanation for the findings is that only certain components of PLH-Teens are working to achieve positive participant outcomes. It is possible that the PLH-FAT-T Short Form is not capturing the important aspects of facilitator delivery when measuring competent adherence to delivery on the subscales (skills and home activity or role-play). As parenting programs have complex theories of change, there are a variety of potential mechanisms that may be working individually or in combination to produce participant outcomes (Leijten et al., 2022; Melendez-Torres et al., 2019). Future research might consider pairing an investigation of competent adherence with a component analysis.

#### **7.5.4 Limitations**

This paper has several limitations. The data used were not collected as part of a randomized trial. Instead, the pre-post surveys and facilitator assessments were collected during program scale-up via routine service delivery in low-income, rural and semi-urban settings in Tanzania throughout the height of the COVID-19 pandemic. This context presented challenges to implementing partners which meant that the data collected were ‘flawed, uncertain, proximate, and sparse’ (Wolpert & Rutter, 2018). As a result, assumptions and compromises were required when cleaning and merging facilitator data with pre-post surveys. Although 95 assessments were conducted, only 24 could be merged with 3,057 pre-post surveys. This small sample of assessments limited the predictive power of the analyses by providing little variability in PLH-FAT-T Short Form scores. There were also several limitations with the PLH-FAT-T Short Form assessment procedure. As assessments were conducted in-person, facilitators may have been affected by reactivity bias (Girard & Cohn, 2016). Furthermore, since it was logistically challenging to assess each facilitator during program scale-up in the pandemic, the assessments conducted only captured one participant group at one timepoint. The analysis assumed that the PLH-FAT-T Short Form assessments captured were reflective of each facilitator’s overall delivery. This assumption is supported by the findings of Shenderovich’s paper of PLH-Teens in South Africa, which found that facilitator competent adherence did not change much over time (2019). However, facilitator delivery may improve (e.g., Forgatch & DeGarmo, 2011) or decline (e.g., Chiapa et al., 2015) over time.

While this study has limitations, it made the best of the data available to examine the potential role that facilitator competent adherence had on the outcomes of the parents and adolescents who participated in PLH-Teens at scale in Tanzania.

## **7.6 Conclusion**

This study is the first investigation of its kind and contributes to the literature by examining the role of competent adherence on parenting program participant outcomes using data collected from in a LMIC during routine delivery at scale in a low-income community setting. Implementation science has an important role to play in determining how to implement and maximize the effectiveness of parenting interventions in practice for the benefit of vulnerable children, youth, and families. An important consideration is the degree to which interventions should be delivered with fidelity to their theory of change, particularly by those who implement interventions in practice. The value of knowing the role that facilitator competent adherence plays in participant outcomes is only heightened as more parenting programs and other behavioral interventions are delivered at scale in low-income settings.

To contribute to this literature, this paper investigated the predictive validity of an observational measure of facilitator competent adherence, the PLH-FAT-T Short Form, used in PLH-Teens – a program aiming to reduce violence against children and child behavioral and emotional problems. Similar to the findings of some other studies on competent adherence and its association with program outcomes, this paper does not provide a clear answer on whether, and to what extent, facilitator competent adherence impacts participant outcomes. There are a variety of potential explanations for the inconsistent results which suggest avenues for future research. Exploration of these avenues

would benefit the ongoing delivery and dissemination of PLH programs as well as the broader parenting program literature as these communities strive to maximize the ability of vulnerable children and families globally to benefit from evidence-based parenting programs.



## **8. Overall Discussion**

This chapter provides an overview of the dissertation (8.1); summarises the findings of each paper (8.2); reflects on the insights generated from the three studies (8.3); explains the limitations of and challenges experienced conducting the research (8.4); outlines the strengths of the research (8.5); discusses implications for future research and practice (8.6); makes recommendations for future research (8.7); and outlines strategies used to disseminate dissertation results (8.8).

### **8.1 Overview of Dissertation**

The dissertation contributes to understanding of the measurement of facilitator competent adherence in parenting programmes aiming to reduce violence against children and child behaviour problems. The dissertation also contributes to understanding of whether, and to what extent, facilitator competent adherence plays a role in the achievement of intended parent and child outcomes in both the parenting programme literature and PLH-Teens. The dissertation had three objectives. The first objective was to synthesise the overall relationship between facilitator competent adherence and the outcomes of parents and children in the parenting programme literature. The second objective was to examine whether the observational measure of facilitator competent adherence used in the implementation of PLH-Teens in Tanzania is reliable and valid for use in research and practice and to determine the level of competent adherence with which community facilitators delivered the programme. The third objective was to determine the predictive validity of the observational measure of competent adherence used in PLH-Teens by examining whether competence adherence is associated with parent and

adolescent outcomes. These objectives were achieved by answering four research questions via three papers:

1. What is the evidence on the relationship between observational measures of facilitator competent adherence and parenting programme outcomes? (Paper 1)
2. What are the psychometric properties of the observational measure used to assess the competent adherence of PLH-Teens facilitators during routine service delivery at scale in Tanzania? (Paper 2)
3. What is the level of competent adherence with which facilitators deliver PLH-Teens during routine service delivery at scale in Tanzania? (Paper 2)
4. What is the predictive validity of the observational measure used to assess the competent adherence of PLH-Teens facilitators during routine service delivery at scale in Tanzania? (Paper 3)

To address research question 1, a systematic review and synthesis without meta-analysis of 18 studies found in the parenting programme literature was conducted (Paper 1). To address research questions 2-4, a psychometric evaluation and multi-level Poisson regressions were performed using secondary data from the 2020-2021 implementation of PLH-Teens in Tanzania (Papers 2 and 3).

## **8.2 Summary of the Findings of Each Paper**

### **8.2.1 Paper 1: The Relationship between Competent Adherence and Outcomes**

Paper 1 reports on the first systematic review examining the relationship between observational measures of facilitator competent adherence and outcomes among studies reporting on parenting programmes aiming to reduce violence against children and/or

improve child behaviour problems. To compile the evidence, a synthesis without meta-analysis was conducted of the 18 identified studies, which found that higher levels of facilitator competent adherence were beneficial for some parent and child outcomes as most studies reported that competent adherence was positively associated with at least one participant outcome. The review also found that the relationship between facilitator competent adherence and outcomes was inconsistent (with some positive, some insignificant, and some mixed associations). While 13 studies reported at least one statistically significant positive relationship between facilitator competent adherence and outcomes, eight studies reported mixed findings across outcomes and four studies did not find any statistically significant association. These results suggest that there is a lack of clarity on the role that facilitator competent adherence plays in parenting programme participant outcomes.

### **8.2.2 Paper 2: Psychometric Properties of the PLH-FAT-T and Level of Competent Adherence Achieved by PLH-Teens Facilitators**

In Paper 2, the psychometric properties of the PLH-FAT-T were evaluated using data from the large-scale implementation of the PLH-Teens programme in Tanzania. Two types of measure validity (content validity and construct validity) and three types of measure reliability (intra-rater reliability, inter-rater reliability, and internal consistency) were examined. Regarding content validity, based on insights generated from the study of the PLH-FAT-YC in Southeastern Europe and input from stakeholders in Tanzania, revisions were made to the PLH-FAT-T to make it more understandable and relevant to those who were conducting assessments. Subsequent analyses of intra- and inter-rater reliability found poor to moderate results (Koo & Li, 2016). Although the intra-rater

reliability of the lead assessors was above 70%, assessors were in the poor to moderate range for ICCs (Aspland & Gardner, 2003; Koo & Li, 2016). The inter-rater reliability of the lead assessors was near the 70% quality threshold but still in the poor range for ICCs (Aspland & Gardner, 2003; Koo & Li, 2016). However, analyses showed that the PLH-FAT-T had strong internal consistency, as evidenced by Cronbach's alphas and omegas consistently close to or above 0.70 (Terwee et al., 2007). Through iterative exploratory factor analyses, a shortened PLH-FAT-T was produced. This shortened version, the PLH-FAT-T Short Form, had 19 fewer items and showed improved intra- and inter-rater reliability. The stronger internal consistency as well as improved construct validity following iterative exploratory factor analyses suggest that the PLH-FAT-T Short Form is performing at an acceptable level on these indices. Together, the evaluation of the psychometric properties of the PLH-FAT-T Short Form suggests that the tool has poor to moderate reliability and sufficient validity. The findings indicate that further work is needed to strengthen the reliability and validity of the PLH-FAT-T Short Form.

Paper 2 also investigated the level of competent adherence achieved by community facilitators delivering PLH-Teens to parents and adolescents at scale in semi-urban and rural Tanzania. To calculate the level of competent adherence, facilitator assessments collected by implementing partners during programme delivery were analysed. Using the PLH-FAT-T Short Form, individual facilitators achieved an average score of 82.3%. Facilitators achieved higher scores on items that assessed their competence (average of 85.9%) than on items that assessed their adherence (average of 75.0%). Overall, these findings indicate that facilitators appear to be able to deliver PLH-Teens to a high level of competent adherence.

### **8.2.3 Paper 3: Predictive Validity of the PLH-FAT-T**

In Paper 3, the predictive validity of the PLH-FAT-T Short Form was examined by investigating the relationship between competent adherence and the outcomes of parents and adolescents who participated in PLH-Teens. A total of 12 parent-reported and 10 adolescent-reported outcomes were investigated, including the primary outcome of child maltreatment. Multi-level Poisson regressions with an interaction term as well as fixed and random effects were conducted to explore the relationship between competent adherence and outcomes. A total of 24 of the 95 PLH-FAT-T Short Form assessments could be linked with pre-post surveys from 3,057 families. Analyses found that competent adherence had a positive association with two of the 12 parent-reported outcomes (positive parental involvement, parental depression) and seven of the 10 adolescent-reported outcomes (child maltreatment, child conduct problems, child emotional problems, poor supervision, parental support of education, sexual health communication, school violence). Competent adherence had a negative association with five parent-reported outcomes (poor supervision, acceptability of corporal punishment, parenting stress, financial insecurity, IPV perpetration). There were no significant associations between competent adherence and five parent-reported outcomes (child maltreatment, child conduct problems, parental support of education, sexual health communication, IPV victimisation) and three adolescent-reported outcomes (positive parental involvement, acceptability of corporal punishment, adolescent depression). Due to the mix of positive, negative, and insignificant associations found between competent adherence and outcomes, the PLH-FAT-T Short Form was found to demonstrate varying degrees of predictive validity across multiple outcomes. Although higher levels of competent adherence were positively associated with the primary outcome

based on adolescent-reports, the PLH-FAT-T's overall predictive validity was not consistent. As a result, competent adherence may be beneficial for participants to some extent.

### **8.3 Discussion of Findings**

In the context of the broader parenting programme and implementation science literature, the findings from this dissertation generated three main observations. First, while facilitator competent adherence appears beneficial to the outcomes of children and parents to some extent, the relationship remains inconsistent. Second, although the PLH-FAT-T Short Form has stronger psychometric properties than most observational measures of competent adherence found in the parenting programme literature, it requires further work to improve its reliability and validity. This suggests that the extent to which PLH researchers and implementers can rely on the PLH-FAT-T Short Form may be limited. Third, community facilitators with minimal background in and training on parenting programmes appear to be able to deliver PLH-Teens to a high level of competent adherence. This suggests that evidence-based parenting programmes can be delivered to a high level of quality in low-income routine service settings at scale.

#### **8.3.1 The Role of Facilitator Competent Adherence is Inconsistent**

The evidence from Papers 1 and 3 suggests that the relationship between facilitator competent adherence and participant outcomes is inconsistent in the parenting programme literature and in PLH-Teens. Several systematic reviews and meta-analyses in the broader implementation science literature indicate that the findings of Paper 1 and Paper 3 are not exceptional. Other intervention fields are struggling with similar conceptual and methodological issues as there are other papers and reviews that do not support the

hypotheses outlined in implementation science theories and models. Although there are several reviews from the parenting programme literature (Furlong et al., 2012), health intervention literature (Durlak & DuPre, 2008), and educational intervention literature (Hill & Erickson, 2019; Naylor et al., 2015) that suggest that better facilitator delivery is associated with better participant outcomes, some systematic reviews and meta-analyses from other fields have found insignificant (Webb et al., 2010) or small effect sizes (Collyer et al., 2019; Rapley & Loades, 2019). A meta-analysis on 36 studies of adult cognitive behavioural therapies by Webb et al. (2010) analysed associations between facilitator adherence and outcomes as well as facilitator competence and outcomes finding insignificant overall effects. The authors theorised that these results may have been due to the large amount of heterogeneity estimated (Webb et al., 2010). Similarly, a systematic review of five studies on child cognitive behavioural therapies by Rapley and Loades (2019) found small to no effects on the relationship between facilitator competent adherence and outcomes. In their discussion, the authors deliberated on whether these results were due to a range of methodological issues associated with studying the relationship as well as the potential role of other implementation variables (Rapley & Loades, 2019). A meta-analysis of 35 studies on child psychotherapies by Collyer et al. (2019) found that therapist competent adherence had a small overall effect on participant outcomes. The authors questioned whether other intervention factors were playing a more important role than competent adherence (Collyer et al., 2019). Thus, researchers in the wider intervention science literature are considering a range of possible explanations for mixed and inconsistent results on the relationship between competent adherence and outcomes.

While the findings from both Papers 1 and 3 showed mixed and inconsistent associations between competent adherence and outcomes, Paper 3 is unlike many other studies in the parenting programme literature as studies rarely report negative associations. As discussed in Paper 3 (7.5.3), there are several potential explanations for the mixed results: the PLH-FAT-T Short Form may have poor predictive validity; the degree to which PLH-FAT-T assessments can be relied on may be limited due to the tool having poor to moderate intra- and inter-rater reliability; the methodological challenges of studying implementation fidelity in the context of programme delivery at scale in a low-income setting may have compromised the ability to detect the actual relationship between competent adherence and outcomes; competent adherence may not relate to outcomes in a positive linear fashion (e.g., Hogue et al., 2008); competent adherence may be interacting with other implementation variables, such as responsive adaptations and participant engagement, to produce outcomes (Berkel et al., 2011; McHugh et al., 2009; Smith et al., 2013); competent adherence may play a less significant, or even negative role, in participant outcomes; and not all programme components are active ingredients in achieving participant outcomes so the activities and skills measured by the tool may not provide an accurate representation of the role of competent adherence.

The mixed and inconsistent relationship between competent adherence and outcomes found in Papers 1 and 3, evidence from some reviews in the implementation science literature, and the negative associations found in Paper 3 suggest that the relationship between facilitator competent adherence and participant outcomes may be more complicated than hypothesised. As a result, there is a need to further investigate the theorised relationship between facilitator competent adherence and outcomes outlined in



seminal implementation science theories and models to fully illuminate the inner workings of the ‘black box’ of evidence-based interventions (Astbury & Leeuw, 2010).

### **8.3.2 The PLH-FAT-T Short Form Requires Improvement**

Paper 2 found that while the PLH-FAT-T Short Form is performing well in the context of the measures of facilitator competent adherence used in the broader parenting programme literature, the tool requires additional research to strengthen its reliability and validity. Based on COSMIN standards, the weakest psychometric properties of the PLH-FAT-T Short Form were its intra- and inter-reliability (Mokkink et al., 2016). However, the intra- and inter-reliability of the PLH-FAT-T Short Form was stronger than other similar observational measures. In the parenting programme literature, two studies report on intra-rater reliability (Martin et al., 2021b). A study of the Parent Programme Implementation Checklist used in Triple P found a mean percentage agreement of 73.0% (range: 60.0-79.0%) and ICCs from -0.14 to 0.94 (Bywater et al., 2019). A study of the Leadership Observation Tool used in Incredible Years found a mean percentage agreement of 87.0% and a mean ICC of 0.92 (Eames et al., 2008). Of the forty-one studies in the parenting programme literature reporting on inter-rater reliability, 17 studies reported on percentage agreements which ranged from 50.0 to 100.0% and 21 studies reported on ICCs which ranged -0.03 to 0.96 (Martin et al., 2021b).

The conclusion set out in Paper 2 that the reliability and validity of the PLH-FAT-T Short Form require strengthening suggests that the extent to which PLH can rely on the results of PLH-FAT-T Short Form assessments may be limited. However, in comparison to the reliability and validity of other measures of competent adherence found in the parenting programme literature that have had their reliability and validity examined, the PLH-FAT-T

Short Form performs well. As a result, the extent to which the parenting programme community can rely on other measures of competent adherence is also limited. Further, a substantial number of measures used to investigate the relationship between competent adherence and outcomes have not been psychometrically evaluated. A third of the studies included in Paper 1 used measures of competent adherence that had not been the subject of a psychometric evaluation. The PLH-FAT-T Short Form's strong performance in the context of the wider literature and that many measures used in the literature have not been psychometrically validated supports the conclusion that the PLH-FAT-T Short Form should not be abandoned. Instead, further research and revision is recommended to improve its reliability and validity.

### **8.3.3 Community Facilitators Delivered PLH-Teens to a High Level of Competent Adherence**

Paper 2 found that community facilitators delivered PLH-Teens in a low-income, routine service delivery setting to a similar level of competent adherence to facilitators in randomised controlled trials. Paper 2 found that the average level of competent adherence achieved by PLH-Teens community facilitators was 82.3% whereas Paper 1 found that the average level of competent adherence achieved was 69.5%. These findings suggest that facilitators in LMICs may be able to achieve levels of facilitator competent adherence comparable to those achieved during trials. However, the parenting programmes, delivery contexts, and measurement tools used in Papers 1 and 3 may be too different to warrant comparison. As a result, caution is required in comparing the level of competent adherence found in Paper 1 with the level found in Paper 3. Related to community facilitators in particular, the community facilitators who delivered PLH-Teens in Tanzania appear to have

achieved a similar level of competent adherence to community facilitators who delivered PLH-Teens in South Africa during the first randomised trial of the programme. The average level of competent adherence achieved by facilitators during the delivery of PLH-Teens at scale in Tanzania was 82.3% whereas the average level of competent adherence achieved by facilitators during the randomised trial was 83.0%, which study did not use the PLH-FAT-T (Shenderovich et al., 2019). These findings suggest that community facilitator competent adherence has the potential to remain stable as parenting programmes are transported from small-scale delivery via randomised trials to large-scale delivery via community settings. Here too, caution in interpretation is necessary due to differences in the context of delivery and the measurement tools used.

#### **8.4 Limitations and Challenges**

In addition to the limitations and challenges discussed in each individual paper, the dissertation has four overall limitations and challenges. First, the dissertation focused on observational measures of facilitator competent adherence. Second, Papers 2 and 3 relied on secondary data collected by implementing partners in a challenging context – routine service delivery in a low-income setting at scale during the COVID-19 pandemic. The context presented practical challenges including constraints that limited the quantity of the data collected. Third, there are several threats to the internal validity of the PLH-FAT-T Short Form as well as the assessments of family outcomes, which have implications for the interpretation of results. Fourth, there are several threats to the external validity of the PLH-FAT-T Short Form and assessments of family outcomes, which have implications for the generalisability of results to other facilitators and families.

#### **8.4.1 Focus on Observational Measures of Facilitator Competent Adherence**

The dissertation limited its scope to the measurement and role of competent adherence only with respect to measures that use observational approaches. Although this means that the dissertation focused on the gold standard form of fidelity measurement (Eames et al., 2008), the research did not examine the measurement and role of non-observational approaches, such as audio recordings or self-reports. As a result, when synthesising the evidence on the relationship between competent adherence and outcomes (Paper 1), only a subset of the literature on measures of facilitator competent adherence was examined. Thus, in interpreting the results herein, the findings should not be generalised to all measures of competent adherence.

#### **8.4.2 Reliance on Secondary Data Collected in a Challenging Context**

The analyses performed in Papers 2 and 3 relied on secondary data collected by implementing organisations in rural and semi-urban Tanzania during the pandemic. Since implementing organisations were moving forward with the delivery of PLH-Teens regardless of researcher involvement, FAIR study researchers were not responsible for and did not have any decision-making authority regarding what quantitative data was collected or how it was collected. As a result, the FAIR data used herein were subject to the resources, commitments, and decisions of the local implementing organisations. Implementer decisions regarding data were balanced with their objectives, timelines, budget, and other realities of delivering the intervention to 75,061 beneficiaries. In particular, the COVID-19 pandemic impacted decision-making in ways that made the conduct of facilitator assessments less of a priority. This shift in priorities took place

because local implementing organisations needed to complete delivery of the programme during the pandemic within established resources and by established deadlines.

Overall, the practicalities of programme delivery in Tanzania and during the pandemic meant that it was not possible to collect and send videos of programme delivery for use in facilitator assessments; it was difficult to establish reliability among assessors; a small quantity of facilitator assessments were conducted by assessors; the quality of the secondary data collected was limited; and FAIR researchers were unable to travel to Tanzania to support training for and monitoring of data collection. Each of these practical challenges are discussed in the following paragraphs.

#### *Inability to Collect and Send Videos*

It was not possible for video recordings of programme delivery to be reliably collected and transmitted so that recordings could be used to conduct facilitator assessments. As limited internet connectivity meant that videos took too much internet bandwidth to upload and share, only a few videos were collected for the purposes of establishing intra-rater reliability. Using video recordings of programme delivery is the ideal way to assess facilitator delivery because this method minimises facilitator reactivity bias; eliminates the need for assessors to travel to and from programme sessions which saves time and other resources; and allows assessors to conduct assessments on a flexible timeline (see [3.1.2](#)). Additionally, video recordings support the use of rigorous psychometric approaches because they can be repurposed to evaluate intra- and inter-rater reliability. In the FAIR study context, even if video recordings could have been collected, implementers and researchers had insufficient funds to purchase enough video cameras for

the 69 assessors. As a result, PLH-FAT-T assessments needed to be conducted by observing facilitator delivery in-person.

### *Challenges Establishing Reliability*

Two practical challenges may have influenced the level of intra- and inter-rater reliability achieved. First, a lack of sufficient translation resources may have prevented assessors from understanding PLH-FAT-T Short Form items in the same way. As the PLH-FAT-T materials created via the content validity process (e.g., the training documents, item assessment criteria) were in English, they had to be translated into Kiswahili for use by assessors. As back translation was not possible due to funding limitations, the Kiswahili version of the tool may not have faithfully replicated the English version.

Second, assessors may not have received sufficient training. Assessors received two days of training on how to conduct PLH-FAT-T assessments, which only allowed for one practice assessment. As assessors had little or no prior experience conducting facilitator assessments, more training may have increased their familiarity with the tool and the assessment procedure as well as supported their ability to consistently code items. The implementation of many competent adherence measures in high-income countries involves extensive training of assessors who already possess considerable prior experience with assessments due to their substantial training in relevant disciplines (e.g., psychology). For example, assessors of the Brief Strategic Family Therapy programme facilitators in the United States receive 80 hours of assessor training (Robbins et al., 2011).

### *Quantity of Secondary Data Collected*

Although the FAIR study expected to receive at least one assessment for each of the 444 facilitators who delivered PLH-Teens, substantially fewer assessments were conducted

and captured by implementing organisations ( $N=95$ ). The small sample size of facilitator assessments impacted the analyses in Papers 2 and 3. For Paper 2, a portion of the assessment data could not be set aside to conduct a confirmatory factor analysis following the iterative exploratory factor analyses. Being able to conduct a confirmatory factor analysis would have allowed the performance of the PLH-FAT-T Short Form to be tested further to determine if the findings of the exploratory factor analyses were replicable (Furr, 2017). For Paper 3, the size of the dataset reduced the power of the analyses, which may have introduced a predictive bias.

Only 95 facilitator assessments were gathered due to the challenges associated with conducting assessments during in-person programme delivery. Meetings and semi-structured interviews with coaches and other programme staff involved in the facilitator assessment process found five interconnected barriers to conducting in-person assessments (Martin et al., forthcoming). First, assessors had a heavy workload and a multiplicity of roles which limited the time they had available to conduct assessments. The main role of the assessors was to provide supervision to facilitators during the delivery of PLH-Teens. As there were 444 facilitators and 69 coaches, each coach had approximately six or seven facilitators to supervise and assess. Many coaches were also acting as programme facilitators due to tight programme budgets and a shortage of facilitators. Further, most coaches were also working full-time in other jobs.

Second, the delivery of PLH-Teens in schools and community centres, some of which are in remote communities, meant that some coaches had to travel long distances to conduct assessments and experienced a variety of transportation-related challenges. The substantial time coaches spent traveling to and from programme sessions was made

difficult by poor road conditions in rural Tanzania (e.g., unpaved roads which became muddy in the rainy season), further lengthening travel times. A cost analysis performed as part of the FAIR study found that coaches reported travelling up to 71 minutes to and from each programme session (Calderon et al., forthcoming).

Third, coaches experienced challenges attending programme sessions and therefore being able to assess all their facilitators. As PLH-Teens is a family programme, sessions were often held after school. This scheduling meant that numerous facilitator pairs were conducting sessions simultaneously. As coaches could only attend one session at a time, concurrent sessions significantly reduced the number of different sessions a coach could observe each week.

Fourth, two adaptations made by implementing organisations to meet delivery deadlines reduced the length of the programme and the length of time over which the programme was delivered. These adaptations limited the opportunities coaches had to conduct assessments (Shenderovich et al., forthcoming). In some waves of programme delivery, implementers combined programme sessions (e.g., sessions 7 and 8). Implementing organisations also scheduled several sessions per week for each participant group (i.e., participant groups received two sessions or more per week) when the programme model indicates that 14 weekly sessions are to be held. While the reduction in programme sessions and compressed delivery schedule affected the opportunities assessors had to conduct assessments, these decisions may also have impacted both facilitators' ability to deliver the programme and the achievement of outcomes.

Fifth, pauses in programme delivery due to COVID-19 posed challenges for the scheduling of assessments. For instance, when the pandemic began during the first wave of



programme delivery, indefinite school closure halted programme delivery at session 10 of 14. Due to the unpredictable nature of the pandemic, it was not known when schools would reopen, and the programme would re-start. As a result, coaches could not create a schedule for future facilitator assessment.

#### *Quality of Secondary Data Collected*

The practicalities of programme delivery at scale during COVID-19 negatively impacted the quality and quantity of the secondary data collected and processed by implementing organisations. With the programme being delivered to 75,061 beneficiaries, there was a substantial number of surveys, forms, assessments, and other programme data produced and processed by implementing organisations. Further, much of the documentation produced, including facilitator assessments, needed to be physically transported from programme sessions to implementing organisation offices where the data was entered into spreadsheets by their data clerks. Flaws in the monitoring and evaluation system resulted in loss of programme documents, missing data, data entry errors, and issues with identification numbers which were found during data cleaning. Issues with the monitoring and evaluation system used to collect data were exacerbated by pandemic-related programme pauses (e.g., office closings and re-openings).

#### *Conducting an International Multi-Sectoral Study Online during COVID-19*

The pandemic also affected the quantity and quality of the data collected and processed by implementing organisations for additional reasons. Due to the pandemic, members of the FAIR research team were not permitted to travel to Tanzania to observe and advise on monitoring and evaluation processes including data collection, transport, entry, and storage. Engaging with implementers in-person would have provided the team

with a fuller understanding of the monitoring and evaluation system being used by local implementing organisations. All inter-organisation study-related communication happened via Zoom meetings, WhatsApp calls and messages, or email communications.

Communication was difficult due to poor internet connectivity in Tanzania which meant that calls frequently dropped, there were long periods during calls when participants waited for individuals to re-join, and meetings often needed to be rescheduled. Additionally, poor sound quality made it difficult to understand what participants were saying. The exclusive use of these modes of communication limited the ability of the research team to identify where implementing organisations were experiencing data collection and processing challenges and therefore needed further support, such as regarding the development and implementation of a rigorous process for linking facilitator data with parent and adolescent data via identification numbers.

#### **8.4.3 Issues of Internal Validity**

Implementer and researcher decisions regarding the PLH-FAT-T Short Form and family outcome assessments may have impacted the accuracy of the results, or internal validity, due to the potential introduction of multiple forms of bias. Four threats to the accuracy of PLH-FAT-T results and three threats to the accuracy of family outcome assessments are discussed.

##### *Threats to the Accuracy of PLH-FAT-T Results*

First, reactivity bias may have impacted PLH-FAT-T Short Form assessment results. Due to the presence of their coach during assessments, facilitators may have delivered the programme with a different level of quality than usual (Girard & Cohn, 2016). Facilitator knowledge that they were being assessed could have biased their performance

either positively or negatively. For instance, facilitators may have been more diligent in implementing the programme when being observed. Thus, reactivity bias may have resulted in an over or underestimation of the level of competent adherence achieved by facilitators during normal programme delivery.

Second, that coaches selected which facilitators to assess and when may have impacted the PLH-FAT-T Short Form results. With little time and resources to conduct assessments, there are several possibilities regarding how coaches selected which facilitators to assess. Coaches may have decided to assess their best facilitators thereby leading to an overestimation of facilitator competent adherence. Alternatively, coaches may have decided to assess facilitators needing the most support, thereby leading to an underestimation of facilitator competent adherence. It is also possible that coaches decided which facilitators to assess based on proximity (i.e., session timing and location). Although selection of facilitators based on proximity may have introduced bias, in such a case the overall result is expected to be random.

Third, the analyses in Papers 2 and 3 relied on the assumption that assessing facilitator competent adherence at one timepoint accurately represented facilitator delivery over time. While there is evidence in the broader literature that facilitator delivery can improve over time (e.g., Forgatch & DeGarmo, 2011) or decline over time (e.g., Chiapa et al., 2015), a study of the delivery of PLH-Teens in South Africa found little change in facilitator competent adherence over the course of 14 sessions (Shenderovich et al., 2019).

Fourth, the decision to use a three-point Likert scale (*0=inadequate, 1=good, 2=excellent*) limited the variability for examination in the analyses. The narrow scale was used in response to insights derived from a psychometric evaluation of the PLH-YC

programme in Southeastern Europe that the previously employed four-point Likert scale (rated on a scale of 0 to 3) was challenging to use as it was difficult for assessors to distinguish between points 1 and 2 on the scale. However, the three-point scale used in the PLH-FAT-T Short Form reduced the amount of variability in responses available for analysis in Paper 2. Consequently, there were fewer patterns in the data to examine (Furr, 2017). Some issues related to the use of the three-point Likert scale could be managed. For instance, with fewer options, it is more likely that assessors will select the same point by chance. This issue was addressed by calculating ICCs.

#### *Threats to the Accuracy of Family Outcome Assessments*

Participants may have intentionally or unintentionally provided inaccurate or biased responses in self-report assessments of parent and adolescent outcomes. For instance, parents and adolescents may have based their post-test responses on what they learned was socially desirable during the programme (Furr, 2017; Ranjan & George, 2014). As a result, Paper 3 results may not provide an accurate representation of the relationship between competent adherence and outcomes.

The pre-post outcomes may have been affected by the decision of implementing organisations to use abbreviated versions of the instruments recommended by CWBSA and PLH. As implementing organisations only selected some items from each validated instrument to capture participant outcomes, the shortened instruments may not have provided an accurate overall reflection of the outcomes achieved.

Finally, since randomisation to intervention was not possible, the assessment results of family outcomes may not represent the true impact of the programme on parents and adolescents. When pre-post surveys are not conducted as part of a randomised trial,

researchers cannot determine with certainty whether the observed changes in participant outcomes from pre- to post-test reflect programme delivery alone, other factors influencing participant outcomes (e.g., other policies or services), or a combination.

#### **8.4.4 Issues of External Validity**

Implementer and researcher decisions regarding the PLH-FAT-T Short Form and family outcome assessments may have impacted the generalisability of the results. Two threats to the generalisability of the PLH-FAT-T Short Form and three threats to the generalisability of the family outcome assessments results are discussed.

##### *Threats to the Generalisability of PLH-FAT-T Short Form Results*

The generalisability of the results from Papers 2 and 3 to the entire sample of facilitators who delivered PLH-Teens in Tanzania is limited. Analyses in both papers incorporated a small sample of facilitators involved in programme delivery. Paper 2 included assessment results from 95 of the 444 facilitators. As few facilitator-related demographic details were collected by implementing partners, it could not be determined whether the 95 facilitators who were assessed are representative of the 444 facilitators. Paper 3 included 24 of the 95 facilitators who received assessments. While the level of competent adherence achieved by the group of 24 facilitators was comparable to that of the group of 95 facilitators, the 24 facilitators may not be representative of the 444 facilitators.

The generalisability of the PLH-FAT-T Short Form results beyond the context of the delivery of PLH-Teens in Tanzania during the pandemic is also limited. Contextual and cultural factors may have influenced the interpretation of PLH-FAT-T items by assessors. As a result, the context of this particular PLH-Teens delivery in Tanzania may mean that

the findings of Paper 2 and 3 would not be the same for other facilitators delivering PLH or other parenting programmes in different contexts and cultures.

#### *Threats to the Generalisability of Family Outcome Assessments*

The representativeness of the pre-post survey results may be limited for three reasons. First, while data from the subset of 3,057 families could be linked to facilitator data, it is not known whether this subset of families is representative of the families who participated in PLH-Teens. However, analyses of family demographics as well as pre-post outcomes from programme beneficiaries found similar responses among the subset and the families who completed pre-post surveys. This suggests that the subset of families included in Paper 3 were representative of the larger population of families who received PLH-Teens.

Second, the family outcome assessments were collected in the Tanzanian context so may not be generalisable to other countries and cultures where PLH programmes are delivered. As a result, it cannot be assumed that the findings of Paper 3 would be the same in other countries or cultural contexts.

Third, the data used were collected during a particular and unusual temporal context – the COVID-19 pandemic. Due to implementing organisation decisions made because of the impact of COVID, the delivery of PLH-Teens had a different structure than delivery of the programme at other times. Thus, family outcome assessments may have been affected by this context. As a result, caution should be used in generalising the findings.

### **8.5 Strengths**

The dissertation has four key strengths. First, it made the best of a small quantity of poor quality data collected in a ‘real-world’ setting. Second, it investigated the

measurement and role of the gold standard approach to fidelity measurement in an unusual and challenging context. Third, it provides timely and useful evidence for the community of researchers and practitioners studying and implementing PLH programmes. Fourth, it contributes to several gaps in the implementation science and parenting programme literature.

### **8.5.1 Use of ‘FUPS’ Data**

The dissertation used a small quantity of poor quality data collected by implementing organisations to examine the measurement and role of facilitator competent adherence in a unique context – the large-scale delivery of a parenting programme via routine service in a low-income setting during a global pandemic. By using the secondary data provided by implementing organisations, the dissertation made the best of the ‘flawed, uncertain, proximate and sparse’ (FUPS) data available (Wolpert & Rutter, 2018). Data collected in routine delivery settings are often flawed (e.g., missingness), uncertain (e.g., rationale for measures used), proximate (e.g., proxies for other concepts), and sparse (e.g., small amount of data collected) (Wolpert & Rutter, 2018). In the FAIR study, the facilitator assessments collected by implementing organisations were flawed (e.g., low quality due to issues with participant and facilitator IDs) and sparse (e.g., low quantity in that few assessments were conducted and could be linked to family data). Although FUPS data presents many issues for rigorous research, they provide valuable evidence about interventions in challenging contexts that can be used to inform future research and practice (Wolpert & Rutter, 2018). In using FUPS data from the delivery of PLH-Teens in Tanzania, the dissertation contributes to the PLH, parenting programme, and implementation science literature by providing two analyses which are the first of their

kind. Although limited, the data provided by implementing organisations were a rich and unique source of information on the measurement and role of competent adherence in a low-income setting at scale. The data were also informative as to the extent to which observational methods could be used to conduct facilitator assessments in practice.

### **8.5.2 Examination of Gold Standard Measurement Approach**

The dissertation evaluates the gold standard approach to fidelity measurement – observational assessments. Although time consuming and resource intensive, observational approaches are considered the best form of fidelity measurement as they are thought to provide the least biased estimates of facilitator delivery (Eames et al., 2008). By investigating the observational measurement of facilitator competent adherence, the dissertation provides evidence on a rigorous form of competent adherence assessment and its role in family outcomes. In analysing observational assessments of facilitator delivery collected via routine service delivery in a low-income setting at scale, the dissertation also provides information on the extent to which the most rigorous form of assessments can be collected in such a context. Further, by conducting a psychometric evaluation of the tool used to assess facilitator competent adherence in PLH-Teens, the dissertation contributes to the limited number of studies on the reliability and validity of observational measures (Martin et al., 2021b).

### **8.5.3 Timely Evidence for PLH Programmes**

The dissertation provides timely evidence that can be used to enhance the delivery of PLH programmes as they continue to be implemented in real-world contexts. The delivery of PLH programmes is expanding rapidly worldwide, including in routine delivery settings (e.g., Cluver et al., 2018; Gardner et al., forthcoming-b; Lachman et al., 2021;



Ward et al., 2020). PLH programmes have now been delivered in 35 LMICs. Yet, there is little research on the implementation of the programme by facilitators. There is one study on implementation fidelity from the original trial of PLH-Teens in South Africa (Shenderovich et al., 2019), but this study did not use the PLH-FAT-T. There is also one evaluation of the content validity and reliability of the measure of facilitator competent adherence used to assess facilitators delivering PLH for Young Children in Moldova, North Macedonia, and Romania (Martin et al., 2022a). As the PLH-FAT-T is being used in the delivery of PLH-Teens in 19 countries, the dissertation advances understanding regarding the level of competent adherence PLH-Teens facilitators can achieve; the degree to which the PLH-FAT-T Short Form is accurately capturing competent adherence (Fan & Randall, 2018; Mokkink et al., 2010a; Thorkildsen, 2010); whether the tool can be reasonably relied on to assess, certify, and provide feedback to PLH-Teens facilitators in routine service delivery; the potential role competent adherence plays in PLH participant outcomes; and directions for future research and practice for PLH programmes.

#### **8.5.4 Contribution to Implementation Science and Parenting Programme Literature**

The dissertation contributes to three key gaps in the implementation science and parenting programme literature. First, there is limited literature on and no prior studies synthesising the relationship between observational facilitator competent adherence and the outcomes of parenting programmes aiming to reduce violence against children and/or child behaviour problems. As a result, it was not known whether the inconsistent and insignificant findings found in several individual studies (e.g., Cantu et al., 2010; Shenderovich et al., 2019) were exceptional. The findings from Paper 1 suggest that the relationship between competent adherence and outcomes is inconsistent in the parenting

programme literature and therefore provides evidence that can be used to consider and evaluate seminal implementation science theories and models in future research (Berkel et al., 2011; Carroll et al., 2007).

Second, few studies have reported on measures of facilitator competent adherence and their psychometric properties, particularly in LMICs (Martin et al., 2021b). Knowledge of the reliability and validity of competent adherence measures is vital for their use in practice as well as for making evidence-based decisions regarding programme improvements (Ruud et al., 2020; Stirman, 2020).

Third, there is limited evidence on the extent to which parenting programmes – and other behavioural interventions – can be delivered with competent adherence at scale, in LMICs, and in routine delivery settings (Smith et al., 2019). As a result, Paper 2 increases knowledge about the level of competent adherence achieved by community facilitators with little or no prior experience with parenting programmes can deliver these programmes in a real-world setting.

## **8.6 Implications for Future Research and Practice**

The dissertation's findings have implications for the way the parenting programme literature conceptualises the role of facilitator competent adherence; decision-making regarding who delivers parenting programmes in community settings and at scale; understanding the level of competent adherence with which parenting programmes can be delivered in the 'real-world'; decision-making about the amount and type of training provided to those who assess facilitator delivery; and the future of how facilitator competent adherence is measured in PLH programmes. Following a discussion of these

implications, the subsequent section outlines recommendations for future research and practice.

### **8.6.1 Conceptualisation of the Role of Facilitator Competent Adherence**

Findings from the dissertation have theoretical implications for conceptualisation of the role of competent adherence. The findings of Papers 1 and 3 contrast with seminal implementation science theories and models which hypothesise that higher levels of competent adherence are associated with greater improvements in participant outcomes. As the findings on the relationship in the broader parenting programme literature are inconsistent and mixed, the role of competent adherence has yet to be firmly established. As a result, further investigation is required to understand the mechanisms at play in achieving participant outcomes.

### **8.6.2 Who Delivers Parenting Programmes**

The finding that community facilitators with minimal background in and training on parenting programmes appear to have delivered PLH-Teens to a high level of competent adherence has implications for decision-making about who delivers parenting programmes in future. Many evidence-based interventions, including parenting programmes, were designed in high-income countries to be delivered by professionals or researchers with a substantial amount of training related to parenting programmes (Olds et al., 2002; Webster-Stratton & McCoy, 2015). In the parenting programme literature, studies on measures of facilitator competent adherence were largely found to assess small samples of highly educated and experienced facilitators (Martin et al., 2021b). Often, the use of such facilitators to deliver interventions is neither practical nor cost-effective in LMICs where there is frequently an insufficient availability of professional, highly trained staff (Scott et

al., 2008; Scott & Gardner, 2015; Tomlinson et al., 2017). Lay workers, community health workers, paraprofessionals, and others without specific professional or formal education on the topic of the intervention are increasingly being used as facilitators in LMICs (Ayala et al., 2010; Barnett et al., 2018; Viswanathan et al., 2010). Knowledge that community facilitators with little training and little to no background experience in parenting programmes may be able to achieve high levels of competent adherence provides valuable information for the future delivery and scale-up of parenting programmes. This knowledge will be particularly helpful for those who are implementing and scaling up parenting programmes in low-income settings and are considering the use of community facilitators instead of facilitators with substantial backgrounds in parenting programmes.

### **8.6.3 Competent Adherence in Real-World Settings**

The finding that PLH-Teens appears to have been delivered to a high level of competent adherence in a low-income setting at scale during a pandemic has implications for expectations regarding what can be achieved in ‘real-world’ settings. There is concern and evidence that the quality of programme implementation declines as the scale of delivery increases (e.g., Araujo et al., 2021; Institute of Medicine & National Research Council, 2014). However, Paper 2 demonstrates that a parenting programme, and potentially other behavioural interventions, can be delivered in low-income community settings to a high level of competent adherence. This finding bodes well for the future delivery of parenting programmes in low-income community settings at scale as it suggests that high levels of competent adherence are still possible without the substantial resources

and support dedicated to competent adherence monitoring and evaluation in high-income countries.

The finding that PLH-Teens facilitators appeared to have delivered the programme to a high level of competent adherence via routine service delivery at scale may mean that the strategies employed to promote implementation fidelity in PLH are effective, transportable, and sustainable. Strategies used to promote competent adherence in PLH programmes include a straightforward programme manual; experiential and non-didactic PLH-Teens facilitator training; and supportive supervision meetings held between coaches and facilitators.

#### **8.6.4 Training for Assessors**

The finding that the reliability of the PLH-FAT-T Short Form needs strengthening has implications for how assessors are trained. As discussed in the limitations section of this chapter ([8.4.2](#)), assessor reliability may have been poor due to issues of language translation and suboptimal training. As a result, future practice would benefit from ensuring that translation is as accurate as possible and that assessors receive the training necessary to understand the purpose of conducting assessments and have the opportunity to conduct a sufficient number of practice assessments.

#### **8.6.5 Future of Competent Adherence Measurement in PLH**

The findings also have implications for the future measurement of competent adherence in PLH programmes. During the delivery of PLH-Teens in Tanzania, it was difficult for assessors to conduct observational facilitator assessments in practice. The limited quantity of facilitator assessments collected during the delivery of PLH-Teens in Tanzania as well as information gathered during interviews and meetings with coaches on

the barriers encountered provides evidence that the existing tool is not a sufficiently practical and sustainable approach in low-income routine service delivery settings. While some of the challenges experienced with the collection of facilitator assessments may have been unique to COVID-19, the pandemic was only one element complicating the facilitator assessment process in practice. The challenges experienced by assessors and the finding that the tool's reliability and predictive validity needs strengthening suggest that substantial attention and research are required to simplify and refine the process of collecting facilitator assessments if PLH continues to use an observational measure to assess facilitator competent adherence. As a result, the dissertation's findings may prompt PLH researchers and practitioners to consider other methods for assessing facilitators.

## **8.7 Recommendations for Future Research and Practice**

The findings suggest several ways forward for future research and practice related to the measurement and role of facilitator competent adherence in both PLH programmes and the parenting programme community: increase the number of evaluations on measure reliability and validity; enhance study reporting; improve study rigour; investigate the mechanisms working to achieve participant outcomes; explore different assessment approaches to determine the optimal balance of measure feasibility with measure reliability and validity; and reflect on whether observational assessments are worth dedicating substantial time and other resources to in community settings. Although the recommendations outlined in this section are focused on the parenting programme literature, they may also be useful to those measuring and investigating the role of facilitator competent adherence in the broader implementation science literature.

### **8.7.1 Pursue Research on Measure Reliability and Validity**

More research is recommended on the reliability and validity of facilitator competent adherence measures. Paper 1 found that not all studies used measures which have been the subject of psychometric evaluation (see [Paper 1's Supplementary Files](#)). Thus, findings in the literature about the quality of facilitator competent adherence and its impact on participants are potentially based on poor quality measures. Of the 18 studies included in Paper 1, six reported on measures that had not been the subject of any psychometric testing, seven reported on internal consistency (four met COSMIN standards), 11 reported on inter-rater reliability (seven met COSMIN standards), none reported on measure intra-rater reliability, and six studies reported on construct validity (four met COSMIN standards). Investing time and other resources to strengthen and evaluate measures may contribute to a better understanding of the role of competent adherence by eliminating the possibility that the inconsistent evidence found on the relationship between competent adherence and outcomes is due to poor measurement.

The findings of Paper 2 suggest that the PLH-FAT-T Short Form requires further research to strengthen its reliability and validity. Three avenues of additional research are recommended to strengthen the tool. First, as intra- and inter-rater reliability were the PLH-FAT-T Short Form's weakest psychometric properties, it is recommended that future research test whether assessor reliability is improved using approaches such as optimal translation and increased assessor training, including further opportunities for practice assessments. Second, as the transferability of the PLH-FAT-T Short Form to other countries, cultures, and contexts may be limited, it is recommended that the replicability of

the findings of Papers 2 and 3 be tested in other settings at scale. Third, it is recommended that confirmatory factor analyses be performed on the shortened version of the tool.

### **8.7.2 Enhance Study Reporting**

Based on the findings of Paper 1, it is recommended that future studies enhance the quality of reporting on study methods and results. Many of the studies identified in Paper 1 did not sufficiently report on salient study details (e.g., facilitator sample size). A lack of relevant detail as to the methods used by studies was one of two reasons a meta-analysis could not be conducted. To enable a meta-analysis in future, studies could adopt approaches including the pre-registration of study plans and the reporting of all key study details. Further, due to the poor quality of reporting, the literature would benefit from reporting guidelines to support researchers to consistently document critical details on measure characteristics and analysis results. Such reporting guidelines are in progress (Martin et al., in submission-b).

### **8.7.3 Improve Study Rigour**

It is recommended that future research focus greater attention and dedicate greater resources to rigorous evaluation of the relationship between competent adherence and outcomes. As Papers 1 and 3 found inconsistent relationships between competent adherence and outcomes, rigorous evaluation would help eliminate methodological issues as a potential explanation. Rigorous evaluation would also support a future meta-analysis. As part of rigorous evaluation, studies should account for clustering at the facilitator- and parent-levels, carefully select covariates, and place more emphasis on collecting larger quantities of high quality facilitator data that can be linked with participant data.



To collect large samples of facilitator assessments in low-income community settings, it is recommended that researchers work closely with implementers. As discussed in the limitations section of this chapter (8.4.2), the challenges of routine service delivery at scale in a low-income setting during a pandemic meant that the quantity and quality of data collected was limited. Since the collection and management of data requires a detailed and sustained coordination effort by and among all stakeholders (Feely et al., 2018), future researchers might benefit from being more involved with implementers in data planning, collecting, and monitoring. Additionally, during the study planning phase, researchers should determine whether implementing organisations are familiar with the agreed-upon data collection and management procedures. If not, it is recommended that training be designed and delivered to those who will have such responsibilities.

#### **8.7.4 Investigate Mechanisms Achieving Participant Outcomes**

To enhance clarity on the relationship between competent adherence and outcomes, three avenues of future research would support further understanding of the mechanisms working to achieve positive parenting programme outcomes. First, research might explore whether competent adherence plays an indirect role in participant outcomes by examining whether implementation variables (e.g., participant responsiveness) or specific participant outcomes serve as intermediaries between facilitator competent adherence and primary outcomes (e.g., Smith et al., 2013). Second, research might investigate whether competent adherence works in combination with other implementation variables, such as fidelity-consistent adaptations, to achieve participant outcomes (e.g., Skivington et al., 2021; Stirman et al., 2015). Third, as parenting programmes have complex theories of change, an examination of which programme components are working to produce participant outcomes

may support understanding of which aspect of delivery measures should assess (Leijten et al., 2022; Melendez-Torres et al., 2019).

### **8.7.5 Explore How to Balance Rigor and Practicality**

It is recommended that future research pursue four lines of inquiry to investigate how to create rigorous assessments of facilitator competent adherence which are simple to collect in practice. First, it is recommended that researchers investigate simplified versions of existing facilitator assessment tools to see if the revised tools can be conducted in practice and generate reliable data. A study by Suhrheinrich et al. (2020) on the delivery of an intervention for patients with autism spectrum disorder compared a three-point Likert scale version of a tool (used to assess therapist fidelity) to the original five-point version of the tool. The study found that the three-point version could be completed by coders with high levels of concordance (94.0-100.0%) with the five-point version (Suhrheinrich et al., 2020). Additionally, the three-point scale was found to be more feasible for coders (Suhrheinrich et al., 2020). Examining the rigor of simplified tools could help ensure that measures to be used in real-world settings are both practical to implement and scientifically sound.

Second, it is recommended that researchers test whether measures can be completed reliably using simpler, non-observational assessment methods. For instance, self-report and audio-based measures should be examined to determine if they can be completed reliably in the field in comparison to observational measures. A study by Breitenstein et al. (2010a) on the Chicago Parent Program investigated the reliability of self-report assessments and observational assessments finding 85% agreement between these two coder types. A study by Tiwari et al. (2021) on the SafeCare home visiting

programme found 72.0% agreement between the reliability of audio and video assessments. The studies by Breitenstein and Tiwari suggest that self-report and audio-based measures may provide sufficiently reliable metrics of facilitator delivery. Given that these methods are more practical and less resource intensive than observational methods and are therefore more likely to be completed in routine delivery, the potential of sufficiently reliable non-observational approaches are worth exploring. Non-observational measures are especially worth exploring in settings where video recordings are not feasible, or assessors are required to travel long distances to observe facilitator delivery.

Third, it is recommended that researchers conduct further research to determine whether assessing facilitators at one timepoint is an accurate indicator of their overall delivery or whether facilitator delivery improves or declines over time. In PLH-Teens, Shenderovich et al. (2019) examined whether competent adherence changed over time and found facilitators to be consistent in their delivery over time. In the Parent Management Training-Oregon model programme, Forgatch and DeGarmo (2011) examined competent adherence at three timepoints over 18 months following facilitator training and found that facilitator competent adherence improved over the three timepoints but plateaued near the end. The latter findings suggest that facilitators go through a period of improvement before they settle into their full ability to deliver the programme (Forgatch & DeGarmo, 2011). In the Family Check-Up® programme, Chiapa et al. (2015) measured facilitator competent adherence over four years finding a large decline in competent adherence over time. Given these conflicting findings, further research on how facilitator competent adherence varies over time would support decision-making regarding when and how frequently competent adherence should be assessed.

Fourth, it is recommended that researchers consider and evaluate new approaches to measure facilitator competent adherence, such as using computer-based assessments. Berkel, Smith, et al. (under review) are testing the use of natural language processing and machine learning to measure facilitator competent adherence in the Family Check-Up ® programme. A study by Gallo et al. (2015) on the Familias Unidas programme found high levels of reliability when comparing facilitator assessments conducted by computers to those conducted by humans (Kappa score: 0.83). The authors suggest that the result may mean that the human resources required for the assessment of competent adherence could be reduced with technology-based approaches (Gallo et al., 2015). As computer-conducted assessments could save considerable time and other resources, it is worth expanding research on whether automated methods are as reliable and valid as methods conducted by humans. However, computer-based approaches may not be feasible in low-resource contexts, such as Tanzania where collecting and sharing videos was a challenge.

#### **8.7.6 Reflect on Observational Measurement**

It is recommended that researchers and practitioners consider whether observational assessments are a prudent use of substantial time and other resources during the delivery of interventions in low-income community settings at scale. Although observational approaches provide the least biased and most comprehensive assessments of facilitator delivery, they are time consuming, resource intensive, and difficult to implement in practice (Breitenstein et al., 2010b; Feely et al., 2018). Many researchers and practitioners in the parenting programme and broader intervention science literature acknowledge that observational approaches require many steps and present considerable practical challenges (Berkel et al., 2019; Stirman, 2020). Observational measures have been implemented with

some success via randomised trials, yet their transferability to ‘real-world’ settings is undetermined. The number of assessments collected during the three waves of programme delivery in Tanzania is evidence that in-person assessment measures may not be feasible or practical in low-income community settings at scale. Without clear evidence that competent adherence is associated with participant outcomes as theorised, arguments in support of dedicating substantial time and other resources to observational assessments are further weakened in these settings. As a result, it is recommended that the PLH community reflect on whether two types of competent adherence measures are appropriate – a rigorous observational measure for use in randomised evaluations and a more feasible, non-observational measure for use in practice. As it was a Paper 2 finding that the PLH-FAT-T Short Form items were understandable, useful, and meaningful to implementers, the PLH-FAT-T Short Form could be repurposed for use in two ways – observationally in trials and non-observationally in practice. For use in both settings, it is recommended that the PLH-FAT-T Short Form continue to be strengthened and examined in relation to participant outcomes. For use in practice, it is recommended that the further strengthened PLH-FAT-T Short Form items be completed using less burdensome self-report or audio-based methods to generate useful feedback to coaches and facilitators delivering PLH-Teens. Although it was worthwhile to investigate the measurement and role of the observational tool currently being used to assess PLH facilitator competent adherence, the dissertation’s findings support reconsideration of whether observational approaches are the most appropriate use of time and other resources in all evaluations and in all settings.

## 8.8 Dissemination Strategies

A variety of strategies have been and are being employed to ensure that the dissertation's findings are used to improve the implementation and scale-up of parenting programmes aiming to reduce violence against children and child behaviour problems in Tanzania and other LMICs. Specifically, I have been collaborating and engaging with key stakeholders involved in parenting programmes in Tanzania to end violence against children. For instance, I co-presented the results of our mixed methods evaluation to approximately 30 Tanzanian government and non-governmental officials (February 2022). Attendees at this meeting included staff from the Prime Minister's Office, Ministry of Education, UNICEF Tanzania, and Ministry of Community Development, Gender, and Children. I co-created a policy brief that was shared with attendees to report on the main results and learnings, which has been published (Martin et al., 2022b). As staff from CWBSA are involved in the collection of PLH facilitator assessments around the world, I participated in meetings with representatives from their organisation to discuss the findings of Paper 2 (April 2022). The purpose of the meeting was to inform the future conduct and collection of assessments. I plan to have further meetings with partners from CWBSA to reflect and collaborate on the findings of Paper 3 as well as on the recommendations stemming from the dissertation. Additionally, along with other members of the FAIR research team, I was invited to contribute emerging findings from the FAIR study to a team of researchers who wrote a report to inform the World Health Organization's parenting programme guidelines (Gardner et al., 2023; WHO, 2023). Finally, I presented the plans for and/or findings of the three papers on several occasions including in the Department of Social Policy and Intervention at Oxford; at an invited lecture at the Faculty of Health

Sciences at Simon Fraser University in Canada (April 2022); at the Comparative and International Education Society 66th Annual Conference (April 2022, invited panel presentation); to colleagues in the Department of Psychology at the University of Cape Town (July 2022); at the Sexual Violence Research Initiative Forum (September 2022); and at the 8<sup>th</sup> International Congress of Clinical and Health Psychology in Children and Adolescents (November 2022, invited panel presentation). I will present the findings at the Society for Prevention Research Annual Meeting (May 2023, symposium presentation). I also had the opportunity to present the findings of Paper 1 at the 2021 Society for Prevention Research Annual meeting. Following the presentation, a seminal academic in the field (Professor Breitenstein, Ohio State University) proposed that we collaborate on the recommendations stemming from this paper. As a result, I have been working with implementation science researchers from the United States, United Kingdom, and Uganda on a commentary advocating that parenting researchers devote greater attention to examining facilitator delivery as well as a Delphi study to develop a reporting guideline on facilitator delivery. The commentary along with the first draft of the reporting guideline is now in submission to an academic journal.

## 9. Conclusion

There is now substantial evidence demonstrating the effectiveness of parenting programmes in reducing violence against children and child behaviour problems. This evidence suggests that parenting programmes could have a positive global impact on family health and well-being if implemented in community settings on a large-scale (e.g., Gardner et al., 2023; WHO, 2023). The considerable challenges faced by families in LMICs indicate that parenting programmes would be of particular value to families in these contexts.

Although the evidence supporting the effectiveness of parenting programmes in LMICs is growing, little is known on the extent to which parenting programmes can be delivered with competent adherence in these settings including at scale. Little is also known regarding how to measure competent adherence and its role in achieving participant outcomes in both PLH and the broader parenting programme literature. Implementation science therefore has much to contribute to the dissemination and scale-up of parenting programmes.

This dissertation sought to better understand the measurement and role of facilitator competent adherence in achieving intended parent and child outcomes in the parenting programme literature as well as within PLH-Teens. The dissertation synthesised the relationship between facilitator competent adherence and outcomes reported in 18 studies from the parenting programme literature; conducted a psychometric evaluation of the measure used to examine the competent adherence of PLH-Teens facilitators; reported on the competent adherence of community facilitators who delivered PLH-Teens to 75,061 participants during the 2020-2021 scale-up of PLH-Teens in Tanzania; and used multi-level Poisson regressions to examine the relationship between facilitator competent adherence



and the outcomes of the adolescent girls and parents who participated in PLH-Teens in Tanzania.

Overall, the dissertation found that facilitator competent adherence appears beneficial for outcomes, to some extent, in both the broader parenting programme literature and delivery of PLH-Teens in Tanzania; the PLH-FAT-T Short Form shows promise but needs further work to increase its reliability and validity; and PLH-Teens appeared to have been delivered to a high level of competent adherence by community facilitators with minimal background in and training on delivering parenting programmes. These findings have implications for how competent adherence is conceptualised; understanding about the level of competent adherence that can be achieved in practice; how assessors of facilitator delivery are trained; and the future of competent adherence measurement in PLH programmes. As a result, it was recommended that future studies conduct evaluations on the reliability and validity of measures of competent adherence; the parenting programme community enhance study reporting and rigour; future studies investigate the relationship between competent adherence and other implementation fidelity variables; future studies explore how to balance measure feasibility and rigour; and the PLH community reflect on whether observational assessments are appropriate in all evaluations and in all settings.

As parenting programmes continue to be delivered and scaled worldwide, the findings and recommendations of this dissertation are intended to be used to maximise the ability of vulnerable children and families globally to benefit from evidence-based parenting programmes – particularly in the 35 LMICs where PLH programmes are delivered.

## Appendices

### Appendix 1: Study Protocols

Please see the following published protocols outlining plans for the FAIR study and the systematic review included in Paper 1:

**Martin, M.**, Lachman, J., Wamoyi, J., Shenderovich, Y., Wambura, M., Mgunga, S., ... & Manjengenga, N. (2021). A mixed methods evaluation of the large-scale implementation of a school-and community-based parenting program to reduce violence against children in Tanzania: A study protocol. *Implementation Science Communications*, 2(1), 1-13. <https://doi.org/10.1186/s43058-021-00154-5>.

**Martin, M.**, Gardner, F., Lachman, J., & Steele, B. (2020). Protocol - Measures of facilitator competence and/or adherence used in parenting programmes: A three-part systematic review and meta-analysis. PROSPERO, International Prospective Register of Systematic Reviews.  
[https://www.crd.york.ac.uk/prospERO/display\\_record.php?ID=CRD42020167872](https://www.crd.york.ac.uk/prospERO/display_record.php?ID=CRD42020167872)

## **Appendix 2: Glossary**

**Child Behaviour Problems:** Conduct or emotional problems characterised by externalising (e.g., aggression, antisocial behaviour) or internalising (e.g., anxiety, depression) behaviours by children (Campbell et al., 2000; Eisenberg et al., 2001; Liu et al., 2011; McMahon & Frick, 2019).

**Competent Adherence:** The accuracy and style with which a facilitator delivers programme components (Breitenstein et al., 2010a; Fixsen et al., 2005; Forgatch et al., 2005).

**Facilitator:** A practitioner or lay person who delivers a programme to participants (Fixsen et al., 2005). A facilitator may also be referred to as a ‘purveyor’ (Fixsen et al., 2005, p. 537), ‘clinician’, ‘therapist’, or ‘practitioner’.

**Facilitator Assessment Tool:** The observational implementation fidelity measure used in Parenting for Lifelong Health to assess facilitator competent adherence.

**Furaha Teens:** The adaptation of Parenting for Lifelong Health for Parents and Adolescents (PLH-Teens) delivered in the Tanzanian context. The programme title translates to ‘Happy Teens’ in Kiswahili.

**Implementation Science:** A field that examines how interventions unfold in practice and uses this information to improve intervention implementation and outcomes (Bhattacharyya et al., 2009; Mihalic, 2004; Peters et al., 2013).

**Implementation Fidelity:** The extent to which an intervention is implemented as intended by programme developers and as outlined in its logic model or programme manual (Bumbarger & Perkins, 2008; Dane & Schneider, 1998; Dusenbury et al., 2003).

**Parenting Programmes:** Interventions that seek to support parents/caregivers – those responsible for the care and/or upbringing of a child between the ages of 0 and 18 irrespective of biological relationships – to acquire the knowledge and skills that will enable them to maintain and improve the health and well-being of their children (Barlow & Coren, 2018).

**Parenting for Lifelong Health:** A suite of group-based parenting programmes aiming to reduce familial violence against children and child behavioural and emotional problems (Lachman et al., 2016). The programmes are targeted at parents with children across the development spectrum: PLH for Babies (prenatal-6 months), PLH for Toddlers (10-60 months), PLH for Young Children (PLH-YC) (2-9 years), and PLH for Teens (PLH-Teens) (10-17 years). PLH programmes were developed and evaluated by researchers and practitioners from several universities and non-profit organisations, including the University of Oxford, the University of Cape Town, the University of Stellenbosch, Clowns Without Borders South Africa, UNICEF, and the World Health Organization. PLH programmes are being implemented in over 35 LMICs in sub-Saharan Africa, Southeast Asia, Latin America, and Eastern Europe. To date, PLH programmes have reached hundreds of thousands of beneficiaries.

**Reliability:** The degree to which a measure produces consistent outputs across multiple measurements under various conditions (Fan & Randall, 2018; Mokkink et al., 2010a)

**Scale-up:** Defined as “deliberate efforts to increase the impact of health innovations successfully tested...so as to benefit more people and foster the development of sustainable policies and programs” (Cash, 2011, p. 3).

**Validity:** The degree to which a measure accurately reflects its intended constructs

(Mokkink et al., 2010a; Thorkildsen, 2010).

**Violence Against Children:** According to the World Health Organization (2022),

“violence against children includes all forms of violence against young people under 18 years old, whether perpetrated by parents or other caregivers, peers, romantic partners, or strangers” (p.1).

### **Appendix 3: List of Acronyms**

AIECP – ‘Accept-Explore-Connect-Practice’

COSMIN – Consensus-Based Standards for the Selection of Health Measurement Instruments

CWBSA – Clowns Without Borders South Africa

DREAMS – Determined, Resilient, Empowered, AIDS-free, Mentored, and Safe Initiative run by USAID

FAIR – Furaha Adolescent Implementation Research study

ICC – Intra-class correlation

KMO – Kaiser-Meyer-Olkin statistic

LMIC – Low- and middle-income country

Pact – Pact Tanzania

PLH – Parenting for Lifelong Health

PLH-YC – Parenting for Lifelong Health for Young Children

PLH-Teens – Parenting for Lifelong Health for Parents and Adolescents

PLH-FAT-YC – Parenting for Lifelong Health for Young Children Facilitator Assessment Tool

PLH-FAT-T – Parenting for Lifelong Health for Parents and Adolescents Facilitator Assessment Tool

PLH-FAT-T Short Form – A shortened version of the PLH-FAT-T produced via Paper 2

RE-AIM – Reach, Efficacy, Adoption, Implementation and Maintenance Framework

SWiM – Synthesis Without Meta-Analysis

UNICEF – United Nations Children’s Fund

USAID – United States Agency for International Development

#### Appendix 4: PLH-FAT-YC

##### Parenting for Lifelong Health for Young Children -Facilitator Assessment Tool (PLH-YC-FAT)

Assessor Name:			
Facilitator Name:		Facilitator ID:	
Assessment Date:		Session Number and Date:	
Video File Name:		Session/Video Length:	
Number of Enrolled Parents:		Number of Parents in Attendance:	
Facilitator Age:		Facilitator Gender:	
Has the facilitator been assessed before (Y/N) ?		If yes, how many times has the facilitator been assessed previously?	
Co-facilitator name:		Facilitator Condition	

##### ACTIVITY SUBSCALE

HOME ACTIVITY DISCUSSION	Inadequate 0	Needs Improvement 1	Good 2	Excellent 3
<b>SESSION NUMBER:</b> _____ <b>START TIME ON VIDEO:</b> _____ <b>END TIME ON VIDEO:</b> _____				
1. Remind parents of the core home activity for the previous week at the beginning of the activity				

2. Review core building blocks from previous session with parents at the beginning of the activity				
3. Parents share experiences of how home activity went during the week				
4. Keep parents focused on core home activity				
5. Help parents connect experiences to the core building blocks				
6. Explore at least one specific challenge experienced by a parent regarding the main home activity				
7. Explore solutions to challenge shared				
8. Help parents choose an appropriate and specific solution				
9. At least one parent practice solutions to challenges <b>OR</b> ways to improve parenting skills				
10. Debrief with the parents after practicing solutions to challenges				
11. Praise and encourage parents to try solution at home				
12. Thank and praise parents for sharing experiences (at the end of the home activity discussion)				
Comments/Notes:				



**FREQUENCY SUBSCALE**

Please count the number of discrete times the facilitator used reflexive statements and praise (specific/unspecific) during the **FIRST 20 MINUTES OF THE HOME ACTIVITY**

<b>Behaviour</b>	<b>Frequency</b>	
<b>Reflexive Statements</b>		
<b>Praise</b>	<b>Unspecific:</b>	<b>Specific</b>

<b>ILLUSTRATED STORY DISCUSSION</b>	<b>Inadequate 0</b>	<b>Needs Improvement 1</b>	<b>Good 2</b>	<b>Excellent 3</b>
<b>SESSION NUMBER:</b> _____ <b>START TIME ON VIDEO:</b> _____ <b>END TIME ON VIDEO:</b> _____				
1. Read through story with parents				
2. Explore actions, behaviours, and emotions				
3. Make sure that the questions in the manual have been covered				
4. Identify core building blocks connected to the story				
Comments/Notes:				

<b>GROUP PRACTICE</b>	<b>Inadequate 0</b>	<b>Needs Improvement 1</b>	<b>Good 2</b>	<b>Excellent 3</b>
<b>SESSION NUMBER:</b> _____ <b>START TIME ON VIDEO:</b> _____ <b>END TIME ON VIDEO:</b> _____				
<b>BIG GROUP PRACTICE</b>				
1. Establish roles for big group practice (e.g. parent and child)				
2. Set up scenario and use space appropriately				
3. Describe exactly what “parent” and “child” will be doing during the group practice				
4. Give support to parents during group practice (shadow)				
5. Debrief with ‘parent’ about experiences and feelings				
6. Debrief with ‘child’ about their experience and feelings				
7. Thank and praise parents who practiced in big group				

SMALL GROUP PRACTICE				
8. Practice in pairs while supporting around room				
9. Debrief with parents after practicing in pairs				
10. Thank and praise parents				
Comments/Notes:				

#### SKILLS SUBSCALE

MODELLING BEHAVIOUR	Inadequate 0	Needs Improvement 1	Good 2	Excellent 3
1. Give lots of positive reinforcement and specific praise to parents				
2. Give positive, specific, and realistic instructions				
3. Maintain commitments to time management principles				
4. Model behaviours with co-facilitator				
5. Demonstrate respectful behaviour towards parents				
Comments/Notes:				

<b>ACCEPT-EXPLORE- CONNECT-PRACTICE</b>		<b>Inadequate 0</b>	<b>Needs Improvement 1</b>	<b>Good 2</b>	<b>Excellent 3</b>
<b>ACCEPT</b>					
1. Uses body language to show acceptance					
2. Accepts parent responses verbally					
3. Openness					
4. Reflexive statements					
<b>EXPLORE</b>					
5. Explore experience/opinion of parent in detail using open-ended questions					
6. Explore thoughts and feelings					
7. Explore child perspective					
<b>CONNECT</b>					
8. Connect experiences to the building blocks from session					
9. Connects individual experiences to universal principles					
<b>PRACTICE</b>					
10. Identify opportunities to practice skills (in addition to					

the structured group practice)				
Comments:				
<b>COLLABORATIVE LEADERSHIP</b>	<b>Inadequate 0</b>	<b>Needs Improvement 1</b>	<b>Good 2</b>	<b>Excellent 3</b>
1. Arrange room in a way that encourages equal and active participation				
2. Facilitator is situated within the group, is at the level of the parents, and in a different place than the co-facilitator				
3. Parents appear engaged in session				
4. Parents appear comfortable and satisfied				
5. Parents share views and opinions				
6. Parent-facilitator participation ratio				
7. Assure equal and active participation among parents				
8. Targeted engagement of quiet or non-participating parents				
9. Limiting parent responses				
10. Keep parents on topic of discussion				

11. Demonstrate knowledge of session content				
12. Level of self-confidence with session content				
13. Help parents generate their own ideas regarding principles or solutions to challenges				
14. Help parents assess consequences of proposed solutions				
15. Allow parents to choose their own solutions to challenges				
16. Make sure that solutions are positive, specific, and realistic				
17. Maintain leadership and control of the group				
18. Work well with the co-facilitator				
Comments:				

## OVERALL ASSESSMENT

Activities assessment		Skills assessment	
Total score on core activities (A)		Total score core facilitation skills (C)	
Total possible score (B)	<b>78</b>	Total possible score (D)	<b>99</b>
Total percent score core activities = (A/B) x 100 %	%	Total percent score core skills = (C/D) x 100 %	%
<b>What are the Facilitator's strengths?</b>			
<b>What does the Facilitator need to improve?</b>			
<b>Recommendations:</b>			

## Appendix 5: Summary of Changes to PLH-FAT-YC

Summary of Recommendations and Changes to FAT																	
<u>Recommendation to Improve the FAT</u>	<u>Stakeholder Group</u>	<u>Changes Made or Example of Changes Made</u>															
Break up complex items into separate, simple items	CWBSA Trainers	They recommended that the item, “Did the facilitator accept participant responses verbally and physically?” be divided into four items to capture whether the facilitator demonstrated physical acceptance (e.g., nodding), verbal acceptance (e.g., “mhm”), openness (e.g., “Interesting suggestion!”), and use of a reflexive statement (e.g., “Am I understanding you to say that you will schedule daily time to play with your child?”).															
Use specific definitions for each item and Likert point	CWBSA Trainers	<table><tr><th colspan="5">COLLABORATIVE LEADERSHIP</th></tr><tr><th></th><th>0 Inadequate</th><th>1 Needs Improvement</th><th>2 Good</th><th>3 Excellent</th></tr><tr><td>Did the facilitator...  1. Arrange room in a way that encourages equal and active participation</td><td>Chairs are not set in a circle or semi-circle but rather in lecture format. The flipchart, house of support, and/or illustrated stories are not visible to any of the participants.</td><td>Chairs are set in a circle or semi-circle but there are many empty chairs. The flipchart, house of support, and/or illustrated stories are not visible to all of the participants. No mixture of participants in terms of age and gender (e.g., teens and parents are mixed for PLH Teens).</td><td>Chairs are set in a circle or semi-circle, everyone can see each other. Flipchart, house of support, and/or illustrated stories are visible to all participants (if appropriate). Some empty chairs. Some mixture of participants in terms of age and gender (e.g., teens and parents are mixed for PLH Teens).</td><td>Chairs are set in a circle or semi-circle, everyone can see each other. Flipchart, house of support, and/or illustrated stories is visible to all participants (if appropriate). Every chair is full. Good mixture of participants in terms of age and gender (e.g., teens and parents are mixed for PLH Teens).</td></tr></table>	COLLABORATIVE LEADERSHIP						0 Inadequate	1 Needs Improvement	2 Good	3 Excellent	Did the facilitator...  1. Arrange room in a way that encourages equal and active participation	Chairs are not set in a circle or semi-circle but rather in lecture format. The flipchart, house of support, and/or illustrated stories are not visible to any of the participants.	Chairs are set in a circle or semi-circle but there are many empty chairs. The flipchart, house of support, and/or illustrated stories are not visible to all of the participants. No mixture of participants in terms of age and gender (e.g., teens and parents are mixed for PLH Teens).	Chairs are set in a circle or semi-circle, everyone can see each other. Flipchart, house of support, and/or illustrated stories are visible to all participants (if appropriate). Some empty chairs. Some mixture of participants in terms of age and gender (e.g., teens and parents are mixed for PLH Teens).	Chairs are set in a circle or semi-circle, everyone can see each other. Flipchart, house of support, and/or illustrated stories is visible to all participants (if appropriate). Every chair is full. Good mixture of participants in terms of age and gender (e.g., teens and parents are mixed for PLH Teens).
COLLABORATIVE LEADERSHIP																	
	0 Inadequate	1 Needs Improvement	2 Good	3 Excellent													
Did the facilitator...  1. Arrange room in a way that encourages equal and active participation	Chairs are not set in a circle or semi-circle but rather in lecture format. The flipchart, house of support, and/or illustrated stories are not visible to any of the participants.	Chairs are set in a circle or semi-circle but there are many empty chairs. The flipchart, house of support, and/or illustrated stories are not visible to all of the participants. No mixture of participants in terms of age and gender (e.g., teens and parents are mixed for PLH Teens).	Chairs are set in a circle or semi-circle, everyone can see each other. Flipchart, house of support, and/or illustrated stories are visible to all participants (if appropriate). Some empty chairs. Some mixture of participants in terms of age and gender (e.g., teens and parents are mixed for PLH Teens).	Chairs are set in a circle or semi-circle, everyone can see each other. Flipchart, house of support, and/or illustrated stories is visible to all participants (if appropriate). Every chair is full. Good mixture of participants in terms of age and gender (e.g., teens and parents are mixed for PLH Teens).													
Add items to capture missing activities and skills	CWBSA Trainers	“Did the facilitator identify core building blocks connected to the story?” was added to the illustrated story items in the Activities Subscale.															
Remove redundant items	CWBSA Trainers	The trainers recommended deleting the item, “Did the facilitator provide frequent praise throughout the discussion?” since praise was already incorporated into many questions.															
Create the Frequency Subscale	Parenting Programme Experts	Three additional items were added to the FAT: “Please record the number of discrete times the facilitator used reflexive statements and praise (specific/unspecific) during first twenty minutes of the home activity discussion: (1) reflexive statements, (2) specific praise, and (3) unspecific praise.”															
Changes to item wording	Assessors	The item “Accepts parent responses physically” was changed to “Uses body language to show acceptance”.															



Proposed examples to include in item definitions

Assessors

MODELLING BEHAVIOUR				
Did the facilitator...	0 Inadequate	1 Needs Improvement	2 Good	3 Excellent
1. Give positive reinforcement and specific praise to parents	The facilitator <u>never</u> or <u>almost never</u> praised parents.	The facilitator <u>rarely</u> praised parents or it was <u>rarely</u> specific or genuine. Example: "Thank you, Patty."	The facilitator <u>often</u> praised parents. Praise was <u>often</u> specific and genuine. Example: "Thank you, Patty, for sharing about your experience at home."	The facilitator <u>almost always</u> praised parents throughout the session. Praise was <u>almost always</u> specific, genuine, and reinforced the core building blocks. Example: "Well done, Patty, for involving your child when establishing Household Rules."

## Appendix 6: Ethics Forms and Approvals

### Department of Social Policy and Intervention DREC Checklist

Name	Mackenzie Martin
Student status	PRS
Supervisor	Professor Frances Gardner and Dr Jamie Lachman
Project start and end dates	2020-2021
Title	The Role of Programme Facilitator Characteristics and Competent Adherence in the Scale-Up of the Parenting for Lifelong Health-Teens Programme in Tanzania
Brief description	The proposed DPhil project is composed of four studies using data from a systematic review of the literature and the 2020-2021 scale-up of Parenting for Lifelong Health (PLH)-Teens to 50,000 parent-teen dyads (N=100,000) by school teachers and community volunteers (N=444) in Tanzania. PLH-Teens is a low-cost open-access parenting programme aiming to reduce CBP and VAC in low- and middle-income countries with substantial emerging evidence underpinning its effectiveness. The first paper will report on associations between levels of facilitator competent adherence assessed using observational measures of facilitator competent adherence and parent and child outcomes using data from a systematic review, which has received approval from the Department (SPI_DREC_20_008). The subsequent three papers will use data from the Department's FAIR study, of which I am the research manager. The data collection and analyses proposed for this study have been approved by a CUREC 2a (R64777/RE001). The FAIR study has also received ethics approval in Tanzania (NIMR/HQ/R.8a/Vol.IX/3459) and by partner organizations (NIMR/HQ/R.8a/Vol.IX/2902 and RE002 HEY BABY).
Methods [please tick all that apply]	<p>If you have ticked boxes 1, 2, 3 or 4 only, please ask your supervisor to sign below to confirm, and submit this form to <a href="mailto:ethics@spi.ox.ac.uk">ethics@spi.ox.ac.uk</a></p> <p>Date.....2021-06-07.....</p> <p>Signed by student.....<i>Mackenzie Martin</i>.....</p> <p>Signed by supervisor.....<i>[Signature]</i>.....</p> <p>If you have ticked boxes 5, 6 or 7 you may need to fill in a CUREC 1a checklist [available from <a href="http://www.admin.ox.ac.uk/curec/apply/ssb-idrec-process/">http://www.admin.ox.ac.uk/curec/apply/ssb-idrec-process/</a>] and send an electronic copy to <a href="mailto:ethics@spi.ox.ac.uk">ethics@spi.ox.ac.uk</a>. Please read the FAQ here: <a href="http://www.admin.ox.ac.uk/curec/faqs-glossary/faqs/">http://www.admin.ox.ac.uk/curec/faqs-glossary/faqs/</a> to determine whether your project requires a CUREC before filling one out.</p> <p>Please make sure you download the CUREC checklist from the above website each time you fill it in, in order to make sure you are using the most up-to-date version.</p> <p>Please note that CUREC 1a will request further information and give guidance as to whether a CUREC 2 should also be submitted.</p> <p>You MUST allow enough time for your research to be ethically approved. If a CUREC/2 application is required this can take up to sixty days.</p>
<p>1. Literature review <input type="checkbox"/></p> <p>2. Systematic review <input checked="" type="checkbox"/></p> <p>3. Theoretical papers <input type="checkbox"/></p> <p>4. Documentary Analysis <input type="checkbox"/></p> <p>5. Secondary analysis <input checked="" type="checkbox"/></p> <p>6. Survey <input checked="" type="checkbox"/></p> <p>7. Qualitative <input checked="" type="checkbox"/></p> <p>I confirm that I am using data that was collected before the project was formulated, is fully anonymised and is publicly available. (Checking this box means you will not need to complete a CUREC) <input type="checkbox"/></p>	



Department of Social Policy and Intervention  
University of Oxford

Barnett House, 32 Wellington Square,  
Oxford, OX1 2ER, United Kingdom  
[www.spi.ox.ac.uk](http://www.spi.ox.ac.uk)

Mackenzie Martin  
Department of Social Policy and Intervention  
University of Oxford

SPI\_DREC\_20-21\_026

8<sup>th</sup> June 2021

Dear Mackenzie

**The Role of Programme Facilitator Characteristics and Competent Adherence in the Scale-Up of the Parenting for Lifelong Health-Teens Programme in Tanzania**

Your application for research ethics approval in connection with your research project has been considered by the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for Ethical Approval.

I am pleased to inform you that, on the basis of the information provided, the proposed research has been judged as meeting appropriate ethical standards and DREC approval has been granted.

If any revisions to your research methodology are made subsequent to this approval, these must be detailed in writing and submitted to DREC immediately.

Yours sincerely,

A handwritten signature in black ink, appearing to be 'J. Lachman', written over a horizontal line.

Dr Jamie Lachman  
Chair of DREC

SOCIAL SCIENCES & HUMANITIES INTERDIVISIONAL RESEARCH ETHICS COMMITTEE

Research Services, University of Oxford, Wellington Square, Oxford OX1 2JD  
Tel: +44(0)1865 616576 Fax: +44(0)1865 280467  
[ethics@socsci.ox.ac.uk](mailto:ethics@socsci.ox.ac.uk)



11 September 2020

Dr Jamie Lachman  
Department of Social Policy and Intervention

Dear Jamie

**Research Ethics Approval (CUREC 2)**

**Ref No: R64777/RE001**

**Title: Furaha Adolescent Implementation Research (FAIR) Study**

The above application has been considered on behalf of the Social Sciences and Humanities Interdivisional Research Ethics Committee (IDREC) in accordance with the procedures laid down by the University for ethical approval of all research involving human participants.

I am pleased to inform you that, on the basis of the information provided to the IDREC, the proposed research has been judged as meeting appropriate ethical standards, and accordingly approval has been granted.

Should there be any subsequent changes to the project that raise ethical issues not covered in the original application you should submit details to the IDREC for consideration:

<https://researchsupport.admin.ox.ac.uk/governance/ethics/apply/sshidrec#collapse394916>.

Please note that your study may be selected for review by the SSH IDREC during an annual audit. You may also be required to submit a brief annual progress report on each anniversary of study approval, until the study is completed.

Yours sincerely,

A handwritten signature in black ink, appearing to read "J Blaikie".

Jennifer Blaikie  
Research Ethics Manager

cc: Olivia Thornton



THE UNITED REPUBLIC  
OF TANZANIA



National Institute for Medical Research  
3 Barack Obama Drive  
P.O. Box 9653  
11101 Dar es Salaam  
Tel: 255 22 2121400  
Fax: 255 22 2121360  
E-mail: [nimrethics@gmail.com](mailto:nimrethics@gmail.com)

NIMR/HQ/R.8a/Vol. IX/3459

Ministry of Health, Community  
Development, Gender, Elderly & Children  
University of Dodoma, College of  
Business Studies and Law  
Building No. 11  
P.O. Box 743  
40478 Dodoma

30<sup>th</sup> June 2020

Dr Joyce Wamoyi  
Principal Research Scientist  
National Institute for Medical Research - Mwanza Centre  
P O BOX 1462  
Mwanza

RE: ETHICAL CLEARANCE CERTIFICATE FOR CONDUCTING  
MEDICAL RESEARCH IN TANZANIA


This is to certify that the research entitled: Furaha Adolescent Implementation Research Study (FAIR) (Wamoyi J. et al), has been granted ethical clearance to be conducted in Tanzania.

The Principal Investigator of the study must ensure that the following conditions are fulfilled:


1. Progress report is submitted to the Ministry of Health, Community Development, Gender, Elderly & Children and the National Institute for Medical Research, Regional and District Medical Officers after every six months.
2. Permission to publish the results is obtained from National Institute for Medical Research.
3. Copies of final publications are made available to the Ministry of Health, Community Development, Gender, Elderly & Children and the National Institute for Medical Research.
4. Any researcher, who contravenes or fails to comply with these conditions, shall be guilty of an offence and shall be liable on conviction to a fine as per NIMR Act No. 23 of 1979, PART III Section 10(2).
5. Sites: Kagera, Mbeya and Shinyanga regions.

Approval is valid for one year. 30<sup>th</sup> June 2020 to 29<sup>th</sup> June 2021.

Name: Prof. Yunus Daud Mgaya

  
Signature  
CHAIRPERSON  
MEDICAL RESEARCH  
COORDINATING COMMITTEE

Name: Prof. Abel Nkono Makubi

  
Signature  
CHIEF MEDICAL OFFICER  
MINISTRY OF HEALTH, COMMUNITY  
DEVELOPMENT, GENDER, ELDERLY &  
CHILDREN

CC: Director, Health Services-TAMISEMI, Dodoma  
RMO of Kagera, Mbeya and Shinyanga regions.  
DMO/DED of respective districts.



THE UNITED REPUBLIC  
OF TANZANIA



National Institute for Medical Research  
3 Barack Obama Drive  
P.O. Box 9653  
11101 Dar es Salaam  
Tel: 255 22 2121400  
Fax: 255 22 2121360  
Email: [nimrethics@gmail.com](mailto:nimrethics@gmail.com)

Permanent Secretary (Health)  
Ministry of Health, Community  
Development, Gender, Elderly & Children  
Government City Mtumba, Health Road  
P.O. Box 743  
40478 Dodoma

NIMR/HQ/R.8c/Vol. I/1634

11<sup>th</sup> December, 2020

Dr. Amon Exavery  
Pact Tanzania  
C/o Dr. Naftali Ng'ondi  
MoHCDGEC  
P O Box 32652  
Dar es Salaam


**RE: APPROVAL FOR EXTENSION OF ETHICAL CLEARANCE**

This letter is to confirm that your application for extension on the already approved proposal: Inspiring an evidence-base for parenting programmes to end child abuse: including Tanzania data in a multi-country secondary analysis (Exavery A. et al) whose local investigator is Dr. Naftali Ng'ondi of MoHCDGEC, has been approved.

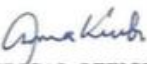
The extension approval is based on the progress report dated 18<sup>th</sup> December, 2020 on the project, Ref. NIMR/HQ/R.8a/Vol. IX/2902, dated 03<sup>rd</sup> October, 2018. Extension approval is valid until 02<sup>nd</sup> October, 2021.

The Principal Investigator must ensure that other conditions of approval remain as per ethical clearance letter. The PI should ensure that progress and final reports are submitted in a timely manner.

Name: Prof. Yunus Daud Mgaya

  
Signature  
CHAIRPERSON  
MEDICAL RESEARCH  
COORDINATING COMMITTEE

Name: Prof. Abel Nkono Makubi

  
Signature  
CHIEF MEDICAL OFFICER  
MINISTRY OF HEALTH, COMMUNITY  
DEVELOPMENT, GENDER, ELDERLY  
& CHILDREN





Department of Social Policy and Intervention  
University of Oxford

Barnett House, 32 Wellington Square,  
Oxford, OX1 2ER, United Kingdom  
[www.spi.ox.ac.uk](http://www.spi.ox.ac.uk)

Professor Lucie Cluver  
Department of Social Policy and Intervention  
University of Oxford

SPICUREC1a\_\_20\_015

30<sup>th</sup> July 2020

Dear Lucie,

**Parenting for Lifelong Health: Scale-Up of Parenting Evaluation Research (PLH-SUPER)**

Your application for research ethics approval in connection with your research project has been considered by the Departmental Research Ethics Committee (DREC) in accordance with the procedures laid down by the University for Ethical Approval.

I am pleased to inform you that, on the basis of the information provided, the proposed research has been judged as meeting appropriate ethical standards and DREC approval has been granted.

If any revisions to your research methodology are made subsequent to this approval, these must be detailed in writing and submitted to DREC immediately.

Yours sincerely,

A handwritten signature in cursive script, appearing to read 'Jane Barlow'.

Professor Jane Barlow  
Research Director and Acting Chair of DREC

## **Appendix 7: Paper 1 Supplementary Files**

### **Supplementary File 1 – Details of the Martin et al. 2021 review**

#### *Search Strategy*

To develop the search strategy, relevant literature was consulted to determine the variety of terms used to capture similar concepts. Both grey and academic literature were referenced including several systematic reviews. To illustrate, early childhood development and parenting intervention terms were included from each of Barlow's and Wight's reviews (Barlow et al., 2017; Gardner et al., forthcoming-a). Next, the terms were tested in various electronic databases and then pooled (*Table 1*). To increase search sensitivity, the words listed in each column of the table were combined with the Boolean operator "OR" (Reznowski, 2011). Further, terms with a variety of endings or that could be pluralised were truncated using "\*". Variations in the spelling of terms were explored. To increase search specificity, the terms in each column were then combined using the Boolean operator "AND" (Reznowski, 2011). The search was limited to titles and abstracts to increase search specificity, however the breadth and variety of the included terms as well as the number of databases used allowed for greater search sensitivity (Rose, 2016; Watson & Richardson, 1999). Achieving a balance between search sensitivity and specificity is delicate; if a search is more sensitive, it increases the time spent screening irrelevant articles but it also increases the likelihood of capturing more relevant articles whereas if a search is more specific, it reduces the time spent screening irrelevant articles but it also reduces the likelihood of capturing more relevant articles (Taylor et al., 2007).



**Table 3**  
*Search Terms*

	<u>POPULATION</u>		<u>INTERVENTION</u>	<u>OUTCOME</u>	
<u>Parents</u>	<u>Children</u>	<u>Facilitator</u>	<u>Programme</u>	<u>Measure</u>	<u>Fidelity</u>
Parent*	Child*	Facilitator*	Training	Scale*	Competen*
Caregiver*	Kid*	Practitioner*	Program*	Sub-scale*	Quality
Guardian*	Adolesc*	Therapist*	Intervention*	Tool*	Adheren*
Carer*	Teen*	Clinician*	Treatment	Measure*	Fidelity*
	Youth*	Teacher*	Trial*	Instrument*	Integrity
	Baby	Worker*	Prevention	Report*	Compliance
	Babies	Provider*		Index*	
	Toddler*	Leader*		Checklist*	
	Neonate*	Specialist*		Test*	
	Infant*	Professional*			
	Newborn	Coordinator*			
	Juvenile*	Administrator*			
	Minor*	Counsellor*			
	Early child*	Counselor*			
	ECD	Implementer*			
		Coach*			
		Instructor*			
		Trainer*			
		Mentor*			
		Educator*			

The resulting search strategy was:

(parent\* OR caregiver\* OR guardian\* OR carer\*.ab) AND (training OR program\* OR intervention\* OR treatment OR trial\* or prevention.ab) AND (competen\* OR quality OR adheren\* OR fidelity\* OR integrity OR compliance.ab) AND (child\* OR kid\* OR adolesc\* OR teen\* OR youth\* OR baby OR babies OR toddler\* OR neonate\* OR infant\* OR newborn OR juvenile\* OR minor\* OR early child\* OR ECD.ab) AND (facilitator\* OR practitioner\* OR therapist\* OR clinician\* OR teacher\* OR worker\* OR provider\* OR leader\* OR specialist\* OR professional\* OR coordinator\* OR administrator\* OR counsellor\* OR counselor\* OR implementer\* OR coach\* OR instructor\* OR trainer\* OR mentor\* OR educator\*.ab) AND (scale\* OR sub-scale\* OR tool\* OR measure\* OR instrument\* OR report\* OR index\* OR checklist\* OR test\*.ab)

### *Electronic Bibliographic Databases*

The search strategy was implemented in the following electronic bibliographic databases: Applied Social Sciences Index and Abstracts, Cochrane Database of Systematic Reviews, Cochrane Central Register of Controlled Trials (CENTRAL), EconLIT, PsycINFO, EBSCO combined search (CINAHL, ERIC, MEDLINE), Global Health, The International Bibliography of the Social Sciences (IBSS), Social Science Premium Collection, and ProQuest Dissertations and Theses. These databases were selected based on the recommendations of the Cochrane Handbook, the relevant scope of the databases, and other reviews of parenting programmes (Higgins & Green, 2011). For instance, some of these databases were selected as a review by Wight and colleagues, an ongoing systematic review of parenting programmes with aims similar to those of interest in the thesis, established the ability of the databases to capture relevant parenting programme literature (Gardner et al., forthcoming-a). For the purposes of Paper 1, the searches were updated in August 2021.

### *Additional Information Sources*

The review also drew upon information from other sources. First, the articles in Wight's review were requested from and shared by the authors. The titles and abstracts of the articles shared were then hand searched and data was extracted from relevant full-text articles. Articles from this review were included to ensure representation of parenting programmes from LMICs where there are typically fewer parenting programmes (Knerr et al., 2013). Second, backward reference searching was conducted using the reference lists of all included articles. Third, Google Scholar was used to conduct forward reference searching. Backward and forward citation tracking helps surface additional articles not

captured by the original search and that use different terminology (Higgins & Green, 2011). Fourth, parenting programme experts were contacted via email and asked to share relevant published or ongoing studies.

### *Study Selection*

After conducting the electronic bibliographic searches, the articles that surfaced in each database were saved and uploaded to Raayan – a web-based software to support systematic reviewers screen article titles and abstracts and remove duplicates (Ouzzani et al., 2016). This software was chosen as it is free to use, removes articles with ease, has simple ‘include’ and ‘exclude’ buttons, is capable of inviting collaborators to the review, and is able to quickly compare inclusion and exclusion decisions between reviewers (Ouzzani et al., 2016).

Next, each title and abstract was reviewed for relevance. When screening titles and abstracts, the reviewers used a screening form developed to simplify and expedite the decision-making. The form is comprised of ‘yes’ and ‘no’ questions wherein a single ‘no’ results in an article’s exclusion. The articles identified for inclusion were then reviewed in their entirety to screen for relevance. When screening full-text articles, the reviewers used another screening form.

### *Inter-Coder Reliability*

To ensure replicability, a number of steps were taken to establish inter-coder reliability at each stage of study inclusion: title/abstract screening, full-text screening, and data extraction (Belur et al., 2018). To facilitate reliability, a second coder was recruited and trained by this author (main coder). As recommended in the literature, the main coder reviewed the inclusion and exclusion criteria with the second coder and then they practiced

screening the titles and abstracts of 30 articles (Lombard et al., 2017). To ensure replicability at the full-text screening stage, the main coder trained the second coder in full-text inclusion and then they practiced screening three articles together. Finally, at the data extraction stage, the main coder trained the second coder on the process and then the coders practiced data extraction from another three articles together. Following training, the second coder independently coded a random selection of 10% of the articles at each of the title and abstract, full-text, and data extraction stages (Lombard et al., 2017). Once reliability was established at each stage, the main coder completed the screening of the remaining articles. Establishing reliability at each stage was done to confirm study replicability as inclusion decisions at each step require different considerations and thus should be examined separately (Belur et al., 2018). Percentage agreement, the degree of agreement between two coders, calculated by dividing the number of instances of agreement by the total number of instances (McHugh, 2012), was used to determine inter-coder reliability. As is recommended in the literature, the standard of inter-rater reliability sought was 80% (Belur et al., 2018; McHugh, 2012). Any discrepancies were resolved through discussion and consensus. Percentage agreements between coders ranged from 92.8-94.4% and were thus sufficiently high.

#### *Data Extraction and Analysis*

Once each full-text article was reviewed and a decision was made to include it, data was extracted from the article using a data extraction spreadsheet. Upon reviewing studies relevant for inclusion in Paper 1, a meta-analysis was deemed methodologically unfeasible. As a result, a synthesis without meta-analysis was performed based on the Synthesis Without Meta-Analysis (SWiM) guidelines outlined by Campbell and colleagues

(2020). These guidelines specify nine key categories of information that should be provided when a quantitative synthesis is not possible. The guidelines specify that analyses should provide a rationale of and description for the choice of studies; outline the metric(s) used to report the results; describe the methods used to synthesise the results; list the criteria used to evaluate and synthesise the results; comment on the certainty of the evidence; explore the heterogeneity of the results; present the results in tables including an assessment of risk of bias; describe the results based on each outcome and the certainty of the findings; and delineate limitations of the synthesis (Campbell et al., 2020). In addition to drawing on SWiM guidelines where possible and appropriate, Paper 1 findings were reported following the PRISMA guidelines (Liberati et al., 2009)

#### *Protocol Registration*

The systematic review was registered on the National Institute for Health Research's International Prospective Register of Systematic Reviews (PROSPERO: CRD42020167872). Paper 1 deviated from the protocol in that it was not possible to conduct a meta-analysis due to study heterogeneity.

## Supplementary File 2 – SWiM Checklist

SWiM reporting item	Item description	Page in manuscript where item is reported	Other*
<i>Methods</i>			
<b>1</b> Grouping studies for synthesis	1a) Provide a description of, and rationale for, the groups used in the synthesis (e.g., groupings of populations, interventions, outcomes, study design)	Table 1, page 6-7	
	1b) Detail and provide rationale for any changes made subsequent to the protocol in the groups used in the synthesis	Not applicable	
<b>2</b> Describe the standardised metric and transformation methods used	Describe the standardised metric for each outcome. Explain why the metric(s) was chosen, and describe any methods used to transform the intervention effects, as reported in the study, to the standardised metric, citing any methodological guidance consulted	Not worth calculating (explained on page 8)	
<b>3</b> Describe the synthesis methods	Describe and justify the methods used to synthesise the effects for each outcome when it was not possible to undertake a meta-analysis of effect estimates	Page 7-8	
<b>4</b> Criteria used to prioritise results for summary and synthesis	Where applicable, provide the criteria used, with supporting justification, to select the particular studies, or a particular study, for the main synthesis or to draw conclusions from the synthesis (e.g., based on study design, risk of bias assessments, directness in relation to the review question)	Table 1, page 6-8	
<b>5</b> Investigation of heterogeneity	State the method(s) used to examine heterogeneity in reported effects when it was not possible to undertake a meta-analysis of effect estimates and its extensions to investigate heterogeneity	Not applicable	

in reported effects			
<b>6</b> Certainty of evidence	Describe the methods used to assess certainty of the synthesis findings	Page 11-13	
<b>7</b> Data presentation methods	Describe the graphical and tabular methods used to present the effects (e.g., tables, forest plots, harvest plots). Specify key study characteristics (e.g., study design, risk of bias) used to order the studies, in the text and any tables or graphs, clearly referencing the studies included	Table 2, Online Resource 4	
<i>Results</i>			
<b>8</b> Reporting results	For each comparison and outcome, provide a description of the synthesised findings, and the certainty of the findings. Describe the result in language that is consistent with the question the synthesis addresses, and indicate which studies contribute to the synthesis	Page 12-15, Table 2, Online Resource 4	
<i>Discussion</i>			
<b>9</b> Limitations of the synthesis	Report the limitations of the synthesis methods used and/or the groupings used in the synthesis, and how these affect the conclusions that can be drawn in relation to the original review question	Page 17-20	

### Supplementary File 3 – PRISMA Checklist

Section and Topic	Item #	Checklist item	Location where item is reported
<b>TITLE</b>			
Title	1	Identify the report as a systematic review.	Title, page 1
<b>INTRODUCTION</b>			
Rationale	3	Describe the rationale for the review in the context of existing knowledge.	Page 4-6
Objectives	4	Provide an explicit statement of the objective(s) or question(s) the review addresses.	Page 7
<b>METHODS</b>			
Eligibility criteria	5	Specify the inclusion and exclusion criteria for the review and how studies were grouped for the syntheses.	Table 1, page 6-8
Information sources	6	Specify all databases, registers, websites, organisations, reference lists and other sources searched or consulted to identify studies. Specify the date when each source was last searched or consulted.	Page 7-8, Online Resource 1
Search strategy	7	Present the full search strategies for all databases, registers and websites, including any filters and limits used.	Online Resource 1
Selection process	8	Specify the methods used to decide whether a study met the inclusion criteria of the review, including how many reviewers screened each record and each report retrieved, whether they worked independently, and if applicable, details of automation tools used in the process.	Table 1, Page 6-8
Data collection process	9	Specify the methods used to collect data from reports, including how many reviewers collected data from each report, whether they worked independently, any processes for obtaining or confirming data from study investigators, and if applicable, details of automation tools used in the process.	Page 6-9
Data items	10a	List and define all outcomes for which data were sought. Specify whether all results that were compatible with each outcome domain in each study were sought (e.g. for all measures, time points, analyses), and if not, the methods used to decide which results to collect.	Page 6
	10b	List and define all other variables for which data were sought (e.g. participant and intervention characteristics, funding sources). Describe any assumptions made about any missing or unclear information.	Page 6
Study risk of bias assessment	11	Specify the methods used to assess risk of bias in the included studies, including details of the tool(s) used, how many reviewers assessed each study and whether they worked independently, and if applicable, details of automation tools used in the process.	Page 17-18
Effect measures	12	Specify for each outcome the effect measure(s) (e.g. risk ratio, mean difference) used in the synthesis or presentation of results.	Table 2 and Online Resource 4
Synthesis methods	13a	Describe the processes used to decide which studies were eligible for each synthesis (e.g. tabulating the study intervention characteristics and comparing against the planned groups for each synthesis (item #5)).	Page 7-8 and Online Resource 1



Section and Topic	Item #	Checklist item	Location where item is reported
	13b	Describe any methods required to prepare the data for presentation or synthesis, such as handling of missing summary statistics, or data conversions.	Page 8-9
	13c	Describe any methods used to tabulate or visually display results of individual studies and syntheses.	Page 8-9
	13d	Describe any methods used to synthesize results and provide a rationale for the choice(s). If meta-analysis was performed, describe the model(s), method(s) to identify the presence and extent of statistical heterogeneity, and software package(s) used.	Page 9
	13e	Describe any methods used to explore possible causes of heterogeneity among study results (e.g. subgroup analysis, meta-regression).	Not possible
	13f	Describe any sensitivity analyses conducted to assess robustness of the synthesized results.	Not applicable
Reporting bias assessment	14	Describe any methods used to assess risk of bias due to missing results in a synthesis (arising from reporting biases).	Page 17
Certainty assessment	15	Describe any methods used to assess certainty (or confidence) in the body of evidence for an outcome.	Not applicable
<b>RESULTS</b>			
Study selection	16a	Describe the results of the search and selection process, from the number of records identified in the search to the number of studies included in the review, ideally using a flow diagram.	Figure 1
	16b	Cite studies that might appear to meet the inclusion criteria, but which were excluded, and explain why they were excluded.	Not applicable
Study characteristics	17	Cite each included study and present its characteristics.	Table 2 and Online Resource 4
Risk of bias in studies	18	Present assessments of risk of bias for each included study.	Not applicable
Results of individual studies	19	For all outcomes, present, for each study: (a) summary statistics for each group (where appropriate) and (b) an effect estimate and its precision (e.g. confidence/credible interval), ideally using structured tables or plots.	Table 2 (modified)
Results of syntheses	20a	For each synthesis, briefly summarise the characteristics and risk of bias among contributing studies.	Not applicable
	20b	Present results of all statistical syntheses conducted. If meta-analysis was done, present for each the summary estimate and its precision (e.g. confidence/credible interval) and measures of statistical heterogeneity. If comparing groups, describe the direction of the effect.	Not applicable
	20c	Present results of all investigations of possible causes of heterogeneity among study results.	Not applicable
	20d	Present results of all sensitivity analyses conducted to assess the robustness of the synthesized results.	Not applicable

Section and Topic	Item #	Checklist item	Location where item is reported
Reporting biases	21	Present assessments of risk of bias due to missing results (arising from reporting biases) for each synthesis assessed.	Not applicable
Certainty of evidence	22	Present assessments of certainty (or confidence) in the body of evidence for each outcome assessed.	Page 11-14
<b>DISCUSSION</b>			
Discussion	23a	Provide a general interpretation of the results in the context of other evidence.	Page 15
	23b	Discuss any limitations of the evidence included in the review.	Page 16-18
	23c	Discuss any limitations of the review processes used.	Page 18-19
	23d	Discuss implications of the results for practice, policy, and future research.	Page 19-20
<b>OTHER INFORMATION</b>			
Registration and protocol	24a	Provide registration information for the review, including register name and registration number, or state that the review was not registered.	Abstract and Declarations
	24b	Indicate where the review protocol can be accessed, or state that a protocol was not prepared.	Abstract and Declarations
	24c	Describe and explain any amendments to information provided at registration or in the protocol.	Not applicable
Support	25	Describe sources of financial or non-financial support for the review, and the role of the funders or sponsors in the review.	Declarations
Competing interests	26	Declare any competing interests of review authors.	Declarations
Availability of data, code and other materials	27	Report which of the following are publicly available and where they can be found: template data collection forms; data extracted from included studies; data used for all analyses; analytic code; any other materials used in the review.	Online Resource 1 and Declarations (link to OSF page)

### Supplementary File 4 - Extra Study Data

<b>Table 4</b> <i>Location, Gender, Ethnic Make-up of Study Participants</i>				
<u>Paper</u>	<u>Program</u>	<u>Targeted</u>	<u>Female Participants</u>	<u>Ethnicity</u>
(Cantu et al., 2010)	Strengthening Families Program in the United States	Caregivers and children	Caregivers: 59.00% Children: 47.00%	68.00% European American, 11.00% Latino/a, 11.00% American Indigenous, 2.00% African American
(Chiapa et al., 2015) *data from Smith, 2013	Family Check-Up in the United States	Caregivers and children	Caregivers: 100.00% Children: 49.00%	Not reported
(Eames et al., 2010)	Incredible Years BASIC program in the United Kingdom	Caregivers only	Not reported	Not reported
(Forgatch et al., 2005)	Parent Management Training Oregon Model in the United States	Caregivers only	Caregivers: 50.00% Children: 25.00%	88.2% European American, 1.8% Latino/a, 0.00% Indigenous American, 0.00% African American, 10.00% multiracial or other
(Forgatch & DeGarmo, 2011)	Parent Management Training Oregon Model in Norway	Caregivers only	Not reported	Not reported
(Giannotta et al., 2019)	Incredible Years in Sweden	Caregivers only	Caregivers: 85.00% Children: not reported	Not reported
(Hukkelberg & Ogden, 2013)	Parent Management Training Oregon Model in Norway	Caregivers only	Caregivers: not reported Children: 26.00%	Not reported
(Hogue et al., 2008)	Multi-dimensional Family Therapy in the United States	Caregivers and children	Caregivers: not reported Children: 19.00%	70.00% African American, 20% European American, 10% Hispanic American

(Maaskant et al., 2016)	Parent Management Training Oregon Model in the Netherlands	Caregivers only	Caregivers: not reported Children: 54.00%	Not reported
(Rendu, 2004)	BASIC Parent-Training Program in the United Kingdom	Caregivers only	Caregivers: not reported Children: 27.40%	Not reported
(Robbins et al., 2011)	Brief Strategy Family Therapy in the United States	Caregivers and children	Caregivers: not reported Children: 21.46%	44.38% Hispanic American, 30.83% European American, 22.90% African American
(Roggman et al., 2016)	Early Head Start in the United States	Caregivers only	Caregivers: 100.0% Children: 58.00%	84.00% European American
(Satterfield, 2013)	Functional Family Therapy in Ireland	Caregivers and Children	Caregivers: not reported Children: 35.00%	93.30% White Irish, 1.70% White Romanian, 1.70% White Zimbabwean, 1.70% biracial Irish
(Scott et al., 2008)	Incredible Years in the United Kingdom	Caregivers only	Not reported	20% ethnic minority
(Smith et al., 2013)	Family Check-Up in the United States	Caregivers and children	Caregivers: not reported Children: 49.00%	51.00% European American, 30.00% African American, 12.00% biracial, 7.00% Hispanic American, 1.00% Indigenous American
(Snider, 2019)	Parent-Child Interaction Therapy in the United States	Caregivers and children	Caregivers: 90.00% Children: 25.00%	84.00% European American
(St. George et al., 2016)	Family Unidas in the United States	Caregivers and children	Caregivers: not reported Children: 48.00%	At least one caregiver per family identified as Hispanic American
(Thijssen et al., 2017)	Parent Management Training Oregon Model in the Netherlands	Caregivers only	Not reported	Not reported

**Table 5***Summary of Study Data*

<u>Paper</u>	<u>Program</u>	<u>Fidelity Domain Measured</u>	<u>Method</u>	<u>Sample Size</u>	<u>Associated Outcomes</u>	<u>Accounted for Clustering</u>	<u>Control Variables</u>	<u>Adjusted for Multiple Comparisons</u>	<u>Results</u>	<u>Findings</u>	<u>Summary</u>
(Cantu et al., 2010) or S1	Strengthening Families Program in the United States	Composite measure of adherence	Regressions (unadjusted and multi-level); continuous approach	Facilitator = 47 Parents = 96	Parenting skills and behaviours (parent-reported; Intervention Targeted Parenting Attitude and Behaviour Scale)	Yes (program level)	Yes (in multi-level model including participant characteristics, pre-test scores, program characteristic, facilitator characteristics)	No	Parenting skills and behaviours: B= 2.45, SE= 2.92, t=0.84, p=0.45	Adherence was not related to program outcomes	<b>Parenting:</b> ↔
(Chiapa et al., 2015) or S2 *data from Smith, 2013	Family Check-Up in the United States	Competent Adherence (composite)	Latent growth curve modelling; continuous approach	Facilitator = 79 Families = 79	Child behaviour (parent- and teacher-reported; Child Behaviour Checklist)	No	Yes (income, site, caregiver depression, therapist transfer, child behaviour problems at baseline)	No	Fidelity latent growth curve slope to child behaviour with covariates not included: b= -0.69, B= -0.95, posterior SD = 0.333, <-0.01  With covariates included: $\beta$ = -0.95, p =	Decline in competent adherence associated with less behaviour change (drift)	<b>Child behaviour:</b> ↑

(Eames et al., 2010) or S3	Incredible Years BASIC program in the United Kingdom	Two dimensions of competent adherence (facilitator praise and reflective statements)	Regression and one-way ANOVA s; categorical approach	Facilitator = Not reported Families = 104	Parental praise and reflexive statements (observer-reported; Dyadic Parent-Child Interaction Coding System)	No	Yes (baseline DPICs)	No	0.003, 95% CI (-2.11, -0.22)	Facilitator reflective and parent reflective: B=0.20, p <0.05, R <sup>2</sup> =0.04	Facilitator competent adherence associated with better parenting skills	<b>Parenting:</b> ↑
									Facilitator praise and parent praise: B=0.35, p < 0.01, R <sup>2</sup> =0.12			
									ANOVA praise: (F[2, 150]=10.90, p < 0.01)			
									ANOVA reflective: (F[2,150]=4.81, p < 0.01)			
(Forgatch et al., 2005) or S4	Parent Management Training Oregon Model in the United States	Competent adherence (composite)	Multivariate SEM/partial analysis; continuous approach	Facilitator = 4 Families = 20	Parenting skills and behaviours (observer-reported; Family and Peer	Yes (within-couple dependence)	Yes (baseline status, child age, child gender, encouragement sessions, discipline sessions)	No		Change in maternal parenting: B=0.51, p <0.05, R <sup>2</sup> =0.53	Higher competent adherence predicted improved parenting	<b>Parenting:</b> ↑

					Process Code)				Change in step-father parenting: B=0.49, p < 0.05, R <sup>2</sup> =0.66		
(Forgatch & DeGarmo, 2011) or S5	Parent Management Training Oregon Model in Norway	Competent adherence (composite )	Bivariate correlations and SEM/path analysis; continuous approach	Facilitator = 35  Families = 242	Parenting skills and behaviours (observer- reported; Family and Peer Process Code)	No	Yes (pre- and post- treatment, child age, child gender, encourage sessions, discipline sessions)	No	Parenting skills and behaviours: B= 0.17(0.23) with p<0.05 and R <sup>2</sup> =0.40(0.4 3) with multi-level parameters adjusted for clustering in brackets  FIMP encouragem ent and change in mother's parenting: 0.05, p>0.05  FIMP discipline and change in mother's parenting: 0.07, p>0.05	Higher competent adherence predicted improved parenting	<b>Parenting (path analysis): ↑</b>  <b>Parenting (correlations): ⇔</b>

									FIMP encouragem ent and change in father's parenting: 0.07, p>0.05		
									FIMP discipline and change in father's parenting: 0.10, p>0.05		
(Giannott a et al., 2019) or S6	Incredible Years in Sweden	Competent Adherence	Multi- level linear regressio n	Facilita tor = 111  Parents = 535  Childre n = 749	Parenting skills and behaviours (parent- reported; Parenting Sense of Competenc e Scale) and child behaviour (parent- reported; Eyberg Child Behaviour Inventory and Swanson Nolan and Pelham-IV)	Yes (maximum likelihood robust multi-level modelling to take parenting and groups into account)	Yes (child age, parent age, parent involvement, parent attendance)	No	Parent outcomes: (1) angry outbursts = 0.01(0.05) with 0.00(0.01) residual, (2) harsh parenting = 0.02(0.13) with 0.00(0.02) residual, (3) attempted understandin g = - 0.01(.01) with 0.00(0.01) residual, (4) praise = - 0.04(0.18)	Competent adherence was not associated with parent and child outcomes	<b>Parenting:</b> ↔  <b>Child behaviour:</b> ↔



with  
 .00(0.03)  
 residual, (5)  
 reward = -  
 .08(0.11)  
 with  
 0.01(0.13)  
 residual, (6)  
 competence  
 = -  
 0.02(0.03)  
 with .00(.01)  
 residual

Child  
 outcomes:  
 (1) ECBI = -  
 0.01(0.02)  
 with  
 0.01(0.02)  
 residual, (2)  
 ECBI  
 problem =  
 0.01(0.01)  
 with  
 0.01(0.01)  
 residual, (3)  
 inattention =  
 -0.01(0.02)  
 with  
 0.00(0.01)  
 residual, (4)  
 hyperactivity  
 = -  
 0.01(0.02)  
 with  
 0.00(0.01)  
 residual, (5)

									oppositional defiance = 0.02(0.02) with 0.00(0.01) residual		
(Hukkelberg & Ogden, 2013) or S7	Parent Management Training Oregon Model in Norway	Competent adherence	SEM/path analysis	Facilitator = 134  Families = 331	Child behaviour (teacher-reports on Child Behaviour Checklist and parent-reports on Parent Daily Report)	No	Yes	No	Parent report: B= 0.18, p < 0.01, R <sup>2</sup> =0.49  Teacher report: B= -0.02, p > 0.32, R <sup>2</sup> =0.41, p > 0.05	Competent adherence predicted reductions in behaviour issues (parent-report only); found alliance and competent adherence to be independent from each other	<b>Child Behaviour (parent-report): ↑</b>  <b>Child Behaviour (teacher-report): ⇔</b>
(Hogue et al., 2008) or S8	Multi-dimensional Family Therapy in the United States	Competent and adherence measured separately	Latent growth curve modelling with pseudo z-tests; continuous approach	Facilitator = 5  Families = 36	Child behaviour (parent- and youth-reported; revised Child Behaviour Checklist)	Yes (accounted for nesting)	Yes (therapeutic alliance, outcomes, treatment condition and interaction terms)	No	<u>Main Effects:</u>  Parent-reported externalizing and adherence mean slope = -2.07, SE=0.87,	Better adherence predicted greater reductions in parent-reported externalizing but not youth-reported	<b><u>Adherence:</u></b>  <b>Parent-reported externalizing (linear): ↑</b>  <b>Youth-reported externalizing</b>

pseudo-z = -2/36, p <0.05, 95% CI: -2.94, -1/18, d=0.37	externalizing behaviour	<b>g (linear):</b> ⇔
Parent-reported internalizing and adherence mean slope: 0.27, SE=0.84, p>0.05	Competence did not predict either internalizing or externalizing behaviour as reported by parents and youth	<b>Parent-reported internalizing (linear):</b> ⇔ <b>Parent-internalizing (curvilinear):</b> ↑
Parent-reported externalizing and competence mean slope: 0.53, SE=0.55, p>0.05	Some evidence of a curvilinear relationship for adherence-internalizing	<b>Competence:</b> <b>Parent-reported internalizing (linear):</b> ⇔
Parent-reported internalizing and competence mean slope: -0.44, SE=0.59, p>0.05	relationship, but not reported for externalizing despite this being a main outcome	<b>Parent- and youth-reported externalizing (linear):</b> ⇔ <b>Externalizing curvilinear not reported</b>
Youth-reported	Curvilinear relationship	

externalizing and adherence mean slope: -0.32, SE=0.63,  $p>0.05$  ps between competence and outcomes are not explored

Youth-reported externalizing and competence mean slope: -0.09, SE=0.54,  $p>0.05$

Curvilinear Relationship  
s:

Internalizing and adherence mean slope = -1.50, pseudo-z = 2.46,  $p<0.05$ , 95% CI: -2.11, 0.89,  $d=0.40$

<u>Baseline to Post-test:</u> Parenting stress: $\beta=0.27$ , SE=	Higher competent adherence associated with better	<b><u>Post-Test:</u></b> <b>Stress:</b> ↑ <b>Warmth:</b> ↔ <b>Responsiveness:</b> ↔
---	---	--

(Maaskant et al., 2016) or S9	Parent Management Training Oregon Model in	Competent adherence	Multi-level regression; continuous	Facilitator = Not reported	Parenting stress (parent-reported; Parenting	Unclear – likely partially (multi-level modeling	Yes (baseline outcome levels)	No
-------------------------------	--	---------------------	------------------------------------	----------------------------	--	--	-------------------------------	----

the Netherlands	us approach	Families = 86	Stress Index); parenting behaviours (parent-reported; Parenting Behaviour Questionnaire); and child behaviour (parent-reported; Child Behaviour Checklist)	accounted for repeated measures and multiple respondents per family but not for multiple families per therapist/facilitator; number of facilitators not clear)	0.07, $p < 0.01$ at baseline and $\beta=0.28$ , $SE=0.09$ , $p < 0.03$ at follow-up	improvements in some parenting dimensions but not others	<b>Explaining:</b> $\uparrow$ <b>Autonomy:</b> $\uparrow$ <b>Strictness:</b> $\Leftrightarrow$ <b>Discipline:</b> $\Leftrightarrow$ <b>Child Behaviour:</b> $\Leftrightarrow$ <b>Follow-Up:</b> <b>Stress:</b> $\uparrow$ <b>Warmth:</b> $\Leftrightarrow$ <b>Responsiveness:</b> $\uparrow$ <b>Explaining:</b> $\uparrow$ <b>Autonomy:</b> $\uparrow$ <b>Strictness:</b> $\Leftrightarrow$ <b>Discipline:</b> $\Leftrightarrow$ <b>Child Behaviour:</b> $\Leftrightarrow$
					Parenting warmth: $\beta = -0.02$ , $SE=0.09$ , $p < 0.86$ at baseline and $\beta=0.15$ , $SE=0.09$ , $p < 0.17$ at follow-up		
					Parenting responsiveness: $\beta = 0.13$ , $SE=0.12$ , $p < 0.31$ at baseline and $\beta=0.23$ , $SE=0.09$ , $p < 0.05$ at follow-up		
					Parent explaining: $\beta=0.25$ , $SE=0.09$ , $p < 0.03$ at baseline and $\beta=0.26$ , $SE=0.11$ , $p < 0.03$ at follow-up		

0.05 at  
follow-up

Parenting  
autonomy:  
 $\beta=0.34$ ,  
 $SE=0.14$ ,  $p$   
 $<0.04$  at  
baseline and  
 $\beta=0.32$ ,  
 $SE=0.11$ ,  $p <$   
 $0.03$  at  
follow-up

Parent  
strictness:  
 $\beta=0.01$ ,  
 $SE=0.09$ ,  $p <$   
 $0.95$  at  
baseline and  
 $\beta=0.12$ ,  
 $SE=0.11$ ,  $p <$   
 $0.36$  at  
follow-up

Parent  
discipline:  
 $\beta=0.17$ ,  
 $SE=0.13$ ,  $p <$   
 $0.31$  at  
baseline and  
 $\beta=0.04$ ,  
 $SE=0.11$ ,  $p$   
 $<0.69$  at  
follow-up

Child  
behaviour:

$\beta=0.22$ ,  
SE=0.10,  $p <$   
0.07 at  
baseline and  
 $\beta=0.21$ ,  
SE=0.13,  $p$   
<0.17

Baseline to  
Follow-Up:

Parent

stress:

B=0.28,

SE=0.09,

$p < 0.03$

Parent

warmth:

B=0.15,

SE=0.09,

$p < 0.17$

Parent

responsiveness:

B=0.23,

SE=0.09,

$p < 0.05$

Parent

autonomy:

B=0.32,

SE=0.11,

$p < 0.05$

Parent

strictness:

B=0.12,

SE=0.11,

$p < 0.69$

									Parent discipline: B=0.04, SE=0.11, p<0.69 Child behaviour: B=0.21, SE=0.13, p<0.17		
(Rendu, 2004) or S10	BASIC Parent-Training Program in the United Kingdom	Competent adherence (measured in terms of 'group facilitation' and 'practicalities')	Linear regression; continuous approach	Facilitator = 13 Families = 84	Child behaviour (parent-reported; Parents Accounts of Child Symptoms Interview)	No	Yes (model 1: pre-treatment child behaviour and model 2: pre-treatment child behaviour, child age, hyperactivity)	No	Group facilitation model 1: B=-0.19, SE= 0.08, Beta= -0.24, p = 0.01  Group facilitation model 2: B=-0.14, SE=0.08, Beta=.17, p = .09  Practicalities : B=-0.08, SE= 0.10, Beta= -0.09, p= 0.40  Practicalities model 2: B= -0.07, SE=	Some facilitator competent adherence dimensions associated	<b>Group facilitation and child behaviour: ⇔ and ↑</b>  <b>Practicalities and child behaviour: ⇔</b>



(Robbins et al., 2011) or S11	Brief Strategy Family Therapy in the United States	Competent adherence (broken down into four behaviours)	Latent growth curve modelling; continuous approach	Facilitator = 5 Families = 480	Family functioning (parent- and youth-reported; Family Environmental Scale) and child behaviour (adolescent drug use; self-report measure)	Yes (nesting at family-level)	Yes (baseline outcomes)	No	0.10, Beta=0.07, p=0.45	One of the four behaviours was related to outcomes ("joining")	Some aspects of competent adherence associated with outcomes	<b>Family functioning:</b> ↑ <b>Drug use:</b> ↑ <b>Some outcomes not reported</b>
									Family functioning: b = 0.053, SE=0.019, p <0.005, standardized coefficient gamma=0.203			
(Roggman et al., 2016) or S12	Early Head Start in the United States	Competent adherence	Regression; continuous approach	Facilitator = Not reported	Parenting behaviour and skills (observer-reported;	No	Yes (intervention site)	No	Adolescent drug use: b = .121, SE = .051, p < .018	The results of the other three behaviours are not reported	Competent adherence associated with better parent and	<b>Parenting:</b> ↑
									Parenting behaviour and skills: B=1.22, SE=0.62,			

				Families = 71	Home Observation Measure of the Environment) and child development (Peabody Picture Vocabulary Test-III)				$\beta=0.52, p < 0.001$	child outcomes	<b>Child academics:</b> ↑
									Child vocabulary: $B=6.88, SE=3.16, \beta=0.33 p < 0.05$		
(Satterfield, 2013) or S13	Functional Family Therapy in Ireland	Competent adherence	Regression; continuous approach	Facilitator = 9 Families = 60	Child behaviour (parent- and youth-reports; Strengths and Difficulties Questionnaire)	No	Yes (pre-test adolescent functioning)	No	Youth-report: $B=-.410, p < 0.642$  Parent-report $B=-2.87, p < .000$	Competent adherence predicted behaviour reductions from parent perspective but not teen perspective	<b>Youth reported behaviour:</b> ↔  <b>Parenting reported behaviour:</b> ↑
(Scott et al., 2008) or S14	Incredible Years in the United Kingdom	Competent adherence	Regression; continuous approach	Facilitator = 13 Families = 73	Child behaviour (parent-reported; Parent Account of Child Symptoms Interview)	Yes (accounted for nesting using multi-level modelling)	Yes (child age, attendance, pre-treatment scores)	No	$B=-0.25, SE=.09, p < 0.05$  A one-point increase in skill was associated with a 0.58 SD increase in child behaviour	Greater competent adherence predicted better behaviour	<b>Child behavior:</b> ↑

(Smith et al., 2013) or S15	Family Check-Up in the United States	Competent adherence	SEM/path analysis and correlations; continuous approach	Facilitator = Not reported	Parent skills and behaviours (observer-reported; HOME; Relationships Process Code; Coders Impression Inventory) and child behaviour (parent-reported; Child Behaviour Checklist)	No	Yes (baseline parenting behaviour and skills, baseline child problem behaviour)	No	Correlation between competent adherence and parent behaviour and skills at age 2: $r=0.14$ , not significant Correlation between competent adherence and parent behaviour and skills at age 3: $r=0.05$ , not significant  Correlation between competent adherence and child behaviour at age 2: $r=0.14$ , not significant  Correlation between competent adherence and child behaviour age 3:	Greater competent adherence not directly but indirectly associated with better improvements in parenting and behaviour	<b>Parenting:</b> ⇔ <b>Child behaviour:</b> ⇔
-----------------------------	--------------------------------------	---------------------	---	----------------------------	--	----	---	----	--	--	--

$r=0.09$ , not significant

Path analysis of competent adherence to parent behaviour and skills at age 3:  $B=-0.05$ , posterior  $SD=0.05$ ,  $\beta=-0.10$ , 95% CI[-0.328, 0.091],  $p>0.01$

Indirect effects (competent adherence -> engagement -> positive behaviour support -> problem behaviour:  $B=-0.24$ , posterior  $SD=0.19$ , 95% CI=-.664-3.019,  $p>0.05$

(Snider, 2019) or S16	Parent-Child Interaction Therapy in the United States	Competent adherence	Hierarchical linear modelling; continuous approach	Facilitator = 17 Families = 32	Parenting behaviour and skills (parent-reported; Alabama Parenting Questionnaire) and child behaviour (parent-reported; Eyberg Child Behaviour Inventory)	No	Yes (early attribution but had no impact in model)	No	Fixed effects  ECBI Intensity Adherence b=6.70, SE=3.28, p=.069 Competence b= -3.97, SE= 2.55, p=.150  ECBI Problem Adherence b=2.07, SE=0.95, p=.054 Competence b=-1.03, SE= 0.71, p=.179  APQ-9 Problem Adherence b=-0.13, SE= 0.09, p=.187 Competence b= -0.05, SE= 0.07, p=.525	Competent adherence not associated with child behaviour or parenting	<b>Child behaviour:</b> ⇔ <b>Parenting:</b> ⇔
(St. George et	Family Unidas in	Competence and adherence	SEM/path analysis;	Facilitator = Not	Child behaviour (youth-	No	Partly (facilitator variables,	No	Adherence and family functioning:	Higher competence	

al., 2016) or S17	the United States	measured separately	continuous approach	reported Families = 367	reported; self-reported substance abuse) and family functioning (latent construct created from items from the Parent Relationship with Peer Group Scale, Parenting Practices Scale, and Family Relations Scale; parent-reported)			organizational variables, family demographics, attendance, pre-test family functioning, pre-test adolescent substance use included in other parts of the structural equation model)	not associated and do not provide results  Adherence and substance use: $b = -0.28$ , $p < 0.50$  Competence and family functioning: not associated and do not provide results  Competence and substance use: $b = -.999$ , $p < 0.05$ , 95% CI $[-1.682, -0.315]$	associated with reductions in substance abuse but no relationship between adherence or competence and family functioning and no relationship between adherence and substance use	<b>Family functioning: <math>\Leftrightarrow</math> (but data not provided)</b>  <b>Adherence and substance use: <math>\Leftrightarrow</math></b>  <b>Competence and substance use: <math>\Uparrow</math></b>
(Thijssen et al., 2017) or S18	Parent Management Training Oregon Model in the	Competent adherence	Pearson correlation; continuous approach	Facilitator = 25 Families = 86	Child behaviour (parent-reported; Child Behaviour Checklist	No	No	No	Child behaviour (CBCL): $r = -0.18$ (T1), $r = -0.05$ (T2), $r = -0.26$ (T3)	Associations were not significant but sub-constructs of	<b>Child behaviour: <math>\Leftrightarrow</math></b>  <b>Parent stress T1 and T2: <math>\Leftrightarrow</math></b>

Netherlands	and Parent Daily Report), parenting stress (parent-reported; Nijmeegse Ouderlijke Stress Index and Symptom Checklist Revised), and parenting skills and behaviours (parent-reported; Caregiver Wish List)	Child behaviour (PDR): $r=-0.02$ (T1), $r=0.09$ (T2), $r=0.03$ (T3)  Parent stress (NOSI): $r=-0.22$ (T1), $r=-0.13$ (T2), $r=-0.32$ ( $p<0.05$ )  Parent Stress SCL-R: $r=-0.12$ (T1), $r=-0.13$ (T2), $r=-0.21$ (T3) Parenting practices (CWL): $r=0.22$ (T1), $r=0.05$ (T2), $r=0.21$ (T3)	facilitator competent adherence were; association between facilitator competent adherence and parenting stress was significant at one time point	<b>Parent stress T3: ↑</b>
-------------	---	--	--	----------------------------

<b>Table 6</b> <i>Psychometric Evidence of Measures in Included Studies</i>	
<u>Paper</u>	<u>Psychometric Evidence</u> <u>(see Martin et al., 2021)</u>
(Cantu et al., 2010)	No
(Chiapa et al., 2015)	Yes
(Eames et al., 2010)	No
(Forgatch et al., 2005)	Yes
(Forgatch & DeGarmo, 2011)	Yes
(Giannotta et al., 2019)	Yes
(Hukkelberg & Ogden, 2013)	Yes
(Hogue et al., 2008)	No
(Maaskant et al., 2016)	No
(Rendu, 2004)	Yes
(Robbins et al., 2011)	No
(Roggman et al., 2016)	Yes
(Satterfield, 2013)	No
(Scott et al., 2008)	Yes
(Smith et al., 2013)	Yes
(Snider, 2019)	Yes
(St. George et al., 2016)	Yes
(Thijssen et al., 2017)	No



<b>Table 7</b> <i>Study Risk of Bias, Quality, and Measure Practicality</i>						
<u>Study</u>	<u>Study Risk of Bias and Quality Checklist</u>			<u>Measure Practicality Checklist</u>		
	<u>Sampling</u>	<u>Assessors</u>	<u>Reactivity</u>	<u>Feasibility and Sustainability of Training</u>	<u>Utility</u>	<u>Availability</u>
Cantu 2010; S1	+	+	?	?	+	-
Chiapa 2015; S2	?	?	?	?	+	+
Eames 2010; S3	-	?	+	+	+	+
Forgatch 2005; S4	-	-	+	+	+	+
Forgatch 2011; S5	-	+	-	?	+	+
Giannotta 2019; S6	+	+	-	+	+	+
Hukkelberg 2013; S7	-	+	?	?	?	-
Hogue 2008; S8	+	+	+	+	+	-
Maaskant 2016; S9	-	+	-	?	?	?
Rendu 2004; S10	-	+	+	?	+	+
Robbins 2010; S11	+	+	?	+	+	+
Roggman 2016; S12	+	+	?	+	+	+
Satterfield 2013; S13	+	-	?	+	+	-
Scott 2008; S14	+	+	+	-	?	-
Smith 2013; S15	+	+	+	+	+	+
Snider 2019; S16	+	+	+	?	+	+
St. George 2016; S17	+	+	+	?	+	+
Thijssen 2017; S18	?	?	+	?	+	+

## Appendix 8: Paper 2 Supplementary Files

### Supplementary File 1 – Summary of Recommendations and Changes to PLH-FAT-T

<b>Table 11</b>											
<i>Summary of Recommendations and Changes to PLH-FAT-T</i>											
<u>Recommendation to</u>	<u>Stakeholder Group</u>	<u>Changes Made or Example of Changes Made</u>									
<u>Improve the PLH-FAT-T</u>											
Both facilitators assessed at the same time	Insights from assessors in Southeastern Europe and CWBSA Trainers	One facilitator assessed on the home activity discussion and the other facilitator assessed on the role-play discussion									
Change Likert scale from four-points to three-points	Insights from assessors in Southeastern Europe, Lead Assessors and Trainers from CWBSA	<table border="1"> <tr> <th>RATING</th><th>0 Inadequate</th><th>1 Good</th><th>2 Exceeds Expectations</th></tr> <tr> <th>DESCRIPTION</th><td>The facilitator demonstrates little or no evidence that activity has been done, or the activity is done poorly.</td><td>The facilitator does the activity but not as instructed and is not at a sufficient level of quality for competent delivery.</td><td>The facilitator does the activity at a high level of competency.</td></tr> </table>		RATING	0 Inadequate	1 Good	2 Exceeds Expectations	DESCRIPTION	The facilitator demonstrates little or no evidence that activity has been done, or the activity is done poorly.	The facilitator does the activity but not as instructed and is not at a sufficient level of quality for competent delivery.	The facilitator does the activity at a high level of competency.
RATING	0 Inadequate	1 Good	2 Exceeds Expectations								
DESCRIPTION	The facilitator demonstrates little or no evidence that activity has been done, or the activity is done poorly.	The facilitator does the activity but not as instructed and is not at a sufficient level of quality for competent delivery.	The facilitator does the activity at a high level of competency.								
Use specific definitions for each item and Likert point	Insights from psychometric evaluation in Southeastern Europe, Lead Assessors and Trainers from CWBSA	<table border="1"> <tr> <th>HOME ACTIVITY DISCUSSION Did the Facilitator ...</th><th>0 Inadequate</th><th>1 Good</th><th>2 Excellent</th></tr> <tr> <td>1. Remind the participants of the core home activity for the previous week at the beginning of the discussion</td><td>The facilitator did not remind the participants about the previous home activity.</td><td>The facilitator asked or reminded the participants about many of the home activities, but could have been more specific and/or more succinct regarding what the main home activity was from the previous session.</td><td>The facilitator asked or reminded participants of the main home activity in a clear, specific, and succinct manner.</td></tr> </table>		HOME ACTIVITY DISCUSSION Did the Facilitator ...	0 Inadequate	1 Good	2 Excellent	1. Remind the participants of the core home activity for the previous week at the beginning of the discussion	The facilitator did not remind the participants about the previous home activity.	The facilitator asked or reminded the participants about many of the home activities, but could have been more specific and/or more succinct regarding what the main home activity was from the previous session.	The facilitator asked or reminded participants of the main home activity in a clear, specific, and succinct manner.
HOME ACTIVITY DISCUSSION Did the Facilitator ...	0 Inadequate	1 Good	2 Excellent								
1. Remind the participants of the core home activity for the previous week at the beginning of the discussion	The facilitator did not remind the participants about the previous home activity.	The facilitator asked or reminded the participants about many of the home activities, but could have been more specific and/or more succinct regarding what the main home activity was from the previous session.	The facilitator asked or reminded participants of the main home activity in a clear, specific, and succinct manner.								
Add items, remove items, revise item wording, revise item definitions items to suit the Tanzanian context	Lead Assessors from CWBSA	For the following item, “Thanks and praises the participants for sharing their ideas”, the definition of “Good” on the Likert scale was changed from “The facilitator gave general praise” to “The facilitator gave general praise and/or did not summarise the key principles or building blocks from the home practice”.									

**Supplementary File 2 - Revised PLH-FAT-T used in Tanzania**

**Parenting for Lifelong Health for Adolescents and Teens-Facilitator Assessment Tool  
(PLH-FAT-T)**

Assessor Name:			
Facilitator 1 Name:		Facilitator 1 ID:	
Facilitator 2 Name:		Facilitator 2 ID:	
Assessment Date:		Session Number and Date:	
Video File Name (if applicable):		Session/Video Length:	
Number of Enrolled Parents:		Number of Parents in Attendance:	
Number of Enrolled Teens:		Number of Teens in Attendance:	
Facilitator 1 Age and Gender:		Facilitator 2 Age and Gender:	
Has the facilitators been assessed before (Y/N) ?	Facilitator 1:  Facilitator 2:	If yes, how many times have the facilitators been assessed previously?	Facilitator 1:  Facilitator 2:

## SECTION ONE | ACTIVITY ASSESSMENT

RATING	0 Inadequate	1 Good	2 Excellent
DESCRIPTION	The facilitator demonstrates little or no evidence that activity has been done, or the activity is done poorly.	The facilitator does the activity but not as instructed and is not at a sufficient level of quality for competent delivery.	The facilitator does the activity at a high level of competency.

HOME ACTIVITY DISCUSSION <i>Assessment of Facilitator 1</i> Did the facilitator...	0 Inadequate	1 Good	2 Excellent
1. Reminds the participants of the core home activity for the previous week at the beginning of the discussion	0	1	2
2. Reviews the core building blocks from previous session with participants at the beginning of the discussion	0	1	2
3. Allows participants to share their experiences of how home activity went after the previous session	0	1	2
4. Involves teens in the discussion about the home activity	0	1	2
5. Keeps participants focused on the core home activity	0	1	2
6. Helps participants connect their experiences to the core building blocks	0	1	2
7. Identifies at least one specific challenge experienced by a participant regarding the main home activity	0	1	2
8. Explores solutions to challenges shared and help participants choose a specific solution	0	1	2
9. Practices the chosen solution with parents and teens	0	1	2
10. Debriefs with the participants after practicing and encourages them to try the solution at home	0	1	2
11. Thanks and praises participants for sharing experiences (at the end of the home activity discussion)	0	1	2
Comments/Notes:			

<b>ROLE-PLAY AND ROLE-PLAY DISCUSSION</b> <i>Assessment of Facilitator 2</i> <b>Did the facilitator...</b>		<b>0</b> <b>Inadequate</b>	<b>1</b> <b>Good</b>	<b>2</b> <b>Excellent</b>
1.	Provides introductory context to the role-play	0	1	2
2.	Acts out the role-play following the steps for leading the role-play (fidelity)	0	1	2
3.	Acts out the role-play following tips for leading the role-play (quality)	0	1	2
4.	Discusses the role-play with participants after facilitators acted it out	0	1	2
5.	Explores possible solutions for negative role-plays	0	1	2
6.	Gets participants to act out a positive scenario that changes a negative role-play into a positive one	0	1	2
7.	Gives support to participants during the positive role-play (shadowing)	0	1	2
8.	Debriefs with participants about experiences and feelings after acting out the positive scenario	0	1	2
9.	Discusses with participants about how the role-play relates to their lives	0	1	2
10.	Connects the role-play to the building blocks of the session	0	1	2
11.	Thanks and praises the participants for sharing their ideas and solutions	0	1	2
<b>Comments/Notes:</b>				
<b>ASSESSMENT AREA</b>		<b>MAXIMUM SCORE</b>		<b>ASSESSMENT SCORE</b>
<b>Facilitator 1 overall score on home activity discussion</b>		<b>22</b>		
<b>Facilitator 2 overall score on role-play &amp; role-play discussion</b>		<b>22</b>		
<b>Total combined activities score</b>		<b>44</b>		

## SECTION TWO | FACILITATION SKILLS ASSESSMENT

RATING	0 Inadequate	1 Good	2 Excellent
DESCRIPTION	There is little or no evidence that the facilitator has exhibited those competencies during the session being observed. Skills and abilities have not been demonstrated at required level.	The facilitator exhibits some of the required skills and demonstrates general understanding of what is expected but requires coaching and support to grow.	The facilitator demonstrates exceptional application of required skills. The application of required skill is consistent and demonstrates high level capabilities.

MODELLING BEHAVIOUR			
ASSESSMENT OF FACILITATOR 1 Did the facilitator...	0 Inadequate	1 Good	2 Excellent
1. Gives positive reinforcement and specific praise to parents and teens	0	1	2
2. Gives positive, specific, and realistic instructions	0	1	2
3. Maintains commitments to time management principles	0	1	2
4. Models behaviours with co-facilitator	0	1	2
5. Demonstrates respectful behaviour towards participants	0	1	2
Overall score on Modelling Behaviour (out of 10)			
Comments/Notes:			
Accept-Explore-Connect-Practice (AECPP) Did the facilitator...	0 Inadequate	1 Good	2 Excellent
1. Uses body language to show acceptance	0	1	2
2. Accepts parent responses verbally	0	1	2
3. Explores participant experiences and opinions using open-ended questions	0	1	2
4. Explores thoughts and feelings	0	1	2
5. Explores the perspective of the teen	0	1	2

6. Explores the perspective of the parent	0	1	2
7. Connects experiences to the building blocks the from session	0	1	2
8. Identifies opportunities to practice skills (in addition to the structured group practice)	0	1	2
<b>Comments/Notes:</b>			

<b>COLLABORATIVE LEADERSHIP SKILLS</b> <b>Did the facilitator...</b>	<b>0</b> <b>Inadequate</b>	<b>1</b> <b>Good</b>	<b>2</b> <b>Excellent</b>
1. Arranges the room in a way that encourages equal and active participation	0	1	2
2. Facilitator is situated within the group, is at the level of the participants, and in a different place than the co-facilitator	0	1	2
3. Participants appear comfortable and engaged in the session	0	1	2
4. Participant-facilitator speaking ratio	0	1	2
5. Assures equal and active participation among participants	0	1	2
6. Engages quiet or non-participating parents and teens	0	1	2
7. Limits participant responses	0	1	2
8. Keeps participants focused on the topic of discussion	0	1	2
9. Demonstrates knowledge of session content	0	1	2
10. Delivers the session with confidence	0	1	2
11. Helps participants generate their own ideas regarding principles or solutions to challenges	0	1	2
12. Helps participants assess positive and negative consequences to proposed solutions	0	1	2
13. Ensures that solutions are positive, specific, and realistic	0	1	2
14. Maintains leadership and control of the group	0	1	2
15. Works well with the co-facilitator	0	1	2
<b>Comments/Notes:</b>			

ASSESSMENT AREA	MAXIMUM SCORE	ASSESSMENT SCORE
Overall score on modelling behaviour	10	
Overall score on using AECP	16	
Overall score on collaborative leadership	30	
Total score facilitation skills (FACILITATOR 2)	56	

<b>MODELLING BEHAVIOUR</b> <b>ASSESSMENT OF FACILITATOR 2</b> <b>Did the facilitator...</b>	<b>0</b> <b>Inadequate</b>	<b>1</b> <b>Good</b>	<b>2</b> <b>Excellent</b>
1. Gives positive reinforcement and specific praise to parents and teens	0	1	2
2. Gives positive, specific, and realistic instructions	0	1	2
3. Maintains commitments to time management principles	0	1	2
4. Models behaviours with co-facilitator	0	1	2
5. Demonstrates respectful behaviour towards participants	0	1	2
<b>Overall score on Modelling Behaviour (out of 10)</b>			
<b>Comments/Notes:</b>			

<b>Accept-Explore-Connect-Practice (AECP)</b> <b>ASSESSMENT OF FACILITATOR 2</b> <b>Did the facilitator...</b>	<b>0</b> <b>Inadequate</b>	<b>1</b> <b>Good</b>	<b>2</b> <b>Excellent</b>
1. Uses body language to show acceptance	0	1	2
2. Accepts parent responses verbally	0	1	2
3. Explores participant experiences and opinions using open-ended questions	0	1	2
4. Explores thoughts and feelings	0	1	2
5. Explores the perspective of the child	0	1	2



6. Explores the perspective of the parent	0	1	2
7. Connects experiences to the building blocks the from session	0	1	2
8. Identifies opportunities to practice skills (in addition to the structured group practice)	0	1	2
<b>Comments/Notes:</b>			

<b>COLLABORATIVE LEADERSHIP SKILLS</b> <b>ASSESSMENT OF FACILITATOR 2</b> <b>Did the facilitator...</b>	<b>0</b> <b>Inadequate</b>	<b>1</b> <b>Good</b>	<b>2</b> <b>Excellent</b>
16. Arranges the room in a way that encourages equal and active participation	0	1	2
1. Facilitator is situated within the group, is at the level of the participants, and in a different place than the co-facilitator	0	1	2
2. Participants appear comfortable and engaged in the session	0	1	2
3. Participant-facilitator speaking ratio	0	1	2
4. Assures equal and active participation among participants	0	1	2
5. Engages quiet or non-participating parents and teens	0	1	2
6. Limits participant responses	0	1	2
7. Keeps participants focused on the topic of discussion	0	1	2
8. Demonstrates knowledge of session content	0	1	2
9. Delivers the session with confidence	0	1	2
10. Helps participants generate their own ideas regarding principles or solutions to challenges	0	1	2
11. Helps parents and teens assess positive and negative consequences to proposed solutions	0	1	2
12. Ensures that solutions are positive, specific, and realistic	0	1	2
13. Maintains leadership and control of the group	0	1	2
14. Works well with the co-facilitator	0	1	2

**Comments/Notes:**

ASSESSMENT AREA	MAXIMUM SCORE	ASSESSMENT SCORE
Overall score on modelling behaviour	10	
Overall score on using AECP	16	
Overall score on collaborative leadership	30	
Total score facilitation skills (FACILITATOR 2)	56	

### OVERALL ASSESSMENT FACILITATOR 1

Activities assessment		Skills assessment	
Total score on core activity (A)	<b>20</b>	Total score core facilitation skills (C)	<b>50</b>
Total possible score (B)	<b>22</b>	Total possible score (D)	<b>56</b>
Total percent score core activities = (A/B) x 100 %	<b>%</b>	Total percent score core skills = (C/D) x 100 %	<b>%</b>
<b>What are the Facilitator's strengths?</b>			
<b>What does the Facilitator need to improve?</b>			
<b>Recommendations:</b>			

## OVERALL ASSESSMENT FACILITATOR 2

Activities assessment		Skills assessment	
Total score on core activity (A)		Total score core facilitation skills (C)	
Total possible score (B)	<b>22</b>	Total possible score (D)	<b>56</b>
Total percent score core activities = (A/B) x 100 %	%	Total percent score core skills = (C/D) x 100 %	%
<b>What are the Facilitator's strengths?</b>			
<b>What does the Facilitator need to improve?</b>			
<b>Recommendations:</b>			

### Supplementary File 3 – PLH-FAT-T Descriptive Statistics

**Table 12**

*Descriptive Statistics for the PLH-FAT-T*

<u>Home Activity Items (N=43)</u>							
<u>Item</u>	<u>Mean</u>	<u>SD</u>	<u>Median</u>	<u>“0” Inadequate N (%)</u>	<u>“1” Good N (%)</u>	<u>“2” Excellent N (%)</u>	<u>Missing N (%)</u>
HA_1	1.65	0.53	2.00	1 (2.3)	13 (30.2)	29 (67.4)	0
HA_2	1.45	0.77	2.00	7 (16.7)	9 (21.4)	26 (61.9)	1 (2.3)
HA_3	1.67	0.47	2.00	0	14 (32.6)	29 (67.4)	0
HA_4	1.52	0.67	2.00	4 (9.5)	12 (28.6)	26 (61.9)	1 (2.3)
HA_5	1.55	0.60	2.00	2 (4.8)	15 (35.7)	25 (59.5)	1 (2.3)
HA_6	1.46	0.60	2.00	2 (4.9)	18 (43.9)	21 (51.2)	2 (4.7)
HA_7	1.38	0.73	2.00	6 (14.3)	14 (33.3)	22 (52.4)	1 (2.3)
HA_8	1.24	0.76	1.00	8 (19.0)	16 (38.1)	18 (42.9)	1 (2.3)
HA_9	1.29	0.74	1.00	7 (16.7)	16 (38.1)	19 (45.2)	1 (2.3)
HA_10	1.45	0.67	2.00	4 (9.5)	15 (35.7)	23 (54.8)	1 (2.3)
HA_11	1.81	0.40	2.00	0	8 (19.0)	34 (81.0)	1 (2.3)
<i>Note: Each item has a different number of missing values, therefore influencing the comparability of the percentages in the tables.</i>							

<u>Role-play Items (N=43)</u>							
<u>Item</u>	<u>Mean</u>	<u>SD</u>	<u>Median</u>	<u>“0”</u> <u>Inadequate</u> <u>N (%)</u>	<u>“1”</u> <u>Good</u> <u>N (%)</u>	<u>“2”</u> <u>Excellent</u> <u>N (%)</u>	<u>Missing</u> <u>N (%)</u>
RP_1	1.86	0.35	2.00	0	6 (14.0%)	37 (86.0%)	0
RP_2	1.72	0.46	2.00	0	11 (28.2%)	28 (71.8%)	4 (9.3%)
RP_3	1.70	0.47	2.00	0	13 (30.2%)	30 (69.8%)	0
RP_4	1.74	0.44	2.00	3 (7.1%)	20 (47.6%)	19 (45.2%)	1 (2.3%)
RP_5	1.38	0.62	1.00	3 (7.1%)	20 (47.6%)	19 (45.2%)	1 (2.3%)
RP_6	1.39	0.64	1.00	3 (8.3%)	16 (44.4%)	17 (47.2%)	7 (16.3%)
RP_7	1.51	0.66	2.00	3 (8.6%)	11 (31.4%)	21 (60.0%)	8 (18.6%)
RP_8	1.53	0.61	2.00	2 (5.6%)	13 (36.1%)	21 (58.3%)	7 (16.3%)
RP_9	1.29	0.75	1.00	6 (17.1%)	13 (37.1%)	16 (45.7%)	8 (18.6%)
RP_10	1.39	0.69	1.50	4 (11.1%)	14 (38.9%)	18 (50.0%)	7 (16.3%)
RP_11	1.80	0.47	2.00	1 (2.9%)	5 (14.3%)	29 (82.9%)	8 (18.6%)
<i>Note: Each item has a different number of missing values, therefore influencing the comparability of the percentages in the tables.</i>							

<u>Skills Items (N=95)</u>							
<u>Item</u>	<u>Mean</u>	<u>SD</u>	<u>Median</u>	<u>“0”</u> <u>Inadequate</u> <u>N (%)</u>	<u>“1”</u> <u>Good</u> <u>N (%)</u>	<u>“2”</u> <u>Excellent</u> <u>N (%)</u>	<u>Missing</u> <u>N (%)</u>
MB_1	1.73	0.47	2.00	1 (1.1)	22 (24.4)	67 (74.4)	5 (5.3)
MB_2	1.81	0.39	2.00	0	17 (18.9)	73 (81.1)	5 (5.3)
MB_3	1.71	0.48	2.00	1 (1.1)	24 (27.0)	64 (71.9)	6 (6.3)
MB_4	1.56	0.58	2.00	4 (4.5)	31 (34.8)	54 (60.7)	6 (6.3)
MB_5	1.89	0.32	2.00	0	10 (11.2)	79 (88.8)	6 (6.3)
AECP_1	1.74	0.49	2.00	2 (2.4)	18 (21.2)	65 (76.5)	10 (10.5)
AECP_2	1.87	0.36	2.00	0	13 (15.3)	72 (84.7)	10 (10.5)
AECP_3	1.50	0.62	2.00	5 (6.4)	29 (37.2)	44 (56.4)	17 (17.9)
AECP_4	1.53	0.59	2.00	4 (4.7)	32 (37.6)	49 (57.6)	10 (10.5)
AECP_5	1.49	0.63	2.00	6 (7.1)	31 (36.5)	48 (56.5)	10 (10.5)
AECP_6	1.67	0.50	2.00	1 (1.2)	27 (32.1)	56 (66.7)	11 (11.6)
AECP_7	1.48	0.57	2.00	3 (3.5)	39 (45.3)	44 (51.2)	9 (9.5)
AECP_8	1.62	0.56	2.00	3 (3.6)	26 (31.0)	55 (65.5)	11 (11.6)

CL_1	1.83	0.38	2.00	0	14 (16.9)	69 (83.1)	12 (12.6)
CL_2	1.86	0.41	2.00	2 (2.5)	7 (8.6)	72 (88.9)	14 (14.7)
CL_3	1.78	0.42	2.00	0	18 (21.7)	65 (78.3)	12 (12.6)
CL_4	1.76	0.43	2.00	0	20 (24.1)	63 (75.9)	12 (12.6)
CL_5	1.74	0.44	2.00	0	21 (25.9)	60 (74.1)	14 (14.7)
CL_6	1.30	0.70	2.00	11 (13.6)	35 (43.2)	35 (43.2)	14 (14.7)
CL_7	1.53	0.57	2.00	3 (3.7)	32 (39.5)	46 (56.8)	14 (14.7)
CL_8	1.56	0.59	2.00	4 (5.0)	27 (33.8)	49 (61.2)	15 (15.8)
CL_9	1.80	0.43	2.00	1 (1.3)	14 (17.7)	64 (81.0)	16 (16.8)
CL_10	1.84	0.37	2.00	0	13 (16.0)	68 (84.0)	15 (15.8)
CL_11	1.33	0.69	2.00	10 (12.5)	34 (42.5)	36 (45.0)	15 (15.8)
CL_12	1.35	0.68	2.00	9 (11.2)	34 (42.5)	37 (46.2)	15 (15.8)
CL_13	1.36	0.68	2.00	9 (11.2)	33 (41.2)	38 (47.5)	15 (15.8)
CL_14	1.73	0.47	2.00	1 (1.3)	19 (24.1)	59 (74.7)	16 (16.8)
CL_15	1.78	0.45	2.00	1 (1.2)	16 (20.0)	63 (78.8)	15 (15.8)
<i>Note:</i> Each item has a different number of missing values, therefore influencing the comparability of the percentages in the tables.							



### Supplementary File 4 – Iterative Exploratory Factor Analyses

**Table 13**

*Exploratory Factor Analysis of Home Activity Items – Round One*

<u>Home Activity</u> <u>Item</u>	<u>Factor</u> <u>Loading</u>	<u>Item Wording</u> <u>The facilitator...</u>
HA_1	-	Reminds the participants of the core home activity for the previous week at the beginning of the discussion
HA_2	<b>0.782</b>	Reviews the core building blocks from previous session with participants at the beginning of the discussion
HA_3	0.348	Allows participants to share their experiences of how home activity went after the previous session
HA_4	-	Involves teens in the discussion about the home activity
HA_5	0.282	Keeps participants focused on the core home activity
HA_6	<b>0.550</b>	Helps participants connect their experiences to the core building blocks
HA_7	<b>0.761</b>	Identifies at least one specific challenge experienced by a participant regarding the main home activity
HA_8	<b>0.812</b>	Explores solutions to challenges shared and help participants choose a specific solution
HA_9	<b>0.608</b>	Practices the chosen solution with parents and teens
HA_10	<b>0.601</b>	Debriefs with the participants after practicing and encourages them to try the solution at home
HA_11	<b>0.853</b>	Thanks and praises participants for sharing experiences (at the end of the home activity discussion)
SS Loadings		3.81
Proportion Variance		0.35
Cronbach alpha		0.71
Omega		0.74
RMSEA Index		0.12

KMO	0.61
VSS	The Velicer MAP achieves a minimum of 0.05 with 1 factor
<i>Note:</i> Bolded items indicate factor loadings above  0.5 . HA stands for home activity discussion.	

**Table 14**

*Exploratory Factor Analysis of Home Activity Items – Round Two*

<u>Home Activity</u> <u>Item</u>	<u>Factor</u> <u>Loading</u>	<u>Item Wording</u> <u>The facilitator...</u>
HA_2	<b>0.772</b>	Reviews the core building blocks from previous session with participants at the beginning of the discussion
HA_6	<b>0.559</b>	Helps participants connect their experiences to the core building blocks
HA_7	<b>0.775</b>	Identifies at least one specific challenge experienced by a participant regarding the main home activity
HA_8	<b>0.872</b>	Explores solutions to challenges shared and help participants choose a specific solution
HA_9	<b>0.857</b>	Practices the chosen solution with parents and teens
HA_10	<b>0.655</b>	Debriefs with the participants after practicing and encourages them to try the solution at home
HA_11	<b>0.804</b>	Thanks and praises participants for sharing experiences (at the end of the home activity discussion)
SS Loadings		3.69
Proportion Variance		0.53
Cronbach alpha		0.81
Omega		0.82
MSEA Index		0.20
KMO		0.69
VSS		The Velicer MAP achieves a minimum of 0.09 with 1 factor
<i>Note:</i> Bolded items indicate factor loadings above  0.5 . HA stands for home activity discussion.		

**Table 15***Exploratory Factor Analysis of Role-play Items – Round One*

<u>Role-play Item</u>	<u>Factor 1 Loading</u>	<u>Factor 2 Loading</u>	<u>Item Wording</u> <i>The facilitator...</i>
RP_1	<b>0.625</b>	-	Provides introductory context to the role-play
RP_2	<b>0.781</b>	-	Acts out the role-play following the steps for leading the role-play (fidelity)
RP_3	<b>0.934</b>	-	Acts out the role-play following tips for leading the role-play (quality)
RP_4	<b>0.689</b>	-0.185	Discusses the role-play with participants after facilitators acted it out
RP_5	0.328	<b>0.529</b>	Explores possible solutions for negative role-plays
RP_6	-	-	Gets participants to act out a positive scenario that changes a negative role-play into a positive one
RP_7	-	0.363	Gives support to participants during the positive role-play (shadowing)
RP_8	0.138	<b>0.780</b>	Debriefs with participants about experiences and feelings after acting out the positive scenario
RP_9	-	<b>0.557</b>	Discusses with participants about how the role-play relates to their lives
RP_10	-0.193	<b>0.822</b>	Connects the role-play to the building blocks of the session
RP_11	0.316	-	Thanks and praises the participants for sharing their ideas and solutions
SS Loadings	2.616	2.056	
Proportion Variance	0.238	0.187	
Cumulative Variance	0.238	0.425	
Cronbach alpha			0.64
Omega			0.71
RMSEA Index			0.08
KMO			0.53
VSS			The Velicer MAP achieves a minimum of 0.05 with 2 factors

*Note:* Bolded items indicate factor loadings above |0.5|. RP stands for role-play.

**Table 16***Exploratory Factor Analysis of Role-play Items – Round Two*

<u>Role-play Item</u>	<u>Factor 1</u> <u>Loading</u>	<u>Factor 2</u> <u>Loading</u>	<u>Item Wording</u> <u>The facilitator...</u>
RP_1	<b>0.752</b>	-	Provides introductory context to the role-play
RP_2	<b>0.934</b>	-	Acts out the role-play following the steps for leading the role-play (fidelity)
RP_3	<b>0.809</b>	0.171	Acts out the role-play following tips for leading the role-play (quality)
RP_4	<b>0.594</b>	-0.116	Discusses the role-play with participants after facilitators acted it out
RP_5	0.278	<b>0.599</b>	Explores possible solutions for negative role-plays
RP_8	0.181	<b>0.752</b>	Debriefs with participants about experiences and feelings after acting out the positive scenario
RP_9	-0.114	<b>0.607</b>	Discusses with participants about how the role-play relates to their lives
RP_10	-0.230	<b>0.819</b>	Connects the role-play to the building blocks of the session
SS Loadings	2.622	2.010	
Proportion Variance	0.328	0.251	
Cumulative Variance	0.325	0.579	
Cronbach alpha			0.68
Omega			0.76
RMSEA Index			0.20
KMO			0.62
VSS		The Velicer MAP achieves a minimum of 0.07 with 2 factors	

*Note:* Bolded items indicate factor loadings above |0.5|. RP stands for role-play.

**Table 17***Exploratory Factor Analysis of Skills Items – Round One*

<u>Skills Items</u>	<u>Factor Loading</u>	<u>Item Wording</u>
MB_1	<b>0.526</b>	Gives positive reinforcement and specific praise to parents and teens
MB_2	<b>0.738</b>	Gives positive, specific, and realistic instructions
MB_3	<b>0.606</b>	Maintains commitments to time management principles
MB_4	0.420	Models behaviours with co-facilitator
MB_5	<b>0.665</b>	Demonstrates respectful behaviour towards participants
AECP_1	-	Uses body language to show acceptance
AECP_2	<b>0.608</b>	Accepts parent responses verbally
AECP_3	<b>0.651</b>	Explores participant experiences and opinions using open-ended questions
AECP_4	<b>0.584</b>	Explores thoughts and feelings
AECP_5	0.421	Explores the perspective of the teen
AECP_6	0.447	Explores the perspective of the parent
AECP_7	<b>0.561</b>	Connects experiences to the building blocks the from session
AECP_8	<b>0.572</b>	Identifies opportunities to practice skills (in addition to the structured group practice)
CL_1	<b>0.549</b>	Arranges the room in a way that encourages equal and active participation
CL_2	<b>0.841</b>	Facilitator is situated within the group, is at the level of the participants, and in a different place than the co-facilitator

CL_3	<b>0.691</b>	Participants appear comfortable and engaged in the session
CL_4	<b>0.691</b>	Participant-facilitator speaking ratio
CL_5	<b>0.754</b>	Assures equal and active participation among participants
CL_6	0.336	Engages quiet or non-participating parents and teens
CL_7	0.411	Limits participant responses
CL_8	<b>0.579</b>	Keeps participants focused on the topic of discussion
CL_9	<b>0.582</b>	Demonstrates knowledge of session content
CL_10	<b>0.668</b>	Delivers the session with confidence
CL_11	<b>0.527</b>	Helps participants generate their own ideas regarding principles or solutions to challenges
CL_12	0.468	Helps participants assess positive and negative consequences to proposed solutions
CL_13	0.331	Ensures that solutions are positive, specific, and realistic
CL_14	0.241	Maintains leadership and control of the group
CL_15	0.399	Works well with the co-facilitator
SS Loadings		8.721
Proportion Variance		0.311
Cronbach alpha		0.87
Omega		0.87
RMSEA index		0.643
KMO		0.61
VSS		The Velicer MAP achieves a minimum of 0.02 with 1 factor
<i>Note:</i> Bolded items indicate factor loadings above  0.5 . MB stands for modelling behaviour, AECp stands for Accept-Explore-Connect-Practice, and CL stands for collaborative leadership.		

**Table 18***Exploratory Factor Analysis of Skills Items – Round Two*

<u>Skills Items</u>	<u>Factor Loading</u>	<u>Item Wording</u>
MB_1	0.495	Gives positive reinforcement and specific praise to parents and teens
MB_2	<b>0.760</b>	Gives positive, specific, and realistic instructions
MB_3	<b>0.580</b>	Maintains commitments to time management principles
MB_5	<b>0.700</b>	Demonstrates respectful behaviour towards participants
AECP_2	<b>0.628</b>	Accepts parent responses verbally
AECP_3	<b>0.668</b>	Explores participant experiences and opinions using open-ended questions
AECP_4	<b>0.602</b>	Explores thoughts and feelings
AECP_7	<b>0.571</b>	Connects experiences to the building blocks the from session
AECP_8	<b>0.580</b>	Identifies opportunities to practice skills (in addition to the structured group practice)
CL_1	<b>0.566</b>	Arranges the room in a way that encourages equal and active participation
CL_2	<b>0.874</b>	Facilitator is situated within the group, is at the level of the participants, and in a different place than the co-facilitator
CL_3	<b>0.731</b>	Participants appear comfortable and engaged in the session
CL_4	<b>0.750</b>	Participant-facilitator speaking ratio
CL_5	<b>0.751</b>	Assures equal and active participation among participants
CL_8	<b>0.577</b>	Keeps participants focused on the topic of discussion

CL_9	<b>0.613</b>	Demonstrates knowledge of session content
CL_10	<b>0.686</b>	Delivers the session with confidence
CL_11	0.490	Helps participants generate their own ideas regarding principles or solutions to challenges
SS Loadings		7.68
Proportion Variance		0.43
Cronbach alpha		0.87
Omega		0.87
RMSEA index		0.119
KMO		0.70
VSS	The Velicer MAP achieves a minimum of 0.03 with 1 factor	
<i>Note:</i> Bolded items indicate factor loadings above  0.5 . MB stands for modelling behaviour, AECP stands for Accept-Explore-Connect-Practice, and CL stands for collaborative leadership.		



**Table 19***Exploratory Factor Analysis of Skills Items – Round Three*

<u>Skills Items</u>	<u>Factor Loading</u>	<u>Item Wording</u>
MB_2	<b>0.759</b>	Gives positive, specific, and realistic instructions
MB_3	<b>0.588</b>	Maintains commitments to time management principles
MB_5	<b>0.703</b>	Demonstrates respectful behaviour towards participants
AECP_2	<b>0.655</b>	Accepts parent responses verbally
AECP_3	<b>0.648</b>	Explores participant experiences and opinions using open-ended questions
AECP_4	<b>0.579</b>	Explores thoughts and feelings
AECP_7	<b>0.555</b>	Connects experiences to the building blocks the from session
AECP_8	<b>0.575</b>	Identifies opportunities to practice skills (in addition to the structured group practice)
CL_1	<b>0.555</b>	Arranges the room in a way that encourages equal and active participation
CL_2	<b>0.893</b>	Facilitator is situated within the group, is at the level of the participants, and in a different place than the co-facilitator
CL_3	<b>0.735</b>	Participants appear comfortable and engaged in the session
CL_4	<b>0.762</b>	Participant-facilitator speaking ratio
CL_5	<b>0.752</b>	Assures equal and active participation among participants
CL_8	<b>0.551</b>	Keeps participants focused on the topic of discussion
CL_9	<b>0.639</b>	Demonstrates knowledge of session content
CL_10	<b>0.707</b>	Delivers the session with confidence
SS Loadings		7.244
Proportion Variance		0.453

Cronbach alpha	0.87
Omega	0.87
RMSEA index	0.114
KMO	0.78
VSS	The Velicer MAP achieves a minimum of 0.03 with 1 factor
<i>Note:</i> Bolded items indicate factor loadings above  0.5 . MB stands for modelling behaviour, AECP stands for Accept-Explore-Connect-Practice, and CL stands for collaborative leadership.	

## Appendix 9: Paper 3 Supplementary Files

### Supplementary File 1 – Fixed and Random Effects

<b>Table 4</b>				
<i>Intraclass Correlations for Fixed and Random Effects Included in Multi-Level Poisson Models</i>				
<u>Outcome</u>	<u>Participant ID</u>	<u>Facilitator ID</u>	<u>Wave</u>	<u>District</u>
Parent-Reported Child Maltreatment	0.78	0.70	0.24	0.09
Parent-Reported Child Conduct Problems	0.59	0.55	0.18	0.12
Parent-Reported Positive Involvement	0.83	0.89	5.87e-10	0.66
Parent-Reported Poor Supervision	0.84	0.53	0.02	0.27
Parent-Reported Parenting Stress	0.57	0.62	0.00049	0.60
Parent-Reported Acceptability of Corporal Punishment	8.230632e-10	0.42	1.38e-10	0.25
Parent-Reported Depression	1.99e-09	0.24	0.00093	0.12
Parent-Reported Financial Insecurity	0.44	0.60	0.01	0.30

Parent-Reported Sexual Health Communication	0.25	0.22	0.006	0.06
Parent-Reported Parental Support for Education	0.32	0.47	5.44e-10	0.27
Parent-Reported IPV Perpetration	0.93	0.98	3.56e-10	0.13
Parent-Reported IPV Victimization	0.94	0.91	4.29e-10	0.14
Adolescent-Reported Child Maltreatment	0.78	0.78	0.30	0.09
Adolescent-Reported Positive Parental Involvement	0.74	0.65	6.91e-10	0.46
Adolescent-Reported Poor Supervision	0.78	0.61	0.05	0.22
Adolescent-Reported Conduct Problems	0.64	0.42	0.34	0.13
Adolescent-Reported Emotional Problems	0.74	0.94	5.49e-10	0.07

Adolescent-Reported Depression	0.36	0.36	0.02	0.19
Adolescent-Reported Acceptability of Corporal Punishment	9.92e-10	0.53	0.00097	0.22
Adolescent-Reported Sexual Health Communication	0.34	0.29	0.02	0.14
Adolescent-Reported Parental Support of Education	0.73	0.91	7.03e-10	0.61
Adolescent-Reported School Violence	0.95	0.97	7.34-10	0.65

**Supplementary File 2** – Analyses of Associations between Facilitator Competent Adherence and Outcomes

<b>Table 5</b>							
<i>Associations Between Facilitator Competent Adherence and Parent-Reported Outcomes</i>							
<u>Child Maltreatment</u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.54	0.03	p<0.001***	p<0.001***	0.58	0.55	0.62
Index	0.21	0.09	p=0.02	p=0.03	1.23	1.03	1.46
Time*Index	0.05	0.02	p=0.06	p=0.10	1.05	1.00	1.10
<u>Poor Supervision</u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.45	0.04	p<0.001***	p<0.001***	0.63	0.59	0.68
Index	0.01	0.12	p=0.96	p=0.98	1.01	0.79	1.28
Time*Index	0.08	0.03	p=0.001***	p<0.01**	1.08	1.03	1.15
<u>Child Conduct Problems</u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.03	0.03	p=0.31	p=0.40	0.97	0.91	1.03
Index	0.25	0.08	p=0.001***	p<0.01*	1.28	1.10	1.50
Time*Index	0.01	0.03	p=0.63	p=0.73	1.01	0.96	1.07
<u>Parenting Stress</u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.34	0.02	p<0.00***	p<0.001***	0.71	0.68	0.74
Index	0.16	0.08	p=0.04 .	p=0.10	1.17	1.01	1.37
Time*Index	0.10	0.02	p<0.00***	p<0.00***	1.11	1.07	1.15
<u>Parenting Depression</u>							

	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.19	0.02	p<0.001***	p<0.001***	0.82	0.79	0.86
Index	0.11	0.03	p<0.001***	p<0.00***	1.12	1.05	1.19
Time*Index	-0.04	0.01	p<0.001***	p<0.01**	0.96	0.94	0.99
<u>Financial Insecurity</u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.06	0.02	p=0.01**	p=0.02*	0.94	0.90	0.98
Index	-0.04	0.08	p=0.64	p=0.73	0.96	0.82	1.13
Time*Index	0.05	0.01	p<0.001***	p<0.001***	1.05	1.03	1.08
<u>Sexual Health Communication</u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	0.51	0.02	p<0.001***	p<0.001***	1.67	1.61	1.74
Index	0.01	0.04	p=0.89	p=0.97	1.01	0.92	1.10
Time*Index	0.01	0.01	p=0.46	p=0.59	1.01	0.99	1.03
<u>Acceptability of Corporal Punishment ◇</u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.72	0.10	p<0.00***	p<0.00***	0.49	0.40	0.60
Index	0.23	0.19	p=0.23	p=0.35	1.26	0.86	1.83
Time*Index	0.25	0.07	p<0.001***	p<0.01**	1.28	1.11	1.49
<u>Positive Involvement ^</u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.97	0.06	p<0.001***	p<0.001***	0.38	0.34	0.43
Index	-0.07	0.11	p=0.52	p=0.65	0.93	0.76	1.15
Time*Index	0.42	0.05	p<0.001***	p<0.001***	1.52	1.37	1.69
<u>Parental Support of Education ^</u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>

Time	-0.17	0.04	p<0.001***	p<0.001***	0.84	0.78	0.91
Index	-0.04	0.05	p=0.35	p=0.50	0.96	0.87	1.05
Time*Index	0.04	0.03	p=0.31	p=0.50	1.04	0.97	1.11
<u>IPV Perpetration ^</u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	1.01	0.10	p<0.001***	p<0.001***	2.75	2.24	3.38
Index	0.18	0.25	p=0.47	p=0.63	1.20	0.73	1.98
Time*Index	0.47	0.09	p<0.001***	p=0.83	1.61	1.36	1.90
<u>IPV Victimization ^</u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	1.20	0.11	p<0.001***	p<0.001***	3.33	2.68	4.14
Index	0.20	0.21	p=0.33	p=0.50	1.23	0.82	1.84
Time*Index	0.59	0.09	p<0.001***	p<0.001***	1.81	1.52	2.16
Note: For child maltreatment, poor supervision, child conduct problems, parenting stress, parent depression, financial insecurity, and acceptability of corporal punishment, data from N=3,057 parents and N=24 facilitators were used. For positive involvement, parental support of education, IPV perpetration, and IPV victimization, data from N=1,654 parents and N=22 facilitators were used. Signif. codes: ***p<0.001, **p<0.01, *p<0.05. The outcomes with the “^” symbol were run without wave as a fixed effect due to rank deficiency. The outcome with the “◊” was run as a logistic regression as it is binary. IRR is the incidence rate ratio. LL is the lower bound and UL is the upper bound of the 95% confidence interval.							



**Table 6***Associations Between Facilitator Competent Adherence and Adolescent-Reported Outcomes*Child Maltreatment

	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.62	0.03	p<0.001***	p<0.001***	0.54	0.51	0.57
Index	0.37	0.09	p<0.001***	p<0.001***	1.44	1.20	1.73
Time*Index	-0.15	0.02	p<0.001***	p<0.001***	0.86	0.82	0.90

Poor Supervision

	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.41	0.03	p<0.001***	p<0.001***	0.67	0.62	0.71
Index	0.00	0.13	p=0.97	p=1.00	1.00	0.78	1.29
Time*Index	-0.07	0.02	p<0.001***	p<0.01**	0.93	0.89	0.97

Child Conduct Problems

	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.09	0.03	p=0.01**	p=0.01**	0.91	0.85	0.97
Index	0.25	0.06	p<0.001***	p<0.001***	1.28	1.13	1.45
Time*Index	-0.31	0.02	p<0.001***	p<0.001***	0.73	0.70	0.77

Depression

	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.11	0.02	p<0.001***	p<0.001***	0.90	0.86	0.94
Index	0.01	0.06	p=0.84	p=0.90	1.01	0.91	1.13
Time*Index	-0.02	0.01	p=0.25	p=0.33	0.99	0.96	1.01

Sexual Health Communication

	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	0.59	0.02	p<0.001***	p<0.001***	1.80	1.73	1.88
Index	-0.03	0.06	p=0.54	p=0.61	0.97	0.86	1.08

Time*Index	0.09	0.01	p<0.001***	p<0.00***	1.10	1.07	1.12
<u>Acceptability of Corporal Punishment <math>\diamond</math></u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-1.40	0.13	p<0.001***	p<0.001***	0.25	0.19	0.32
Index	0.24	0.24	p=0.32	p=0.48	1.27	0.79	2.04
Time*Index	-0.06	0.08	p=0.47	p=0.57	0.94	0.80	1.11
<u>Positive Parental Involvement <math>\wedge</math></u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.39	0.05	p<0.001***	p<0.001***	0.68	0.62	0.75
Index	-0.05	0.07	p=0.44	p=0.67	0.95	0.83	1.08
Time*Index	-0.01	0.04	p=0.90	p=0.90	1.00	0.92	1.08
<u>Child Emotional Problems <math>\wedge</math></u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	0.53	0.07	p<0.001***	p<0.001***	1.70	1.49	1.94
Index	0.34	0.18	p=0.06	p=0.12	1.41	0.99	2.01
Time*Index	-0.26	0.06	p<0.001***	p<0.001***	0.77	0.69	0.86
<u>Parental Support of Education <math>\wedge</math></u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.74	0.10	p<0.001***	p<0.001***	0.48	0.40	0.58
Index	-0.11	0.14	p=0.44	p=0.67	0.90	0.68	1.18
Time*Index	0.80	0.08	p<0.001***	p<0.001***	2.22	1.90	2.58
<u>School Violence <math>\wedge</math></u>							
	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	0.63	0.08	p<0.001***	p<0.001***	1.87	1.60	2.19
Index	0.39	0.19	p=0.05*	p=0.11	1.47	1.00	2.15
Time*Index	0.53	0.07	p<0.001***	p<0.001***	1.69	1.49	1.92

Note: For child maltreatment, poor supervision, child conduct problems, parenting stress, parent depression, financial insecurity, and acceptability of corporal punishment, data from  $N=3,057$  adolescents and  $N=24$  facilitators were used. For positive involvement, child emotional problems, parental support of education, IPV perpetration, and IPV victimization, data from  $N=1,684$  adolescents and  $N=22$  facilitators were used. Signif. codes: \*\*\* $p<0.001$ , \*\* $p<0.01$ , \* $p<0.05$ . The outcomes with the “^” symbol were run without wave as a fixed effect due to rank deficiency. The outcome with the “◇” was run as a logistic regression as it is binary. IRR is the incidence rate ratio. LL is the lower bound and UL is the upper bound of the 95% confidence interval.

**Table 7**  
*Associations between Parent- and Adolescent-Reported Outcomes and Facilitator Skills*

Parent- and Adolescent-Reported Child Maltreatment

<i>N=23 facilitators N=2868 parents</i>		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>			
Time	-2.11	0.31	$p<0.001^{***}$	$p<0.001^{***}$	0.12	0.07	0.22
Skills	0.09	0.03	$p<0.001^{***}$	$p<0.001^{***}$	1.09	1.03	1.15
Time*Skills	0.06	0.01	$p<0.001^{***}$	$p<0.001^{***}$	1.06	1.04	1.08
<i>N=23 facilitators N=2868 adolescents</i>		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>			
Time	0.29	0.28	$p=0.30$	$p=0.41$	1.33	0.77	2.31
Skills	0.15	0.03	$p<0.001^{***}$	$p<0.001^{***}$	1.17	1.11	1.23
Time*Skills	-0.04	0.01	$p<0.001^{***}$	$p=.001^{***}$	0.96	0.95	0.98

Parent- and Adolescent-Reported Poor Supervision

<i>N=23 facilitators N=2868 parents</i>		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>			
Time	-2.82	0.37	$p<0.001^{***}$	$p<0.001^{***}$	0.06	0.03	0.12
Skills	0.01	0.04	$p=0.76$	$p=0.82$	1.01	0.93	1.10
Time*Skills	0.09	0.01	$p<0.001^{***}$	$p<0.001^{***}$	1.09	1.06	1.12

<u>Parent- and Adolescent-Reported Child Conduct Problems</u>							
N=23 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=2868 adolescents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.99	0.24	p<0.001***	p<0.001***	0.37	0.23	0.59
Skills	-0.01	0.04	p=0.82	p=0.85	0.99	0.91	1.07
Time*Skills	0.02	0.01	p=0.01**	p=0.03*	1.02	1.00	1.04
<u>Adolescent-Reported Emotional Problems ^</u>							
N=23 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=2868 parents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-2.40	0.34	p<0.001***	p<0.001***	0.09	0.05	0.18
Skills	0.08	0.03	p<0.001***	p<0.001***	1.08	1.02	1.14
Time*Skills	0.09	0.01	p<0.001***	p<0.001***	1.09	1.07	1.12
N=23 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=2868 parents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	1.44	0.26	p<0.001***	p<0.001***	4.24	2.55	7.05
Skills	0.07	0.02	p<0.001***	p<0.001***	1.08	1.04	1.12
Time*Skills	-0.06	0.01	p<0.001***	p<0.001***	0.94	0.93	0.96
<u>Parent-Reported Parenting Stress</u>							
N=21 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=1588 adolescents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	1.28	0.77	p=0.10	p=0.17	3.58	0.79	16.33
Skills	0.17	0.05	p<0.001***	p<0.001***	1.19	1.07	1.32
Time*Skills	-0.03	0.03	p=0.38	p=0.51	0.98	0.92	1.03

<i>N</i> =23 facilitators <i>N</i> =2868 adolescents		<u>Std</u> <u>Error</u>	<u>Raw P</u> <u>Value</u>	<u>Adjusted P</u> <u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
	<u>Estimate</u>						
Time	-0.30	0.15	p=0.05*	p=0.10	0.74	0.55	1.00
Skills	0.05	0.03	p=0.08	p=0.13	1.05	0.99	1.11
Time*Skills	0.00	0.01	p=0.89	p=0.45	1.00	0.99	1.01
<u>Parent- and Adolescent-Reported Depression</u>							
<i>N</i> =23 facilitators <i>N</i> =2868 parents		<u>Std</u> <u>Error</u>	<u>Raw P</u> <u>Value</u>	<u>Adjusted P</u> <u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
	<u>Estimate</u>						
Time	0.40	0.12	p<0.001***	p=.002**	1.49	1.17	1.90
Skills	0.05	0.01	p<0.001***	p<0.001***	1.05	1.03	1.07
Time*Skills	-0.02	0.00	p<0.001***	p<0.001***	0.98	0.97	0.99
<i>N</i> =23 facilitators <i>N</i> =2868 adolescents		<u>Std</u> <u>Error</u>	<u>Raw P</u> <u>Value</u>	<u>Adjusted P</u> <u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
	<u>Estimate</u>						
Time	-0.29	0.12	p=0.01**	p=0.03*	0.75	0.59	0.94
Skills	0.01	0.02	p=0.55	p=0.59	1.01	0.97	1.05
Time*Skills	0.01	0.00	p=0.13	p=0.22	1.01	1.00	1.02
<u>Parent-Reported Financial Insecurity</u>							
<i>N</i> =23 facilitators <i>N</i> =2868 parents		<u>Std</u> <u>Error</u>	<u>Raw P</u> <u>Value</u>	<u>Adjusted P</u> <u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
	<u>Estimate</u>						
Time	-0.48	0.09	p<0.001***	p=0.46	0.62	0.52	0.74
Skills	-0.03	0.03	p=0.33	p=0.08	0.97	0.92	1.03
Time*Skills	0.02	0.00	p<0.001***	p<0.001***	1.02	1.01	1.02
<u>Parent- and Adolescent-Reported Sexual Health Communication</u>							

N=23 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=2868 parents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	0.68	0.08	p<0.001***	p<0.001***	1.97	1.67	2.32
Skills	-0.01	0.01	p=0.48	p=0.58	0.99	0.96	1.02
Time*Skills	-0.01	0.00	p=0.06	p=0.10	0.99	0.99	1.00
N=23 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=2868 adolescents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	0.10	0.08	p=0.21	p=0.30	1.11	0.95	1.29
Skills	-0.02	0.02	p=0.44	p=0.05*	0.98	0.95	1.02
Time*Skills	0.02	0.00	p<0.001***	p<0.001***	1.02	1.01	1.03
<u>Parent- and Adolescent-Reported Acceptability of Corporal Punishment ◇</u>							
N=23 facilitators		<u>Std.</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=2868 parents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.99	0.72	p=0.17	p=0.46	0.37	0.09	1.52
Skills	0.08	0.06	p=0.23	p=0.46	1.08	0.95	1.22
Time*Skills	0.02	0.03	p=0.54	p=0.81	1.02	0.97	1.07
N=23 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=2868 adolescents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-1.62	1.23	p=0.19	p=0.23	0.20	0.02	2.20
Skills	0.13	0.08	p=0.10	p=0.23	1.14	0.98	1.34
Time*Skills	0.00	0.05	p=0.93	p=0.93	1.00	0.91	1.09
<u>Parent- and Adolescent-Reported Positive Involvement △</u>							
N=21 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=1558 parents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-5.04	0.75	p<0.001***	p<0.001***	0.01	0.00	0.03
Skills	-0.01	0.03	p=0.78	p=0.82	0.99	0.93	1.06

Time*Skills	0.15	0.03	p<0.001***	p<0.001***	1.16	1.10	1.22
N=21 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=1588 adolescents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	0.70	0.54	p=0.20	p=0.68	2.00	0.69	5.79
Skills	-0.01	0.02	p=0.64	p=0.68	0.99	0.95	1.03
Time*Skills	-0.04	0.02	p=0.05*	p=0.08	0.96	0.92	1.00
<u>Parent- and Adolescent-Reported Parental Support of Education ^</u>							
N=21 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=1558 parents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.01	0.45	p=0.98	p=0.98	0.991	0.408	2.40
Skills	-0.01	0.02	p=0.37	p=0.57	0.986	0.957	1.02
Time*Skills	-0.01	0.02	p=0.71	p=0.79	0.994	0.961	1.03
N=21 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=1588 adolescents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-7.74	1.03	p<0.001***	p<0.001***	0.00	0.00	0.00
Skills	0.00	0.04	p=0.93	p=0.93	1.00	0.92	1.08
Time*Skills	0.25	0.04	p<0.001***	p<0.001***	1.28	1.19	1.37
<u>Parent-Reported IPV Perpetration ^</u>							
N=21 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=1558 parents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-23.69	1.96	p<0.001***	p<0.001***	5.16E-11	1.10E-13	2.41E-09
Skills	0.04	0.10	p=0.66	p=0.77	1.04E+00	8.62E-01	1.27E+00
Time*Skills	0.94	0.08	p<0.001***	p<0.001***	2.56E+00	2.20E+00	2.98E+00
<u>Parent-Reported IPV Victimization ^</u>							

N=21 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=1558 parents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-24.23	1.96	p<0.001***	p<0.001***	3.01E-11	6.46E-13	1.40E-09
Skills	0.04	0.08	p=0.60	p=0.75	1.04E+00	8.94E-01	1.22E+00
Time*Skills	0.96	0.08	p<0.001***	p<0.001***	2.61E+00	2.25E+00	3.03E+00
<u>Adolescent-Reported School Violence ^</u>							
N=21 facilitators		<u>Std</u>	<u>Raw P</u>	<u>Adjusted P</u>			
N=1558 adolescents	<u>Estimate</u>	<u>Error</u>	<u>Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-8.41	1.15	p<0.001***	p<0.001***	0.00	0.00	0.00
Skills	0.18	0.05	p<0.001***	p<0.001***	1.19	1.08	1.31
Time*Skills	0.33	0.04	p<0.001***	p<0.001***	1.39	1.28	1.51
Note: Signif. codes: ***p<0.001, **p<0.01, *p<0.05. The outcomes with the “^” symbol were run without wave as a fixed effect due to rank deficiency. The outcome with the “◇” was run as a logistic regression as it is binary. IRR is the incidence rate ratio. LL is the lower bound and UL is the upper bound of the 95% confidence interval.							

<b>Table 8</b>								
<i>Association between Parent-Reported Outcomes and Facilitator Adherence to Home Activity</i>								
<u>Child Maltreatment</u>								
N=11 facilitators					<u>Adjusted P</u>			
N=1397 parents		<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	0.00	0.08	0.96	9.77E-01	1.00	0.86	1.17	
Home Activity	0.00	0.03	0.89	9.77E-01	1.00	0.95	1.06	
Time*Home Activity	-0.04	0.01	p<0.001***	p<0.001***	0.96	0.94	0.97	
<u>Poor Supervision</u>								
N=11 facilitators					<u>Adjusted P</u>			
N=1397 parents		<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	0.60	0.10	p<0.001***	p<0.001***	1.83	1.50	2.23	
Home Activity	-0.02	0.04	p=0.58	p=0.77	0.98	0.91	1.06	



Time*Home Activity	-0.07	0.01	p<0.001***	p<0.001***	0.93	0.91	0.95
<u>Child Conduct Problems</u>							
N=11 facilitators				<u>Adjusted P</u>			
N=1397 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	0.30	0.09	p<0.001***	p<0.001***	1.34	1.13	1.60
Home Activity	0.00	0.03	p=0.93	p=0.98	1.00	0.94	1.05
Time*Home Activity	-0.01	0.01	p=0.11	p=0.24	0.99	0.97	1.00
<u>Parenting Stress</u>							
N=11 facilitators				<u>Adjusted P</u>			
N=1397 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.10	0.06	p=0.10	0.23	0.91	0.81	1.02
Home Activity	0.04	0.04	p=0.32	0.52	1.04	0.97	1.12
Time*Home Activity	-0.02	0.01	p<0.001***	p<0.001***	0.98	0.97	0.99
<u>Parenting Depression</u>							
N=11 facilitators				<u>Adjusted P</u>			
N=1397 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	0.00	0.06	p=0.98	p=0.98	1.00	0.90	1.11
Home Activity	0.01	0.01	p=0.42	p=0.65	1.01	0.99	1.02
Time*Home Activity	-0.02	0.01	p<0.001***	p<0.001***	0.98	0.97	0.99
<u>Financial Insecurity</u>							
N=11 facilitators				<u>Adjusted P</u>			
N=1397 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.15	0.07	p=0.03*	p=0.08	0.86	0.76	0.99
Home Activity	0.02	0.04	p=0.61	p=0.79	1.02	0.95	1.10
Time*Home Activity	0.04	0.01	p<0.001***	p<0.001***	1.04	1.03	1.05

<u>Sexual Health Communication</u>							
N=11 facilitators							
N=1397 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.09	0.07	p=0.22	p=0.38	0.92	0.80	1.05
Home Activity	0.01	0.02	p=0.71	p=0.86	1.01	0.97	1.04
Time*Home Activity	0.05	0.01	p<0.001***	p<0.001***	1.05	1.04	1.07
<u>Acceptability of Corporal Punishment ◇</u>							
N=11 facilitators							
N=1397 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.98	0.27	p<0.001***	p<0.001***	0.38	0.22	0.64
Home Activity	0.04	0.10	p=0.67	p=0.67	1.04	0.86	1.25
Time*Home Activity	0.07	0.02	p=0.01**	p=0.02*	1.07	1.02	1.12
<u>Positive Involvement △</u>							
N=10 facilitators							
N=767 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.91	0.19	p<0.001***	p<0.001***	0.40	0.28	0.58
Home Activity	0.00	0.05	0.95	p=0.95	1.00	0.91	1.09
Time*Home Activity	0.02	0.02	0.18	p=0.40	2.68	1.21	5.90
<u>Parental Support of Education △</u>							
N=10 facilitators							
N=767 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.06	0.11	p=0.58	p=0.65	0.94	0.75	1.17
Home Activity	0.01	0.02	p=0.55	p=0.11	1.01	0.97	1.05
Time*Home Activity	-0.01	0.01	p=0.27	p=0.54	1.45	1.02	2.06
<u>IPV Perpetration △</u>							

<i>N</i> =10 facilitators <i>N</i> =767 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-2.91	0.54	p<0.001***	p<0.001***	0.05	0.02	0.16
Home Activity	-0.09	0.15	p=0.55	p=0.65	0.91	0.68	1.23
Time*Home Activity	0.79	0.13	p<0.001***	p<0.001***	0.25	0.02	3.55
<u>IPV Victimization ^</u>							
<i>N</i> =10 facilitators <i>N</i> =767 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-2.80	0.50	p<0.001***	p<0.001***	0.06	0.02	0.16
Home Activity	-0.08	0.15	p=0.57	p<0.001***	0.92	0.69	1.23
Time*Home Activity	0.74	0.12	p<0.001***	p<0.001***	0.33	0.02	4.64
Note: Signif. codes: ***p<0.001, **p<0.01, *p<0.05. The outcomes with the “^” symbol were run without wave as a fixed effect due to rank deficiency. The outcome with the “◇” was run as a logistic regression as it is binary.							

<b>Table 9</b> <i>Association between Parent-Reported Outcomes and Facilitator Adherence to Role-play</i>							
<u>Child Maltreatment</u>							
<i>N</i> =13 facilitators <i>N</i> =1660 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-1.24	0.21	p<0.001***	p<0.001***	0.29	0.19	0.44
Role-play	0.06	0.12	p=0.61	p=0.77	1.06	0.85	1.33
Time*Role-play	0.05	0.02	p=0.01**	p=0.02*	1.05	1.01	1.08
<u>Poor Supervision</u>							
<i>N</i> =13 facilitators <i>N</i> =1660 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-3.99	0.39	p<0.001***	p<0.001***	0.02	0.01	0.04
Role-play	0.01	0.12	p=0.94	p=0.94	1.01	0.80	1.27

Time*Role-play	0.23	0.03	p<0.001***	p<0.001***	1.25	1.18	1.33
<u>Child Conduct Problems</u>							
N=13 facilitators							
N=1660 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.27	0.27	p=0.32	p=0.52	0.76	0.45	1.30
Role-play	0.19	0.09	p=0.04	p=0.08	1.21	1.01	1.45
Time*Role-play	-0.01	0.02	p=0.80	p=0.88	0.99	0.96	1.04
<u>Parenting Stress</u>							
N=13 facilitators							
N=1660 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-2.63	0.18	p<0.001***	p<0.001***	0.07	0.05	0.10
Role-play	0.06	0.08	p=0.49	p=0.65	1.06	0.90	1.24
Time*Role-play	0.18	0.01	p<0.001***	p<0.001***	1.20	1.17	1.23
<u>Parenting Depression</u>							
N=13 facilitators							
N=1660 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.83	0.14	p<0.001***	p<0.001***	0.44	0.33	0.57
Role-play	0.04	0.04	p=0.32	p=0.05*	1.04	0.96	1.13
Time*Role-play	0.05	0.01	p<0.001***	p<0.001***	1.05	1.03	1.07
<u>Financial Insecurity</u>							
N=13 facilitators							
N=1660 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	0.12	0.15	p=0.43	0.64	1.13	0.84	1.51
Role-play	-0.05	0.07	p=0.46	0.65	0.95	0.83	1.09
Time*Role-play	-0.03	0.01	p=0.01**	p=0.01**	0.97	0.94	0.99

<u>Sexual Health Communication</u>							
N=13 facilitators							
N=1660 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	0.54	0.12	p<0.001***	p<0.001***	1.71	1.35	2.18
Role-play	0.02	0.04	p=0.65	p=0.80	1.02	0.95	1.09
Time*Role-play	0.00	0.01	p=0.71	p=0.84	1.00	0.98	1.02
<u>Acceptability of Corporal Punishment <math>\diamond</math></u>							
N=13 facilitators							
N=1660 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-3.79	0.70	p<0.001***	p<0.001***	0.02	0.01	0.09
Role-play	-0.02	0.22	p=0.94	p=0.94	0.98	0.63	1.53
Time*Role-play	0.25	0.05	p<0.001***	p<0.001***	1.28	1.15	1.43
<u>Positive Involvement <math>\wedge</math></u>							
N=12 facilitators							
N=887 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-4.45	0.46	p<0.001***	p<0.001***	0.01	0.00	0.03
Role-play	-0.23	0.11	p=0.03 .	p=0.06	0.79	0.64	0.98
Time*Role-play	0.31	0.04	p<0.001***	p<0.001***	1.37	1.26	1.48
<u>Parental Support of Education <math>\wedge</math></u>							
N=12 facilitators							
N=887 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	-0.72	0.23	p<0.001***	p<0.001***	0.49	0.31	0.76
Role-play	-0.08	0.04	p=0.05*	p=0.08	0.92	0.85	1.00
Time*Role-play	0.05	0.02	p=0.02*	p=0.03*	1.05	1.01	1.09

<u>IPV Perpetration ^</u>							
N=12 facilitators							
N=887 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	6.49	0.75	p<0.001***	p<0.001***	658.10	150.17	2884.17
Role-play	0.27	0.26	p=0.31	p=0.34	1.31	0.78	2.19
Time*Role-play	-0.52	0.06	p<0.001***	p<0.001***	0.59	0.53	0.67
<u>IPV Victimization ^</u>							
N=12 facilitators							
N=887 parents	<u>Estimate</u>	<u>Std Error</u>	<u>Raw P Value</u>	<u>Adjusted P Value</u>	<u>IRR</u>	<u>LL</u>	<u>UL</u>
Time	6.06	0.73	p<0.001***	p=0.18	427.31	102.40	1783.04
Role-play	0.27	0.18	p=0.12	p=0.91	1.31	0.93	1.85
Time*Role-play	-0.47	0.06	p<0.001***	p<0.001***	0.63	0.56	0.71
Note: Signif. codes: ***p<0.001, **p<0.01, *p<0.05. The outcomes with the “^” symbol were run without wave as a fixed effect due to rank deficiency. The outcome with the “∇” was run as a logistic regression as it is binary.							

<b>Table 10</b>							
<i>Association between Adolescent-Reported Outcomes and Facilitator Adherence to Home Activity</i>							
<u>Child Maltreatment</u>							
N=11 facilitators							
N=1397 adolescents	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time	-0.24	0.08	p=0.004**	p=0.01**	0.79	0.67	0.93
Home Activity	0.04	0.02	p=0.08	p=0.16	1.04	1.00	1.08
Time*Home Activity	-0.06	0.01	p<0.001***	p<0.001***	0.94	0.93	0.96
<u>Poor Supervision</u>							
N=11 facilitators							
N=1397 adolescents	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time	0.18	0.10	p=0.09	p>1.00	1.19	0.98	1.46

Home Activity	0.03	0.09	p=0.72	p=0.83	1.03	0.87	1.23
Time*Home Activity	-0.03	0.01	p<0.001***	p=.001***	0.97	0.95	0.99
<u>Child Conduct Problems</u>							
N=11 facilitators							
N=1397 adolescents	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time	0.42	0.08	p<0.001***	p<0.001***	1.53	1.30	1.80
Home Activity	0.00	0.03	p=0.90	p=0.93	1.00	0.95	1.05
Time*Home Activity	-0.06	0.01	p<0.001***	p<0.001***	0.94	0.93	0.96
<u>Teen Depression</u>							
N=11 facilitators							
N=1397 adolescents	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time	-0.22	0.08	p=0.006	p=0.02*	0.80	0.69	0.94
Home Activity	0.02	0.02	p=0.45	p=0.56	1.02	0.98	1.06
Time*Home Activity	-0.01	0.01	p=0.15	p=0.24	0.99	0.97	1.00
<u>Sexual Health Communication</u>							
N=11 facilitators							
N=1397 adolescents	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time	-0.04	0.07	p=0.55	p=0.66	0.96	0.84	1.10
Home Activity	-0.02	0.03	p=0.42	p=0.56	0.98	0.92	1.03
Time*Home Activity	0.07	0.01	p<0.001***	p<0.001***	1.07	1.06	1.08
<u>Acceptability of Corporal Punishment ◇</u>							
N=11 facilitators							
N=1397 adolescents	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time	-0.31	0.29	p=0.28	p=0.34	0.73	0.42	1.29
Home Activity	0.01	0.06	p=0.84	p=0.84	1.01	0.89	1.15

Time*Home Activity	-0.17	0.03	p<0.001***	p<0.001***	0.84	0.79	0.90
<u>Positive Involvement ^</u>							
N=10 facilitators	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
N=797 adolescents							
Time	0.34	0.14	p=0.01**	p=0.04*	1.40	1.07	1.84
Home Activity	0.00	0.03	p=0.94	p=0.98	1.00	0.94	1.07
Time*Home Activity	-0.06	0.02	p<0.001***	p<0.001***	0.94	0.91	0.97
<u>Child Emotional Problems ^</u>							
N=10 facilitators	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
N=797 adolescents							
Time	0.57	0.16	p<0.001***	p=0.002**	1.40	1.28	2.42
Home Activity	-0.06	0.06	p=0.33	p=0.49	1.00	0.85	1.06
Time*Home Activity	-0.03	0.02	p=0.17	p=0.35	0.94	0.94	1.01
<u>Parental Support of Education ^</u>							
N=10 facilitators	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
N=797 adolescents							
Time	-2.12	0.27	p<0.001***	p<0.001***	0.12	0.07	0.20
Home Activity	0.00	0.08	p=0.98	p=0.98	1.00	0.86	1.16
Time*Home Activity	0.23	0.03	p<0.001***	p<0.001***	1.26	1.18	1.34
<u>School Violence ^</u>							
N=10 facilitators	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
N=797 adolescents							
Time	-1.24	0.24	p<0.001***	p<0.001***	0.29	0.18	0.47
Home Activity	-0.03	0.07	p=0.67	p=0.79	0.97	0.84	1.12



Time*Home Activity	0.19	0.02	p<0.001***	p<0.001***	1.21	1.15	1.27
Note: Signif. codes: ***p<0.001, **p<0.01, *p<0.05. The outcomes with the “^” symbol were run without wave as a fixed effect due to rank deficiency. The outcome with the “◇” was run as a logistic regression as it is binary.							

<b>Table 11</b>							
<i>Association between Adolescent-Reported Outcomes and Facilitator Adherence to Role-play</i>							
<u>Child Maltreatment</u>							
N=13 facilitators N=1660 adolescents	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time	0.03	0.18	p=0.86	p=0.86	1.03	0.72	1.47
Role-play	0.06	0.14	p=0.64	p=0.69	1.07	0.81	1.40
Time*Role-play	-0.05	0.02	p<0.001***	p=.004**	0.95	0.92	0.98
<u>Poor Supervision</u>							
N=13 facilitators N=1660 adolescents	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time	0.71	0.23	p<0.001***	p=.004**	2.04	1.31	3.18
Role-play	-0.08	0.07	p=0.24	0.39	0.92	0.81	1.06
Time*Role-play	-0.14	0.02	p<0.001***	p<0.001***	0.87	0.83	0.90
<u>Child Conduct Problems</u>							
N=13 facilitators N=1660 adolescents	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time	1.85	0.24	p<0.001***	p<0.001***	6.36	3.99	10.12
Role-play	0.17	0.06	p=0.01**	p=0.02*	1.19	1.05	1.35
Time*Role-play	-0.17	0.02	p<0.001***	p<0.001***	0.84	0.81	0.88
<u>Teen Depression</u>							

N=13 facilitators N=1660 adolescents		Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time		0.36	0.14	p=0.01**	p=0.02*	1.43	1.09	1.87
Role-play		-0.05	0.05	p=0.37	p=0.48	0.95	0.86	1.06
Time*Role-play		-0.02	0.01	p=0.05*	p=0.10	0.98	0.96	1.00
<u>Sexual Health Communication</u>								
N=13 facilitators N=1660 adolescents		Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time		-0.07	0.14	p=0.60	p=0.67	0.93	0.71	1.21
Role-play		-0.02	0.04	p=0.70	p=0.72	0.98	0.90	1.07
Time*Role-play		0.05	0.01	p<0.00***	p<0.00***	1.06	1.03	1.08
<u>Acceptability to Corporal Punishment ◇</u>								
N=13 facilitators N=1660 adolescents		Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time		-2.40	0.83	p<0.001***	p=0.01**	0.09	0.02	0.46
Role-play		-0.13	0.28	p=0.64	p=0.67	0.88	0.50	1.53
Time*Role-play		0.16	0.08	p=0.04*	p=0.09	1.17	1.00	1.36
<u>Positive Involvement ^</u>								
N=12 facilitators N=887 adolescents		Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time		-1.35	0.26	p<0.001***	p<0.001***	0.26	0.16	0.43
Role-play		-0.13	0.06	p=0.03*	p=0.06	0.88	0.78	0.99
Time*Role-play		0.07	0.02	p<0.001***	p<0.001***	1.07	1.02	1.12
<u>Child Emotional Problems ^</u>								

<i>N</i> =12 facilitators <i>N</i> =887 adolescents	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time	6.43	0.80	p<0.001***	p<0.001***	619.27	128.85	2976.35
Role-play	-0.06	0.24	p=0.81	p=0.85	0.94	0.59	1.50
Time*Role-play	-0.51	0.07	p<0.001***	p<0.001***	0.60	0.52	0.69
<u>Parental Support of Education ^</u>							
<i>N</i> =12 facilitators <i>N</i> =887 adolescents	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time	-6.72	0.78	p<0.001***	p<0.001***	0.00	0.00	0.01
Role-play	-0.38	0.17	p=0.02*	p=0.05*	0.69	0.50	0.95
Time*Role-play	0.55	0.07	p<0.001***	p<0.001***	1.73	1.51	1.99
<u>School Violence ^</u>							
<i>N</i> =12 facilitators <i>N</i> =887 adolescents	Estimate	Std Error	Raw P Value	Adjusted P Value	IRR	LL	UL
Time	-1.96	0.73	p=0.01**	p=0.01**	0.14	0.03	0.59
Role-play	-0.19	0.28	p=0.51	p=0.60	0.83	0.48	1.44
Time*Role-play	0.30	0.08	p<0.001***	p<0.001***	1.34	1.15	1.57
Note: Signif. codes: ***p<0.001, **p<0.01, *p<0.05. The outcomes with the “^” symbol were run without wave as a fixed effect due to rank deficiency. The outcome with the “◇” was run as a logistic regression as it is binary.							

### Supplementary File 3 – Summary of Pre-Post Surveys Collected by Implementing Organizations in Tanzania

<b>Table 12</b> <i>Summary of Pre-Post Surveys</i>						
<u>Outcome</u>	<u>Type</u>	<u>Measure</u>	<u>Items</u>	<u>Unit Measure</u>	<u>Report</u>	<u>Scale Reliability</u> (Lachman et al., forthcoming)
Child Maltreatment	Primary	ISPCAN Child Abuse Screening Tools-Trial Version (Meinck et al., 2018)	Frequency of disciplining child by spanking, slapping or hitting with your hand  Frequency of disciplining child with an object like a stick or a belt  Frequency of saying things to child that upset them  Frequency of shouting, yelling, or screaming at the child	0 to 7 and 8 for more than or equal to 8 in the past 4 weeks	Parent and adolescent	Parent-report: $\alpha=0.65, \omega=0.65$  Adolescent-report: $\alpha=0.64 \omega=0.64$
Child Conduct Problems	Secondary	Strengths and Difficulties Questionnaire (Goodman, 1997)	Child often fights with other children or bullies them  Child often has temper tantrums or hot tempers	0 = Not true, 1 = Somewhat true, 2 = Very true	Parent and adolescent	Parent-report: $\alpha=0.70 \omega=0.75$

			Child often lies or cheats			Adolescent-report: $\alpha=0.68$ $\omega=0.74$
			Child steals from home, school or elsewhere			
			Child is generally obedient and does what adults request			
Child Emotional Problems	Secondary	Strengths and Difficulties Questionnaire (Goodman, 1997)	Child gets a lot of headaches, stomach aches or sickness	0 = Not true, 1 = Somewhat true, 2 = Very true	Adolescent	$\alpha=0.90$ $\omega=0.90$
			Child worries a lot			
			Child feels nervous in new situations and often loses their confidence			
			Child has many fears and are easily scared			
Positive Parental Involvement	Secondary	Alabama Parenting Questionnaire (Frick, 1991)	Frequency of having a friendly talk with child	0 = Never, 1 = Almost never, 2 =	Parent and adolescent	Parent-report: $\alpha=0.95$ $\omega=0.95$
			Frequency of caregiver getting involved in activities child likes	Sometimes, 3 = Often, 4 = Always		Adolescent-report: $\alpha=0.94$ $\omega=0.94$

			Frequency of talking to the child about their friends			
Poor Supervision	Secondary	Alabama Parenting Questionnaire (Frick, 1991)	<p>Child stays out in the evening past time he/she is supposed to be home</p> <p>Child goes out without a set time to be home</p> <p>Child goes out after dark without an adult Accompanying</p>	<p>0 = Never, 1 = Almost never, 2 = Sometimes, 3 = Often, 4 = Always</p>	Parent and adolescent	<p>Parent-report: <math>\alpha=0.81</math> <math>\omega=0.81</math></p> <p>Adolescent-report: <math>\alpha=0.77</math> <math>\omega=0.77</math></p>
Parenting Stress	Secondary	Parental Stress Scale (Berry & Jones, 1995)	<p>Caring for child takes more time and energy than you have to give</p> <p>Child is a major source of stress in your life</p>	<p>0 = Strongly disagree, 1 = Disagree, 2 = Neutral, 3 = Agree, 4 = Strongly agree</p>	Parent	$r=0.70$
Acceptability of Corporal Punishment	Secondary	Multiple Indicator Cluster Survey (UNICEF, 2022)	In order to bring up, raise, or educate a child properly, a child needs to be physically punished	<p>0 = Strongly disagree, Disagree, and Not sure; 1 = Agree and Strongly agree</p>	Parent and adolescent	N/A

Depression	Secondary	Centre for Epidemiologic Studies Depression Scale (CES-D 10) (Irwin et al., 1999)	<p>Frequency of feeling depressed</p> <p>Frequency of feeling that everything was an effort</p> <p>Frequency of feeling hopeful</p>	<p>0 = Rarely or none of the time, 1 = Some or a little of the time, 2 = Occasionally or a moderate amount of time, 3 = Most of the time</p>	Parent and adolescent	<p>Parent-report: <math>\alpha=0.57</math> <math>\omega=0.67</math></p> <p>Adolescent-report: <math>\alpha=0.69</math> <math>\omega=0.71</math></p>
Financial insecurity	Secondary	Family Financial Coping Scale (Shenderovich et al., 2020)	<p>How many times the family ran out of money for something important like food or fuel for cooking</p> <p>How many times the caregiver worried about money</p>	0 = Never, 1 = Rarely, 2 = Sometimes, 3 = Often	Parent and adolescent	$r=0.79$
Parental Support of Education	Secondary	Parental Support for School Scale (Ceballo et al., 2014)	<p>Praises child for working hard at school</p> <p>Supports child's schoolwork in any way they can</p>	1 = Never, 2 = Hardly ever, 3 = Sometimes, 4 = Most of the time, 5 = Almost every day	Parent and adolescent	<p>Parent-report: <math>r=0.90</math></p> <p>Adolescent-report: <math>r=0.90</math></p>

Intimate Partner Violence Perpetration and Victimization	Secondary	Revised Conflict Tactics Scale Short Form (Straus et al., 1996)	<p>How often their partner or any other adult insult or shout or yell in the past month/30 days</p> <p>How often they insulted or shouted or yelled or swore at their partner or another adult</p> <p>How often their partner or any other adult hit, pushed, shoved, or slapped them</p> <p>How often they hit, pushed, shoved, or slapped their partner or any other adult</p>	0 to 7 and 8 for more than or equal to 8 in the past 4 weeks	Parent	<p>IPV Victimization: <math>r=0.57</math></p> <p>IPV Perpetration: <math>r=0.58</math></p>
Sexual Health Communication	Secondary	Risk Avoidance Planning Scale (Cluver et al., 2018)	<p>Caregiver talked about puberty and issues growing up with child</p> <p>Caregiver talked about safe sex and contraception options</p>	0 = No, I find it too hard to talk about this, 1 = We have not made plans yet but I would like to	Parent and adolescent	<p>Parent-report: <math>\alpha=0.91</math> <math>\omega=0.91</math></p> <p>Adolescent-report: <math>\alpha=0.90</math> <math>\omega=0.90</math></p>



			Caregiver talked about having sexual relationship with an older man or women	talk about it, 2 = We have discussed this together		
School Violence	Secondary	Created based on ISPCAN Child Abuse Screening Tools-Trial Version	How often they were bullied by their school peers, such as by being called names, being excluded, being threatened, or being physically attacked	0 to 7 and 8 for more than or equal to 8 in the past 4 weeks	Adolescent	$\alpha=0.79$ $\omega=0.79$
			How often a teacher or any other adult at their school hit them with a hand or object like a stick or a belt in the last four weeks			
			How often a teacher or other adult in their school disciplined them by shouting, yelling, or screaming			

## References

- Achen, C. H. (2005). Let's put garbage-can regressions and garbage-can probits where they belong. *Conflict Management and Peace Science*, 22(4), 327-339.
- Allen, J. D., Linnan, L. A., & Emmons, K. M. (2012). Fidelity and its relationship to implementation effectiveness, adaptation, and dissemination. In R. C. Brownson, G. A. Colditz, & E. K. Proctor (Eds.), *Dissemination and Implementation Research in Health: Translating Science to Practice*. Oxford University Press.  
<https://doi.org/10.1093/acprof:oso/9780199751877.003.0014>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3.
- Araujo, M. C., Rubio-Codina, M., & Schady, N. (2021). 70 to 700 to 70,000: Lessons from the Jamaica Experiment. In *The Scale-Up Effect in Early Childhood and Public Policy* (pp. 211-232). Routledge.
- Aspland, H., & Gardner, F. (2003). Observational measures of parent-child interaction: An introductory review. *Child and Adolescent Mental Health*, 8(3), 136-143.
- Astbury, B., & Leeuw, F. L. (2010). Unpacking black boxes: Mechanisms and theory building in evaluation. *American Journal of Evaluation*, 31(3), 363-381.
- Ayala, G. X., Vaz, L., Earp, J. A., Elder, J. P., & Cherrington, A. (2010). Outcome effectiveness of the lay health advisor model among Latinos in the United States: An examination by role. *Health Education Research*, 25(5), 815-840.
- Bandura, A., & Walters, R. H. (1977). *Social learning theory*. New York: General Learning Press.
- Baraldi, A. N., & Enders, C. K. (2010). An introduction to modern missing data analyses. *Journal of School Psychology*, 48(1), 5-37.  
<https://doi.org/https://doi.org/10.1016/j.jsp.2009.10.001>
- Barchard, K. (2010). Internal consistency reliability. In N. Salkind (Ed.), *Encyclopedia of Research Design*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412961288>
- Barlow, J., & Coren, E. (2018). The effectiveness of parenting programs: A review of Campbell reviews. *Research on Social Work Practice*, 28(1), 99-102.
- Barlow, J., Johnston, I., Kendrick, D., Polnay, L., & Stewart-Brown, S. (2006a). Individual and group-based parenting programmes for the treatment of physical child abuse and neglect. *Cochrane Database of Systematic Reviews*(3).

- Barlow, J., Johnston, I., Kendrick, D., Polnay, L., & Stewart-Brown, S. (2006b). Individual and group-based parenting programmes for the treatment of physical child abuse and neglect. *Cochrane Database of Systematic Reviews* (3).
- Barlow, J., Midgley, N., & Sled, M. (2017). Systematic review of the effectiveness of interventions targeting parents and infants/toddlers in improving parental reflective functioning. *National Institute for Health Research PROSPERO*.  
[https://www.crd.york.ac.uk/prospéro/display\\_record.php?ID=CRD42017062418](https://www.crd.york.ac.uk/prospéro/display_record.php?ID=CRD42017062418)
- Barnett, M. L., Lau, A. S., & Miranda, J. (2018). Lay health worker involvement in evidence-based treatment delivery: A conceptual model to address disparities in care. *Annual Review of Clinical Psychology*, 14, 185-208.
- Bartlett, M. S. (1954). A note on the multiplying factors for various  $\chi^2$  approximations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 296-298.
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1), 1-48.  
<https://doi.org/10.18637/jss.v067.i01>
- Belur, J., Tompson, L., Thornton, A., & Simon, M. (2018). Interrater reliability in systematic review methodology: Exploring variation in coder decision-making. *Sociological Methods & Research*, 0(0), 0049124118799372.  
<https://doi.org/10.1177/0049124118799372>
- Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 57(1), 289-300.
- Berkel, C., Gallo, C. G., Sandler, I. N., Mauricio, A. M., Smith, J. D., & Brown, C. H. (2019). Redesigning implementation measurement for monitoring and quality improvement in community delivery settings. *The Journal of Primary Prevention*, 40(1), 111-127.
- Berkel, C., Mauricio, A. M., Schoenfelder, E., & Sandler, I. N. (2011). Putting the pieces together: An integrated model of program implementation. *Prevention Science*, 12(1), 23-33.
- Berry, J. O., & Jones, W. H. (1995). The Parental Stress Scale: Initial psychometric evidence. *Journal of Social and Personal Relationships*, 12(3), 463-472.
- Bhattacharyya, O., Reeves, S., & Zwarenstein, M. (2009). What is implementation research? Rationale, concepts, and practices. *Research on Social Work Practice*, 19(5), 491-502.

- Bond, G. R., Evans, L., Salyers, M. P., Williams, J., & Kim, H.-W. (2000). Measurement of fidelity in psychiatric rehabilitation. *Mental Health Services Research*, 2(2), 75-87.
- Bornstein, R. (2004). *The SAGE Encyclopedia of Social Science Research Methods*.  
<https://doi.org/10.4135/9781412950589>
- Botvin, G. J. (2004). Advancing prevention science and practice: Challenges, critical issues, and future directions. *Prevention Science*, 5(1), 69-72.
- Bowlby, J. (1974). *Attachment and loss* (Vol. 1). London: Hogarth Press and Institute of Psycho-Analysis.
- Breitenstein, S. M., Fogg, L., Garvey, C., Hill, C., Resnick, B., & Gross, D. (2010a). Measuring implementation fidelity in a community-based parenting intervention. *Nursing Research*, 59(3), 158.
- Breitenstein, S. M., Gross, D., Garvey, C. A., Hill, C., Fogg, L., & Resnick, B. (2010b). Implementation fidelity in community-based interventions. *Research in Nursing & Health*, 33(2), 164-173.
- Bruton, A., Conway, J. H., & Holgate, S. T. (2000). Reliability: What is it, and how is it measured? *Physiotherapy*, 86(2), 94-99.  
[https://doi.org/https://doi.org/10.1016/S0031-9406\(05\)61211-4](https://doi.org/https://doi.org/10.1016/S0031-9406(05)61211-4)
- Buchanan-Pascall, S., Gray, K. M., Gordon, M., & Melvin, G. A. (2018). Systematic review and meta-analysis of parent group interventions for primary school children aged 4–12 years with externalizing and/or internalizing problems. *Child Psychiatry & Human Development*, 49(2), 244-267.
- Bumbarger, B., & Perkins, D. F. (2008). After randomised trials: Issues related to dissemination of evidence-based interventions. *Journal of Children's Services*, 3(2), 55-64.
- Burkey, M. D., Hosein, M., Morton, I., Purgato, M., Adi, A., Kurzrok, M., Kohrt, B. A., & Tol, W. A. (2018). Psychosocial interventions for disruptive behaviour problems in children in low-and middle-income countries: A systematic review and meta-analysis. *Journal of Child Psychology and Psychiatry*, 59(9), 982-993.
- Bywater, T., Gridley, N., Berry, V., Blower, S., & Tobin, K. (2019). The parent programme implementation checklist (PPIC): The development and testing of an objective measure of skills and fidelity for the delivery of parent programmes. *Child Care in Practice*, 25(3), 281-309.

- Calderon, F., Martin, M., Han, Q., Wamoyi, J., Ndyetabura, E., Mgunga, S., Manjengenja, N., & Lachman, J. M. (forthcoming). Costs and cost-effectiveness of a parenting programme to prevent violence against children at scale in Tanzania.
- Campbell, M., McKenzie, J. E., Sowden, A., Katikireddi, S. V., Brennan, S. E., Ellis, S., Hartmann-Boyce, J., Ryan, R., Shepperd, S., & Thomas, J. (2020). Synthesis without meta-analysis (SWiM) in systematic reviews: Reporting guideline. *BMJ*, 368.
- Campbell, S. B., Shaw, D. S., & Gilliom, M. (2000). Early externalizing behavior problems: Toddlers and preschoolers at risk for later maladjustment. *Development and Psychopathology*, 12(3), 467-488.
- Cantu, A. M., Hill, L. G., & Becker, L. G. (2010). Implementation quality of a family-focused preventive intervention in a community-based dissemination. *Journal of Children's Services*, 5(4), 18.
- Carrasco, J. (2022). iccCounts: Intraclass correlation coefficient for count data. <https://CRAN.R-project.org/package=iccCounts>
- Carroll, C., Patterson, M., Wood, S., Booth, A., Rick, J., & Balain, S. (2007). A conceptual framework for implementation fidelity. *Implementation Science*, 2(1), 40.
- Carroll, J. E., Gruenewald, T. L., Taylor, S. E., Janicki-Deverts, D., Matthews, K. A., & Seeman, T. E. (2013). Childhood abuse, parental warmth, and adult multisystem biological risk in the Coronary Artery Risk Development in Young Adults study. *Proceedings of the National Academy of Sciences*, 110(42), 17149-17153.
- Cash, R. A. (2011). *From One to Many: Scaling Up Health Programs in Low Income Countries*. University Press. <https://books.google.ca/books?id=GgfHZwEACAAJ>
- Casillas, K. L., Fauchier, A., Derkash, B. T., & Garrido, E. F. (2016, 2016/03/01/). Implementation of evidence-based home visiting programs aimed at reducing child maltreatment: A meta-analytic review. *Child Abuse & Neglect*, 53, 64-80. <https://doi.org/https://doi.org/10.1016/j.chiabu.2015.10.009>
- Ceballo, R., Maurizi, L. K., Suarez, G. A., & Aretakis, M. T. (2014). Gift and sacrifice: Parental involvement in Latino adolescents' education. *Cultural Diversity and Ethnic Minority Psychology*, 20(1), 116.
- Chapman, J. E., & Schoenwald, S. K. (2011). Ethnic similarity, therapist adherence, and long-term multisystemic therapy outcomes. *Journal of Emotional and Behavioral Disorders*, 19(1), 3-16.

- Chen, M., & Chan, K. (2016). Effects of parenting programs on child maltreatment prevention: A meta-analysis. *Trauma, Violence, & Abuse, 17*(1), 88-104. <https://doi.org/https://doi.org/10.1177/1524838014566>
- Chen, P., & Krauss, A. (2004). *The SAGE Encyclopedia of Social Science Research Methods*. <https://doi.org/10.4135/9781412950589>
- Chen, S. Y., Feng, Z., & Yi, X. (2017, Jun). A general introduction to adjustment for multiple comparisons. *Journal of Thoracic Disease, 9*(6), 1725-1729. <https://doi.org/10.21037/jtd.2017.05.34>
- Chiapa, A., Smith, J. D., Kim, H., Dishion, T. J., Shaw, D. S., & Wilson, M. N. (2015). The trajectory of fidelity in a multiyear trial of the Family Check-up predicts change in child problem behavior. *Journal of Consulting and Clinical Psychology, 83*(5), 1006.
- Cho, D. W. (1981). Inter-rater reliability: Intraclass correlation coefficients. *Educational and Psychological Measurement, 41*(1), 223-226.
- Cluver, L., Meinck, F., Shenderovich, Y., Ward, C. L., Romero, R. H., Redfern, A., Lombard, C., Doubt, J., Steinert, J., Catanho, R., Wittesaele, C., De Stone, S., Salah, N., Mpimpilashe, P., Lachman, J., Loening, H., Gardner, F., Blanc, D., Nocuza, M., & Lechowicz, M. (2016). A parenting programme to prevent abuse of adolescents in South Africa: Study protocol for a randomised controlled trial. *Trials, 17*(1), 328. <https://doi.org/10.1186/s13063-016-1452-8>
- Cluver, L. D., Meinck, F., Steinert, J. I., Shenderovich, Y., Doubt, J., Romero, R. H., Lombard, C. J., Redfern, A., Ward, C. L., & Tsoanyane, S. (2018). Parenting for Lifelong Health: A pragmatic cluster randomised controlled trial of a non-commercialised parenting programme for adolescents and their families in South Africa. *BMJ Global Health, 3*(1), e000539. <https://doi.org/http://dx.doi.org/10.1136/bmjgh-2017-000539>
- Collyer, H., Eisler, I., & Woolgar, M. (2019). Systematic literature review and meta-analysis of the relationship between adherence, competence and outcome in psychotherapy for children and adolescents. *European Child & Adolescent Psychiatry, 1*-15.
- Colman, I., Murray, J., Abbott, R. A., Maughan, B., Kuh, D., Croudace, T. J., & Jones, P. B. (2009). Outcomes of conduct problems in adolescence: 40 year follow-up of national cohort. *BMJ, 338*.
- Conduct Problems Prevention Research Group. (2002). The implementation of the Fast Track Program: An example of a large-scale prevention science efficacy trial. *Journal of Abnormal Child Psychology, 30*, 1-17.

- Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, 10(1), 7.
- Cowell, R. A., Cicchetti, D., Rogosch, F. A., & Toth, S. L. (2015). Childhood maltreatment and its effect on neurocognitive functioning: Timing and chronicity matter. *Development and Psychopathology*, 27(2), 521-533.  
<https://doi.org/10.1017/S0954579415000139>
- Cuartas, J. (2022). Corporal punishment and child development in low- and middle-income countries: Progress, challenges, and directions. *Child Psychiatry & Human Development*. <https://doi.org/10.1007/s10578-022-01362-3>
- Cuartas, J., McCoy, D. C., Rey-Guerra, C., Britto, P. R., Beatriz, E., & Salhi, C. (2019). Early childhood exposure to non-violent discipline and physical and psychological aggression in low-and middle-income countries: National, regional, and global prevalence estimates. *Child Abuse & Neglect*, 92, 93-105.
- CWBSA. (2017). Pilot phase implementation report - transforming the lives of vulnerable children and young people: Tanzania positive parenting component.
- Dane, A. V., & Schneider, B. H. (1998). Program integrity in primary and early secondary prevention: Are implementation effects out of control? *Clinical Psychology Review*, 18(1), 23-45. [https://doi.org/https://doi.org/10.1016/S0272-7358\(97\)00043-3](https://doi.org/https://doi.org/10.1016/S0272-7358(97)00043-3)
- Dobson, D., & Cook, T. J. (1980). Avoiding type III error in program evaluation: Results from a field experiment. *Evaluation and Program Planning*, 3(4), 269-276.  
[https://doi.org/https://doi.org/10.1016/0149-7189\(80\)90042-7](https://doi.org/https://doi.org/10.1016/0149-7189(80)90042-7)
- Durlak, J. A. (1998). Why program implementation is important. *Journal of Prevention & Intervention in the Community*, 17(2), 5-18. [https://doi.org/10.1300/J005v17n02\\_02](https://doi.org/10.1300/J005v17n02_02)
- Durlak, J. A., & DuPre, E. P. (2008). Implementation matters: A review of research on the influence of implementation on program outcomes and the factors affecting implementation. *American Journal of Community Psychology*, 41(3-4), 327-350.
- Dusenbury, L., Brannigan, R., Falco, M., & Hansen, W. B. (2003). A review of research on fidelity of implementation: Implications for drug abuse prevention in school settings. *Health Education Research*, 18(2), 237-256.
- Eames, C., Daley, D., Hutchings, J., Hughes, J., Jones, K., Martin, P., & Bywater, T. (2008). The Leader Observation Tool: A process skills treatment fidelity measure for the Incredible Years parenting programme. *Child: Care, Health and Development*, 34(3), 391-400.

- Eames, C., Daley, D., Hutchings, J., Whitaker, C., Jones, K., Hughes, J., & Bywater, T. (2009). Treatment fidelity as a predictor of behaviour change in parents attending group-based parent training. *Child: Care, Health and Development*, 35(5), 603-612.
- Eames, C., Daley, D., Hutchings, J., Whitaker, C. J., Bywater, T., Jones, K., & Hughes, J. C. (2010). The impact of group leaders' behaviour on parents acquisition of key parenting skills during parent training. *Behaviour Research and Therapy*, 48(12), 1221-1226.
- Eisenberg, N., Cumberland, A., Spinrad, T. L., Fabes, R. A., Shepard, S. A., Reiser, M., Murphy, B. C., Losoya, S. H., & Guthrie, I. K. (2001). The relations of regulation and emotionality to children's externalizing and internalizing problem behavior. *Child Development*, 72(4), 1112-1134.
- Eisner, M. (2014). The South Carolina Triple P System Population trial to prevent child maltreatment: Seven reasons to be sceptical about the study results. *Cambridge: Violence Research Centre, University of Cambridge*.
- Erskine, H. E., Norman, R. E., Ferrari, A. J., Chan, G. C., Copeland, W. E., Whiteford, H. A., & Scott, J. G. (2016). Long-term outcomes of attention-deficit/hyperactivity disorder and conduct disorder: A systematic review and meta-analysis. *Journal of the American Academy of Child & Adolescent Psychiatry*, 55(10), 841-850.
- Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory factor analysis*. Oxford University Press.
- Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, 4(3), 272.
- Fahrmeir, L., Kneib, T., Lang, S., & Marx, B. (2013). Regression: Models, methods, and applications. In *Regression*. Springer.
- Fan, F., & Randall, J. (2018). Reliability. In B. Frey (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications, Inc. <https://doi.org/10.4135/9781506326139>
- Feeley, M., DeRubeis, R. J., & Gelfand, L. A. (1999). The temporal relation of adherence and alliance to symptom change in cognitive therapy for depression. *Journal of Consulting and Clinical Psychology*, 67(4), 578.
- Feely, M., Seay, K. D., Lanier, P., Auslander, W., & Kohl, P. L. (2018). Measuring fidelity in research studies: A field guide to developing a comprehensive fidelity measurement system. *Child and Adolescent Social Work Journal*, 35(2), 139-152.



- Fergusson, D. M., John Horwood, L., & Ridder, E. M. (2005). Show me the child at seven: The consequences of conduct problems in childhood for psychosocial functioning in adulthood. *Journal of Child Psychology and Psychiatry*, 46(8), 837-849.
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. SAGE.
- Fixsen, D. L., Blase, K. A., Naoom, S. F., & Wallace, F. (2009). Core implementation components. *Research on Social Work Practice*, 19(5), 531-540.
- Fixsen, D. L., Naoom, S. F., Blase, K. A., Friedman, R. M., Wallace, F., Burns, B., Carter, W., Paulson, R., Schoenwald, S., & Barwick, M. (2005). Implementation research: A synthesis of the literature. <https://nirn.fpg.unc.edu/resources/implementation-research-synthesis-literature>
- Font, S. A., & Berger, L. M. (2015). Child maltreatment and children's developmental trajectories in early to middle childhood. *Child Development*, 86(2), 536-556. <https://doi.org/10.1111/cdev.12322>
- Forgatch, M. S., & DeGarmo, D. S. (2011). Sustaining fidelity following the nationwide PMTO™ implementation in Norway. *Prevention Science*, 12(3), 235-246.
- Forgatch, M. S., & Kjøbli, J. (2016). Parent Management Training—Oregon model: Adapting intervention with rigorous research. *Family Process*, 55(3), 500-513.
- Forgatch, M. S., Patterson, G. R., & DeGarmo, D. S. (2005). Evaluating fidelity: Predictive validity for a measure of competent adherence to the Oregon model of parent management training. *Behavior Therapy*, 36(1), 3-13.
- Foster, E. M., Jones, D. E., & Conduct Problems Prevention Research Group. (2005). The high costs of aggression: Public expenditures resulting from conduct disorder. *American Journal of Public Health*, 95(10), 1767-1772.
- Fraser, M. W. (1997). *Risk and resilience in childhood: An ecological perspective*. NASW Press Washington, DC.
- Frick, P. J. (1991). The Alabama Parenting Questionnaire. *Unpublished rating scale, University of Alabama*.
- Furlong, M., & McGilloway, S. (2015). Barriers and facilitators to implementing evidence-based parenting programs in disadvantaged settings: A qualitative study. *Journal of Child and Family Studies*, 24(6), 1809-1818.
- Furlong, M., McGilloway, S., Bywater, T., Hutchings, J., Smith, S. M., & Donnelly, M. (2012). Behavioural and cognitive-behavioural group-based parenting programmes

- for early-onset conduct problems in children aged 3 to 12 years. *Campbell Systematic Reviews*, 8(1), 1-239.
- Furlong, M., McGilloway, S., Bywater, T., Hutchings, J., Smith, S. M., & Donnelly, M. (2013). Cochrane review: Behavioural and cognitive-behavioural group-based parenting programmes for early-onset conduct problems in children aged 3 to 12 years. *Evidence-Based Child Health: A Cochrane Review Journal*, 8(2), 318-692.
- Furr, R. M. (2017). *Psychometrics: An Introduction*. SAGE publications.
- Gallo, C., Pantin, H., Villamar, J., Prado, G., Tapia, M., Ogihara, M., Cruden, G., & Brown, C. H. (2015). Blending qualitative and computational linguistics methods for fidelity assessment: Experience with the Familias Unidas preventive intervention. *Administration and Policy in Mental Health and Mental Health Services Research*, 42(5), 574-585.
- Gamer, M., Lemon, J., & Kendall, W. (2017). Package ‘irr’.
- Garbarino, J. (1977). The human ecology of child maltreatment: A conceptual model for research. *Journal of Marriage and the Family*, 721-735.
- Gardner, F. (2000). Methodological issues in the direct observation of parent–child interaction: Do observational findings reflect the natural behavior of participants? *Clinical Child and Family Psychology Review*, 3(3), 185-198.
- Gardner, F., Lachman, J. M., Knerr, W., Melendez-Torres, G. J., Riddell, J., McCoy, A., Fang, Z., & Wight, D. (forthcoming-a). Parenting for preventing risk of child maltreatment in low- and middle-income countries: Updated systematic review and meta-analysis.  
[https://www.crd.york.ac.uk/prospero/display\\_record.php?ID=CRD42018088697](https://www.crd.york.ac.uk/prospero/display_record.php?ID=CRD42018088697)
- Gardner, F., McCoy, A., Lachman, J. M., Melendez-Torres, G. J., Ward, C., Cheah, P., & Topanya, S. (forthcoming-b). Randomized trial of a parenting intervention embedded within the public health system to reduce violence against children in Thailand.
- Gardner, F., Montgomery, P., & Knerr, W. (2016). Transporting evidence-based parenting programs for child problem behavior (age 3–10) between countries: Systematic review and meta-analysis. *Journal of Clinical Child & Adolescent Psychology*, 45(6), 749-762.
- Gardner, F., Shenderovich, Y., McCoy, A., Schafer, M., Martin, M., Janowski, R., Varadan, S., Backhaus, S., Barlow, J., & Ward, C. (2023). World Health Organization guidelines on parenting to prevent child maltreatment and promote positive development in children aged 0-17 Years—Report of the reviews for the

INTEGRATE framework. <https://orca.cardiff.ac.uk/id/eprint/157322/1/who-integrate-reviews-for-who-parenting-guideline-jan-27th-2023.pdf>

- Gershoff, E. T. (2010). More harm than good: A summary of scientific research on the intended and unintended effects of corporal punishment on children. *Law & Contemporary Problems*, 73, 31.
- Gershoff, E. T., Goodman, G. S., Miller-Perrin, C. L., Holden, G. W., Jackson, Y., & Kazdin, A. E. (2018). The strength of the causal evidence against physical punishment of children and its implications for parents, psychologists, and policymakers. *American Psychologist*, 73(5), 626.
- Gershoff, E. T., & Grogan-Kaylor, A. (2016). Spanking and child outcomes: Old controversies and new meta-analyses. *Journal of Family Psychology*, 30(4), 453.
- Giannotta, F., Özdemir, M., & Stattin, H. (2019). The implementation integrity of parenting programs: Which aspects are most important? *Child & Youth Care Forum*, 48(6), 917-933.
- Ginsburg, L. R., Hoben, M., Easterbrook, A., Anderson, R. A., Estabrooks, C. A., & Norton, P. G. (2021). Fidelity is not easy! Challenges and guidelines for assessing fidelity in complex interventions. *Trials*, 22(1), 1-13.
- Girard, J. M., & Cohn, J. F. (2016). A primer on observational measurement. *Assessment*, 23(4), 404-413. <https://doi.org/10.1177/1073191116635807>
- Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. (2003). Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American Journal of Public Health*, 93(8), 1261-1267.
- Glasgow, R. E., & Riley, W. T. (2013). Pragmatic measures: What they are and why we need them. *American Journal of Preventive Medicine*, 45(2), 237-243. <https://doi.org/https://doi.org/10.1016/j.amepre.2013.03.010>
- Glasgow, R. E., Vogt, T. M., & Boles, S. M. (1999). Evaluating the public health impact of health promotion interventions: The RE-AIM framework. *American Journal of Public Health*, 89(9), 1322-1327. <https://doi.org/10.2105/ajph.89.9.1322>
- Goense, P., Boendermaker, L., van Yperen, T., Stams, G.-J., & van Laar, J. (2015). Implementation of treatment integrity procedures. *Zeitschrift für Psychologie*.
- Goodman, R. (1997). The Strengths and Difficulties Questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38(5), 581-586.

- Gottfredson, D. C., Cook, T. D., Gardner, F. E., Gorman-Smith, D., Howe, G. W., Sandler, I. N., & Zafft, K. M. (2015). Standards of evidence for efficacy, effectiveness, and scale-up research in prevention science: Next generation. *Prevention Science*, 16(7), 893-926. <https://doi.org/10.1007/s11121-015-0555-x>
- Gray, G. R., Totsika, V., & Lindsay, G. (2018). Sustained effectiveness of evidence-based parenting programs after the research trial ends. *Frontiers in Psychology*, 9, 2035-2035. <https://doi.org/10.3389/fpsyg.2018.02035>
- Gross, J., & Ligges, U. (2015). nortest: Tests for normality. R package version 1.0-4. <https://CRAN.R-project.org/package=nortest>
- Gushta, M., & Rupp, A. (2010). Reliability. In N. Salkind (Ed.), *Encyclopedia of Research Design*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412961288>
- Gwet, K. L. (2014). Intrarater reliability. *Wiley StatsRef: Statistics Reference Online*.
- Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). *Multivariate Data Analysis* (Vol. 5). Prentice Hall Upper Saddle River, NJ.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, 8(1), 23-34. <https://doi.org/10.20982/tqmp.08.1.p023>
- Hawe, P., Shiell, A., Riley, T., & Gold, L. (2004). Methods for exploring implementation variation and local context within a cluster randomised community intervention trial. *Journal of Epidemiology & Community Health*, 58(9), 788-793.
- Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, 7(3), 238.
- Heilmann, A., Mehay, A., Watt, R. G., Kelly, Y., Durrant, J. E., van Turnhout, J., & Gershoff, E. T. (2021). Physical punishment and child outcomes: A narrative review of prospective studies. *The Lancet*, 398(10297), 355-364.
- Heinl, D., Prinsen, C. A. C., Drucker, A. M., Ofenloch, R., Humphreys, R., Sach, T., Flohr, C., & Apfelbacher, C. (2016). Measurement properties of quality of life measurement instruments for infants, children and adolescents with eczema: Protocol for a systematic review. *Systematic Reviews*, 5, 25-25. <https://doi.org/10.1186/s13643-016-0202-z>
- Herschell, A. D., Quetsch, L. B., & Kolko, D. J. (2019). Measuring adherence to key teaching techniques in an evidence-based treatment: A comparison of caregiver,

- therapist, and behavior observation ratings. *Journal of Emotional and Behavioral Disorders*, 1063426618821901.
- Higgins, J. P., & Green, S. (2011). *Cochrane Handbook for Systematic Reviews of Interventions* (Vol. 4). John Wiley & Sons.
- Hill, H. C., & Erickson, A. (2019). Using implementation fidelity to aid in interpreting program impacts: A brief review. *Educational Researcher*, 48(9), 590-598.
- Hillis, S., Mercy, J., Amobi, A., & Kress, H. (2016). Global prevalence of past-year violence against children: A systematic review and minimum estimates. *Pediatrics*, 137(3), e20154079.
- Hinshaw, S. P., & Lee, S. S. (2003). In *Child Psychopathology* (2 ed.). The Guilford Press.
- Hogue, A., Henderson, C., Dauber, S., Barajas, P., Fried, A., & Liddle, H. (2008). Treatment adherence, competence, and outcome in individual and family therapy for adolescent behavior problems. *Journal of Consulting and Clinical Psychology*, 76(4), 544.
- Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, 30(2), 179-185. <https://doi.org/10.1007/BF02289447>
- Horvath, A. O., Del Re, A., Flückiger, C., & Symonds, D. (2011). Alliance in individual psychotherapy. *Psychotherapy*, 48(1), 9.
- Hukkelberg, S. S., & Ogden, T. (2013). Working alliance and treatment fidelity as predictors of externalizing problem behaviors in parent management training. *Journal of Consulting and Clinical Psychology*, 81(6), 1010.
- Humphreys, L. G., & Ilgen, D. R. (1969). Note on a criterion for the number of common factors. *Educational and Psychological Measurement*, 29(3), 571-578. <https://doi.org/10.1177/001316446902900303>
- Humphreys, L. G., & Montanelli Jr, R. G. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research*, 10(2), 193-205. [https://doi.org/10.1207/s15327906mbr1002\\_5](https://doi.org/10.1207/s15327906mbr1002_5)
- Institute of Medicine. (2009). Defining the scope of prevention: Preventing mental, emotional, and behavioral disorders among young people. <https://www.ncbi.nlm.nih.gov/books/NBK32775/>
- Institute of Medicine & National Research Council. (2014). Strategies for scaling effective family-focused preventive interventions to promote children's cognitive, affective, and behavioral health: Workshop summary.

<https://www.ncbi.nlm.nih.gov/books/NBK230080/#:~:text=real%2Dworld%20setting%20for%20Scaling%20Effective%20Family%2DFocused%20Preventive%20Interventions%20to%20Promote,Health%20to%20explore%20effective%20preventive>

- Irwin, M., Artin, K. H., & Oxman, M. N. (1999). Screening for depression in the older adult: Criterion validity of the 10-item Center for Epidemiological Studies Depression Scale (CES-D). *Archives of Internal Medicine*, 159(15), 1701-1704.
- Jeong, J., Franchett, E. E., Ramos de Oliveira, C. V., Rehmani, K., & Yousafzai, A. K. (2021). Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis. *PLoS Medicine*, 18(5), e1003602.
- Jocson, R. M., Alampay, L. P., Lachman, J. M., Maramba, D. H. A., Melgar, M. E., Ward, C. L., Madrid, B. J., & Gardner, F. (2023). Pre-post mixed methods study of a parent and teen support intervention to prevent violence against adolescents in the philippines. *Journal of Adolescent Health*, 17, 100279.  
<https://doi.org/https://doi.org/10.1016/j.jadohealth.2023.02.027>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31-36.  
<https://doi.org/10.1007/BF02291575>
- Kaminski, J. W., Valle, L. A., Filene, J. H., & Boyle, C. L. (2008). A meta-analytic review of components associated with parent training program effectiveness. *Journal of Abnormal Child Psychology*, 36(4), 567-589.
- Kazdin, A. E. (1997). Parent management training: Evidence, outcomes, and issues. *Journal of the American Academy of Child & Adolescent Psychiatry*, 36(10), 1349-1356.
- Kemp, L. (2016). Adaptation and fidelity: A recipe analogy for achieving both in population scale implementation. *Prevention Science*, 17(4), 429-438.
- Kendall, P. C., & Beidas, R. S. (2007). Smoothing the trail for dissemination of evidence-based practices for youth: Flexibility within fidelity. *Professional Psychology: Research and Practice*, 38(1), 13.
- Kjøbli, J., Hukkelberg, S., & Ogden, T. (2013). A randomized trial of group parent training: Reducing child conduct problems in real-world settings. *Behaviour Research and Therapy*, 51(3), 113-121.
- Kline, P. (2014). *An Easy Guide to Factor Analysis*. Routledge.

- Knerr, W., Gardner, F., & Cluver, L. (2013). Improving positive parenting skills and reducing harsh and abusive parenting in low-and middle-income countries: A systematic review. *Prevention Science*, 14(4), 352-363.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>
- Kumpfer, K., Magalhães, C., & Xie, J. (2017). Cultural adaptation and implementation of family evidence-based interventions with diverse populations. *Prevention Science*, 18(6), 649-659. <https://doi.org/10.1007/s11121-016-0719-3>
- Lachman, J. M., Alampay, L., Alinea, M., Madrid, B., Ward, C. L., & Gardner, F. (2021). Effectiveness of a parenting programme to reduce violence in a cash transfer system in the Philippines: RCT with follow-up. *The Lancet Regional Health - Western Pacific*.
- Lachman, J. M., Martin, M., Wamoyi, J., Shenderovich, Y., Han, Q., Calderon, F., Wambura, M., Mgunga, S., Nydetabura, E., Ally, A., Barankena, A., Exavery, A., & Manjengenga, N. (forthcoming). Reducing family and school-based violence at scale in Tanzania: Results from a large-scale pre-post study of the Furaha Teens parenting program. <https://theevaluationfund.org/wp-content/uploads/2020/10/FAIR-Final-Report.pdf>
- Lachman, J. M., Sherr, L. T., Cluver, L., Ward, C. L., Hutchings, J., & Gardner, F. (2016). Integrating evidence and context to develop a parenting program for low-income families in South Africa. *Journal of Child and Family Studies*, 25(7), 2337-2352.
- Lansford, J. E., & Deater-Deckard, K. (2012). Childrearing discipline and violence in developing countries. *Child Development*, 83(1), 62-75.
- Lau, A. S. (2006). Making the case for selective and directed cultural adaptations of evidence-based treatments: Examples from parent training. *Clinical Psychology: Science and Practice*, 13(4), 295-310.
- Lee, S., & Lee, D. K. (2018). What is the proper way to apply the multiple comparison test? *Korean Journal of Anesthesiology*, 71(5), 353-360. <https://doi.org/10.4097/kja.d.18.00242>
- Leijten, P., Gardner, F., Melendez-Torres, G., Van Aar, J., Hutchings, J., Schulz, S., Knerr, W., & Overbeek, G. (2019). Meta-analyses: Key parenting program components for disruptive child behavior. *Journal of the American Academy of Child & Adolescent Psychiatry*, 58(2), 180-190. <https://doi.org/https://doi.org/10.1016/j.jaac.2018.07.900>

- Leijten, P., Melendez-Torres, G., & Gardner, F. (2022). Research review: The most effective parenting program content for disruptive child behavior—a network meta-analysis. *Journal of Child Psychology and Psychiatry*, 63(2), 132-142.
- Leitão, S. M., Seabra-Santos, M. J., & Gaspar, M. F. (2020). Therapist factors matter: A systematic review of parent interventions directed at children's behavior problems. *Family Process*, 60(1), 84-101.
- Lester, S. N. (2015). *Evaluation of the Parent Centre's positive parenting skills training programme: A randomised controlled trial*. [Master of Philosophy, University of Cape Town].
- Liberati, A., Altman, D. G., Tetzlaff, J., Mulrow, C., Gøtzsche, P. C., Ioannidis, J. P. A., Clarke, M., Devereaux, P. J., Kleijnen, J., & Moher, D. (2009). The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate healthcare interventions: Explanation and elaboration. *BMJ*, 339, b2700. <https://doi.org/10.1136/bmj.b2700>
- Liu, J., Chen, X., & Lewis, G. (2011). Childhood internalizing behaviour: Analysis and implications. *Journal of Psychiatric and Mental Health Nursing*, 18(10), 884-894.
- Lombard, M., Snyder-Duch, J., & Bracken, C. (2017). In *The SAGE Encyclopedia of Communication Research Methods*. SAGE Publications. <https://doi.org/10.4135/9781483381411>
- Maaskant, A. M., van Rooij, F. B., Overbeek, G. J., Oort, F. J., & Hermanns, J. M. (2016). Parent training in foster families with children with behavior problems: Follow-up results from a randomized controlled trial. *Children and Youth Services Review*, 70, 84-94.
- Mair, P. (2018). *Modern Psychometrics with R*. Springer.
- Markus, K., & Smith, K. (2010). Content validity. In N. Salkind (Ed.), *Encyclopedia of Research Design*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412961288>
- Marryat, L., Thompson, L., & Wilson, P. (2017). No evidence of whole population mental health impact of the Triple P parenting programme: Findings from a routine dataset. *BMC pediatrics*, 17(1), 40. <https://doi.org/https://doi.org/10.1186/s12887-017-0800-5>
- Martin, M., Lachman, J., Wamoyi, J., Shenderovich, Y., Wambura, M., Mgunga, S., Ndyetabura, E., Ally, A., Barankena, A., Exavery, A., & Manjengenja, N. (2021a). A mixed methods evaluation of the large-scale implementation of a school- and community-based parenting program to reduce violence against children in



Tanzania: A study protocol. *Implementation Science Communications*, 2(1), 52.  
<https://doi.org/10.1186/s43058-021-00154-5>

- Martin, M., Lachman, J. M., Calderon, F., Han, Q., Shenderovich, Y., Ndyetabura, E., Mwakakeke, S., Chikola, L., Mgunga, S., Manjengenga, N., Exavery, A., Gardner, F., & Wamoyi, J. (in submission-a). Assessing facilitator delivery of a parenting program at scale in Tanzania: A psychometric evaluation of an observational measure of facilitator competent adherence.
- Martin, M., Lachman, J. M., Mgunga, S., Shenderovich, Y., Manjengenga, N., Nydetabura, E., Gardner, F. E., & Wamoyi, J. (forthcoming). Barriers and enablers of assessing parenting program facilitators: A qualitative analysis of the delivery of Parenting for Lifelong Health-Teens at scale in Tanzania.
- Martin, M., Lachman, J. M., Murphy, H., Gardner, F., & Foran, H. (2022a). The development, reliability, and validity of the Facilitator Assessment Tool: An implementation fidelity measure used in Parenting for Lifelong Health for Young Children. *Child: Care, Health and Development*.  
<https://doi.org/https://doi.org/10.1111/cch.13075>
- Martin, M., Lachman, J. M., Shenderovich, Y., Han, Q., Calderon, F., Wambura, M., Mgunga, S., Nydetabura, E., Ally, A., Barankena, A., Exavery, A., Manjengenga, N., & Wamoyi, J. (2022b). Evidence brief: Large-scale implementation of a parenting programme to reduce violence against children in Tanzania  
<https://theevaluationfund.org/wp-content/uploads/2020/10/Large-Scale-Implementation-of-a-Parenting-Programme-to-Reduce-Violence-Against-Children-in-Tanzania.pdf>
- Martin, M., Shenderovich, Y., Caron, E., Smith, J. D., Siu, G., & Breitenstein, S. M. (in submission-b). The case for assessing facilitator fidelity and creating fidelity reporting guidelines: Introducing the Fidelity of Implementation in Parenting Programs guideline.
- Martin, M., Steele, B., Lachman, J. M., & Gardner, F. (2021b). Measures of facilitator competent adherence used in parenting programs and their psychometric properties: A systematic review. *Clinical Child and Family Psychology Review*, 1-20.  
<https://doi.org/https://doi.org/10.1007/s10567-021-00350-8>
- Martin, M., Steele, B., Spreckelsen, T. F., Lachman, J. M., Gardner, F., & Shenderovich, Y. (2023). The association between facilitator competent adherence and outcomes in parenting programs: A systematic review and SWiM analysis. *Prevention Science*, 1-13. <https://doi.org/10.1007/s11121-023-01515-3>

- Mayo-Wilson, E., Dias, S., Mavranezouli, I., Kew, K., Clark, D. M., Ades, A., & Pilling, S. (2014). Psychological and pharmacological interventions for social anxiety disorder in adults: A systematic review and network meta-analysis. *The Lancet Psychiatry*, 1(5), 368-376.
- Mazzucchelli, T. G., & Sanders, M. R. (2010). Facilitating practitioner flexibility within an empirically supported intervention: Lessons from a system of parenting support. *Clinical Psychology: Science and Practice*, 17(3), 238-252.
- McCoy, A., Melendez-Torres, G. J., & Gardner, F. (2020). Parenting interventions to prevent violence against children in low- and middle-income countries in East and Southeast Asia: A systematic review and multi-level meta-analysis. *Child Abuse & Neglect*, 103, 104444. <https://doi.org/https://doi.org/10.1016/j.chiabu.2020.104444>
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, 1(1), 30.
- McHugh, M. L. (2012). Interrater reliability: The kappa statistic. *Biochem Med (Zagreb)*, 22(3), 276-282.
- McHugh, R. K., Murray, H. W., & Barlow, D. H. (2009). Balancing fidelity and adaptation in the dissemination of empirically-supported treatments: The promise of transdiagnostic interventions. *Behaviour Research and Therapy*, 47(11), 946-953.
- McKenzie, J. F., Wood, M. L., Kotecki, J. E., Clark, J. K., & Brey, R. A. (1999). Establishing content validity: Using qualitative and quantitative steps. *American Journal of Health Behavior*.
- McMahon, R. J., & Frick, P. J. (2019). Conduct and oppositional disorders. In M. J. Prinstein, E. A. Youngstrom, E. J. Mash, & R. A. Barkley (Eds.), *Treatment of Disorders in Childhood and Adolescence* (pp. 102-172). The Guilford Press.
- Meinck, F., Boyes, M. E., Cluver, L., Ward, C. L., Schmidt, P., DeStone, S., & Dunne, M. P. (2018). Adaptation and psychometric properties of the ISPCAN Child Abuse Screening Tool for use in trials (ICAST-Trial) among South African adolescents and their primary caregivers. *Child Abuse & Neglect*, 82, 45-58.
- Melendez-Torres, G. J., Leijten, P., & Gardner, F. (2019). What are the optimal combinations of parenting intervention components to reduce physical child abuse recurrence? Reanalysis of a systematic review using qualitative comparative analysis. *Child Abuse Review*, 28(3), 181-197.
- Mercy, J. A., Butchart, A., Rosenberg, M. L., Dahlberg, L., & Harvey, A. (2008). Preventing violence in developing countries: A framework for action. *International Journal of Injury Control and Safety Promotion*, 15(4), 197-208.

- Merrill, L. L., Hervig, L. K., & Milner, J. S. (1996). Childhood parenting experiences, intimate partner conflict resolution, and adult risk for child physical abuse. *Child Abuse & Neglect*, 20(11), 1049-1065.
- Mihalic, S. (2004). The importance of implementation fidelity. *Emotional and Behavioral Disorders in Youth*, 4(4), 83-105.
- Mikton, C., & Butchart, A. (2009). Child maltreatment prevention: A systematic review of reviews. *Bulletin of the World Health Organization*, 87, 353-361.
- Mikton, C., Power, M., Raleva, M., Makoe, M., Al Eissa, M., Cheah, I., Cardia, N., Choo, C., & Almuneef, M. (2013). The assessment of the readiness of five countries to implement child maltreatment prevention programs on a large scale. *Child Abuse & Neglect*, 37(12), 1237-1251.
- Miller, C. J., Barnett, M. L., Baumann, A. A., Gutner, C. A., & Wiltsey-Stirman, S. (2021). The FRAME-IS: A framework for documenting modifications to implementation strategies in healthcare. *Implementation Science*, 16(1), 36. <https://doi.org/10.1186/s13012-021-01105-3>
- Mills, R., Alati, R., Callaghan, M., Najman, J. M., Williams, G. M., Bor, W., & Strathearn, L. (2011). Child abuse and neglect and cognitive function at 14 years of age: Findings from a birth cohort. *Pediatrics*, 127(1), 4. <https://doi.org/10.1542/peds.2009-3479>
- Mokkink, L., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010a). The COSMIN study reached international consensus on taxonomy, terminology, and definitions of measurement properties for health-related patient-reported outcomes. *Journal of Clinical Epidemiology*, 63(7), 737-745.
- Mokkink, L., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. W. (2010b). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research* 19(4), 539-549. <https://doi.org/10.1007/s11136-010-9606-8>
- Mokkink, L. B., Prinsen, C. A. C., Bouter, L. M., Vet, H. C. W. d., & Terwee, C. B. (2016). The Consensus-based Standards for the selection of health Measurement INstruments (COSMIN) and how to select an outcome measurement instrument. *Brazilian Journal of Physical Therapy*, 20(2), 105-113. <https://doi.org/10.1590/bjpt-rbf.2014.0143>

- Mowbray, C. T., Holter, M. C., Teague, G. B., & Bybee, D. (2003). Fidelity criteria: Development, measurement, and validation. *American Journal of Evaluation*, 24(3), 315-340. [https://doi.org/https://doi.org/10.1016/S1098-2140\(03\)00057-2](https://doi.org/https://doi.org/10.1016/S1098-2140(03)00057-2)
- Moylan, C. A., Herrenkohl, T. I., Sousa, C., Tajima, E. A., Herrenkohl, R. C., & Russo, M. J. (2010). The effects of child abuse and exposure to domestic violence on adolescent internalizing and externalizing behavior problems. *Journal of Family Violence*, 25(1), 53-63. <https://doi.org/10.1007/s10896-009-9269-9>
- Multon, K., & Coleman, J. (2018). Inter-rater reliability. In B. Frey (Ed.), *The SAGE Encyclopedia of Educational Research, Measurement, and Evaluation*. SAGE Publications, Inc. <https://doi.org/10.4135/9781506326139>
- Muse, K., & McManus, F. (2013). A systematic review of methods for assessing competence in cognitive-behavioural therapy. *Clinical Psychology Review*, 33(3), 484-499. <https://doi.org/https://doi.org/10.1016/j.cpr.2013.01.010>
- Naylor, P.-J., Nettlefold, L., Race, D., Hoy, C., Ashe, M. C., Wharf Higgins, J., & McKay, H. A. (2015). Implementation of school based physical activity interventions: A systematic review. *Preventive Medicine*, 72, 95-115. <https://doi.org/https://doi.org/10.1016/j.ypmed.2014.12.034>
- Nunnally, J., & Bernstein, I. (1967). McGraw-Hill series in psychology. *Psychometric Theory*. New York, NY, US: McGraw-Hill.
- O'Leary-Kelly, S. W., & Vokurka, R. (1998). The empirical assessment of construct validity. *Journal of Operations Management*, 16(4), 387-405.
- O'Connor, C., Small, S. A., & Cooney, S. M. (2007). Program fidelity and adaptation: Meeting local needs without compromising program effectiveness. *What Works, Wisconsin – Research to Practice Series*(4).
- Olds, D. L., Robinson, J., O'Brien, R., Luckey, D. W., Pettitt, L. M., Henderson, C. R., Ng, R. K., Sheff, K. L., Korfmacher, J., & Hiatt, S. (2002). Home visiting by paraprofessionals and by nurses: A randomized, controlled trial. *Pediatrics*, 110(3), 486-496.
- Olofsson, V., Skoog, T., & Tillfors, M. (2016). Implementing group based parenting programs: A narrative review. *Children and Youth Services Review*, 69, 67-81. <https://doi.org/https://doi.org/10.1016/j.childyouth.2016.07.004>
- Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web and mobile app for systematic reviews. *Systematic Reviews*, 5(1), 210.

- Owen, J., & Hilsenroth, M. J. (2014). Treatment adherence: The importance of therapist flexibility in relation to therapy outcomes. *Journal of Counseling Psychology*, 61(2), 280.
- Palinkas, L. A., Aarons, G. A., Chorpita, B. F., Hoagwood, K., Landsverk, J., & Weisz, J. R. (2009). Cultural exchange and the implementation of evidence-based practices: Two case studies. *Research on Social Work Practice*, 19(5), 602-612.
- Parenting for Lifelong Health. (2018). Parenting for Lifelong Health programme for parents and teens: Family guidebook. [https://cdn.who.int/media/docs/default-source/documents/child-maltreatment/plh-for-teens-family-guidebook-english.pdf?sfvrsn=9f09a19f\\_2&download=true](https://cdn.who.int/media/docs/default-source/documents/child-maltreatment/plh-for-teens-family-guidebook-english.pdf?sfvrsn=9f09a19f_2&download=true)
- Pasalich, D. S., Witkiewitz, K., McMahon, R. J., Pinderhughes, E. E., & Conduct Problems Prevention Research Group. (2016). Indirect effects of the Fast Track intervention on conduct disorder symptoms and callous-unemotional traits: Distinct pathways involving discipline and warmth. *Journal of Abnormal Child Psychology*, 44, 587-597.
- Patterson, G. R., DeBaryshe, B. D., & Ramsey, E. (1989). A developmental perspective on antisocial behavior. *American Psychologist*, 44(2), 329-335. <https://doi.org/10.1037/0003-066X.44.2.329>
- Patterson, G. R., Reid, J. B., & Dishion, T. J. (1992). *Antisocial Boys. OR: Castalia*.
- Pearl, J., Glymour, M., & Jewell, N. P. (2016). *Causal Inference in Statistics: A Primer*. John Wiley & Sons.
- Pelton, L. H. (2015). The continuing role of material factors in child maltreatment and placement. *Child Abuse & Neglect*, 41, 30-39.
- Peng, C., & Mueller, D. (2004). *The SAGE Encyclopedia of Social Science Research Methods*. <https://doi.org/10.4135/9781412950589>
- Peters, D. H., Adam, T., Alonge, O., Agyepong, I. A., & Tran, N. (2013). Implementation research: What it is and how to do it. *BMJ*, 347, f6753. <https://doi.org/10.1136/bmj.f6753>
- Petersilia, J. (1990). Conditions that permit intensive supervision programs to survive. *Crime & Delinquency*, 36(1), 126-145.
- Polanczyk, G. V., Salum, G. A., Sugaya, L. S., Caye, A., & Rohde, L. A. (2015). Annual research review: A meta-analysis of the worldwide prevalence of mental disorders in children and adolescents. *Journal of Child Psychology and Psychiatry*, 56(3), 345-365.

- Prinz, R. J., Sanders, M. R., Shapiro, C. J., Whitaker, D. J., & Lutzker, J. R. (2009). Population-based prevention of child maltreatment: The U.S. Triple P system population trial. *Prevention Science*, 10(1), 1-12. <https://doi.org/10.1007/s11121-009-0123-3>
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health and Mental Health Services Research*, 38(2), 65-76. <https://doi.org/https://doi.org/10.1007/s10488-010-0319-7>
- R Core Team. (2021a). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- R Core Team. (2021b). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Racz, S. J., & McMahon, R. J. (2011). The relationship between parental knowledge and monitoring and child and adolescent conduct problems: A 10-year update. *Clinical Child and Family Psychology Review*, 14, 377-398.
- Ranjan, A., & George, T. S. (2014). Evaluating social desirability bias in assessment. *Indian Journal of Health and Wellbeing*, 5(8), 948.
- Rapley, H. A., & Loades, M. E. (2019). A systematic review exploring therapist competence, adherence, and therapy outcomes in individual CBT for children and young people. *Psychotherapy Research*, 29(8), 1010-1019.
- Rebellon, C. J., & Straus, M. (2017). Corporal punishment and adult antisocial behavior: A comparison of dyadic concordance types and an evaluation of mediating mechanisms in Asia, Europe, and North America. *International Journal of Behavioral Development*, 41(4), 503-513.
- Redfern, A., Cluver, L. D., Casale, M., & Steinert, J. I. (2019). Cost and cost-effectiveness of a parenting programme to prevent violence against adolescents in South Africa. *BMJ Global Health*, 4(3), e001147.
- Reef, J., Diamantopoulou, S., van Meurs, I., Verhulst, F. C., & van der Ende, J. (2011). Developmental trajectories of child to adolescent externalizing behavior and adult DSM-IV disorder: Results of a 24-year longitudinal study. *Social Psychiatry and Psychiatric Epidemiology*, 46(12), 1233-1241.
- Rendu, A. (2004). *Treatment adherence in a behaviourally based parenting program* [Doctor of Philosophy, University College London ].

- Revelle, W. (2017). Procedures for psychological, psychometric, and personality. *R Package Bpsych^, Version 1 (5)*.
- Revelle, W., & Zinbarg, R. E. (2009). Coefficients alpha, beta, omega, and the glb: Comments on Sijtsma. *Psychometrika*, 74(1), 145-154.
- Reznowski, G. (2011). *Literary research and Canadian literature: Strategies and sources* (Vol. 10). Scarecrow Press.
- Rizopoulos, D. (2007). ltm: An R package for latent variable modeling and item response analysis. *Journal of Statistical Software*, 17, 1-25.
- Robbins, M. S., Feaster, D. J., Horigian, V. E., Puccinelli, M. J., Henderson, C., & Szapocznik, J. (2011). Therapist adherence in brief strategic family therapy for adolescent drug abusers. *Journal of Consulting and Clinical Psychology*, 79(1), 43.
- Roggman, L. A., Boyce, L. K., Cook, G. A., & Jump, V. K. (2001). Inside home visits: A collaborative look at process and quality. *Early Childhood Research Quarterly*, 16(1), 53-71.
- Roggman, L. A., Cook, G. A., Innocenti, M. S., Jump Norman, V., Boyce, L. K., Christiansen, K., & Peterson, C. A. (2016). Home visit quality variations in two Early Head Start programs in relation to parenting and child vocabulary outcomes. *Infant Mental Health Journal*, 37(3), 193-207.
- Rohrer, J. M. (2018). Thinking clearly about correlations and causation: Graphical causal models for observational data. *Advances in Methods and Practices in Psychological Science*, 1(1), 27-42.
- Rojas-Andrade, R., & Bahamondes, L. L. (2019). Is implementation fidelity important? A systematic review on school-based mental health programs. *Contemporary School Psychology*, 23(4), 339-350. <https://doi.org/10.1007/s40688-018-0175-0>
- Rose, A. L., Jack, H. E., Wan, C., Toloza, E., Bhattiprolu, K., Ragunathan, M., Schwartz, K. T., & Magidson, J. F. (2022). Implementing task-shared child and adolescent psychological interventions in low-and middle-income countries: A scoping review. *Journal of Clinical Child & Adolescent Psychology*, 1-16.
- Rose, M. (2016). Sensitivity and specificity of electronic databases: The example of searching for evidence on child protection issues related to pregnant women. *Evidence Based Midwifery*.

- Ruud, T., Drake, R., & Bond, G. R. (2020). Measuring fidelity to evidence-based practices: Psychometrics. *Administration and Policy in Mental Health and Mental Health Services Research*, 47(6), 871-873.
- Salkind, N. (2010). *Encyclopedia of Research Design*.  
<https://doi.org/10.4135/9781412961288>
- Sanders, M. R., Kirby, J. N., Tellegen, C. L., & Day, J. J. (2014). The Triple P-Positive Parenting Program: A systematic review and meta-analysis of a multi-level system of parenting support. *Clinical Psychology Review*, 34(4), 337-357.  
<https://doi.org/10.1016/j.cpr.2014.04.003>
- Sanders, M. R., Spry, C. S., Tellegen, C. L., Kirby, J. N., Metzler, C. M., & Prinz, R. J. (2020). Development and validation of fidelity monitoring and enhancement in an evidence-based parenting program. *The Journal of Behavioral Health Services & Research*, 47, 569-580.
- Santacroce, S. J., Maccarelli, L. M., & Grey, M. (2004). Intervention fidelity. *Nursing Research*, 53(1), 63-66.
- Satterfield, L. W. (2013). *The effectiveness of functional family therapy in an irish context: An examination of international implementation* [Doctoral Dissertation, Indiana University].
- Scott, S., Carby, A., & Rendu, A. (2008). *Impact of therapists' skill on effectiveness of parenting groups for child antisocial behavior* [Doctoral Dissertation, King's College London].
- Scott, S., & Gardner, F. (2015). Parenting programs. In *Rutter's Child and Adolescent Psychiatry* (pp. 483-496).
- Scriven, M. (1999). The fine line between evaluation and explanation. *Research on Social Work Practice*, 9(4), 521-524.
- Shapiro, C. J., Prinz, R. J., & Sanders, M. R. (2010). Population-based provider engagement in delivery of evidence-based parenting interventions: Challenges and solutions. *The Journal of Primary Prevention*, 31(4), 223-234.  
<https://doi.org/https://doi.org/10.1007/s10935-010-0210-z>
- Shapiro, S. S., & Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3-4), 591-611. <https://doi.org/10.1093/biomet/52.3-4.591>
- Shenderovich, Y., Eisner, M., Cluver, L., Doubt, J., Berezin, M., Majokweni, S., & Murray, A. L. (2019). Delivering a parenting program in south africa: The impact of



- implementation on outcomes. *Journal of Child and Family Studies*, 28(4), 1005-1017.
- Shenderovich, Y., Lachman, J. M., Ward, C. L., Wessels, I., Gardner, F., Tomlinson, M., Oliver, D., Janowski, R., Martin, M., Okop, K., Sacolo-Gwebu, H., Ngcobo, L. L., Fang, Z., Alampay, L., Baban, A., Baumann, A. A., Benevides de Barros, R., Bojo, S., Butchart, A., ... Cluver, L. . (2021). The science of scale for violence prevention: A new agenda for family strengthening in low- and middle-income countries. *Frontiers in Public Health*.  
<https://doi.org/https://doi.org/10.3389/fpubh.202>
- Shenderovich, Y., Wamoyi, J., Mgunga, S., Martin, M., Nydetabura, E., Manjengenja, N., & Lachman, J. M. (forthcoming). Adapting a violence prevention parenting programme as part of scaling up
- Shenderovich, Y., Ward, C. L., Lachman, J. M., Wessels, I., Sacolo-Gwebu, H., Okop, K., Oliver, D., Ngcobo, L. L., Tomlinson, M., Fang, Z., Janowski, R., Hutchings, J., Gardner, F., & Cluver, L. (2020). Evaluating the dissemination and scale-up of two evidence-based parenting interventions to reduce violence against children: study protocol. *Implementation Science Communications*, 1(1), 109.  
<https://doi.org/10.1186/s43058-020-00086-6>
- Shrout, P. E., & Fleiss, J. L. (1979). Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, 86(2), 420.
- Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment*, 5(4), 299-321.
- Skivington, K., Matthews, L., Simpson, S. A., Craig, P., Baird, J., Blazeby, J. M., Boyd, K. A., Craig, N., French, D. P., & McIntosh, E. (2021). A new framework for developing and evaluating complex interventions: Update of Medical Research Council guidance. *BMJ*, 374.
- Smith, J. D., Dishion, T. J., Shaw, D. S., & Wilson, M. N. (2013). Indirect effects of fidelity to the family check-up on changes in parenting and early childhood problem behaviors. *Journal of Consulting and Clinical Psychology*, 81(6), 962.
- Smith, J. D., Rudo-Stern, J., Dishion, T. J., Stormshak, E. A., Montag, S., Brown, K., Ramos, K., Shaw, D. S., & Wilson, M. N. (2019). Effectiveness and efficiency of observationally assessing fidelity to a family-centered child intervention: A quasi-experimental study. *Journal of Clinical Child & Adolescent Psychology*, 48(1), 16-28.

- Snider, M. D. (2019). *Examining the impact of treatment fidelity on client outcomes in a statewide implementation of parent-child interaction therapy* [Master's Thesis, Western Virginia University].
- Snow, A., Cook, K. F., Lin, P. S., Morgan, R. O., & Magaziner, J. (2005). Proxies and other external raters: Methodological considerations. *Health Services Research, 40*(5p2), 1676-1693.
- St. George, S. M., Huang, S., Vidot, D. C., Smith, J. D., Brown, C. H., & Prado, G. (2016). Factors associated with the implementation of the Familias Unidas intervention in a type 3 translational trial. *Translational Behavioral Medicine, 6*(1), 105-114.
- Stemler, S., & Tsai, J. (2008). Best practices in interrater reliability three common approaches. In *Best Practices in Quantitative Methods*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412995627>
- Sterne, J. A., Hernán, M. A., Reeves, B. C., Savović, J., Berkman, N. D., Viswanathan, M., Henry, D., Altman, D. G., Ansari, M. T., & Boutron, I. (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ, 355*.
- Stirman, S. W. (2020). Commentary: Challenges and opportunities in the assessment of fidelity and related constructs. *Administration and Policy in Mental Health and Mental Health Services Research, 47*(6), 932-934. <https://doi.org/https://doi.org/10.1007/s10488-020-01069-4>
- Stirman, S. W., Gutner, C. A., Crits-Christoph, P., Edmunds, J., Evans, A. C., & Beidas, R. S. (2015). Relationships between clinician-level attributes and fidelity-consistent and fidelity-inconsistent modifications to an evidence-based psychotherapy. *Implementation Science, 10*(1), 1-10.
- Stith, S. M., Liu, T., Davies, L. C., Boykin, E. L., Alder, M. C., Harris, J. M., Som, A., McPherson, M., & Dees, J. (2009). Risk factors in child maltreatment: A meta-analytic review of the literature. *Aggression and Violent Behavior, 14*(1), 13-29.
- Stoltenborgh, M., Bakermans-Kranenburg, M. J., & van Ijzendoorn, M. H. (2013, Mar). The neglect of child neglect: A meta-analytic review of the prevalence of neglect. *Social Psychiatry and Psychiatric Epidemiology, 48*(3), 345-355. <https://doi.org/10.1007/s00127-012-0549-y>
- Stone, A. A., Bachrach, C. A., Jobe, J. B., Kurtzman, H. S., & Cain, V. S. (1999). *The Science of Self-report: Implications for Research and Practice*. Psychology Press.
- Stormshak, E. A., Bierman, K. L., McMahon, R. J., Lengua, L. J., & Conduct Problems Prevention Research Group. (2000). Parenting practices and child disruptive

- behavior problems in early elementary school. *Journal of Clinical Child Psychology*, 29(1), 17-29.
- Straus, M. A., Hamby, S. L., Boney-McCoy, S., & Sugarman, D. B. (1996). The revised Conflict Tactics Scales (CTS2) development and preliminary psychometric data. *Journal of Family Issues*, 17(3), 283-316.
- Suhrheinrich, J., Dickson, K. S., Chan, N., Chan, J. C., Wang, T., & Stahmer, A. C. (2020). Fidelity assessment in community programs: An approach to validating simplified methodology. *Behavior Analysis in Practice*, 13(1), 29-39.
- Taylor, B., Wylie, E., Dempster, M., & Donnelly, M. (2007). Systematically retrieving research: A case study evaluating seven databases. *Research on Social Work Practice*, 17(6), 697-706.
- Terwee, C., Bot, S. D., de Boer, M. R., van der Windt, D. A., Knol, D. L., Dekker, J., Bouter, L. M., & de Vet, H. C. (2007). Quality criteria were proposed for measurement properties of health status questionnaires. *Journal of Clinical Epidemiology*, 60(1), 34-42. <https://doi.org/10.1016/j.jclinepi.2006.03.012>
- Terwee, C., Prinsen, C. A. C., Chiarotto, A., Westerman, M. J., Patrick, D. L., Alonso, J., Bouter, L. M., de Vet, H. C. W., & Mokkink, L. B. (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, 27(5), 1159-1170. <https://doi.org/10.1007/s11136-018-1829-0>
- Thijssen, J., Albrecht, G., Muris, P., & de Ruiter, C. (2017). Treatment fidelity during therapist initial training is related to subsequent effectiveness of Parent Management Training—Oregon model. *Journal of Child and Family Studies*, 26(7), 1991-1999.
- Thomson, H. J., & Thomas, S. (2013). The effect direction plot: Visual display of non-standardised effects across multiple outcome domains. *Research Synthesis Methods*, 4(1), 95-101.
- Thorkildsen, T. (2010). Validity of measurement. In N. Salkind (Ed.), *Encyclopedia of Research Design*. SAGE Publications, Inc. <https://doi.org/10.4135/9781412961288>
- Tiwari, A., Whitaker, D., & Self-Brown, S. (2021). Comparing fidelity monitoring methods in an evidence-based parenting intervention. *Journal of Children's Services*, 16(2).
- Tomlinson, M., Hunt, X., & Rotheram-Borus, M. J. (2018). Diffusing and scaling evidence-based interventions: Eight lessons for early child development from the implementation of perinatal home visiting in South Africa. *Annals of the New York Academy of Sciences*, 1419(1), 218-229. <https://doi.org/https://doi.org/10.1111/nyas.13650>

- Tomlinson, M., Sherr, L., Macedo, A., Hunt, X., & Skeen, S. (2017). Paid staff or volunteers – does it make a difference? The impact of staffing on child outcomes for children attending community-based programmes in South Africa and Malawi. *Global Health Action*, 10(1), 1381462. <https://doi.org/10.1080/16549716.2017.1381462>
- Tommeraas, T., & Ogden, T. (2017). Is there a scale-up penalty? Testing behavioral change in the scaling up of parent management training in Norway. *Administration and Policy in Mental Health and Mental Health Services Research*, 44(2), 203-216.
- UNICEF. (2010). Child disciplinary practices at home. *Evidence from a Range of Low and Middle-Income Countries*. New York: UNICEF.
- UNICEF. (2011). *Violence against children in Tanzania: Findings from a national survey*. <https://evaw-global-database.unwomen.org/fr/countries/africa/united-republic-of-tanzania/2011/violence-against-children-in-tanzania-findings-from-a-national-survey-2009>
- UNICEF. (2022). Multiple Indicator Cluster Surveys. <https://mics.unicef.org>
- United Nations Department of Economic Affairs. (2019). International youth day: Ten key messages. [https://www.un.org/development/desa/youth/wp-content/uploads/sites/21/2019/08/WYP2019\\_10-Key-Messages\\_GZ\\_8AUG19.pdf](https://www.un.org/development/desa/youth/wp-content/uploads/sites/21/2019/08/WYP2019_10-Key-Messages_GZ_8AUG19.pdf)
- Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, 41(3), 321-327. <https://doi.org/10.1007/BF02293557>
- Viswanathan, M., Kraschnewski, J. L., Nishikawa, B., Morgan, L. C., Honeycutt, A. A., Thieda, P., Lohr, K. N., & Jonas, D. E. (2010). Outcomes and costs of community health worker interventions: A systematic review. *Medical Care*, 792-808.
- Walker, S. P., Wachs, T. D., Grantham-McGregor, S., Black, M. M., Nelson, C. A., Huffman, S. L., Baker-Henningham, H., Chang, S. M., Hamadani, J. D., Lozoff, B., Gardner, J. M. M., Powell, C. A., Rahman, A., & Richter, L. (2011). Inequality in early childhood: Risk and protective factors for early child development. *The Lancet*, 378(9799), 1325-1338. [https://doi.org/10.1016/S0140-6736\(11\)60555-2](https://doi.org/10.1016/S0140-6736(11)60555-2)
- Walton, H., Spector, A., Tombor, I., & Michie, S. (2017). Measures of fidelity of delivery of, and engagement with, complex, face-to-face health behaviour change interventions: A systematic review of measure quality. *British Journal of Health Psychology*, 22(4), 872-903. <https://doi.org/10.1111/bjhp.12260>
- Waltz, J., Addis, M. E., Koerner, K., & Jacobson, N. S. (1993). Testing the integrity of a psychotherapy protocol: assessment of adherence and competence. *Journal of*

*Consulting and Clinical Psychology*, 61(4), 620.  
<https://doi.org/https://doi.org/10.1037/0022-006X.61.4.620>

- Ward, C., Sanders, M. R., Gardner, F., Mikton, C., & Dawes, A. (2015). Preventing child maltreatment in low-and middle-income countries: Parent support programs have the potential to buffer the effects of poverty. *Child Abuse & Neglect*.
- Ward, C., Sanders, M. R., Gardner, F., Mikton, C., & Dawes, A. (2016a). Preventing child maltreatment in low-and middle-income countries. *Child Abuse & Neglect*, 54, 97-107.
- Ward, C., Sanders, M. R., Gardner, F., Mikton, C., & Dawes, A. (2016b). Preventing child maltreatment in low-and middle-income countries: What is needed to build the evidence based for effective implementation of parent support programs? *Child Abuse & Neglect*, 54, 97-107.
- Ward, C. L., Wessels, I. M., Lachman, J. M., Hutchings, J., Cluver, L. D., Kassanjee, R., Nhapi, R., Little, F., & Gardner, F. (2020). Parenting for Lifelong Health for Young Children: A randomized controlled trial of a parenting program in South Africa to prevent harsh parenting and child conduct problems. *Journal of Child Psychology and Psychiatry*, 61(4), 503-512.
- Warne, R. T., & Larsen, R. (2014). Evaluating a proposed modification of the Guttman rule for determining the number of factors in an exploratory factor analysis. *Psychological Test and Assessment Modeling*, 56(1), 104.
- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, 44(3), 219-246.  
<https://doi.org/https://doi.org/10.1177/0095798418771807>
- Watson, R. J., & Richardson, P. H. (1999). Accessing the literature on outcome studies in group psychotherapy: The sensitivity and precision of Medline and PsycINFO bibliographic database searching. *Br J Med Psychol*, 72 ( Pt 1), 127-134.  
<https://doi.org/10.1348/000711299159763>
- Webb, C. A., Auerbach, R. P., & DeRubeis, R. J. (2012). Processes of change in CBT of adolescent depression: Review and recommendations. *Journal of Clinical Child & Adolescent Psychology*, 41(5), 654-665.
- Webb, C. A., DeRubeis, R. J., & Barber, J. P. (2010). Therapist adherence/competence and treatment outcome: A meta-analytic review. *Journal of Consulting and Clinical Psychology*, 78(2), 200.

- Webster-Stratton, C. H., Reid, M. J., & Marsenich, L. (2014). Improving therapist fidelity during implementation of evidence-based practices: Incredible Years program. *Psychiatric Services*, 65(6), 789-795.
- Webster-Stratton, C., & McCoy, K. P. (2015). Bringing the Incredible Years® programs to scale. *New Directions for Child and Adolescent Development*, 2015(149), 81-95.
- Wessels, I., & Ward, C. L. (2015). A ‘best buy’ for violence prevention: Evaluating parenting skills programmes. *South African Crime Quarterly*, 54(1), 17-28.
- WHO. (2016). INSPIRE: Seven strategies for ending violence against children. <https://www.who.int/publications-detail/inspire-seven-strategies-for-ending-violence-against-children>
- WHO. (2020). Parenting for Lifelong Health: A suite of parenting programmes to prevent violence. <https://www.who.int/teams/social-determinants-of-health/parenting-for-lifelong-health>
- WHO. (2023). WHO guidelines on parenting interventions to prevent maltreatment and enhance parent-child relationships with children 0-17 years. <https://www.who.int/publications/i/item/9789240065505>
- Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemond, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K., & Yutani, H. (2019). Welcome to the tidyverse. *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>
- Wickham, H., François, R., Henry, L., Müller, K., & D, V. (2023). dplyr: A grammar of data manipulation. <https://CRAN.R-project.org/package=dplyr>
- Widom, C. S. (2017). Long-term impact of childhood abuse and neglect on crime and violence. *Clinical Psychology: Science and Practice*, 24(2), 186.
- Widom, C. S., Czaja, S. J., Bentley, T., & Johnson, M. S. (2012). A prospective investigation of physical health outcomes in abused and neglected children: New findings from a 30-year follow-up. *American Journal of Public Health*, 102(6), 1135-1144. <https://doi.org/10.2105/ajph.2011.300636>
- Wilson, S. J., Lipsey, M. W., & Derzon, J. H. (2003). The effects of school-based intervention programs on aggressive behavior: A meta-analysis. *Journal of Consulting and Clinical Psychology*, 71(1), 136.

Wolpert, M., & Rutter, H. (2018). Using flawed, uncertain, proximate and sparse (FUPS) data in the context of complexity: Learning from the case of child mental health. *BMC Medicine*, 16(1), 82. <https://doi.org/10.1186/s12916-018-1079-6>

World Health Organization. (2022). Violence against children: Key facts. <https://www.who.int/news-room/fact-sheets/detail/violence-against-children>