

RESEARCH ARTICLE

WILEY

The development, reliability, and validity of the Facilitator Assessment Tool: An implementation fidelity measure used in Parenting for Lifelong Health for Young Children

Mackenzie Martin¹  | Jamie M. Lachman^{1,2} | Hugh Murphy³ |
Frances Gardner¹ | Heather Foran³

¹Department of Social Policy and Intervention, University of Oxford, Oxford, UK

²MRC/CSO Social and Public Health Science Unit, University of Glasgow, Glasgow, UK

³Institute for Psychology, Alpen-Adria-University Klagenfurt, Klagenfurt, Austria

Correspondence

Mackenzie Martin, Department of Social Policy and Intervention, University of Oxford, Oxford, UK.

Email: mackenzie.martin@spi.ox.ac.uk

Funding information

This study was funded by the European Union's Horizon 2020 Research Innovation Programme under grant agreement number 779318 and by the Complexity and Relationships in Health Improvement Programmes of the Medical Research Council MRC UK and Chief Scientist Office (MC_UU_12017/14, MC_UU_1201711, and CSO SPHSU11)

Abstract

Background: The Parenting for Lifelong Health for Young Children (PLH-YC) programme aims to reduce violence against children and child behaviour problems among families in low- and middle-income countries (LMICs). Although the programme has been tested in four randomised controlled trials and delivered in over 25 countries, there are gaps in understanding regarding the programme's implementation fidelity and, more generally, concerning the implementation fidelity of parenting programmes in LMICs.

Aims: This study aims to address these gaps by examining the psychometric properties of the PLH-YC-Facilitator Assessment Tool (FAT)—an observational tool used to measure the competent adherence of PLH-YC facilitators. Examining the psychometric properties of the FAT is important in order to determine whether there is an association between facilitator competent adherence and programme outcomes and, if correlated, to improve facilitator performance. It is also important to develop the implementation literature among parenting interventions in LMICs.

Methods: The study examined the content validity, intra-rater reliability, and inter-rater reliability of the FAT. Revision of the tool was based on consultation with programme trainers, experts, and assessors. A training curriculum and assessment manual was created. Assessors were trained in Southeastern Europe and their assessments of facilitator delivery were analysed as part of a large-scale factorial experiment ($N = 79$ facilitators).

Results: The content validity process with PLH-YC trainers, experts, and assessors resulted in substantial improvements to the tool. Analyses of percentage agreements and intraclass correlations found that, even with practical challenges, assessments were completed with adequate yet not strong intra- and inter-rater reliability.

Abbreviations: COSMIN, COnsensus-based standards for the selection of health Measurement INstruments; CWBSA, Clowns Without Borders South Africa; FAT, Facilitator Assessment Tool; ICC, intraclass correlation; IY, incredible years; PLH, Parenting for Lifelong Health; PLH-YC, Parenting for Lifelong Health-Young Children; PMTO, Parent Management Training-Oregon Model; RCTs, randomised controlled trials.

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2022 The Authors. *Child: Care, Health and Development* published by John Wiley & Sons Ltd.

Conclusions: This study contributes to the literature on the implementation of parenting programmes in LMICs. The study found that the FAT appears to capture its intended constructs and can be used with an acceptable degree of consistency. Further research on the tool's reliability and validity—specifically, its internal consistency, construct validity, and predictive validity—is recommended.

KEYWORDS

behaviour, fidelity, implementation, parenting

1 | INTRODUCTION

1.1 | Parenting for Lifelong Health

There is considerable evidence that parenting programmes increase positive parenting and parent-child relationships as well as reduce child maltreatment and behaviour problems (e.g., Chen & Chan, 2016; Furlong et al., 2013). One such programme for families with children across the developmental spectrum is the Parenting for Lifelong Health for Young Children programme (PLH-YC), which was originally developed in South Africa by individuals from the Universities of Bangor, Cape Town, and Oxford in collaboration with the World Health Organization, UNICEF, and Clowns Without Borders South Africa along with input from parents and practitioners (Lachman, Sherr, et al., 2016). PLH-YC is a group-based parenting programme for parents of children ages 2 to 9 years rooted in social learning theory and behaviour change principles (e.g., goal setting and discussing progress on goals) (Lachman, Sherr, et al., 2016; Michie et al., 2015; Ward et al., 2014). The programme ranges from five to 12 sessions in length and uses participatory, nondidactic, and strengths-based approaches to empower parents to develop skills in fostering positive relationships, handling conflict and emotions, and employing effective disciplinary approaches (Lachman, Sherr, et al., 2016). Parenting for Lifelong Health (PLH) programmes are now being implemented in over 25 countries in Africa, Asia, and Southeastern Europe. To date, four randomised controlled trials (RCTs) have examined the effectiveness of PLH-YC—two in South Africa, one in the Philippines, and one in Thailand (Gardner et al., forthcoming; Lachman et al., 2017, 2021; Ward et al., 2020). These studies found improved positive parenting and child behaviour, while reducing family violence. Further information about PLH and its growing evidence base is available in numerous published papers, protocols, and resources (e.g., Martin, Lachman, et al., 2021; Shenderovich et al., 2020; World Health Organisation [WHO], 2020). As the evidence supporting parenting programmes such as PLH is positive and substantial, studying their implementation fidelity represents an important way to explore which programme elements are correlated with outcomes and how to improve delivery.

Key messages

- The Facilitator Assessment Tool (FAT) is an observational implementation measure used to assess the competent adherence with which programme facilitators deliver the Parenting for Lifelong Health-Young Children (PLH-YC) programme.
- As PLH-YC is being delivered widely in many countries, it is necessary to establish the tool's psychometric properties to provide confidence that the competent adherence of facilitators is consistently and accurately assessed.
- This study analysed the content validity, intra-rater reliability, and inter-rater reliability to provide initial evidence of the tool's ability to capture its intended constructs and be used consistently by and between assessors.
- Further analyses of the FAT's reliability and validity (internal consistency, construct validity, and predictive validity) are warranted and would provide additional evidence as to the strength of this tool for measuring facilitator competent adherence.

1.2 | Measures of competent adherence

Many parenting programmes have developed programme-specific tools to measure implementation fidelity (e.g., Martin, Steele, et al., 2021). This paper focuses on a measure developed for and used in PLH-YC to assess two aspects of implementation fidelity—competence and adherence, which together capture the skill and diligence with which facilitators deliver intervention components (Fixsen et al., 2005). Assessing competent adherence has numerous research and practical benefits, including providing an objective assessment of the extent to which a particular parenting programme is implemented as planned. These assessments allow researchers to examine whether facilitator delivery is associated with parent and/or child outcomes, understand what makes an effective facilitator, and indicate where to target programme improvements (Forgatch et al., 2005).

A systematic review by Martin, Steele, et al. (2021) identified 65 measures of competent adherence employed in 63 parenting programmes. Parenting programmes differ according to content and delivery, and this variety is reflected in the number of fidelity tools available. Measures also vary in other respects: Some use observational, nonobservational, or a combination of methods; some are structured with dichotomous or Likert scale items; and all capture one or more aspects of competent adherence.

1.3 | PLH facilitator training and assessment

Facilitators—typically community members and professionals including teachers, psychologists, and social workers—receive PLH-YC training through a 5-day workshop (approximately 30 h). Training is provided by Clowns Without Borders South Africa (CWBSA), a nonprofit serving as a capacity building agency for the dissemination of PLH. In training, facilitators learn to implement programme activities and skills, including using group discussions on parenting challenges, illustrated comics to identify and model parenting skills, and group activities to practice skills (Lachman, Cluver, et al., 2016). Following training, facilitators deliver PLH-YC and receive regular supervision from coaches. The PLH-YC Facilitator Assessment Tool (FAT) was developed by study investigators and programme developers to assess the quality with which facilitators deliver programme-specific activities and techniques.

1.4 | PLH facilitator assessment procedure

The FAT assessment procedure is similar to other measures of facilitator quality of delivery, including measures used by Parent Management Training-Oregon Model (PMTO) (Holtrop et al., 2020; Knutson et al., 2019) and Incredible Years (IY) (Eames et al., 2008). Like the PMTO and IY tools, the FAT uses observational methods (live or video-recorded) to assess facilitator delivery of entire programme sessions. Observational methods are used since these are considered more objective than nonobservational methods such as facilitator self-reports, which could be prone to social desirability bias (Eames et al., 2008; Stone et al., 1999). However, observational assessments are often more resource-intensive and may be susceptible to reactivity bias (Girard & Cohn, 2016). Training of FAT assessors is approximately 14 h and includes theoretical and practical components following a training curriculum, assessor manual, and coding matrix. As the training is conducted in low-income settings, fewer financial resources are dedicated to training assessors than in high-income settings.

1.5 | FAT

The FAT is composed of two subscales (50 items). The Activities Subscale assesses facilitator adherence to three key PLH-YC activities (24 items)—home activity discussion (10 items; e.g., “identify specific

challenges when shared by at least one parent”), illustrated story discussion (seven items; e.g., “discuss possible solutions for negative stories”), and group practice activity (seven items; e.g., “debrief with the participants about experiences and feelings”). The Skills Subscale assesses facilitator competence in delivering key PLH-YC process skills (26 items)—modelling parenting behaviours (seven items, e.g., “give positive, specific, and realistic instructions”), demonstrating collaborative facilitation (seven items; e.g., “accept participant responses verbally by reflecting back what the participant says”), encouraging participation (seven items; e.g., “participants appear comfortable and involved in the session”), and utilising leadership skills (six items, e.g., “use open-ended questions during group discussions”). Items are rated using a 4-point Likert scale (0 = *inadequate*, 1 = *needs improvement*, 2 = *good*, 3 = *excellent*). Items are summed to produce an impression score represented as a percentage.

1.6 | Current study

Although there is growing evidence of the effectiveness of PLH-YC, there are gaps in understanding regarding its implementation quality and, more specifically, concerning facilitator competent adherence. As there was not a suitable measure of competent adherence available in the literature, this study describes the development of the tool used to assess PLH-YC facilitators and provides preliminary psychometric evidence on the tool using data from the delivery of PLH-YC in North Macedonia, the Republic of Moldova (“Moldova”), and Romania as part of the RISE study. RISE is a collaboration seeking to evaluate the effectiveness and costs of PLH-YC (Frantz et al., 2019; Lachman et al., 2019). As a result, this study aimed to answer the following research question: what is the content validity, intra-rater reliability, and inter-rater reliability of the FAT as a measure of PLH-YC facilitator competent adherence in Southeastern Europe? As the first study on the FAT’s psychometric properties, these indices were selected to examine whether the tool shows promise for further psychometric analyses or needs revision.

2 | METHODS

This paper examined the content validity, intra-rater reliability, and inter-rater reliability of the FAT in Southeastern Europe using COSMIN recommendations (Mokkink et al., 2010). Content validity was examined by consulting with three stakeholder groups and revising the FAT accordingly. Intra- and inter-rater reliability was examined by determining the degree to which assessors used the FAT consistently over time as well as with each other.

2.1 | Participants

This study involved three stakeholder groups. First, during the content validity process, we consulted with eight certified trainers from

CWBSA with at least 2 years of experience conducting FAT assessments in numerous countries. Second, we sought the advice of three parenting programme experts. These experts were consulted due to their extensive knowledge of PLH-YC and on conducting facilitator assessments in both research and practice. Third, those who conducted facilitator assessments (“assessors”) supported our evaluation of all three psychometric properties. The assessors were 11 trained coaches involved in RISE ($n = 5$ in Moldova, $n = 3$ in North Macedonia, and $n = 3$ in Romania). The Moldovan assessors had a range of experiences and backgrounds supporting vulnerable families, including teaching and family therapy. The North Macedonian assessors were later career psychologists with experience in conducting assessments similar to the FAT. The Romanian assessors were early career psychologists based at a local university. The Moldovan and Romanian assessors also had prior experience as PLH-YC facilitators.

2.2 | Procedure and analytic strategy

2.2.1 | Content validity

The FAT's content validity—the extent the tool appears to capture the intended constructs—was examined by gathering and synthesising the perspectives of the three aforementioned stakeholder groups on the measure's comprehensiveness, relevance, and comprehensibility (Mokkink et al., 2010; Terwee et al., 2018). First, we held a content validity workshop with CWBSA trainers, wherein detailed field notes were taken. The trainers were asked to describe their use of the FAT; share their perspectives on the tool's comprehensiveness, comprehensibility, and relevance; and suggest revisions to improve the tool's utility and accuracy. Their feedback was used to modify the FAT and create an initial training curriculum and manual. Second, we consulted the three parenting programme experts, who provided feedback on an updated version of the FAT and training manual. Third, during assessor training, RISE study assessors provided input on the FAT's comprehensiveness and relevance. It is important to note that the assessor training was not delivered to completion in North Macedonia and Romania due to scheduling conflicts, which may have compromised assessment reliability.

2.2.2 | Intra-rater reliability

Intra-rater reliability, or assessor consistency, was examined by having each assessor observe a video recording of a facilitator delivering the programme twice with assessments conducted 3 weeks apart (Gwet, 2014; Heindel et al., 2016). As a result, a video from one facilitator per country was selected. It was only possible to assess one facilitator per country due to time and resource constraints. FAT scores were compared by calculating percentage agreements and intra-class correlations (ICCs) for each assessor and subscale (Margolin et al., 1998). Percentage agreements were selected because they indicate the ratio of instances wherein assessors chose the same rating.

Agreement levels above 70% were acceptable (Aspland & Gardner, 2003). ICCs were also examined to take chance agreement and correlation into account (Bruton et al., 2000; Koo & Li, 2016). A two-way mixed-effects model with an absolute agreement definition and single-rater type was used (McGraw & Wong, 1996; Shrout & Fleiss, 1979). ICCs were interpreted within a 95% confidence interval where ICCs under 0.50 were considered poor, between 0.50 and 0.75 were moderate, between 0.75 and 0.90 were good, and above 0.90 were excellent (Koo & Li, 2016). Percentage agreements and ICCs were calculated using the “irr package” in R (Gamer et al., 2017). As all items of the FAT should have been completed, mean imputation was used to take missing data into account when less than 10% of the data was missing (Watkins, 2018).

2.2.3 | Inter-rater reliability

Inter-rater reliability, the degree to which different coders similarly assess facilitator delivery (Chen & Krauss, 2004; Cho, 1981), was examined by having assessors observe video recordings of the same three facilitators selected out of 31 possible facilitators in Moldova, 16 in North Macedonia, and 32 in Romania (Hallgren, 2012). Thus, videos from a total of nine facilitators were used. It was only possible to assess three facilitators per country due to time and resource constraints. The data were analysed using the same methods as the assessments of intra-rater reliability.

3 | RESULTS

3.1 | Content validity

The content validity consultations produced recommendations to improve the FAT. CWBSA trainers made four recommendations: break up complex items into separate, simple items; use specific definitions for each item and Likert point; add items to capture missing activities and skills; and remove redundant items.

After revisions based on trainer feedback, the FAT was shared with the parenting programme experts. Rooted in evidence linking praise and reflexive statements to participant outcomes (Eames et al., 2010), the experts recommended adding three items to measure the frequency of specific and unspecific praise (i.e., expressing approval and appreciation) and reflexive statements (i.e., reiterating participant contributions).

The further revised tool was then shared with PLH-YC coaches during assessment training. Initially, assessors recommended changes to item wording and suggested examples to include in the definitions. After using the tool, the assessors provided further insight, indicating which items were difficult to understand and how they could be improved. This process helped ensure clear differences between points on the Likert scale.

Following assessor input, the revised FAT was finalised, resulting in 62 items with 26 items in the Activities Subscale, 33 items in the

TABLE 1 Summary of recommendations and changes to FAT

Recommendation to improve the FAT	Stakeholder group	Changes made or example of changes made															
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 border="1"> <thead> <tr> <th colspan="5">COLLABORATIVE LEADERSHIP</th> </tr> <tr> <th>Did the facilitator...</th> <th>0 Inadequate</th> <th>1 Needs Improvement</th> <th>2 Good</th> <th>3 Excellent</th> </tr> </thead> <tbody> <tr> <td>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> </tbody> </table>	COLLABORATIVE LEADERSHIP					Did the facilitator...	0 Inadequate	1 Needs Improvement	2 Good	3 Excellent	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																	
Did the facilitator...	0 Inadequate	1 Needs Improvement	2 Good	3 Excellent													
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	RISE Assessors	The item “Accepts parent responses physically” was changed to “Uses body language to show acceptance.”															
Proposed examples to include in item definitions	RISE Assessors	<table border="1"> <thead> <tr> <th colspan="5">MODELLING BEHAVIOUR</th> </tr> <tr> <th>Did the facilitator...</th> <th>0 Inadequate</th> <th>1 Needs Improvement</th> <th>2 Good</th> <th>3 Excellent</th> </tr> </thead> <tbody> <tr> <td>1. Give positive reinforcement and specific praise to parents</td> <td>The facilitator never or almost never praised parents.</td> <td>The facilitator rarely praised parents or it was rarely specific or genuine. Example: “Thank you, Patty.”</td> <td>The facilitator often praised parents. Praise was often specific and genuine. Example: “Thank you, Patty, for sharing about your experience at home.”</td> <td>The facilitator almost always praised parents throughout the session. Praise was almost always specific, genuine, and reinforced the core building blocks. Example: “Well done, Patty, for involving your child when establishing Household Rules.”</td> </tr> </tbody> </table>	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 never or almost never praised parents.	The facilitator rarely praised parents or it was rarely specific or genuine. Example: “Thank you, Patty.”	The facilitator often praised parents. Praise was often specific and genuine. Example: “Thank you, Patty, for sharing about your experience at home.”	The facilitator almost always praised parents throughout the session. Praise was almost always specific, genuine, and reinforced the core building blocks. Example: “Well done, Patty, for involving your child when establishing Household Rules.”
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 never or almost never praised parents.	The facilitator rarely praised parents or it was rarely specific or genuine. Example: “Thank you, Patty.”	The facilitator often praised parents. Praise was often specific and genuine. Example: “Thank you, Patty, for sharing about your experience at home.”	The facilitator almost always praised parents throughout the session. Praise was almost always specific, genuine, and reinforced the core building blocks. Example: “Well done, Patty, for involving your child when establishing Household Rules.”													

TABLE 2 Assessor-level intra-rater reliability percentage agreements

Assessor and country	Total percentage agreement (activities and skills subscale, N = 59 items)	Activities subscale (N = 26 items) percentage agreement	Skills subscale (N = 33 items) percentage agreement	Frequency subscale (N = 3 items) percentage agreement
Moldova	57.6–89.9% (mean: 75.3%)	46.2–92.3% (mean: 74.6%)	45.5–90.9% (mean: 75.8%)	0.0–100.0% (mean: 60%)
North Macedonia	78.0–88.1% (mean: 83.1%)	80.8–92.3% (mean: 87.2%)	75.8–84.8% (mean: 79.8%)	0.0–100.0% (mean: 44.4%)
Romania	66.1–91.5% (mean: 76.8%)	69.2–92.3% (mean: 80.8%)	63.6–90.9% (mean: 73.7%)	66.7–100.0% (mean: 88.9%)

Note: Timepoint 1: December 2019, Timepoint 2: January 2020.

Skills Subscale, and three items in the Frequency Subscale. A summary of the recommendations and changes made is provided in Table 1.

3.2 | Intra- and inter-rater reliability

In terms of intra-rater reliability, the overall percentage agreements across the three countries ranged from 57.6% to 91.5% with ICCs of 0.52–0.94 (Tables 2 and 3). Subscale percentage agreements ranged

from 46.2% to 92.3% with ICCs of 0.32–0.96 on the Activities Subscale, 45.5% to 90.9% with ICCs of 0.13–0.93 on the Skills Subscale, and 0.0% to 100.0% on the Frequency Subscale with ICCs of –0.79–1.00. At the country-level, overall percentage agreements ranged from 57.6% to 89.9% with ICCs of 0.52–0.94 in Moldova, 78.0% to 88.1% with ICCs of 0.82–0.90 in North Macedonia, and 66.1% to 91.5% with ICCs of 0.79–0.94 in Romania. There was no missing intra-rater reliability data.

In terms of inter-rater reliability, the overall percentage agreements ranged from 18.1% to 74.0% with ICCs of 0.49–0.91

TABLE 3 Assessor-level intra-rater reliability ICCs

Country	Total score	Activity	Skills	Frequency
Moldova	0.52–0.94 (mean: 0.78)	0.32–0.94 (mean: 0.79)	0.13–0.93 (mean: 0.70)	–0.13–1.00 (mean: 0.75)
North Macedonia	0.82–0.90 (mean: 0.86)	0.90–0.96 (mean: 0.93)	0.58–0.68 (mean: 0.62)	–0.79–1.00 (mean: 0.30)
Romania	0.79–0.94 (mean: 0.85)	0.81–0.95 (mean: 0.86)	0.79–0.93 (mean: 0.84)	0.80–1.00 (mean: 0.93)

Note: Model: two-way mixed effects, type: single rater, definition: absolute agreement.

TABLE 4 Moldova inter-rater reliability data by assessment

Facilitator assessment	Total score		Activity		Skills		Frequency	
	Percentage agreement	ICCs (95% CI)	Percentage agreement	ICCs (95% CI)	Percentage agreement	ICCs (95% CI)	Percentage agreements	ICCs (95% CI)
Moldova	18.1%	0.62 95% CI [0.53, 0.70]	29.5%	0.71 95% CI [0.62, 0.79]	9.1%	0.37 95% CI [0.25–0.49]	0.0%	0.52 95% CI [0.15, 0.89]
1	30.5%	0.77 95% CI [0.68–0.84]	50.0%	0.85 95% CI [0.76, 0.92]	15.2%	0.32 95% CI [0.17, 0.51]	MISSING	MISSING
2	15.3%	0.56 95% CI [0.40, 0.70]	23.1%	0.62 95% CI [0.43, 0.78]	9.1%	0.35 95% CI [0.17, 0.54]	0.0%	0.45 95% CI [–0.09, 0.98]
3	8.5%	0.49 95% CI [0.34, 0.64]	15.4%	0.57 95% CI [0.38, 0.75]	3.0%	0.35 95% CI [0.18, 0.54]	0.0%	0.63 95% CI [0.16, 0.99]

Note: ICC = intra-class correlation, 95% CI = 95% confidence interval; model: two-way mixed effects, type: single rater, definition: absolute agreement, five assessors and three facilitator assessments were used, missing: frequency data for assessment 1.

TABLE 5 North Macedonia inter-rater reliability data by assessment

Facilitator assessment	Total score		Activity		Skills		Frequency	
	Percentage agreement	ICCs (95% CI)	Percentage agreement	ICCs (95% CI)	Percentage agreement	ICCs (95% CI)	Percentage agreements	ICCs (95% CI)
Overall	74.0%	0.91 [0.88, 0.93]	71.8%	0.93 [0.90, 0.95]	75.8%	0.86 [0.82–0.90]	11.1%	0.95 [0.85, 0.99]
1	71.2%	0.90 [0.84, 0.93]	69.2%	0.93 [0.86–0.97]	72.7%	0.84 [0.74–0.91]	33.3%	0.80 [0.21, 0.99]
2	71.2%	0.90 [0.84, 0.93]	69.2%	0.92 [0.86, 0.96]	72.7%	0.84 [0.74, 0.91]	0.0%	0.99 [0.90, 1.00]
3	79.7%	0.93 [0.89, 0.95]	76.9%	0.94 [0.90, 0.97]	81.8%	0.85 [0.75, 0.92]	0.0%	0.78 [0.20, 0.99]

Note: ICC = intra-class correlation, 95% CI = 95% confidence interval; model: two-way mixed effects, type: single rater, definition: absolute agreement; three assessors and three facilitator assessments were used.

(Table 4–6). The overall percentage agreement across the three assessments was 18.1% (activities: 29.5%, skills: 9.1%, frequency: 0.0%) in Moldova with an ICC of 0.62 (activities: 0.71, skills: 0.37, frequency: 0.52), 74% (activities: 71.8%, skills: 75.8%, frequency: 11.1%) in North Macedonia with an ICC of 0.91 (activities: 0.93, skills: 0.86,

and frequency: 0.95), and 32.8% (activities: 38.7%, skills: 33.3%, frequency: 22.2%) in Romania with an ICC of 0.49 (activities: 0.65, skills: 0.50, frequency: 0.84). There was no missing inter-rater reliability data in Macedonia and a low amount of missing data in Moldova (0.11%) and Romania (4.30%). Assessors indicated that some data were

TABLE 6 Romania inter-rater reliability data

Facilitator assessment	Total score		Activity		Skills		Frequency	
	Percentage agreement	ICCs (95% CI)	Percentage agreement	ICCs (95% CI)	Percentage agreement	ICCs (95% CI)	Percentage agreements	ICCs (95% CI)
Overall	32.8%	0.49 [0.39, 0.59]	39.7%	0.65 [0.53, 0.75]	33.3%	0.50 [0.36, 0.63]	22.2%	0.84 [0.59, 0.96]
1	42.4%	0.64 [0.51, 0.75]	46.2%	0.72 [0.54, 0.85]	39.4%	0.57 [0.38, 0.74]	0.0%	0.71 [0.06, 0.99]
2	25.4%	0.47 [0.28, 0.63]	30.8%	0.55 [0.29, 0.75]	21.2%	0.40 [0.18, 0.62]	33.3%	0.90 [0.50, 1.00]
3	40.7%	0.62 [0.39, 0.77]	42.3%	0.71 [0.41, 0.86]	39.4%	0.54 [0.30, 0.73]	33.3%	0.92 [0.51, 1.00]

Note: ICC = intra-class correlation, 95% CI = 95% confidence interval; model: two-way mixed effects, type: single rater, definition: absolute agreement; three assessors and three facilitator assessments were used; missing: 24 items.

missing as it was not possible to assess some items (e.g., an opportunity did not arise for the facilitator to deliver a programme component). These items should have been scored as each FAT item captures an aspect of PLH-YC that should be delivered.

4 | DISCUSSION

This study analysed three psychometric properties of the FAT. The content validity process resulted in a revised FAT that was more understandable, specific, and practical due to stakeholder recommendations to the FAT's items and assessment procedure. The analysis of assessor intra-rater reliability found somewhat acceptable levels of consistency (overall percentage agreements ranged from 57.6% to 91.5% with ICCs of 0.52–0.94 and assessor inter-rater reliability ranged from 18.1% to 74.0% with ICCs of 0.49–0.91). Although assessors did not always achieve consensus, the ICCs were, with few exceptions, larger than the percentage agreements and most exceeded the suggested cut-off of ICC > 0.50 for moderate levels of reliability (Stemler & Tsai, 2008). This finding suggests that assessors were largely consistent in their application of the measure and its items, yet it was still difficult for assessors to achieve consensus on many occasions. It was particularly difficult for assessors from all countries to achieve intra- and inter-rater reliability on the Frequency Subscale. Negative ICCs were found in multiple instances, indicating more within variance than between. As a result, this subscale was not rated reliably and should be dropped or modified.

4.1 | Contextual factors

The intra- and inter-rater reliability results are encouraging in light of challenges encountered during assessor training and varying levels of assessor experience. The training was not delivered to completion in North Macedonia and Romania due to scheduling conflicts, resulting in approximately 80% of the training being delivered in each of these contexts. In North Macedonia and Romania, three and two assessors

respectively missed several hours of training due to other organisational commitments scheduled at the same time. Aside from Romania, the training relied on translation into local languages, which may have led to an imprecise understanding of assessment procedures. Further, there may be different levels of reliability across countries due to differences in assessment experience. Results may have been stronger in North Macedonia as these assessors were experienced psychologists accustomed to completing assessments similar to the FAT. In contrast, the Moldovan assessors, where reliability was lower, were less experienced practitioners from non-psychology backgrounds. Thus, future training should take prior experience into consideration and provide additional support for those with less experience. Furthermore, future research should explore what assessor training, characteristics, skills, and ongoing supports are required for reliably conducting facilitator assessments.

4.2 | Limitations and strengths

In addition to challenges delivering the training, this study had limitations. First, the study used a small sample (Koo & Li, 2016), necessitating caution in interpreting study results. Second, due to the study's limited scope, we used a purposive selection of sessions for assessment instead of random selection. This may have resulted in the assessment of nonrepresentative facilitator delivery (Walton et al., 2017). Third, missing data may have compromised results. Future revisions of the FAT should ensure that assessors complete all items using strategies such as highlighting the importance of completing all items in assessor training and establishing monitoring and evaluation processes wherein FAT forms are checked for completion.

Despite limitations, this study makes an important contribution to our understanding of the psychometric properties of an assessment tool widely used to assess facilitators of an evidence-based parenting programme in LMICs. While valuable in all contexts, the need for practical, reliable, and valid assessment tools in low-income contexts is particularly heightened. The study found sufficient intra- and inter-rater reliability despite challenges encountered during training.

However, the findings suggest that more attention should be paid to how reliability can be strengthened and explore whether those with certain backgrounds and skills would be better suited to conducting reliable assessments. For instance, future assessor training could require assessors to achieve an acceptable minimum level of intra- and inter-rater reliability prior to conducting FAT assessments in practice. Further, two common issues in studying observational measures were avoided. First, facilitator reactivity to assessment was minimised because all programme sessions were recorded, thereby decreasing the likelihood that facilitators performed differently for assessments (Kazdin, 1982). Second, assessors did not evaluate facilitators they supervised to ensure assessor independence from the results (Walton et al., 2017).

5 | CONCLUSION

The results of this study suggest that the Facilitator Assessment Tool appears to capture the competent adherence with which facilitators deliver Parenting for Lifelong Health for Young Children. Future research is necessary to strengthen the reliability of the measurement due to results suggesting sufficient though not high levels of intra- and inter-rater reliability. We also recommend further research on the FAT's psychometric properties. These include research on its internal consistency to determine whether Cronbach's alphas and omegas indicate that items are appropriately associated with each other, construct validity to explore whether the tool is measuring its intended constructs via exploratory factor analyses, and predictive validity to examine whether higher facilitator competent adherence scores are associated with improved family outcomes (Barchard, 2010; Markus & Lin, 2010; Mislevy & Rupp, 2010). Relatedly, given that over 3000 facilitators are delivering PLH programmes internationally across more than 25 countries, we recommend that the FAT's psychometric properties be examined across multiple contexts. We also recommend that similar analyses are conducted of the PLH programme for adolescents, which has similar delivery methods but involves different activities from PLH-YC. In summary, this study was an important first step in ensuring reliable assessments of programme delivery—a critical factor when monitoring the dissemination and scale-up of evidence-based interventions.

ACKNOWLEDGEMENTS

This study would not have been possible without the involvement and contributions of the three stakeholder groups consulted—trainers from Clowns Without Borders South Africa, parenting programme experts, and assessors and facilitators from North Macedonia, Romania, and Moldova.

CONFLICT OF INTEREST

MM, HM, and HMF declare that they have no competing interests. JML and FG are co-developers of PLH for Young Children (licensed under a Creative Commons 4.0 Non-commercial No Derivatives license) and, with colleagues, co-founders of the Parenting for Lifelong

Health initiative. JML receives occasional fees for providing training and supervision to facilitators and coaches. JML and FG have participated (and are participating) in a number of research studies on the programme as investigators and the University of Oxford receives research funding for these studies. Conflict is avoided by declaring potential conflicts and by conducting and disseminating rigorous, transparent, and impartial evaluation research on both PLH and other similar parenting programmes.

ETHICS STATEMENT

The research received approval from the University of Oxford's Department of Social Policy and Intervention Research Ethics Committee (Ref: SPIC1a_20_0004), the University of Klagenfurt Ethics Board of the Institute of Psychology (Ref: 2018-21), as well as by local country institutional review boards in North Macedonia (Ref: 03-1475-2), the Republic of Moldova (Ref: 43-56/12.04.018), and Romania (Ref: 322/1.03.2019).

DATA AVAILABILITY STATEMENT

The Parenting for Lifelong Health for Young Children-Facilitator Assessment Tool is included as supplementary material.

ORCID

Mackenzie Martin  <https://orcid.org/0000-0002-3202-3800>

REFERENCES

- Aspland, H., & Gardner, F. (2003). Observational measures of parent-child interaction: An introductory review. *Child and Adolescent Mental Health, 8*(3), 136–143. <https://doi.org/10.1111/1475-3588.00061>
- Barchard, K. (2010). Internal consistency reliability. In N. Salkind (Ed.), *Encyclopedia of research design*. SAGE Publications. <https://doi.org/10.4135/9781412961288>
- 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/10.1016/S0031-9406\(05\)61211-4](https://doi.org/10.1016/S0031-9406(05)61211-4)
- 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/10.1177/1524838014566718>
- Chen, P., & Krauss, A. (2004). *The SAGE encyclopedia of social science research methods*. SAGE Publications. <https://doi.org/10.4135/9781412950589>
- Cho, D. W. (1981). Inter-rater reliability: Intraclass correlation coefficients. *Educational and Psychological Measurement, 41*(1), 223–226. <https://doi.org/10.1177/001316448104100127>
- Eames, C., Daley, D., Hutchings, J., Hughes, J. C., 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. <https://doi.org/10.1111/j.1365-2214.2008.00828.x>
- 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. <https://doi.org/10.1016/j.brat.2010.07.011>
- Fixsen, D., Naoom, S., Blase, K., Friedman, R., & Wallace, F. (2005). *Implementation research: A synthesis of the literature*. University of South Florida, Louis de la Parte Florida Mental Health Institute, National Implementation Research Network. <https://nirn.fpg.unc.edu/sites/nirn.fpg.unc.edu/files/resources/NIRN-MonographFull-01-2005.pdf>

- 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. [https://doi.org/10.1016/S0005-7894\(05\)80049-8](https://doi.org/10.1016/S0005-7894(05)80049-8)
- Frantz, I., Foran, H. M., Lachman, J. M., Jansen, E., Hutchings, J., Băban, A., Fang, X., Gardner, F., Lesco, G., Raleva, M., Ward, C. L., Williams, M. E., & Heinrichs, N. (2019). Prevention of child mental health problems in Southeastern Europe: A multicentre sequential study to adapt, optimise and test the parenting programme 'Parenting for Lifelong Health for Young Children', protocol for stage 1, the feasibility study. *BMJ Open, 9*(1), e026684. <https://doi.org/10.1136/bmjopen-2018-026684>
- 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. <https://doi.org/10.1002/14651858.CD008225.pub2>
- Gamer, M., Lemon, J., Gamer, M. M., & Kendall, W. (2017). Package 'irr'.
- Gardner, F., McCoy, A., Lachman, J. M., Melendez-Torres, G. J., Ward, C., Cheah, P., & Topanya, S. (forthcoming). Randomized trial of a parenting intervention embedded within the public health system to reduce violence against children in Thailand.
- Girard, J. M., & Cohn, J. F. (2016). A primer on observational measurement. *Assessment, 23*(4), 404–413. <https://doi.org/10.1177/1073191116635807>
- Gwet, K. L. (2014). *Intrarater reliability*. Wiley StatsRef: Statistics Reference Online.
- Hallgren, K. A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorial in Quantitative Methods for Psychology, 8*(1), 23–34. <https://doi.org/10.20982/tqmp.08.1.p023>
- 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>
- Holtrop, K., Miller, D., Durtschi, J., & Forgatch, M. (2020). Development and evaluation of a component level implementation fidelity rating system for the GenerationPMTO intervention. *Prevention Science, 22*, 288–298. <https://doi.org/10.1007/s1121-020-01177-5>
- Kazdin, A. E. (1982). Observer effects: Reactivity of direct observation. *New Directions for Methodology of Social & Behavioral Science*.
- Knutson, N. M., Forgatch, M. S., Rains, L. A., Sigmarsdóttir, M., & Domenech Rodríguez, M. M. (2019). *Fidelity of implementation rating system (FIMP): The manual for GenerationPMTO* (3d ed.). [Unpublished training manual]. Implementation Sciences International, Inc.
- 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>
- 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, 17*, 100279. <https://doi.org/10.1016/j.lanwpc.2021.100279>
- Lachman, J. M., Cluver, L., Kelly, J., Ward, C. L., Hutchings, J., & Gardner, F. (2016). Process evaluation of a parenting program for low-income families in South Africa. *Research on Social Work Practice, 28*, 188–202. <https://doi.org/10.1177/1049731516645665>
- Lachman, J. M., Cluver, L., Ward, C. L., Hutchings, J., Mlotshwa, S., Wessels, I., & Gardner, F. (2017). Randomized controlled trial of a parenting program to reduce the risk of child maltreatment in South Africa. *Child Abuse & Neglect, 72*, 338–351. <https://doi.org/10.1016/j.chiabu.2017.08.014>
- Lachman, J. M., Heinrichs, N., Jansen, E., Brühl, A., Taut, D., Fang, X., Gardner, F., Hutchings, J., Ward, C. L., Williams, M. E., Raleva, M., Băban, A., Lesco, G., & Foran, H. M. (2019). Preventing child mental health problems through parenting interventions in Southeastern Europe (RISE): Protocol for a multi-country cluster randomized factorial study. *Contemporary Clinical Trials, 86*, 105855. <https://doi.org/10.1016/j.cct.2019.105855>
- 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. <https://doi.org/10.1007/s10826-016-0389-6>
- Margolin, G., Oliver, P. H., Gordis, E. B., O'hearn, H. G., Medina, A. M., Ghosh, C. M., & Morland, L. (1998). The nuts and bolts of behavioral observation of marital and family interaction. *Clinical Child and Family Psychology Review, 1*(4), 195–213. <https://doi.org/10.1023/A:1022608117322>
- Markus, K., & Lin, C. (2010). Construct validity. In N. Salkind (Ed.). *Encyclopedia of research design*. SAGE Publications. <https://doi.org/10.4135/9781412961288>
- Martin, M., Lachman, J., Wamoyi, J., Shenderovich, Y., Wambura, M., Mgunga, S., Ndyetabura, E., Ally, A., Barankena, A., Exavery, A., & Manjengenja, 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), 52. <https://doi.org/10.1186/s43058-021-00154-5>
- Martin, M., Steele, B., Lachman, J. M., & Gardner, F. (2021). Measures of facilitator competent adherence used in parenting programs and their psychometric properties: A systematic review. *Clinical Child and Family Psychology Review, 24*, 834–853.
- McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods, 1*(1), 30–46. <https://doi.org/10.1037/1082.989X.1.1.30>
- Michie, S., Wood, C. E., Johnston, M., Abraham, C., Francis, J. J., & Hardeman, W. (2015). Behaviour change techniques: The development and evaluation of a taxonomic method for reporting and describing behaviour change interventions (a suite of five studies involving consensus methods, randomised controlled trials and analysis of qualitative data). *Health Technology Assessment (Winchester), 19*(99), 1–188. <https://doi.org/10.3310/hta19990>
- Mislevy, J., & Rupp, A. (2010). Predictive validity. In N. Salkind (Ed.). *Encyclopedia of research design*. SAGE Publications. <https://doi.org/10.4135/9781412961288>
- Mokkink, L., Terwee, C. B., Patrick, D. L., Alonso, J., Stratford, P. W., Knol, D. L., Bouter, L. M., & de Vet, H. C. (2010). 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. <https://doi.org/10.1016/j.clinepi.2010.02.006>
- 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), 1–11. <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), 420–428. <https://doi.org/10.1037/0033-2909.86.2.420>
- Stemler, S., & Tsai, J. (2008). Best practices in interrater reliability: Three common approaches. In J. Osborne (Ed.). *Best practices in quantitative*

- methods (Series ed.). SAGE Publications. <https://doi.org/10.4135/9781412995627>
- 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. <https://doi.org/10.4324/9781410601261>
- 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>
- 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>
- Ward, C. L., Mikton, C., Cluver, L., Cooper, P., Gardner, F., Hutchings, J., McLaren Lachman, J., Muray, L., Tomlinson, M., Wessels, I. M. (2014). Parenting for Lifelong Health: From South Africa to other low-and middle-income countries. *Early childhood matters: Responsive parenting: A strategy to prevent violence*. <https://cafo.org/wp-content/uploads/2015/08/Parenting-for-Lifelong-Health1.pdf>
- Ward, C. L., Wessels, I. M., Lachman, J. M., Hutchings, J., Cluver, L. D., Kassarjee, 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.
- Watkins, M. W. (2018). Exploratory factor analysis: A guide to best practice. *Journal of Black Psychology*, 44(3), 219–246.
- 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>

How to cite this article: Martin, M., Lachman, J. M., Murphy, H., Gardner, F., & Foran, H. (2023). 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*, 49(3), 591–604. <https://doi.org/10.1111/cch.13075>

APPENDIX A: FACILITATOR ASSESSMENT TOOL

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

Assessor name:

Facilitator name:

Assessment date:

Video file name:

Number of enrolled parents:

Facilitator age:

Has the facilitator been assessed before (Y/N)?

Co-facilitator name:

Facilitator ID:

Session number and date:

Session/video length:

Number of parents in attendance:

Facilitator gender:

If yes, how many times has the facilitator been assessed previously?

Facilitator condition

ACTIVITY SUB-SCALE

HOME ACTIVITY

DISCUSSION

SESSION NUMBER: _____

START TIME ON VIDEO: _____

END TIME ON VIDEO: _____

Inadequate 0

Needs improvement 1

Good 2

Excellent 3

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 **OR** 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 SUB-SCALE

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

Behaviour	Frequency			
Reflexive Statements				
Praise	Unspecific:			Specific
ILLUSTRATED STORY DISCUSSION				
SESSION NUMBER: _____				
START TIME ON VIDEO: _____				
END TIME ON VIDEO: _____				
	Inadequate 0	Needs Improvement 1	Good 2	Excellent 3
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:				
GROUP PRACTICE				
SESSION NUMBER: _____				
START TIME ON VIDEO: _____				
END TIME ON VIDEO: _____				
	Inadequate 0	Needs improvement 1	Good 2	Excellent 3
BIG GROUP PRACTICE				
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 SUB-SCALE

	Inadequate 0	Needs Improvement 1	Meets Expectations 2	Exceeds Expectations 3
MODELLING BEHAVIOUR				
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:				
ACCEPT-EXPLORE-CONNECT-PRACTICE				
ACCEPT				
1. Uses body language to show acceptance				
2. Accepts parent responses verbally				
3. Openness				
4. Reflexive statements				
EXPLORE				
5. Explore experience/opinion of parent in detail using open-ended questions				
6. Explore thoughts and feelings				
7. Explore child perspective				
CONNECT				
8. Connect experiences to the building blocks from session				
9. Connects individual experiences to universal principles				
PRACTICE				
10. Identify opportunities to practice skills (in addition to the structured group practice)				
Comments:				
COLLABORATIVE LEADERSHIP				
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				

(Continues)

COLLABORATIVE LEADERSHIP	Inadequate 0	Needs Improvement 1	Meets Expectations 2	Exceeds Expectations 3
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)	78	Total possible score (D)	99
Total percent score core activities = $(A/B) \times 100\%$	%	Total percent score core skills = $(C/D) \times 100\%$	%
What are the facilitator's strengths?			
What does the facilitator need to improve?			
Recommendations:			