UvA-DARE (Digital Academic Repository)

Assessment of service provider competency for child and adolescent psychological treatments and psychosocial services in global mental health: Evaluation of feasibility and reliability of the WeACT tool in Gaza, Palestine

Jordans, M.J.D.; Coetzee, A.; Steen, H.F.; Koppenol-Gonzalez, G.V.; Galayini, H.; Diab, S.Y.; Aisha, S.A.; Kohrt, B.A.

Link to publication

Interventions
Original Research Paper

CAMBRIDGE
UNIVERSITY PRESS

Assessment of service provider competency for child and adolescent psychological treatments and psychosocial services in global mental health: evaluation of feasibility and reliability of the WeACT tool in Gaza, Palestine

M. J. D. Jordans[1,2] [iD], A. Coetzee[1,2], H. F. Steen[1], G. V. Koppenol-Gonzalez[1], H. Galayini[3], S. Y. Diab[3], S. A. Aisha[3] and B. A. Kohrt[4] [iD]

[1]Research and Development Department, War Child Holland, Amsterdam, The Netherlands; [2]Amsterdam Institute of Social Science Research, University of Amsterdam, Amsterdam, The Netherlands; [3]War Child Holland, Occupied Palestinian Territory, Gaza, State of Palestine and [4]Department of Psychiatry, George Washington University, Washington, DC, USA

Abstract

Background. There is a scarcity of evaluated tools to assess whether non-specialist providers achieve minimum levels of competency to effectively and safely deliver psychological interventions in low- and middle-income countries. The objective of this study was to evaluate the reliability and utility of the newly developed Working with children – Assessment of Competencies Tool (WeACT) to assess service providers' competencies in Gaza, Palestine.

Methods. The study evaluated; (1) psychometric properties of the WeACT based on observed role-plays by trainers/supervisors ($N = 8$); (2) sensitivity to change among service provider competencies ($N = 25$) using pre-and-post training WeACT scores on standardized role-plays; (3) in-service competencies among experienced service providers ($N = 64$) using standardized role-plays.

Results. We demonstrated moderate interrater reliability [intraclass correlation coefficient, single measures, ICC = 0.68 (95% CI 0.48–0.86)] after practice, with high internal consistency ($\alpha = 0.94$). WeACT assessments provided clinically relevant information on achieved levels of competencies (55% of the competencies were scored as adequate pre-training; 71% post-training; 62% in-service). Pre-post training assessment saw significant improvement in competencies ($W = -3.64$; $p < 0.001$).

Conclusion. This study demonstrated positive results on the reliability and utility of the WeACT, with sufficient inter-rater agreement, excellent internal consistency, sensitivity to assess change, and providing insight needs for remedial training. The WeACT holds promise as a tool for monitoring quality of care when implementing evidence-based care at scale.

Introduction

In low- and middle-income countries (LMIC), and especially in humanitarian settings, there is a serious lack of specialized service providers to attend to the increasing mental health and psychosocial needs of children and adolescents (Patel *et al.*, 2018). This has resulted in adopting a task-sharing approach, which entails that non-specialists perform certain service provision tasks that are traditionally reserved for specialists (Patel *et al.*, 2007). Given that training of non-professionals of psychological interventions is commonly one or a couple of weeks at most, it is essential to establish the quality and evidence for task-shared services (Singla *et al.*, 2017; Kohrt *et al.*, 2018a). The evidence of psychological interventions following a task-shifting approach has been synthesized in several recent reviews (Chowdhary *et al.*, 2014; Singla *et al.*, 2017; Purgato *et al.*, 2018). The evidence base for interventions for children is still slim, with only substantial evidence for the treatment of PTSD in humanitarian settings and conduct disorders across all LMIC (Barbui *et al.*, 2020). Even less is known about potential contributors to, and mechanisms of, change. Moreover, when evidence of effectiveness has been established and the evidence-base for task-shifted interventions is stronger, the subsequent question is whether the quality of care (including the assessment of potential harm) be maintained outside the confines of controlled research projects.

Following the paradigm of evidence-based task-shifted care, scaling of services requires a cadre of non-specialists that have the skills and competencies to deliver the empirically supported interventions with sufficient quality in real-world settings (Jordans and Kohrt, 2020). This calls for the importance of the measurement of service providers' skills and competencies. Provider competence can be defined as the extent to which a service provider has the

CrossMark

knowledge and skill required to deliver an intervention to the standard needed for it to achieve its expected effects (Fairburn and Cooper, 2011). Brief and validated tools to measure competencies for interventions in LMIC are scarce (Kohrt *et al.*, 2015*b*). To respond to this need, the 18-item ENhancing Assessment of Common Therapeutic factors (ENACT) rating scale was developed for the assessment of common competencies across psychological treatments (Kohrt *et al.*, 2015*a*). The ENACT was developed for adults interventions and focusing specifically on mental health interventions. Based on the ENACT, we developed the Working with children-Assessment of Competencies Tool (WeACT) instrument with the aim of assessing common competencies of service providers working with child and adolescent interventions across different sectors (mental health and psychosocial support, child protection and education). The aim of the tool is to assess whether the level of competencies of service providers is adequate, as well as to detect potential harm, when working with children and adolescents. The aim of the current study is to evaluate the feasibility, utility, and performance of the WeACT instrument when applied to the evaluation of training of facilitators for, and implementation of, a psychosocial intervention in Palestine.

Methods

Setting

The study was conducted in Gaza strip, Palestine, which is characterized by decades of violence and economic hardship. The Palestinian territories, which include the Gaza strip, have been under Israeli occupation since 1967. Children and adolescents are exposed to high levels of traumatic events and/or have chosen to participate in resistance activities, and as a result, experience high levels of psychological distress and child maltreatment (Punamaki *et al.*, 2005; Qouta *et al.*, 2008).

War Child Holland is an international organization that has been providing services in Palestine since 2004. This study was implemented with facilitators of the IDEAL intervention, a life-skills group intervention being implemented by War Child Holland in Gaza strip of Palestine. The intervention has been implemented in Gaza for several years, and was adapted to the cultural context at the time it was introduced there. IDEAL is a psychosocial group intervention for adolescents between 11 and 15 years, consisting of 16 sessions, organized around the following themes; Identity and Assessment; Dealing with Emotions; Peer Relations; Relations with Adults; Rights and Responsibilities; Prejudice and Stigmatization and Future. Facilitators for the IDEAL intervention received a 1-week workshop-based training, which includes topics such as child wellbeing and development, the aims and structure of I-Deal, group facilitation and management skills, child safeguarding, as well as the practice of all sessions (Miller *et al.*, 2020).

Design

The current study consisted of multiple components to evaluate the feasibility, utility, and performance of the WeACT instrument. The study included three study components assessing; (1) psychometric properties [interrater reliability (IRR) and internal consistency] of WeACT scores rated using videotaped role-plays developed for the purpose of assessing the competencies included in the WeACT; (2) WeACT's sensitivity to change through

measurement of facilitators' competencies before and after a manualized psychosocial intervention training; and (3) in-service competencies among a separate group experienced facilitators. For components 2 and 3, WeACT was scored by the same trained raters who observed standardized role-plays of facilitators in mock sessions with trained youth actors.

Sample

For the evaluation of psychometric properties (e.g. inter-rater reliability), we recruited trainers/supervisors to be trained using WeACT for competency rating; the trainers/supervisors were from within the implementing organizations with over 3 years of experience and familiarity with the IDEAL intervention ($N = 8$; 75% female; mean age = 29.5; mean years of supervising experience = 1.5). For the evaluation of sensitivity to change, new facilitators who were being trained in the IDEAL intervention ($N = 25$; 76% female; mean age = 24.6) were selected at random from a group of trainees recruited to be trained as psychosocial service providers for ongoing programs by War Child. Inclusion criteria for this training were candidates with limited prior experience in delivering psychosocial services, aged over 21 and affinity to work with children. For the assessment of in-service competencies, respondents were a different group of IDEAL facilitators ($N = 64$; 64% female; mean age = 28.4), currently involved in implementing the intervention in ongoing programs in the Gaza strip. The facilitators were selected based on having over 1 year of experience in implementing IDEAL or other psychosocial interventions (average years of experience = 5.0).

Instruments

The WeACT was developed in the context of a larger research program that aims to develop an evidence-based system of care including mental health and psychosocial, child protection and educational services (Jordans *et al.*, 2018). After a multi-step process of development and field testing, which involved consultation with Palestinian and global experts, a final consolidated version including 14 unique items was generated. The final tool incorporated a four-tiered response system 1 (potential harm), 2 (absence of competency), 3 (done partially), and 4 (mastery) (see Supplementary material for full version), with a short description of observable behaviors associated to each score. The WeACT has some items that overlap with the ENACT (e.g. empathy, verbal and non-verbal communication), after which it was modeled, though the content of these items has been tailored toward interactions with children and adolescents. Additionally, there are items that are specific for facilitators implementing interventions with children and adolescents (e.g. collaboration with children's caregivers, meaningful participation, and detection of child abuse). The WeACT is developed to be used in enhancing the quality of care, especially for task-shifted interventions (i.e. including mental health and psychosocial, child protection and educational services) across LMIC settings.

Procedures

We developed standardized role-plays and videos, with pre-designed scripts indicating the level of competencies of the responses (fixed for videos and variable for role-plays). Actors from a theatre group in Gaza were trained in conducting the role-plays, taking the role of a group of children and a caregiver, and

Table 1. Internal consistency and inter-rater reliability of the WeACT

	Time point 1 ($N = 8$)		Time point 2 ($N = 6$)	
	ICC (95% CI)	Cronbach's α	ICC (95% CI)	Cronbach's α
Total video	0.468 (0.263–0.720)	0.912	0.681 (0.479–0.858)	0.937
Group video	0.510 (0.257–0.810)	0.934	0.569 (0.302–0.845)	0.909
Individual video	0.407 (0.129–0.866)	0.888	0.753 (0.444–0.964)	0.951

ICC, intraclass correlation coefficient; CI, confidence interval.

for the videos an experienced IDEAL facilitator for taking the role of service provider. Separate role-plays were developed for group and individual settings, in order to demonstrate different sets of competencies, covering demonstration of all 14 WeACT competencies. (1) For the assessment of the psychometrics (IRR and alpha), the trainers/supervisors (i.e. 'raters') received (a) 3-day competency rater training in the rationale and use of the WeACT; (b) live role-play experience assessments using the tool on assessing competencies of a group of experienced psychosocial service providers; and (c) 1-day recap training. The raters viewed the videotaped role-play and rated the 14 WeACT items at two time-points. First, immediately after the training (time-point 1), and again after the live role-plays and recap training (time point 2) (see below under 3). We opted for using standardized role-plays instead of a video-taped session because the former allows for control of the content, and in turn ensure that behaviors relevant to all of the WeACT competencies are covered, while keeping it relatively brief, something that cannot be assured with a non-scripted video-taped session. (2) The group of newly recruited facilitators received a 5-day training for the IDEAL intervention, following a standard training curriculum. Before and after, the training time was added for the assessment of competencies through conducting structured role-plays with the trained theatre group rated by the same trainers/supervisors (i.e. 'raters') (post time 1 IRR evaluation). (3) Similarly, the separate group of experienced IDEAL facilitators followed the same procedure, now through a 1-day workshop.

Analyses

We assessed the IRR for the trainers/supervisors at the two time points, after the WeACT training and after the 1-day recap training, using standardized videotaped role-plays. We used their WeACT scoring to calculate the intraclass correlation coefficient (ICC), assuming a two-way mixed-effects model, single measures, and absolute agreement. We chose to use single measures ICC because in real-world settings it will be more feasible if only one trainer or supervisor were rating a provider. Average measures ICC scores would be higher, but would require multiple raters. Therefore, single measures ICC provides the most conservative estimate for pragmatic implementation. The internal consistency was calculated by means of Cronbach's α. We conducted descriptive analyses of the facilitators' level of competencies (pre and post training among new facilitators; in-service among experienced facilitators), presented as the percentage of scores of levels 3 and 4 combined (per item and total). We ran Wilcoxon signed-rank (W) tests to evaluate pre-to-post training changes in levels of competencies for the newly trained facilitators. The levels 1–4 were analyzed as four separate variables reflecting the count of WeACT items per facilitator scored as potentially harmful,

absence of competency, partial competency, and mastery. Additionally, the levels 1 and 2, and the levels 3 and 4 were analyzed as two separate variables reflecting the count of WeACT items per facilitator scored as not competent (levels 1 and 2 combined) and competent (levels 3 and 4 combined). Both approaches lead to outcome scores of counts ranging from 0 to 14.

Ethics

Ethical approval was obtained from the Palestinian Health Research Council (PHRC/HC/424/18; 8 October 2018). Informed written consent was obtained from all participants. None of the participants received compensation for participation in the study. Referral services were available in case interviewed staff or actors were distressed during the study activities.

Results

The psychometric properties, IRR and Cronbach's α, of the WeACT are summarized in Table 1. The IRR ranged from ICC = 0.407 (95% CI 0.13–0.87, $N = 8$ raters) for the video using an individual client in the role-play at time one to ICC = 0.753 (95% CI 0.44–0.96, $N = 6$ raters) for the same video at time two. The IRR for the first timepoint was significantly influenced by one rater with clearly deviating scores – when excluding this rater (CR04), the ICCs improved slightly (ICC = 0.555 for total, 95% CI 0.35–0.78; ICC = 0.623 for group video, 95% CI 0.37–0.87; ICC = 0.472 for individual video, 95% CI 0.16–0.89 – results not included in Table 1). A second rater was also excluded from the second timepoint as they could not come to the recap training. Internal consistency was above $\alpha = 0.85$ for all assessments.

Figure 1 shows the WeACT item-by-item changes from before to after the training for the new facilitators, based on proportion scores of facilitators showing not-competent levels (levels 1 and 2) v. competent levels (levels 3 and 4). On all but two competencies (86%), we saw an improvement over time, with group- and behavior-management items reducing slightly. For eight out of the 14 competencies, more than 80% of the facilitators showed adequate competency after receiving the training, up from five out of 14 before the training. For four out of 14 competencies, fewer than 50% of the participants demonstrated adequacy after training (seven out of 14 before).

The before to after training comparison in overall adequate competency, i.e. the shift in the number of items reflecting competent levels (levels 3, 4) from before to after training, showed that before training the average number of items with a score of at least 3 was 7.8 and after training this average increased to 10 items ($W = -3.64$, $p < 0.001$). Additional testing of the distribution of the 1–4 scores as four separate variables showed that this overall difference could be mainly attributed to a shift from scores 2 (absence
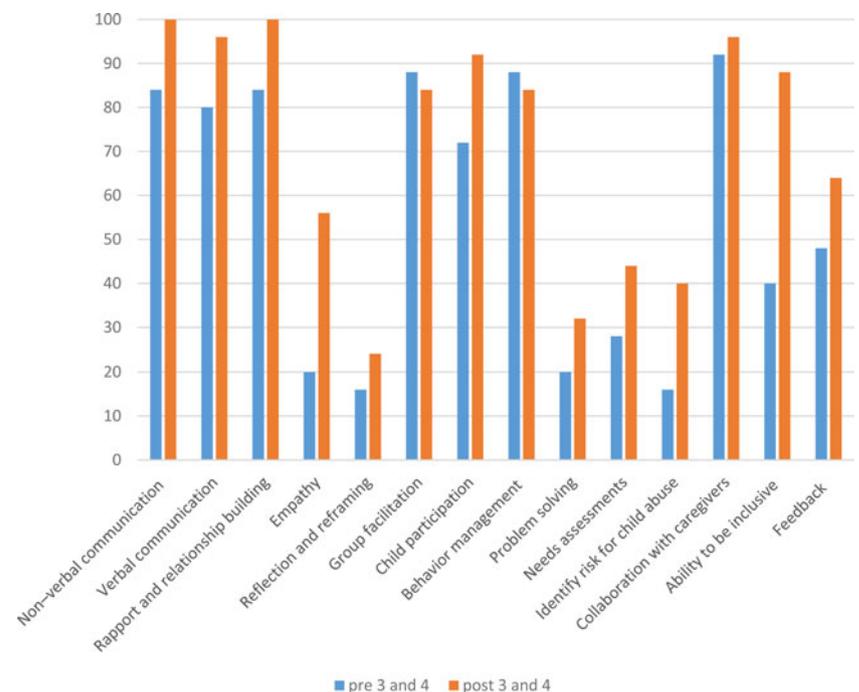
Fig. 1. Results of before and after training assessment of competencies among new service providers (*N* = 25).
Note: % of facilitators scoring ⩾3 per competency item.

■ pre 3 and 4   ■ post 3 and 4

of competency) to 3 (partial competency) from before to after the training. Before the training, the averages were 5.8 items with a score 2 and 7.6 items with a score 3 (*W* = −1.78, *p* = 0.08), and after the training, this shifted to 4 items with a score 2 and 9.4 items with a score 3 (*W* = −3.95, *p* < 0.001). The total number of level 1 scores (potentially harmful) reduced from 10 (3%) before to 0 after the training. When comparing the proportion of observations scored as adequate competency (scored as level 3 or 4) combined for all competencies, we saw an increase from 55% (*N* = 194 scores) to 71% (*N* = 250 scores) from before to after training.

The pattern of scoring for each of the WeACT items among a group of facilitators that have several years of experience in implementing psychosocial interventions, and that were assessed in the midst of implementing the IDEAL intervention, is presented in Fig. 2. Within this sample on three out of the 14 indicators (21%), 80% were assessed as competent, with two items only just below. On five out of the 14 items (36%), there were fewer than 50% of facilitators being rated as competent. These were the same five items that were scored with lower levels of competence in pre-post evaluations of the new facilitators. The proportion of observations scored as adequate (scored as level 3 or 4) combined for all competencies was 62% (*N* = 559 scores).

Discussion

In recent years, we have seen increasing attention for the mental health needs in LMIC (Patel *et al.*, 2018). This includes increasing availability of evidence-based psychological treatments, as well as addressing social determinants of mental health (Lund *et al.*, 2018), which for children includes interventions that address the school and family setting and child protection needs. As evidence for such interventions in LMIC is accruing, attention needs to shift to how empirically supported care can be implemented at scale with sufficient quality by non-specialist providers. In line with the evidence-based care paradigm, we argue that monitoring

levels of competencies, adherence and attendance of the implementation of manualized services, are indicators of quality of care that in turn contributes to positive outcomes (Jordans *et al.*, 2018; Jordans and Kohrt, 2020). A recent systematic review and meta-analysis shows that there is a small but significant association between adherence and outcomes in psychotherapy for children and adolescents, yet such association is not found for competence (Collyer *et al.*, 2019). While seemingly contradictory to our hypothesis, it is important to note that the studies included in the review were mostly not designed to evaluate that association and they all assessed treatment-specific competencies, in addition few used standardized role-plays, and most used linear correlation to test the association (Ottman *et al.*, 2020). There are compelling studies that argue for the importance of common factors in psychotherapy (e.g. therapeutic alliance), rather than treatment-specific ones, in contributing to treatment outcomes (Kazdin and Durbin, 2012). With the current study, we are therefore looking into the common, non-specific, competencies of non-specialist service providers – as one aspect of monitoring quality of care especially in LMIC contexts. In a recent systematic review of the literature on interventions addressing the mental health of children in conflict-affected settings, only three of the included publications (13%) address treatment quality in the evaluations, and two specifically address the delivery agent's relationship with the participant as a mechanism contributing to outcomes (Jordans *et al.*, 2016). In an effort to systematically integrate a focus on the quality of care when implementing evidence-based services for children, we developed the WeACT instrument that assesses service providers' common competencies across different sectors, based on the rationale and process for the development of the ENACT for adult-focused mental health services (Kohrt *et al.*, 2015a).

The current study evaluated the feasibility and utility of the WeACT when applied to the evaluation of training of facilitators for a psychosocial intervention in Palestine. First, the psychometric properties range from adequate to good. We demonstrated that
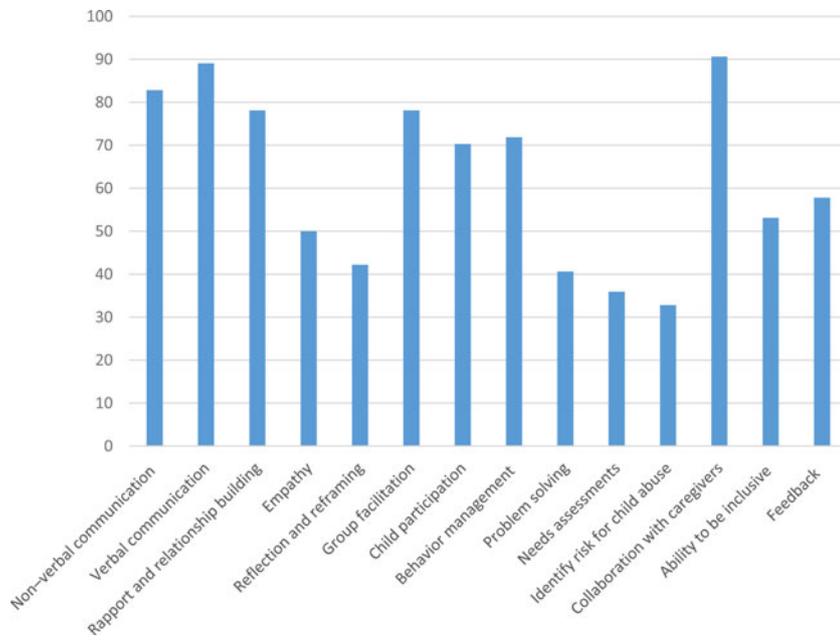
Fig. 2. Results of in-service assessment of competencies among experienced service providers ($N = 64$).
Note: % of facilitators scoring $\geqslant 3$ per competency item.

the group of supervisors using the WeACT to rate standardized video-taped role-plays achieved moderate interrater reliability after some practice of using the tool [based on the interpretation of ICCs by Koo and Li (2016)]. IRR is important for widespread use of the instrument by a group of supervisors. The internal consistency was excellent. Second, we were able to show the instruments' sensitivity to assessing change over time. Based on overall instrument scores, as well as item-level patterns of change, we saw significant changes in the WeACT scores on observed standardized role-plays when comparing before and after training scores. It is important to note that the extent of change, or the lack of change on some of the items, is not necessarily an indicator of the quality of the instrument, but rather a reflection of the change in competencies achieved through the training. Overall, facilitators scored less on the absence of competency and more on adequate competency after compared to before the training. Third, the utility is established by the range and pattern of WeACT scores demonstrated in the pre-post assessment, as well as during the in-service assessment of competencies of the service providers. A (consistent) pattern of high or low scores on specific competencies provides inputs for where more training or supervision is needed to attain adequate competency or mastery. Similarly, demonstrating the lack of change on a number of competence items (i.e. empathy, reframing, problem solving, needs assessment, detecting child abuse) in the pre-post assessment is exactly revealing the purpose and utility – these would be areas of competence development that require more attention. A lack of improvement on some of these competencies can be explained by the fact that the standard IDEAL training did not explicitly address these competencies (e.g. problem solving, needs assessment) and/or that these competencies require more specific training (e.g. reframing, empathy). Indeed, we argue that training courses for specific interventions should be preceded by a foundational training on the common competencies, ahead of learning the intervention-specific competencies. Moreover, in an effort to reduce the risk of doing harm, the utility is further shown by an analysis of the number of times participants were rated as a level 1 on any of the competencies, which reduced to zero after receiving training.

The primary application of WeACT is to maintain or improve the quality of care for child-focused services in LMIC and humanitarian settings, especially among service providers that often receive brief training to take on task-sharing roles. The WeACT can be used in different ways. (1) Selection of potential trainees through observed role-plays to identify candidates with higher entry-level competencies, or indeed to exclude those that show potentially harmful behaviors (level 1 scores). (2) During training or post-training (i.e. supervision) for improvement and remediation of competencies that are scored low. Importantly, this shifts the approach to capacity strengthening from delivery of standard packages to tailor-making capacity strengthening based on ongoing assessments of participants' actual competencies. With nearly 30% of the observation scored as inadequate (levels 1 for harmful or 2 for not done) post training, and an increase of 'only' 16% of observations scored as adequate (levels 3 and 4), there appears to be sufficient scope for capacity strengthening models that emphasize competency acquisition, while respecting the briefness of training as a result of limited resources. The benefits of competency-based capacity strengthening have been demonstrated in LMIC settings (Ameh *et al.*, 2016; McCullough *et al.*, 2018). This shift is aligned with a recently launched WHO initiative called EQUIP (https://www.who.int/mental_health/emergencies/equip/en/), which aims to ensure quality in psychological services through the use of competency tools (Kohrt *et al.*, 2020). (3) Post-training evaluation to assess readiness for implementation ensuring that helpers providing psychological interventions or psychosocial support meet a minimum level of competency. This is especially relevant as previous studies have demonstrated a positive relationship between competence and patient outcomes, albeit for specific conditions (Ginzburg *et al.*, 2012), and the use of competence rating in an applied setting (Strunk *et al.*, 2010). At the same time, the results are mixed for child and adolescent intervention studies (Webb *et al.*, 2010; Collyer *et al.*, 2019). One possible explanation of the absence of a significant association is the minimal variance in levels of competence and adherence demonstrated by service providers in the studies included in these reviews; following this line of thinking, there might be a threshold level of competence

whereby optimal outcomes are achieved, but with little additional intervention gains beyond the threshold level (Ottman *et al.*, 2020). The review by Collyer *et al.* describes a study that demonstrates that after a threshold of 60–80% of adherence had been reached, higher levels did not result in better outcomes (Durlak and DuPre, 2008). In fact, the association might even be curvilinear, which suggests that too much rigidity in obtaining adherence limits the necessity for clinical judgement, and too low levels equate insufficient quality (Barber *et al.*, 2006). Future studies should investigate these premises for common competencies and possibly validate minimum thresholds of competence using the WeACT.

All of the above-mentioned functions are important tools for the monitoring of the quality of child-focused interventions when transitioning from effectiveness studies to scaling up of evidence-based care, especially as no quality assurance frameworks are currently in place in most LMICs implementing task-shared models of care (Kohrt *et al.*, 2020). For future use, a couple of recommendations can be drawn from the current study. Especially when using multiple raters, establishing IRR is essential. This study demonstrates that adequate reliability can be attained, but also that practice in using the tool appeared to be needed to increase the quality of their ratings and achieve sufficient interrater agreement. Consequently, for future use, it is important to invest sufficient time and resources in training and supervising new expert raters, including systematic live practice. Also, working with a theatre group helped in making the role-plays dynamic and authentic, but also required significant preparation time to achieve sufficient standardization.

In order to achieve real-world improvements in the lives and mental health of children in LMIC, evidence-based intervention has to be taken up into routine practice and care. The way to do that remains largely unclear (De Silva and Ryan, 2016). Integrating the monitoring of levels of competency of service providers, combined with other quality indicators such as adherence and attendance, can become important tools to promote that systematic uptake of evidence-based care at scale. This is in line with increasing attention for implementation science to understand and address the barriers to implementing an intervention in routine practice (Proctor *et al.*, 2009).

There are a number of limitations to be mentioned. First, the IRR was assessed using a single video. While the ICC for single measures is an adequate indicator, having raters use different videos would generate more robust evidence. Related, as we only had developed one, we had to use the same video for the second testing of the IRR, which might have introduced a bias as a result of a learning effect. At the same time, we believe the learning effect to be small as significant time had passed between the first and second time-point, and their scoring of, or the correct scores for, the first observations was not discussed among the group of raters. Furthermore, the presence of raters when observing the role-plays may have impacted the performance of the facilitators, either positively or negatively. The sample size of raters ($n = 8$) is comparable to the number of staff who would be expected of organizational staff to be trained to be raters. In most local humanitarian mental health and psychosocial support service organizations, we would anticipate 5–10 as a feasible number of staff to receive WeACT rating training and achieve adequate inter-rater reliability. This is similar to the number of raters trained for organizations conducting ENACT competency rating with adults (Kohrt *et al.*, 2018*b*).

Future studies should expand the research into the WeACT to assess whether the level of competency is associated with, or

mediating, the intervention outcomes. Furthermore, the current study relied entirely on ratings done by supervisors ('experts'). We are interested to evaluate the feasibility and utility of using peer, self, or beneficiary ratings, as these would be more feasible for use at scale (Singla *et al.*, 2014).

Conclusion

There is an urgent need for competent service providers to address the mental health needs of children in LMIC and humanitarian settings. Current programs implementing psychosocial and mental health, child protection, or school-based services rely on briefly trained non-specialists to provide care, without much attention to monitoring the quality of these services. This study evaluated the feasibility and utility of the WeACT instrument, which is developed to assess the levels of common competency among service providers. We demonstrated that the tool, which relies on rating standardized role-plays to rate different levels of competency, can be used by multiple raters with sufficient inter-rater reliability, has excellent internal consistency, is able to pick up on improvement or absence of improvement on specific competencies, and can provide insight into which competencies service providers do well and which ones deserve remedial training. The WeACT holds promise as a tool for monitoring one aspect of quality of care when implementing evidence-based care at scale, though further cross-cultural testing of the tool and comparison with client outcomes are required.

Conflict of interest. None.

Ethical standards. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008. The authors assert that all procedures contributing to this work comply with the ethical standards of the relevant national and institutional guides on the care and use of laboratory animals.

References

Ameh CA, Kerr R, Madaj B, Mdegela M, Kana T, Jones S, Lambert J, Dickinson F, White S and van den Broek N (2016) Knowledge and skills of healthcare providers in sub-Saharan Africa and Asia before and after competency-based training in emergency obstetric and early newborn care. *PLoS ONE* **11**, e0167270.

Barber JP, Gallop R, Crits-Christoph P, Frank A, Thase ME, Weiss RD and Connoly Gibbons MB (2006) The role of therapist adherence, therapist competence, and alliance in predicting outcome of individual drug counseling: results from the National Institute Drug Abuse Collaborative Cocaine Treatment study. *Psychotherapy Research* **16**, 229–240.

Barbui C, Purgato M, Abdulmalik J, Acarturk C, Eaton J, Gastaldon C, Gureje O, Hanlon C, Jordans MJD, Lund C, Nosè M, Ostuzzi G, Papola D, Tedeschi F, Tol WA, Turrini G, Patel V and Thornicroft G (2020) Efficacy of psychosocial interventions for mental health outcomes in low-income and middle-income countries: an umbrella review. *The Lancet Psychiatry* **7**, 162–172.

Chowdhary N, Sikander S, Atif N, Singh N, Ahmad I, Fuhr DC, Rahman A and Patel V (2014) The content and delivery of psychological interventions

for perinatal depression by non-specialist health workers in low and middle income countries: a systematic review. *Best Practice & Research Clinical Obstetrics & Gynaecology* 28, 113–133.

Collyer H, Eisler I and 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* 29, 419–431.

De Silva MJ and Ryan G (2016) Global mental health in 2015: 95% implementation. *The Lancet Psychiatry* 3, 15–17.

Durlak JA and DuPre EP (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, 327–350.

Fairburn CG and Cooper Z (2011) Therapist competence, therapy quality, and therapist training. *Behaviour Research and Therapy* 49, 373–378.

Ginzburg DM, Bohn C, Höfling V, Weck F, Clark DM and Stangier U (2012) Treatment specific competence predicts outcome in cognitive therapy for social anxiety disorder. *Behaviour Research and Therapy* 50, 747–752.

Jordans MJD and Kohrt BA (2020) Scaling up mental health care and psychosocial support in low-resource settings: a roadmap to impact. *Epidemiology and Psychiatric Sciences* 29, e189, 1–7.

Jordans MJD, Pigott H and Tol WA (2016) Interventions for children affected by armed conflict: a systematic review of mental health and psychosocial support in low-and middle-income countries. *Current Psychiatry Reports* 18, 1–15.

Jordans MJD, van den Broek M, Brown F, Coetzee A, Ellermeijer REC, Hartog K, Steen HF and Miller KE (2018) Supporting children affected by War. In A Nickerson, Morina N (eds), *Mental Health in Refugee and Post Conflict Populations: Theory, Research and Clinical Practice*. Amsterdam: Springer, pp. 261–281.

Kazdin AE and Durbin KA (2012) Predictors of child-therapist alliance in cognitive-behavioral treatment of children referred for oppositional and antisocial behavior. *Psychotherapy* 49, 202–217.

Kohrt BA, Asher L, Bhardwaj A, Fazel M, Jordans MJD, Mutamba BB, Nadkarni A, Pedersen GA, Singla DR and Patel V (2018*a*) The role of communities in mental health care in low-and middle-income countries: a meta-review of components and competencies. *International Journal of Environmental Research and Public Health* 15, 1279.

Kohrt BA, Jordans MJD, Rai S, Shrestha P, Luitel NP, Ramaiya MK, Singla DR and Patel V (2015*a*) Therapist competence in global mental health: development of the ENhancing Assessment of Common Therapeutic factors (ENACT) rating scale. *Behaviour Research and Therapy* 69, 11–21.

Kohrt BA, Mutamba BB, Luitel NP, Gwaikolo W, Onyango P, Nakku J, Rose K, Cooper J, Jordans MJD and Baingana F (2018*b*) How competent are non-specialists trained to integrate mental health services in primary care? Global health perspectives from Uganda, Liberia, and Nepal. *International Review of Psychiatry* 30, 182–198.

Kohrt BA, Ramaiya MK, Rai S, Bhardwaj A and Jordans MJD (2015*b*) Development of a scoring system for non-specialist ratings of clinical competence in global mental health: a qualitative process evaluation of the Enhancing Assessment of Common Therapeutic Factors (ENACT) scale. *Global Mental Health* 2, e23, 1–16. doi: 10.1017/gmh.2015.21.

Kohrt BA, Schafer A, Willhoite A, van't Hof E, Pedersen GA, Watts S, Ottman K, Carswell K and van Ommeren M (2020) Ensuring Quality in Psychological Support (WHO EQUIP): developing a competent global workforce. *World Psychiatry* 19, 115.

Koo TK and Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine* 15, 155–163.

Lund C, Brooke-Sumner C, Baingana F, Baron EC, Breuer E, Chandra P, Haushofer J, Herrman H, Jordans M, Kieling C, Medina-Mora ME, Morgan E, Omigbodun O, Tol WA, Patel V and Saxena S (2018) Social determinants of mental disorders and the Sustainable Development Goals: a systematic review of reviews. *The Lancet Psychiatry* 5, 357–369.

McCullough M, Campbell A, Siu A, Durnwald L, Kumar S, Magee WP and Swanson J (2018) Competency-based education in low resource settings: development of a novel surgical training program. *World Journal of Surgery* 42, 646–651.

Miller KE, Koppenol-Gonzalez GV, Jawad A, Steen F, Sassine M and Jordans MJD (2020) A randomised controlled trial of the I-deal life skills intervention with Syrian refugee adolescents in Northern Lebanon. *Intervention* 18, 119.

Ottman KE, Kohrt BA, Pedersen GA and Schafer A (2020) Use of role plays to assess therapist competency and its association with client outcomes in psychological interventions: a scoping review and competency research agenda. *Behaviour Research and Therapy* 130, 103531. doi: 10.1016/j.brat.2019.103531.

Patel V, Araya R, Chatterjee S, Chisholm D, Cohen A, De Silva M, Hosman C, McGuire H, Rojas G and van Ommeren M (2007) Treatment and prevention of mental disorders in low-income and middle-income countries. *Lancet* 370, 991–1005.

Patel V, Saxena S, Lund C, Thornicroft G, Baingana F, Bolton P, Chisholm D, Collins PY, Cooper JL, Eaton J, Hermann H, Herzallah M, Huang Y, Jordans MJD, Kleinman A, Medina Mora ME, Morgan E, Niaz U, Omigbodun O, Prince M, Rahman A, Saraceno B, Sarkar K, De Silva M, Singh I, Stein DJ, Sunkel C and Unutzer J (2018) The Lancet Commission on global mental health and sustainable development. *The Lancet* 392, 1553–1598.

Proctor EK, Landsverk J, Aarons G, Chambers D, Glisson C and Mittman B (2009) Implementation research in mental health services: an emerging science with conceptual, methodological, and training challenges. *Administration and Policy in Mental Health and Mental Health Services Research* 36, 24–34.

Punamaki RL, Komproe I, Qouta S, El Masri M and de Jong JTVM (2005) The deterioration and mobilization effects of trauma on social support: childhood maltreatment and adulthood military violence in a Palestinian community sample. *Child Abuse and Neglect* 29, 351–373.

Purgato M, Gross AL, Betancourt T, Bolton P, Bonetto C, Gastaldon C, Gordon J, O'Callaghan P, Papola D, Peltonen K, Punamaki RL, Richards J, Staples JK, Unterhitzenberger J, van Ommeren MH, de Jong JTVM, Jordans MJD, Tol WA and Barbui C (2018) Focused psychosocial interventions for children in low-resource humanitarian settings: a systematic review and individual participant data meta-analysis. *The Lancet Global Health* 6, e390–e400.

Qouta S, Punamaki RL and El Sarraj E (2008) Child development and family mental health in war and military violence: the Palestinian experience. *International Journal of Behavioral Development* 32, 310–321.

Singla DR, Kohrt BA, Murray LK, Anand A, Chorpita BF and Patel V (2017) Psychological treatments for the world: lessons from low-and middle-income countries. *Annual Review of Clinical Psychology* 13, 149–181.

Singla DR, Weobong B, Nadkarni A, Chowdhary N, Shinde S, Anand A, Fairburn CG, Dimijdan S, Velleman R and Weiss H (2014) Improving the scalability of psychological treatments in developing countries: an evaluation of peer-led therapy quality assessment in Goa, India. *Behaviour Research and Therapy* 60, 53–59.

Strunk DR, Brotman MA, DeRubeis RJ and Hollon SD (2010) Therapist competence in cognitive therapy for depression: predicting subsequent symptom change. *Journal of Consulting and Clinical Psychology* 78, 429.

Webb CA, DeRubeis RJ and Barber JP (2010) Therapist adherence/competence and treatment outcome: a meta-analytic review. *Journal of Consulting and Clinical Psychology* 78, 200–211.