International **R**eview of **R**esearch in **O**pen and **D**istributed **L**earning
Volume 22, Number 3

August – 2021

Development and Evaluation of an Open-Source, Online Training for the Measurement of Adult-Child Responsivity at Home and in Early Childhood Education and Care Settings

Michelle Rodrigues*[1], Alessandra Schneider*[1], Nina Sokolovic[1], Ashley Brunsek[1], Beatriz Oré[2], Michal Perlman[1], and Jennifer M. Jenkins[1]
[1]Department of Applied Psychology and Human Development, University of Toronto, [2]NGO Alto Perú and Independent Consultant, Lima, Peru
* first authorship shared

Abstract

Efforts to monitor and improve responsive caregiving for young children, because of its importance for child development, are part of the United Nations Sustainable Development Goals. Two brief observational measures of responsive caregiving have been developed and validated (Responsive Interactions for Learning—parent [RIFL-P] and educator [RIFL-Ed] versions), with the RIFL-P available in English, Portuguese, and Spanish. The aim of the current study was to present and evaluate two online training programs for the RIFL measures. These distance learning courses were designed as open-source and asynchronous to enable their use in low- and middle-income countries and remote areas. The following course components are used: readings, lectures, observation of interactions on video, coding practice with automated feedback on item coding, and quizzes. Of the 76 trainees who registered for one of the online courses, 58 (76%) completed all theoretical module components. Student performance was generally high. Marks on quizzes ranged between 83%–100%. Ninety percent of those who took the reliability tests passed (40/44). Student satisfaction during and after the course was high. The effective online training programs are available free of charge and the RIFL suite of measures is efficient to implement. Implications for research and practice are discussed.

Keywords: responsive caregiving, parental sensitivity, online learning, observational measurement, low- and middle-income countries

Introduction

Responsive caregiving, defined as sensitivity and stimulation, is one of the cornerstones of nurturing care and a prerequisite for achieving positive developmental outcomes for young children (Black et al., 2017; Britto et al., 2017; Jeong et al., 2021). This specific type of caregiving has reached global attention with directed efforts on behalf of international agencies and governments to implement programs to increase this aspect of caregiving (Santos et al., 2020; World Health Organization [WHO], United Nations Children's Fund [UNICEF], & World Bank Group, 2018). Responsive caregiving has been found to be important in home and educational contexts (Madigan et al., 2019; Vermeer et al., 2016). Despite international acceptance of the importance of responsive caregiver-child interactions, there is a clear need to refine and standardize measurements for this aspect of caregiving (Jeong et al., 2018, 2021). Proxy indicators (e.g., parental mental health, childcare availability, frequency of proxy activities with children; UNICEF & Countdown to 2030, 2020; Pierce, 2021) were initially used to assess the construct, but our group has developed efficient (8 minutes) and psychometrically strong instruments that can be used at the population level in home and educational contexts (Pauker et al., 2018; Prime et al., 2015; Schneider et al., 2021; Sokolovic et al., 2021a, 2021b).

In the current study, based on our measures, we examined whether we could develop asynchronous, online courses based on video recording examples of responsive interactions to teach professionals (with diverse cultural and linguistic backgrounds) how to reliably code responsive interactions. Evidence from the teachers' education field shows that video examples can be an effective way to teach and improve students' coding reliability and content accuracy (Prusak et al., 2010). This study represents a novel contribution to researchers, policymakers, and program leaders in charge of implementing the Nurturing Care Framework (WHO, UNICEF, & World Bank Group, 2018) in national and global spheres, because it presents and evaluates open-source training for a reliable and valid measure of responsive parenting.

This is particularly relevant for low- and middle-income countries (LMICs) where most of the world's children live. This aspect of caregiving is modifiable; however, there is an urgent need for efficient, psychometrically sound measurement of caregiving outcomes (including responsivity) that could be used at the population level across cultures (Jeong et al., 2018). Responsive interactions can only be reliably and validly measured through individual-level assessments that use observational methods (Lotzin et al., 2015). Most coding schemes typically require extensive training, and are complex, time-consuming, and expensive to administer and code (Bailey et al., 2017). This limits their usefulness for population-based studies.

An observational assessment of interaction quality—the Responsive Interactions for Learning (RIFL) measure (previously called Cognitive Sensitivity)—was originally developed using a Canadian sample of parents (Prime et al., 2015) and early childhood educators (Pauker, et al., 2018; Sokolovic et al., 2021b) interacting with young children. Since its development, the RIFL measures have been successfully adapted and tested in LMICs, including Brazil (Schneider et al., 2021) and Peru. This psychometrically sound measure assesses a person's ability to understand and respond appropriately (incorporating sensitivity and stimulation) to the thoughts and feelings of the person with whom they are interacting. This measure uses thin-slice methodology (popularized by Gladwell, 2005), which involves taking a highly complex psychological phenomenon that has been extensively researched, and operationalizing it in a rating that is

brief and intuitive. Thin-slice ratings have been found to possess similar psychometric properties to much longer, labor-intensive coding schemes (Matias et al., 2014; Pederson et al., 1990; Prime et al., 2014a, 2015).

Description of the RIFL Measures

The RIFL is a unidimensional observational tool that assesses three interconnected caregiving skills, namely clear communication, mind reading, and mutuality building. Clear communication refers to communicating in a way that the interactional partner(s) can understand. It is operationalized as providing verbal and nonverbal directions that are meaningful to the activity, as well as promoting a mutual understanding about the goals and rules of the task. Mind reading denotes understanding partners' thoughts and feelings. It is operationalized through items related to an awareness of what the partner knows or understands, rephrasing to achieve understanding, and responsiveness to subtle requests for help. Finally, mutuality building captures the back-and-forth quality of interactions and includes the caregiver's ability to provide positively-valenced feedback and fostering turn-taking within the interaction. The version used to assess interactions between early childhood educators (ECEs) and multiple children includes additional items that capture an educator's ability to meet the needs of multiple children simultaneously.

For the parent (RIFL-P) and sibling (RIFL-S) versions of the measure, two people (e.g., a parent and a child ranging from 18 months to school age or two siblings) are asked to work together for 5 minutes to build a block structure, copying a design they are shown. The complexity of the design varies to ensure it is adequately challenging for different developmental levels. For 18-month-old children, a shape and color sorter is used. For children from 2.6 years of age and older, a Lego model is built, with each person only allowed to touch 2 colors. For the educator (RIFL-Ed) version, the educator is asked to lead either a structured or naturalistic activity with a group of children.

In both cases, interactions are video-recorded and trained coders later observe the 5-minute video. Coders view the video only once and then rate each of the 11 (parent, sibling versions) or 15 items (educator version) on a scale ranging from 1 (*not at all true*) to 5 (*very true*). A mean of the 11 or 15 items is calculated, yielding a composite score of responsive interactions that can range from 1 (*very low responsivity*) to 5 (*very high responsivity*). Most notably, viewing the 5-minute video and reliably carrying out the coding results in a psychometrically sound assessment of responsivity achieved in 8 (parents, siblings) or 10 minutes (RIFL-Ed). Other observational measures of responsivity, both in parents (e.g., PICCOLO; Roggman et al., 2013) and educators (e.g., CLASS; La Paro et al., 2009), take over an hour (Matias et al., 2014; Pederson et al., 1990).

The RIFL-P and RIFL-S have strong psychometric properties across languages. Specifically, in Canadian samples, scores on the RIFL-P have been found to correlate with other parental sensitivity measures, to be inversely associated with contextual risk, and to relate to child outcomes including receptive vocabulary, executive functioning, theory of mind, and academic achievement (Prime et al., 2014a, 2014b, 2015; Sokolovic et al., 2021a). The Brazilian-Portuguese version of the RIFL-P demonstrates high reliability (internal consistency $\alpha = .94$; inter- and intra-rater *r*'s between .83 and .94) and validity (correlations with the PICCOLO parenting measure *r*'s between .32 and .47; correlations with children's cognition, language, and behavior *r*'s between .17 and .29; Schneider et al., 2021). The Spanish version also shows good reliability (internal consistency $\alpha = .97$; inter-rater $r = .87$) and validity (correlations with parenting measures of

autonomy support [Whipple et al., 2011] r = .70, and parental control r = -.47). The RIFL-Ed has also shown good reliability and validity (Pauker et al., 2018; Sokolovic et al., 2021b); notably, scores are associated with popular, validated measures of classroom quality such as the CLASS. No studies linking RIFL-Ed scores to child outcomes have been completed to date.

Open-Source, Online Training of RIFL Coding

Our research team developed multiple password-protected, open-source online courses to train new coders on the different RIFL measures, with the goal of providing a tool that could expand our ability to assess responsivity efficiently at a population level, especially in LMICs. Training for the RIFL-P is currently available in English, Portuguese, and Spanish, and training for the RIFL-Ed is available in English.

The course was designed based on findings from pedagogical research over the last half decade. Hattie (2008) meta-analytically synthesized the instructional methods from over 50,000 empirical studies to identify the most effective methods for student learning. These included learning goals that are explicit, narrow, and well-articulated; success criteria for students; multiple teaching strategies that triangulate the learning goal; and provision of feedback. Quality feedback relies on teachers being continuously aware of their students' learning status and providing directed and brief feedback (González et al., 2017; Molin et al., 2020). These findings are based on face-to-face delivery models, although those from online delivery suggest similar processes of design (Davis et al., 2018).

Systematic reviews and meta-analyses have demonstrated the advantages and disadvantages of online learning (Davis et al., 2018; Hrastinski, 2008; Means et al., 2009; Watts, 2016). The issues relate to maintaining student engagement, prevention of dropout, the provision of interactive elements to the learning, and the type of content to be learned. An early meta-analysis (Means et al., 2009) found that students who took all or part of their class online performed better, on average, than those taking the same course through traditional face-to-face instruction. The effect was strongest when the online learners were able to engage with course materials for longer periods of time. Findings with respect to synchronous versus asynchronous are similar. Asynchronous learners show more directed engagement with course content and deeper reflection of course issues. Synchronous learners experience less isolation, and receive more problem solving which may help them to persist with content (Hrastinski, 2008; Watts, 2016); however, it comes at the expense of achieving the narrow learning goal. Of course, the major advantages of asynchronous, online delivery include timing flexibility, geographical scope, and equalization of learning opportunities (Barteit et al., 2020; Chang et al., 2014).

The present asynchronous, online course was designed as a cost-effective, convenient way to provide training on responsive interactions, with the flexibility needed for uptake in a range of countries and time zones, in both urban and rural settings. It includes pre-recorded lectures, video clips of adult-child interactions, observational exercises with automated feedback, and reading materials. Videos in the English course are from North American samples, while videos in the Portuguese and Spanish versions of the course display Brazilian and Peruvian parents, respectively. Students are given explicit descriptions for each item on the scale, as well as criteria for how to score them along the entire range of the scale. A reliability test is given after the course has been completed, with the option of additional reliability testing if the coder does

not pass the first round. There is also a module for rater drift that allows coders to recalibrate their coding every 10 videos.

In line with Hattie (2008), learning goals are explicit, narrow, and well-articulated; students are aware of and receive immediate feedback about whether they have been successful in achieving the learning goal. Multimedia presentation is used to encourage learning through modalities of text, verbal presentation, and observation, following face-to-face and online empirical evidence of learning (Davis et al., 2018; Hattie, 2008). In the current study, primary and secondary outcomes were articulated for the different versions of the course. The primary outcome was the achievement of reliability, which captures the accuracy with which trainees are able to identify the quality of caregiving observed in different videos. This provides a strong measure of the learning outcomes intended for the course. The secondary outcome was related to trainee engagement with the materials and satisfaction with the courses. This data was collected from end-of-course surveys given to trainees.

Materials and Methods

Course Descriptions

Both the RIFL-P and RIFL-Ed courses are based on coders observing many video clips of caregiver-child interactions. The RIFL-P shows interactions between one parent and one child, while the RIFL-Ed shows one educator interacting with multiple children. Videos of parent-child interactions were obtained in Canada and the U.S.A. (English versions), Brazil (Portuguese version), and Peru (Spanish version). Children's parents and educators consented to their interactions being available on a password protected site for educational purposes. The course completion times range from 6–8 hours for the RIFL-P course (Modules 1–4, one coding practice assignment, one reliability test) and 8–10 hours for the RIFL-Ed course (Modules 1–4, two coding practice assignments, two reliability tests). Learning goals and course components are outlined in Table 1. They involved lectures, observations of interactions on video, coding practice with automated feedback on item coding, and quizzes. Short video clips of caregiver-child interactions were presented with annotations highlighting the presence/absence of specific behaviors related to responsive caregiving. Practice coding assignments included automated feedback. That is, when the trainee rated an item, a pop-up window provided them with feedback on the accuracy of their coding as well as the expert coder's rationale for the item, which was determined by two or three independent coders. Two reliability tests are offered after course completion, and the agreement between the expert coder and the student coder is examined through Pearson Correlation (automatically done within the online platform). If the first test is passed at $r = .8$ or higher (Stemler, 2004), the student is deemed reliable and receives a certificate of completion. If the participant is not successful on the first reliability test, they are required to review parts of the course, engage in an additional coding practice, and take a second reliability test. The two reliability tests reproduce the previous and successful structure of the face-to-face RIFL training.

Table 1

Outline of Course Components

Component	Learning goal(s)	Instructional methods	Duration
Introductory Module	Understand the key behaviors that constitute responsive caregiving	• Didactic lecture • Read a short story	20 min
Module 1	Understand why responsive caregiving is important for children's development	• Didactic lecture • Quiz • Brief written reflection	40 min
Module 2	Code items from the first half of the scale	• Didactic lecture • Videos of poor/strong examples for each item • Quiz	50 min
Module 3	Code items from the second half of the scale	• Didactic lecture • Videos of poor/strong examples for each item • Quiz	30 min
Module 4	Understand key coding tips and the process of reliability testing	• Didactic lecture • Read "Best Practices for Coding"	20 min
Coding Practice #1	Develop coding competencies	• Code 5 videos • Write explanations for each item score • Automated feedback	60 min
Reliability Test #1	Demonstrate agreement with expert coder	• Code 10 videos (no feedback)	90 min
If Pearson $r \geq 0.8$ on reliability test #1, pass and course completed. If $r < 0.8$, continue to next module.			
Coding Practice #2	Refine coding competencies	• Code 5 videos • Write explanations for each item score • Automated feedback	60 min
Reliability Test #2	Demonstrate agreement with expert coder	• Code 10 videos (no feedback)	90 min
If Pearson $r \geq 0.8$, pass and course completed. If $r < 0.8$, complete Practice and Test #3 (RIFL-Ed only).			

Monitoring Drift	Maintain reliability over time	• Code one video for every 10 completed on own project.	N/A
If total score is within 0.5 points of the expert score, reliability is maintained. If not, instructed to review Module 2 & 3 and Coding Practice #1.			

Procedure

The commencement date for the training courses for the RIFL-P in English, Portuguese, and Spanish were as follows: October 2018, June 2019, and June 2020. Although the course is now available in Spanish, no evaluations were carried out on the Spanish version of the course (because of the pandemic). The training course for the educator measure (RIFL-Ed) began in January 2020.

Evaluations were carried out during and after the courses. During the courses, at the end of each module, students provided feedback by answering four questions (rated on a 5-point scale) regarding their satisfaction with the module (overall satisfaction, usefulness of content, clarity, and mode of delivery). As the correlation between items within the modules was high (mean r = .6), we created a mean composite. Assessing satisfaction at the end of each module led to the inclusion of everyone who had taken the module (see Figure 1), allowing for high representativeness of these ratings.

A post-course anonymous survey was designed to assess participants' satisfaction with different course components, ask participants to contrast their experience with other face-to-face coding trainings in which they may have previously participated, determine whether they had used the measure after completing the training, and obtain feedback for improvement. Closed-ended questions were used to assess satisfaction (on a 5-point scale, from 1 = *strongly disagree* to 5 = *strongly agree*), as well as previous experiences and use of the measure (yes/no questions). Open-ended questions were used to understand challenges and recommendations for improvement with the course experience; we used inductive coding to aggregate these comments. The survey took less than 10 minutes to complete. Participants received a $20 (in Canadian dollars) gift card as compensation for their time. All procedures were approved by the University of Toronto Research Ethics Board.

Sample

Requests for use of the RIFL-P and RIFL-Ed measures led to the development of the online courses. Trainees included research assistants (undergraduate and graduate students in psychology and education), academic principal investigators, and professionals working in hospital and government settings. Trainees have been from a range of countries: Canada, the United States, United Kingdom, Israel, China, Peru, and Brazil.

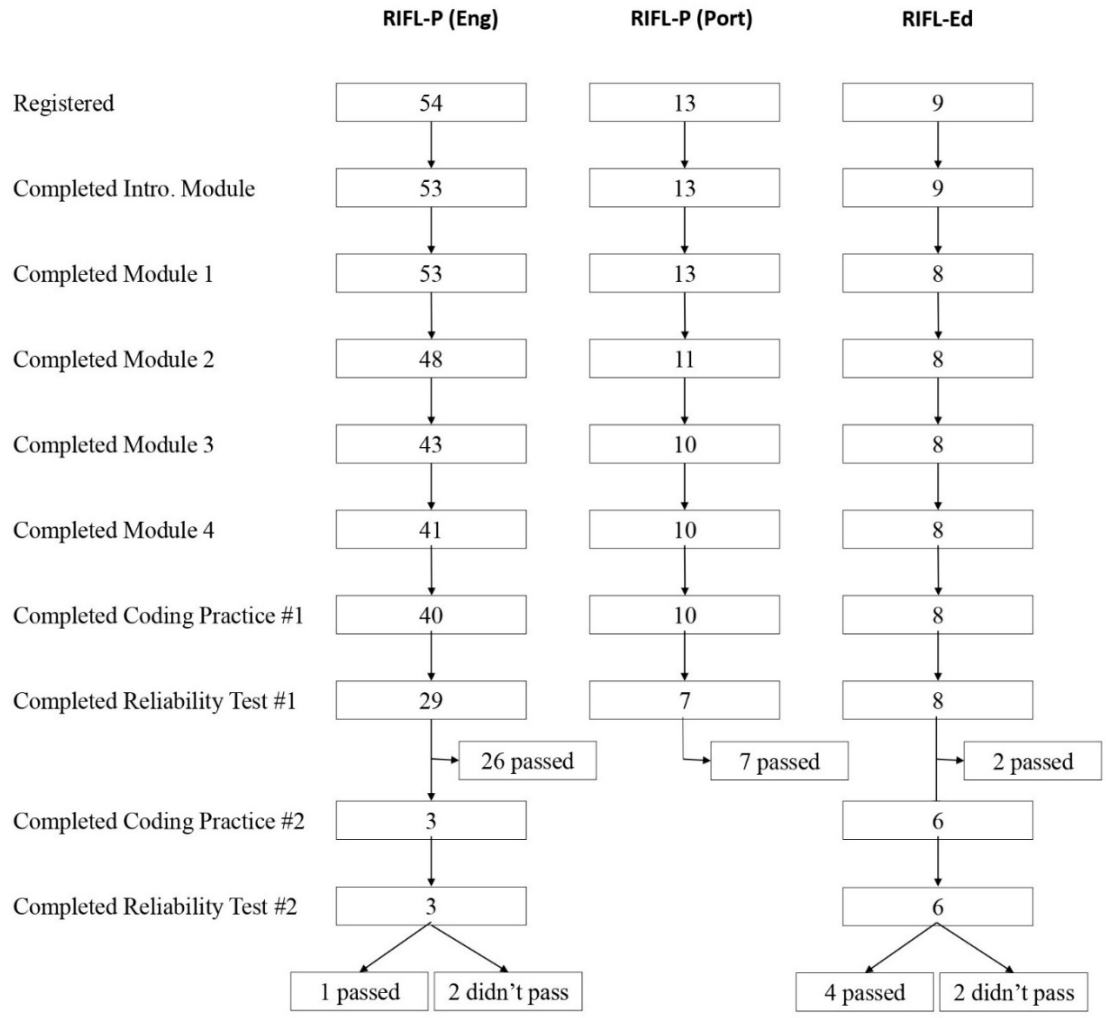
Results

Course Completion

The sign-up and completion rates for the three courses are presented in Figure 1. Access to the course was given to all professionals who expressed interest. While some requested it because they wanted to use the RIFL instrument in their research or professional practice (and thus achieve reliability), others were simply curious about online reliability training, learning about observational coding, etc. Unfortunately, we did not track these different motivations, but it is possible to see a substantial dropout (18/76 = 24%) from initial log-on to Coding Practice #1 completion.

Figure 1

Completion and Pass Rates



Note. RIFL-P (Eng.); RIFL-P (Port.); and RIFL-Ed = Responsive Interactions For Learning, English Version for Parents; Portuguese Version for Parents; and English Version for Educators.

Primary Outcome: Student Performance

Performance on module quizzes was high, with accuracy ranging from 83% to 100% (see Table 2). These quizzes involved simple, factual, multiple-choice or true-false questions about the material that was covered in the preceding online lecture. The high accuracy indicates that participants were actively paying attention to and understanding the material presented in the online lectures.

Table 2

Module Quiz Performance: Accuracy Rates

Quiz	RIFL-P (Eng.) % accurate	RIFL-P (Port.) % accurate	RIFL-Ed % accurate
Quiz 1	96	95	99
Quiz 2	96	83	98
Quiz 3	95	100	96

Note. RIFL-P (Eng.); RIFL-P (Port.); and RIFL-Ed = Responsive Interactions For Learning, English Version for Parents; Portuguese Version for Parents; and English Version for Educators.

The English version of RIFL-P course (N = 29) had a 93% pass rate (27/29), while the Portuguese version (N = 7) had a 100% pass rate. All but two participants across both RIFL-P courses (34/36) passed the reliability tests. The RIFL-Ed course (N = 8) had a 75% pass rate. Of those who passed, one third did so on the first reliability test and two thirds on the second reliability test.

Secondary Outcome: During Course Satisfaction

Satisfaction for all modules for the parent course was high and ranged between 4.6 to 4.94 out of 5, with little difference in ratings across modules. Satisfaction for the educator course was also high, and ranged between 4.38 to 4.79 out of 5. See Table 3 for satisfaction rates for all modules across the various courses.

Table 3

During Course Satisfaction Ratings

Module	RIFL-P (Eng.)	RIFL-P (Port.)	RIFL-Ed
Intro. Module	4.60 (0.72)	4.62 (0.74)	4.60 (0.53)
Module 1	4.83 (0.44)	4.94 (0.24)	4.67 (0.51)
Module 2	4.77 (0.66)	4.61 (0.69)	4.46 (0.79)
Module 3	4.89 (0.32)	4.80 (0.41)	4.38 (0.70)
Module 4	4.89 (0.37)	4.60 (0.67)	4.79 (0.41)

Note. Values are reported as mean (*SD*) on a 5-point Likert scale. Intro. Module = Introductory Module; RIFL-P (Eng.); RIFL-P (Port.); and RIFL-Ed = Responsive Interactions For Learning, English Version for Parents; Portuguese Version for Parents; and English Version for Educators.

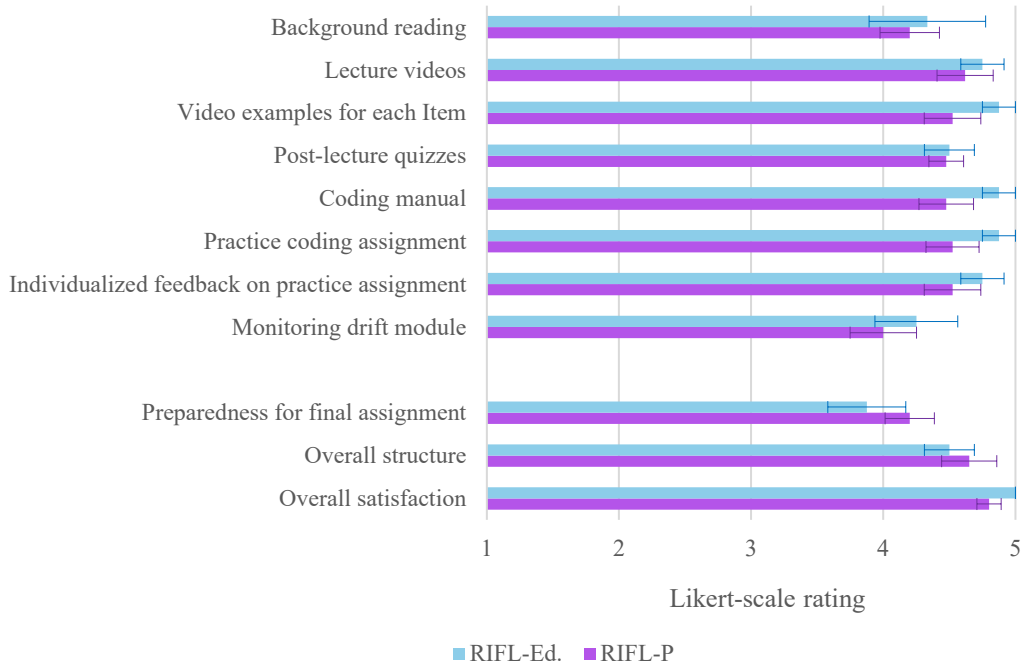
Values are the mean across four questions: overall satisfaction, usefulness of content, clarity of presentation, and mode of delivery. The consistent satisfaction across all modules suggest that all the course content was equally valuable to participants and there was not repetition or fatigue over time.

Satisfaction Post Course

Twenty-one participants (of 29; 72% response rate) completed the survey about the RIFL-P and eight participants (100% response rate) completed the survey about the RIFL-Ed. Results can be seen in Figure 2. A single anonymous link was sent to all RIFL-P course participants and we were unable to disaggregate those who completed the English vs. Portuguese versions of the course. Overall, post-course satisfaction was high (4.80 for the RIFL-P, 5.00 for the RIFL-Ed, on five-point scales). Participants seemed to especially value the lecture videos (4.62 and 4.75 for the RIFL-P and RIFL-Ed, respectively), video examples for each item (4.52, 4.88), coding manual (4.48, 4.88), coding practice (4.52, 4.88), and automated individualized feedback (4.52, 4.75). The background reading (4.20, 4.33) and monitoring drift modules (4.00, 4.25) were rated as less helpful, on average, and individuals did not feel fully prepared for the first reliability trial (4.20, 3.88). The majority of participants in both courses thought all course components were necessary and would not recommend removing or shortening any section.

Figure 2

Survey Results: Overall Retrospective Satisfaction



Note: Error bars show standard errors.

Eight participants who completed the RIFL-P course had also previously been trained in a different coding measure that required them to achieve interrater reliability. More than half of participants said they were able to grasp the theoretical construct and learn to code more quickly in this course compared to their other course, while most others said it was about the same in both courses. One participant said it was easier to learn when training was delivered face-to-face. For the RIFL-Ed, only two participants had previous

interrater reliability training experience—one said they learned faster in the RIFL-Ed course, while the other said the ease and rate of learning was similar in both courses.

Eleven participants who completed the RIFL-P course used the measure to code dyadic interactions in their own research projects, which required coding of between 20 and 4,000 videos. Two participants who completed the RIFL-Ed began using the measure in the short time between completing the course and completing the satisfaction survey.

Themes from the open-ended comments were as follows: requests for more videos that illustrate the midpoint of the scales (RIFL-P), more practice videos before the reliability test (RIFL-P & RIFL-Ed), shortened introductory lectures (RIFL-P & RIFL-Ed), and an expert explaining their coding of all items in a 5-minute video (RIFL-Ed). Other challenges that were noted were the inability to ask questions to obtain clarification (RIFL-Ed) and the need for increased age and ethnicity variation in taped examples (RIFL-P).

Discussion

Although research has consistently shown the importance of responsive caregiving for children's cognitive and socioemotional development (Britto et al., 2017; Jeong et al., 2021; Scherer et al., 2019), there remains a gap in ways to assess this aspect of caregiving at the population level. Having psychometrically strong, quick to train in and administer, and widely accessible measures of responsive caregiving is essential for monitoring, evaluating, and improving programs and policies designed to improve child outcomes. The aim of the current study was to evaluate whether it is possible to train students, researchers, and practitioners to reliably assess responsivity in parent-child and educator-child interactions (using the RIFL measures) in an asynchronous online course.

The high pass rates for the RIFL-P (English and Portuguese versions) course reveal that people can effectively learn how to reliably code responsivity in parent-child interactions using an online training model. Indeed, it is notable that most participants passed the reliability test on their first attempt, learning how to code responsivity in less than 10 hours. Pass rates for the RIFL-Ed course were also high, but in contrast to the RIFL-P course, the majority of participants required two reliability tests before being deemed reliable. These results were not surprising given the increased complexity of learning how to code interactions in which one educator is displaying different behaviors towards multiple children with varying cognitive and socioemotional skills, compared to dyadic interactions between one parent and one child.

The high pass rates across the RIFL-P and RIFL-Ed may in part be attributable to our choice to design the online platforms based on findings from the literature on effective teaching via face-to-face and online delivery models (Davis et al., 2018; Hattie, 2008). For instance, learning goals for each course component were explicit and narrow, multiple teaching strategies were incorporated into each module, video clips with annotations illustrated the learning goals, and practice coding assignments provided immediate feedback on the learning goals. Importantly, for all courses, an effort was made for the courses to be culturally appropriate and diverse with videos obtained from Brazil, Peru, Canada, and U.S.A., and from different

socioeconomic strata. Capturing illustrative parental and educator behavior across countries and social strata was our most significant challenge, and we continue to refine content as new videos become available.

In addition to the high success rates, participants reported being very satisfied with their overall training experience. Across both courses, participants were satisfied with the multimedia design of the course and found the various aspects such as lectures, videos, and feedback helpful to their learning. Participants provided meaningful feedback during the course surveys, such as displaying interactions that include children in the middle childhood period, illustrating the mid-points of the scale, and providing an additional set of optional videos to review prior to completing the reliability tests. These suggestions are currently being incorporated into the existing courses as we expand our library of available videos.

Given the effectiveness and feasibility of the current courses, we can conclude that online asynchronous training may be the cheapest, most equitable and efficient approach to global reliability training for observational assessments. Achieving reliability on an observational instrument appears to be an apt fit for asynchronous, online teaching, particularly when the course content is focused and detailed (Chang et al., 2014; Hrastinski, 2008; Watts, 2016).

For both the English and Portuguese versions of the RIFL-P course, the average completion time ranged from 6 to 8 hours. While participants took longer, on average, to complete the RIFL-Ed course (8 to 10 hours), these results were not surprising given the additional coding practice and reliability test required for participants to pass the course. In person trainings can often be quite lengthy, with many responsivity measures requiring multiple days of training (e.g., PICCOLO, CLASS), resulting in large labor costs associated with compensating both trainers and trainees. Furthermore, in-person reliability trainings require trainers and coders to be in the same place, often resulting in large travel costs. The online courses presented in this paper reduce these costs and barriers by providing a quick and effective manner to train coders remotely, giving coders flexibility to do so in their own time, and with the only expense being compensation for the trainee's time.

The RIFL-P and RIFL-Ed measures are psychometrically robust, quick to train in, and easy to administer, which allows them to be used at a population level. Indeed, with free, online training available in multiple languages, researchers and practitioners worldwide can learn to use and apply these measures. For instance, responsivity can be assessed and used as a marker to identify families with children at risk for developmental difficulties, for targeted prevention or intervention efforts. The RIFL measures can also provide an efficient manner to monitor, improve, and evaluate programs designed to increase responsive caregiving. Indeed, the RIFL-P is currently being used to evaluate a national home-visiting program and in large, longitudinal cohort studies in Brazil (Hallal et al., 2018). Finally, having parallel measures that capture the same construct across different caregivers in young children's lives is another advantage of the RIFL measures.

Completion rates for courses have been found to be lower in asynchronous online training than in face-to-face environments (Khalil & Ebner, 2014; Paton et al., 2018). It is notable that only 76% (58/76) of the trainees who signed up for the RIFL training courses completed the coding practice #1, and only 58% (44/76) went on to the reliability test. While this is likely to be, in part, a reflection of trainee motivations to take the course, it also likely reflects the challenge of keeping students engaged during asynchronous

learning (Davis et al., 2018). From students' satisfaction ratings both during and post completion of the course, as well as their grades on quizzes, it is clear that the course design suited student learning needs; however, there was still significant dropout. Given that this is a ubiquitous finding in online learning, remedial suggestions such as building the interactive element with synchronous or asynchronous discussion boards and adding an element of competition as in a gaming framework (Burgos et al., 2018; Davis et al., 2018) may also help improve completion rates in the RIFL. Future research should evaluate such social components to reliability training on a measure to reduce dropout.

Several limitations of this study should be noted. First, the sample size was small, particularly for the RIFL-Ed (recently live) and therefore, it is important to continue to monitor completion rates as well as participant satisfaction. This information will guide continuous improvement of the online courses. Second, the post-course survey results were not representative of the population that began the course, and because the survey was anonymous, data cannot be linked with the course results and satisfaction rates. Finally, the predictive validity of the RIFL-Ed measure, which was developed more recently, has yet to be tested.

Conclusion

The RIFL measures and online training are particularly timely due to the unprecedented global attention on the topic of responsive caregiving, as well as the current trend of exploring technology-based platforms for massive online training. The RIFL measures, because of their efficiency, advance the assessment of caregiver responsivity, while the development of an open-source, online training builds capacity in LMICs and remote settings. Helping children to survive and thrive relies on our ability to efficiently train a workforce to measure (and eventually improve) responsive caregiving.

Acknowledgments

We are grateful to the families (in Brazil, Canada, Peru, and the United States of America) and educators (in Canada) who gave their time so generously and consented to have their filmed interactions used for educational purposes. We extend our gratitude to the partner institutions of the video library (in Canada, the George Brown College, Children's Services, City of Toronto; in Brazil, the Postgraduate Program in Epidemiology at the Federal University of Pelotas; in Peru: the Grupo de Análisis para el Desarrollo-GRADE, and Fundación Baltazar y Nicolas; and in the United States of America, the Tufts University). We thank Daniela Rodrigues for her technical assistance and the Education Commons team at the Ontario Institute for Studies in Education/University of Toronto. We also extend our gratitude to the Bernard van Leer Foundation (grant agreement LAT-2017-098) and the Social Sciences and Humanities Research Council of Canada (grant number 890-2015-2031) for supporting this project.

References

Bailey, H. N., Bernier, A., Bouvette-Turcot, A. A., Tarabulsy, G. M., Pederson, D. R., & Becker-Stoll, F. (2017). Deconstructing maternal sensitivity: Predictive relations to mother-child attachment in home and laboratory settings. *Social Development, 26*(4), 679–693. https://doi.org/10.1111/sode.12220

Barteit, S., Guzek, D., Jahn, A., Bärnighausen, T., Jorge, M. M., & Neuhann, F. (2020). Evaluation of e-learning for medical education in low-and middle-income countries: A systematic review. *Computers & Education, 145,* 103726. https://doi.org/10.1016/j.compedu.2019.103726

Black, M. M., Walker, S. P., Fernald, L. C. H., Andersen, C. T., DiGirolamo, A. M., Lu, C., McCoy, D. C., Fink, G., Shawar, Y. R., Shiffman, J., Devercelli, A. E., Wodon, Q. T., Vargas-Barón, E., & Grantham-McGregor, S. (2017). Early childhood development coming of age: Science through the life course. *The Lancet, 389*(10064), 77–90. https://doi.org/10.1016/S0140-6736(16)31389-7

Britto, P. R., Lye, S. J., Proulx, K., Yousafzai, A. K., Matthews, S. G., Vaivada, T., Perez-Escamilla, R., Rao, N., Ip, P., Fernald, L. C., MacMillan, H., Hanson, M., Wachs, T. D., Yao, H., Yoshikawa, H., Cerezo, A., Leckman, J. F., & Bhutta, Z. A. (2017). Nurturing care: Promoting early childhood development. *The Lancet*, *389*(10064), 91–102. https://doi.org/10.1016/S0140-6736(16)31390-3

Burgos, C., Campanario, M. L., de la Peña, D., Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers & Electrical Engineering, 66*, 541–556. https://doi.org/10.1016/j.compeleceng.2017.03.005

Chang, T. P., Pham, P. K., Sobolewski, B., Doughty, C. B., Jamal, N., Kwan, K. Y, Little, K., Brenkert, T. E., & Mathison, D. J. (2014). Pediatric emergency medicine asynchronous e-learning: A multicenter randomized controlled Solomon four-group study. *Academic Emergency Medicine, 21*(8), 912–919. https://doi.org/10.1111/acem.12434

Davis, D., Chen, G., Hauff, C., & Houben, G. J. (2018). Activating learning at scale: A review of innovations in online learning strategies. *Computers & Education, 125,* 327–344. https://doi.org/10.1016/j.compedu.2018.05.019

Gladwell, M. (2005). *Blink: The power of thinking without thinking*. Little, Brown and Company.

González, A., Fernández, M. V. C., & Paoloni, P. V. (2017). Hope and anxiety in physics class: Exploring their motivational antecedents and influence on metacognition and performance. *Journal of Research in Science Teaching*, *54*(5), 558–585. https://doi.org/10.1002/tea.21377

Hallal, P. C., Bertoldi, A. D., Domingues, M. R., da Silveira, M. F., Demarco, F. F., da Silva, I. C., Barros, F. C., Victora, C. G., & Bassani, D. G. (2018). Cohort profile: The 2015 Pelotas (Brazil) birth cohort study. *International Journal of Epidemiology, 47*(4), 1048–1048h. https://doi.org/10.1093/ije/dyx219

Hattie, J. (2008). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.

Hrastinski, S. (2008). Asynchronous and synchronous e-learning. *Educause Quarterly, 31*(4), 51–55. https://er.educause.edu/articles/2008/11/asynchronous-and-synchronous-elearning

Jeong, J., Franchett, E. E., Ramos de Oliveira, C. V., Rehmani, K., & Yousafzai, A. K. (2021). Parenting interventions to promote early child development in the first three years of life: A global systematic review and meta-analysis. *PLoS Medicine, 18*(5), e1003602–e1003602. https://doi.org/10.1371/journal.pmed.1003602

Jeong, J., Pitchik, H., & Yousafzai, A. (2018). Stimulation interventions and parenting in low- and middle-income countries: A meta-analysis. *Pediatrics, 141*(4), e20173510. https://doi.org/10.1542/peds.2017-3510

Khalil, H., & Ebner, M. (2014). MOOCs completion rates and possible methods to improve retention: A literature review. In J. Viteli & M. Leikomaa (Eds.), *Proceedings of EdMedia 2014--World Conference on Educational Media and Technology* (pp. 1305–1313). Association for the Advancement of Computing in Education (AACE). https://www.learntechlib.org/primary/p/147656/

La Paro, K. M., Hamre, B. C., & Pianta, R. C. (2009). *The classroom assessment scoring system, Toddler version*. Teachstone.

Lotzin, A., Lu, X., Kriston, L., Schiborr, J., Musal, T., Romer, G., & Ramsauer, B. (2015). Observational tools for measuring parent–infant interaction: A systematic review. *Clinical Child and Family Psychology Review, 18*(2), 99–132. https://doi.org/10.1007/s10567-015-0180-z

Madigan, S., Prime, H., Graham, S., Rodrigues, M., Anderson, N., Khoury, J., & Jenkins, J. M. (2019). Parenting behavior and child language: A meta-analysis. *Pediatrics*, *144*(4), e20183556. https://doi.org/10.1542/peds.2018-3556

Matias, C., O'Connor, T. G., Futh, A., & Scott, S. (2014). Observational attachment theory-based parenting measures predict children's attachment narratives independently from social learning theory-based measures. *Attachment & Human Development, 16*(1), 77–92. https://doi.org/10.1080/14616734.2013.851333

Means, B., Toyama, Y., Murphy, R, Bakia, M., & Jones, K. (2009). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. U.S. Department of Education, Office of Planning, Evaluation, and Policy Development. https://repository.alt.ac.uk/629/

Molin, F., Haelermans, C., Cabus, S., & Groot, W. (2020). The effect of feedback on metacognition: A randomized experiment using polling technology. *Computers & Education*, *152*, 103885. https://doi.org/10.1016/j.compedu.2020.103885

Paton, R. M., Fluck, A. E., & Scanlan, J. D. (2018). Engagement and retention in VET MOOCs and online courses: A systematic review of literature from 2013 to 2017. *Computers & Education, 125*, 191–201. https://doi.org/10.1016/j.compedu.2018.06.013

Pauker, S., Perlman, M., Prime, H., & Jenkins, J. (2018). Caregiver cognitive sensitivity: Measure development and validation in Early Childhood Education and Care (ECEC) settings. *Early Childhood Research Quarterly, 45*, 45–57. https://doi.org/10.1016/j.ecresq.2018.05.001

Pederson, D. R., Moran, G., Sitko, C., Campbell, K., Ghesquire, K., & Acton, H. (1990). Maternal sensitivity and the security of infant-mother attachment: AQ-sort study. *Child Development, 61*(6), 1974–1983. https://doi.org/10.2307/1130851

Pierce, H. (2021). Nurturing care for early childhood development: Path to improving child outcomes in Africa. *Population Research and Policy Review, 40*(2), 285–307. https://doi.org/10.1007/s11113-020-09581-z

Prime, H., Browne, D. T., Akbari, E., Wade, M., Madigan, S., & Jenkins, J. M. (2015). The development of a measure of maternal cognitive sensitivity appropriate for use in primary care health settings. *Journal of Child Psychology and Psychiatry, 56*(4), 488–495. https://doi.org/10.1111/jcpp.12322

Prime, H., Pauker, S., Plamondon, A., Perlman, M., & Jenkins, J. M. (2014b). Sibship size, sibling cognitive sensitivity, and children's receptive vocabulary. *Pediatrics, 133*(2), e394–e401. https://doi.org/10.1542/peds.2012-2874

Prime, H., Perlman, M., Tackett, J. L., & Jenkins, J. M. (2014a). Cognitive sensitivity in sibling interactions: Development of the construct and comparison of two coding methodologies. *Early Education and Development, 25*(2), 240–258. https://doi.org/10.1080/10409289.2013.821313

Prusak, K., Dye, B., Graham, C., & Graser, S. (2010). Reliability of pre-service physical education teachers' coding of teaching videos using studiocode analysis software. *Journal of Technology and Teacher Education, 18*(1), 131–159. https://www.learntechlib.org/primary/p/29396/

Roggman, L. A., Cook, G. A., Innocenti, M. S., Jump Norman, V., & Christiansen, K. (2013). Parenting interactions with children: Checklist of observations linked to outcomes (PICCOLO) in diverse ethnic groups. *Infant Mental Health Journal, 34*(4), 290–306. https://doi.org/10.1002/imhj.21389

Santos, I., Munhoz, T., Barcelos, R., Blumenberg, C., Bortolotto, C., Matijasevich, A., Santos, H. G. dos, Jr., Marques, L., Correia, L., Souza, M., Lira, P., Pereira, V., Botelho, V., Silva, R., Chicaro, M., & Victora, C. (2020). *Estudo de linha de base da avaliação de impacto do Programa Criança Feliz* [Baseline study of the impact evaluation of the Criança Feliz program]. Ministério da Cidadania, Secretaria de Avaliação e Gestão da Informação. http://www.epidemio-ufpel.org.br/uploads/downloads/82359dcf63d2c7b882037e499a417bfe.pdf

Scherer, E., Hagaman, A., Chung, E., Rahman, A., O'Donnell, K., & Maselko, J. (2019). The relationship between responsive caregiving and child outcomes: Evidence from direct observations of mother-child dyads in Pakistan. *BMC Public Health, 19*(1), 252. https://doi.org/10.1186/s12889-019-6571-1

Schneider, A., Rodrigues, M., Falenchuk, O., Munhoz, T. N., Barros, A. J., Murray, J., Domingues, M. R., & Jenkins, J. M. (2021). Cross-cultural adaptation and validation of the Brazilian Portuguese version of an observational measure for parent–child responsive caregiving. *International Journal of Environmental Research and Public Health, 18*(3), 1246. https://doi.org/10.3390/ijerph18031246

Sokolovic, N., Borairi, S., Rodrigues, M., Perlman, M., & Jenkins, J. M. (2021a). Validating an efficient measure of responsivity in father–child interactions. *Canadian Journal of Behavioral Science/Revue canadienne des sciences du comportement, 53*(1), 84–89. https://doi.org/10.1037/cbs0000185

Sokolovic, N., Brunsek, A., Rodrigues, M., Borairi, S., Jenkins, J. M., & Perlman, M. (2021b). Assessing quality quickly: Validation of the Responsive Interactions for Learning - Educator (RIFL-Ed.) measure. *Early Education and Development*, 1–16. https://doi.org/10.1080/10409289.2021.1922851

Stemler, S. E. (2004). A comparison of consensus, consistency, and measurement approaches to estimating interrater reliability. *Practical Assessment, Research, and Evaluation, 9*(1), 4. https://doi.org/10.7275/96jp-xz07

United Nations Children's Fund [UNICEF] & Countdown to 2030. (2020, November 21). *Thrive. Nurturing care for early childhood development: Country profiles for early childhood development*. https://nurturing-care.org/wp-content/uploads/2020/11/English.pdf

Vermeer, H. J., van IJzendoorn, M. H., Cárcamo, R. A., & Harrison, L. J. (2016). Quality of child care using the environment rating scales: A meta-analysis of international studies. *International Journal of Early Childhood, 48*(1), 33–60. https://doi.org/10.1007/s13158-015-0154-9

Watts, L. (2016). Synchronous and asynchronous communication in distance learning: A review of the literature. *Quarterly Review of Distance Education, 17*(1), 23–32.

Whipple, N., Bernier, A., & Mageau, G. A. (2011). Broadening the study of infant security of attachment: Maternal autonomy-support in the context of infant exploration. *Social Development, 20*(1), 17–32. https://doi.org/10.1111/j.1467-9507.2010.00574.x

World Health Organization [WHO], United Nations Children's Fund [UNICEF], & World Bank Group. (2018). *Nurturing care for early childhood development: A framework for helping children survive and thrive to transform health and human potential.* https://apps.who.int/iris/bitstream/handle/10665/272603/9789241514064-eng.pdf


Athabasca
University


BY