



Does coding internal working models of attachment have to be so hard?

Amanda Venta¹ · Veronica McLaren¹ · Carla Sharp¹ · Anna Abate² · Madeleine Allman¹ · Breana Cervantes¹ · Sophie Kerr¹ · Jessica Hernandez Ortiz¹ · Eric Sumlin¹ · Jesse Walker¹ · Kiana Wall¹

Accepted: 26 May 2022

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

The Child Attachment Interview (CAI) has demonstrated promise in youth, yet widespread use is thwarted by the need for interview transcription, face-to-face training, and reliability certification. The present study sought to examine the empirical basis for these barriers. Thirty-five archival CAIs were re-coded by: (1) expert coders (i.e., trained and reliable) without access to transcripts, (2) trained coders who had not completed reliability training, and (3) novice coders who had no formal training. Agreement with consensus classifications was computed with the expectation of moderate agreement. Results supported coding by experts without transcription of the interview. Near-moderate agreement preliminarily supported the use of trained coders who have not attempted reliability certification with appropriate caveats. While moderate agreement was not achieved for novice raters, findings suggest that self-paced training options for the CAI may hold future promise. These contributions erode a number of significant barriers to the current use of the CAI.

Keywords Child attachment interview · Training · Reliability · Kappa · Internal working model

Public Significance Statement: Attachment security is associated with a multitude of psychiatric constructs and an insecure working model of attachment is a transdiagnostic risk factor for psychopathology in childhood and across the lifespan. Nonetheless, the assessment of internal working models in clinical and research settings is lacking due to the burdens of training, reliability, and coding internal working models. The findings of this study suggest that the use of such instruments need not be so cumbersome, supporting the use of trained coders who have not attempted reliability certification with appropriate caveats. Findings also suggest that self-paced training options may hold future promise.

Does coding internal working models of attachment have to be so hard?

The Child Attachment Interview (CAI) was created to address a measurement gap in the coding of internal working

models of attachment in middle childhood and adolescence. Ample reliability and validity data have been published on the CAI (e.g., [1, 2]), and it has been widely used, with 356 citations of the CAI's development paper [3] and 325 citations of its first psychometric study [1]. The CAI is an interview for use in children 6–18 and trained coders utilize interview narratives and transcripts to produce scores along dimensional scales (i.e., Emotional Openness, Balance, Use of Examples, Idealizing, Dismissal, Resolution of Conflicts, and Coherence) as well as categorical assignments (i.e., Secure, Preoccupied, Dismissing, and Disorganized). A recent narrative review confirmed the CAI's reliability across clinical and non-clinical samples [4] and, indeed, the CAI has played an important role in scientific studies examining youth mental health and family functioning as well as clinical work with youth and families.

Attachment security as measured by the CAI has been repeatedly confirmed as a correlate of important public health outcomes like internalizing disorder [5], externalizing disorder [6], personality disorder [7], thought problems [8], posttraumatic distress [9], self-harm [10], treatment outcome [11], and convergence in youth and parents' reports of psychopathology [12], among other outcomes. For these reasons, researchers have long argued that the assessment

✉ Dr. Amanda Venta Ph.D.
aventa@uh.edu

¹ Department of Psychology, University of Houston, 4849 Calhoun Rd., Room 373, 77204-6022 Houston, TX, USA

² Department of Psychology & Philosophy, Sam Houston State University, Huntsville, TX, USA

of attachment security should be a routine component of clinical practice and research batteries alike. Achieving this goal, however, is precluded by the great time and financial burden of attaining certification to code attachment interviews like the CAI and the reality that interview-based measures appear to assess the construct of attachment security quite differently than more accessible self-report questionnaires [5, 13].

The practical utility of the CAI is limited by a number of factors. First, its use requires a four-day training course delivered face-to-face and completion of a reliability training video set of 30 videos completed after the training. Travel for training courses, as well as the training courses themselves, can be cost-prohibitive, primarily for students, early career researchers, and those from developing nations. Accessibility to training courses like those for the CAI has been further compromised by the COVID-19 pandemic and ongoing travel concerns and restrictions. New data protection regulations in the European Union have rendered some training material unusable, challenging trainers of the CAI to develop new training and reliability coding materials. Second, the process of achieving reliability in coding poses an additional cost and is time consuming, oftentimes occupying months of back-and-forth communication between a trainee and a trainer. Third, administration of the CAI is cumbersome because the interview must be videotaped and fully transcribed prior to coding. Per the CAI manual, the interview lasts between 30 and 45 min [14], though interviews commonly reach one hour in length, requiring between 4 and 9 h of transcription time prior to coding.

All of these issues make the CAI inaccessible to many researchers and clinicians due to the difficulty and expense of acquiring training, the length and additional cost of reliability training, and the burden of video recording and transcribing the interview prior to coding and use. The aim of this study was to determine whether there is empirical justification for the aspects of CAI's training and use that are most cumbersome. Several research questions were posed. First, are transcripts needed to code the CAI or can it be reliably coded with video alone? Removing the need for transcription would greatly increase the usability of the CAI, as it would eliminate the need for hours of labor (and associated cost) to produce the transcript and would allow for the CAI to be coded live, by an individual observing the interview via video feed or two-way mirror. A precedent for real-time coding exists, as demonstrated by Madigan and colleagues' work creating a brief form of the Atypical Maternal Behavior Instrument for Assessment and Classification (AMBIANCE; [15, 16]). Indeed, the modifications to the AMBIANCE made by Madigan's team enhanced the feasibility of training service providers to use the instrument in community settings [16]. In order to answer this

research question, we had three expert (i.e., trained and reliable coders) code CAIs that had been previously consensus coded without utilizing the transcript. Our hypothesis was that agreement, as measured by kappa statistics computed between coders and the consensus coding, would be moderate, indicating that transcripts are not necessary for adequate coding of the CAI. Intraclass coefficients (ICC) were also planned to determine agreement along dimensional scales.

Second, we wanted to examine whether reliability training was necessary for CAI coders who had already attended a structured face-to-face training. While face-to-face training of the CAI is cumbersome, requiring four days of in-person instruction as well as travel to the training location, the reliability process is much longer, often taking months of back-and-forth between trainers and trainees and delaying use of the CAI for interested parties. Further, the CAI developers are currently unable to offer reliability training, rendering coding certification impossible. To date, there has not been an empirical investigation of the added benefit of reliability training over and above face-to-face training alone for the CAI and, thus, it is unknown whether extant barriers to certification are empirically warranted. Again, a precedent for this research question is drawn from work with the AMBIANCE, which indicated that participants in a brief training who did not undergo subsequent reliability training were comparable to expert coders [16]. We sought to compute agreement, as measured by the kappa statistic, between three coders who had attended a face-to-face CAI course but been unable to access reliability training and archival, consensus ratings. We hypothesized that agreement would be moderate, indicating that the reliability training process is not necessary for adequate coding of the CAI for individuals who have undergone training. We also explored agreement on dimensional scales (i.e., ICC).

Finally, we sought to examine whether the CAI's requirement for face-to-face training is empirically based by examining agreement between archival, consensus codes and novice raters who were permitted self-paced study of the CAI manual but had not accessed face-to-face training or reliability training. There are many psychological instruments for which self-paced learning of administration and scoring procedures is deemed sufficient by the measure's developers including, in the attachment domain, the Attachment Q-Set [17]. Measures like this rely on the coder to reflect on their own degree of competence to administer and score the instrument, and, still, adequate psychometric properties are evident in the literature [18]. To our knowledge, no study has endeavored to examine whether face-to-face training for the CAI (or its predecessor, the Adult Attachment Interview, for that matter) is empirically justified. Needless to say, the accessibility of the CAI would be dramatically increased by eliminating the need for face-to-face training.

Given the absence of available evidence on which to base a hypothesis, analyses related to this research question were exploratory and endeavored to compute agreement in dimensional and categorical ratings between archival, consensus codes and novice coders who had access to the CAI manual for self-paced learning but had not attended training or completed the reliability process.

In sum, the current study examined three research questions with the aim of empirically testing whether transcripts, reliability training, and face-to-face training are needed for correct classification of internal working models as conceptualized in the CAI. This study utilized archival CAI videos that had been previously subjected to interrater reliability analyses by expert (i.e., trained and reliable) coders. Only videos where both expert raters agreed on the ultimate classification were selected for inclusion in this study. For the purposes of this study, selected CAI videos were re-coded under three experimental conditions: (1) expert coders (i.e., trained and reliable) without access to CAI transcripts, (2) trained coders who had completed a face-to-face course in the CAI but not completed reliability training, and (3) novice coders who had neither completed a face-to-face course in the CAI nor reliability training but had access to the coding manual. Empirical backing for lifting any one of these three barriers to the CAI's usage would enhance the utility of the CAI, specifically, and the feasibility of measuring internal working models of attachment in youth, generally, in both laboratory and clinical settings.

Method.

Participants.

This study utilized archival videos of CAIs completed with adolescents enrolled in a larger research study [Author self-citation]. A priori power analyses (conducted using <https://cran.r-project.org/web/packages/kappaSize/kappaSize.pdf>) were conducted to compute the sample size required to detect moderate agreement (i.e., $Kappa=0.40$ to 0.60 ; [19]) on a four-category variable (i.e., Secure, Dismissing, Preoccupied, and Disorganized) utilizing two raters and the following proportions for each of the four variable categories: Secure= 0.30 , Dismissing= 0.38 , Preoccupied= 0.14 , and Disorganized= 0.17 . The aforementioned proportions were drawn from [2]. The recommended sample size based on the above procedure was 32 subjects/archival videos. In order to ensure adequate representation of variable categories, videos were randomly selected from $n=621$ in order to mirror the aforementioned proportions. Eligible videos were ones that had previously been subjected to interrater reliability by expert (i.e., trained and reliable) coders and had consensus from both raters regarding the ultimate classification. To that end, videos were selected as follows: Secure with mother 12, Dismissing with mother 13, Preoccupied with mother 4, Disorganized with mother

6, Secure with father 12, Dismissing with father 10, Preoccupied with father 7, and Disorganized with father 6, mirroring the desired distribution across variable categories. The selected videos exceeded the required 32 subjects/archival videos recommended by power analyses, for a total of $N=35$ videos.

For the $N=35$ videos selected for inclusion, participant demographics were computed in order to characterize the sample. Participants were 54.3% female, 80% Caucasian, 8% Asian, and 12% multiracial or other. Age ranged from 13 to 17 years old ($M=15.71$, $SD=1.23$). Participants were recruited upon admission to an inpatient psychiatric hospital and stayed for an average of 35.94 days ($SD=13.85$). Participants had been previously hospitalized an average of 0.94 times ($SD=1.28$) and had seen an average of 2.47 therapists ($SD=1.70$) and 1.41 psychiatrists ($SD=1.37$). 40% of participants were diagnosed with a depressive disorder, 8.6% with bipolar disorder, 40% with an externalizing disorder, and 51.4% with an anxiety disorder.

Measures.

Demographic forms were filled out by parents upon admission. Age, race, gender, medical history, and other demographic information was collected.

The Child Attachment Interview (CAI; [3]) is a semi-structured interview that assesses internal working models of attachment in middle childhood and adolescence. Questions on the CAI include descriptions of the self and descriptions of the relationship between the respondent and his or her caregivers. Additional questions probe experienced during which the respondent may have called on the caregiver for care, understanding, and support, such as during times of conflict, separation, or illness. Adequate psychometric properties for the CAI were documented in the initial publication with a sample of children [1] as well as in subsequent study conducted with adolescents [2]. CAI interviews are rated on the basis of dimensional scales: Emotional openness (Emotion), Balance of positive and negative reference to attachment figures (Balance), Use of examples (Examples), Preoccupied anger (Anger; rated separately for mothers and fathers), Idealization (rated separately for mothers and fathers), Dismissal (rated separately for mothers and fathers), Resolution of conflicts (Conflict), and Overall coherence (Coherence). These subscales are then used to inform a categorical classification of Secure, Dismissing, Preoccupied, or Disorganized separately for each caregiver discussed. As noted, administration and coding of the CAI requires completion of a four-day training course as well as attainment of 85% agreement with the measure's authors on a set of 30 reliability videos. For the current study, correct classifications for the CAI were based on the ratings of trained and reliable coders. Archival CAIs had been

administered, transcribed, and coded by trained research assistants or doctoral students.

Procedures.

This study utilized archival CAI videos from a larger study in which the CAI was one component of a larger battery [20]. Participants were recruited on the day that they were admitted to an inpatient psychiatric hospital by research staff. Parents were approached first to give consent for participation and if provided, adolescents were approached for assent. All assessment procedures were conducted in person and privately by trained research staff within 2 weeks from admission date. All aspects of the study were approved by the appropriate human ethics research committee.

Following selection of $N=35$ videos for inclusion in this study (see Participants), video and transcript (if applicable) data was shared with coders. Coders were divided into three groups: (1) expert coders, (2) coders who had completed a face-to-face course in the CAI but not completed reliability training, and (3) self-trained coders who had neither completed a face-to-face course in the CAI nor reliability training in order to answer several research questions as follows. Coders in group 1 received only access to study videos. Coders in group 2 received access to both videos and transcripts for study videos. Coders in group 3 received access to videos and transcripts for study videos as well as the CAI manual. All coders were blind to consensus codes for study videos. In sum, each of nine coders watched and coded 35 videos each, for more than 300 h of coding time total for this study. Coders sent their codes for study videos to a research coordinator (VM) who compiled them for analyses. Coders were not informed of correct scores nor given the opportunity to change or discuss their codes.

This study was not preregistered. Data and study materials are available via email from the corresponding author.

Data Analytic Plan.

All three research questions were examined by testing coder accuracy relative to “correct” classifications that were based on previously completed interrater reliability data. Kappa and interclass coefficients (ICC) were used to determine coder reliability under different circumstances, as described in each research question. Specifically, agreement between each rater in the experimental groups and the correct classification was computed (as represented by the kappa statistic) in order to provide individual coder data. Then, kappa ratings across the three raters in each experimental group were averaged in order to provide group-level data. Moderate agreement was defined as $Kappa=0.40-0.60$, as recommended by [19]. In order to explore agreement across dimensional ratings, ICCs were computed between the three raters in each experimental group with the original, expert rater. The ICC between the original expert rater and

the second expert rater utilized in the archival interrater reliability analyses was reported for comparison.

Results

Are transcripts needed to code the CAI or can it be reliably coded with video alone?

Three expert (i.e., trained and reliable) coders recoded 35 archival CAIs without access to transcripts. We expected that they would demonstrate moderate agreement with “correct” classifications. Regarding the four-way categorical classification for maternal attachment, Rater 1 achieved 60% correct classification, $Kappa=0.42$, $p<.001$, Rater 2 achieved 74.29% correct classification, $Kappa=0.63$, $p<.00$, and Rater 3 achieved 51.43% correct classification, $Kappa=0.28$, $p=.011$. Overall, the three raters achieved 61.91% correct classification, average $Kappa=0.44$. Thus, moderate agreement was attained at the group level and for all but one rater, when considered individually.

Regarding the four-way categorical classification for paternal attachment, Rater 1 achieved 57.14% correct classification, $Kappa=0.41$, $p<.00$, Rater 2 achieved 74.29% correct classification, $Kappa=0.64$, $p<.001$, and Rater 3 achieved 58.82% correct classification, $Kappa=0.42$, $p<.001$. Overall, the three raters achieved 63.42% correct classification, average $Kappa=0.49$. Thus, moderate agreement was attained at the group level and for all raters when considered individually.

ICCs comparing Raters 1–3’s dimensional scores with those of the original expert (i.e., trained and reliable) coder are presented in Table 1. Agreement ranged between poor and good according to interpretive guidelines [25]. At the group level (average ICC), agreement was moderate.

Is reliability training needed for coding the CAI or is structured face-to-face training sufficient for reliable coding?

Three coders who had attended a face-to-face training on administration and coding of the CAI but were unable to complete the reliability training videos due to suspension of reliability offering by the CAI’s developers recoded 35 archival CAIs with access to transcripts and all other training materials. Their scores were compared with “correct” classifications as determined through previously conducted interrater reliability. Regarding the four-way categorical classification for maternal attachment, Rater 4 achieved 62.86% correct classification, $Kappa=0.47$, $p<.001$, Rater 5 achieved 54.29% correct classification, $Kappa=0.37$, $p<.001$, and Rater 6 achieved 47.06% correct classification, $Kappa=0.22$, $p=.047$. Overall, the three raters achieved 54.74% correct classification, average $Kappa=0.35$. Thus, moderate agreement was not attained at the group level,

Table 1 Intraclass correlations across raters in experimental groups and the original interrater reliability set

Subscale	Expert coders, with reliability certification, without transcripts		Trained coders without reliability certification		Self-taught coders without reliability certification		Control: Two expert coders utilizing transcripts	
	ICC	Interpretation	ICC	Interpretation	ICC	Interpretation	ICC	Interpretation
Emotion	0.42	Poor	0.13	Poor	0.44	Poor	0.64	Moderate
Balance	0.45	Poor	0.43	Poor	0.54	Moderate	0.79	Good
Examples	0.54	Moderate	0.54	Moderate	0.49	Poor	0.73	Moderate
Anger M	0.76	Good	0.72	Moderate	0.70	Moderate	0.80	Good
Anger P	0.75	Good	0.74	Moderate	0.73	Moderate	0.93	Excellent
Idealization M	0.56	Moderate	0.35	Poor	0.31	Poor	0.75	Good
Idealization P	0.34	Poor	0.25	Poor	0.31	Poor	0.68	Moderate
Dismissal M	0.45	Poor	0.52	Moderate	0.51	Moderate	0.71	Moderate
Dismissal P	0.59	Moderate	0.55	Moderate	0.55	Moderate	0.88	Good
Conflict	0.56	Moderate	0.58	Moderate	0.01	Poor	0.68	Moderate
Coherence	0.55	Moderate	0.43	Poor	.37	Poor	0.75	Good
Average	0.54	Moderate	0.48	Poor	0.45	Poor	0.76	Good

Notes. M=maternal, P=Paternal, ICC=Intraclass correlations computed using two-way random effects models and the single measures value is reported. Interpretation was based on [25]'s guidelines where ICC less than 0.5=poor reliability, ICC between 0.5 and 0.75=moderate reliability, ICC between 0.75 and 0.9=good reliability, and ICC greater than 0.90=excellent reliability

though it was close to the benchmark for moderate agreement (0.35 vs. 0.40), nor by two of the three raters in this experimental group.

Regarding the four-way categorical classification of paternal attachment, Rater 4 achieved 71.43% correct classification, Kappa=0.59, $p < .001$, Rater 5 achieved 65.71% correct classification, Kappa=0.53, $p < .001$, and Rater 6 achieved 61.76% correct classification, Kappa=0.47, $p < .001$. Overall, the three raters achieved 66.30% correct classification, average Kappa=0.53. Thus, moderate agreement was attained at the group level and across all three raters.

ICCs comparing Raters 4–6's dimensional scores with those of the original expert (i.e., trained and reliable) coder are presented in Table 1. Agreement ranged between poor and moderate, and, at the group level (average ICC), agreement was Poor but very close to the benchmark for moderate agreement (0.48 vs. 0.50).

Is prior training face-to-face required for adequate coding or is self-paced-study sufficient?

Three coders without prior face-to-face training or reliability training in the CAI recoded 35 archival CAI videos with access to the coding manual and transcripts. Their scores were compared with "correct" classifications as determined through previously conducted interrater reliability analyses. Regarding the four-way categorical classification of maternal attachment, Rater 7 achieved 57.14% correct classification, Kappa=0.31, $p < .001$, Rater 8 achieved 45.71% correct classification, Kappa=0.22, $p = .017$, and Rater 9 achieved

48.57% correct classification, Kappa=0.27, $p = .009$. Overall, the three raters achieved 50.47% correct classification, average Kappa=0.27. Thus, moderate agreement was not attained at the group level nor at the individual level.

Regarding the four-way categorical classification of paternal attachment, Rater 7 achieved 57.14% correct classification, Kappa=0.40, $p < .001$, Rater 8 achieved 51.43% correct classification, Kappa=0.31, $p = .002$, and Rater 9 achieved 54.29% correct classification, Kappa=0.37, $p < .001$. Overall, the three raters achieved 54.29% correct classification, Average Kappa=0.36. Thus, moderate agreement was not achieved at the group level though it was close to the benchmark for moderate agreement (0.36 vs. 0.40) and, at the individual level, was achieved for only one of the three raters.

ICCs comparing Raters 7–9's dimensional scores with those of the original expert (i.e., trained and reliable) coder are presented in Table 1. Agreement ranged between poor and moderate, and, at the group level (average ICC), agreement was Poor but close to the benchmark for moderate agreement (0.45 vs. 0.50).

Discussion

The current study endeavored to test whether three cumbersome aspects of the CAI's use—transcribing, face-to-face training, and reliability training—are empirically warranted in hopes of reducing the burden of use associated with the CAI. We posed three specific research questions. First, we wanted to test whether transcripts are needed by expert (i.e., trained and reliable) raters for accurate coding. Our hypothesis, that moderate agreement would be attained

between expert raters without access to the transcript and correct classifications, was largely supported. At the group level, moderate agreement was attained for the classification of maternal and paternal attachment and also for dimensional ratings. Second, we sought to test whether reliability training is necessary, with the hypothesis that raters who had attended face-to-face training in the CAI but not completed reliability training would be able to achieve moderate agreement with correct classifications. This hypothesis was partially supported. Moderate agreement was achieved for paternal attachment classifications and nearly achieved ($\text{Kappa}=0.35$ vs. 0.40) for maternal attachment classifications and dimensional ratings ($\text{ICC}=0.45$ vs. 0.50). Finally, we sought to examine whether the CAI's requirement for face-to-face training is empirically based by examining agreement between archival, consensus codes and novice raters. Across the board, moderate agreement between these raters and correct classifications was not achieved for maternal attachment classification, paternal attachment classification, or dimensional ratings.

The current findings support our first hypothesis, that access to the transcript is not needed for moderately accurate coding of the CAI. It should be noted that, based on a set of $n=126$ videos double coded for interrater reliability analyses by this research team, fully-trained expert coders utilizing transcripts as well as videos achieved $\text{Kappa}=0.48$ for both maternal and paternal attachment classifications. Therefore, the average $\text{Kappas}=0.44$ and 0.49 observed in this study when participants did not utilize transcripts reflect comparable performance and significant time savings for the research team. Similarly, moderate accuracy persisted when examining agreement in dimensional scales. Eliminating the need for a transcript in coding of the CAI (along categorical and/or dimensional lines) allows for live coding of the CAI by expert (i.e., trained and reliable) coders—a significant innovation in enhancing the clinical and research utility of the CAI.

While findings less strongly supported hypothesis 2—some Kappa ratings were just shy of the benchmark for moderate agreement—agreement rates either achieved or were quite close to demonstrating moderate agreement across categorical and dimensional ratings and are therefore considered as preliminary support for hypothesis 2. Ideally, replication of this study should be undertaken prior to the coding of CAI videos by trained coders who have not completed reliability training. However, reliability training for the CAI has been unavailable for more than two years and remains unavailable due to new data protection restrictions in the European Union that compromised reliability training materials, rendering the CAI unusable for a new generation of trainees who are able to access training but not reliability certification. Based on the preliminary findings of this study,

we believe that trained coders may be used for the coding of the CAI despite not having completed reliability training, noting this caveat in manuscripts and clinical reports that utilize the instrument.

Our findings did not provide strong support for the idea that self-paced training with the CAI manual is sufficient for accurate coding of the CAI. While close-to-moderate agreement was noted at the group level for the categorical classification of paternal attachment and for dimensional ratings, agreement for maternal categorical classifications was poor. Though these findings caution against coding the CAI without attending face-to-face training, they do suggest that self-paced training has future potential. In the current study, the novice coder group had no previous knowledge of the CAI, no direction about how to use the manual, and no exposure to CAI videos or correct classifications prior to completing their own ratings. Their coding performance is surprisingly good given the little information they received prior to coding. We acknowledge that raters' status as clinical psychology trainees may have enhanced their capacity to effectively make use of the manual without instruction; future research may endeavor to create a training manual and video set tailored towards self-paced learning with the CAI and examine the coding accuracy of coders thereafter. Self-paced training would facilitate access to the CAI among researchers and clinicians who cannot travel to training locations, would allow training to continue despite COVID-19 related travel restrictions, and would likely reduce the cost of accessing training, thereby making the measure accessible to a broader audience.

While not related to our primary research questions or hypotheses, the findings of the current study demonstrate interesting individual differences in coding ability across all three groups examined and preliminary evidence of patterns in the interrater reliability of dimensional scales. Regarding individual differences, it is possible that attachment ratings may be influenced by the assessor's own attachment style, thereby accounting for individual differences in coding ability. While this idea has not been directly tested, a recent study using naïve observer methodology found that naïve judges who were higher in attachment avoidance were less confident in their ratings, and judges higher in anxiety had weaker associations between accuracy and confidence [21]. In another set of experiments, it was found that emotional reactions to attachment narratives are associated with differential brain responses [22] and that attachment-dependent content modulated activation of cognitive-emotional brain circuits [23]. Moreover, the universality of attachment concepts has been challenged, with some suggesting that they are culturally influenced and cannot be assumed to hold the same meaning across cultures [24]. Likewise, differences in clinical skill, developmental stage in terms of graduate

training, and exposure to attachment theory and related course content may explain differences in ability that were not assessed in the current study but warrant future study. Examining ICCs for the dimensional CAI scales points to several interesting directions for future research. In the control group—ICCs computed between two expert raters utilizing CAI transcripts to code—ICCs were broadly in the range of what we would expect for strong agreement, while the others groups' ICCs were moderate, suggesting that the gold standard for training—face-to-face training, reliability training, and access to the transcript—is best. Still, when this ideal full package is not feasible, other options may be acceptable depending on the design of the research study or clinical decision-making in question. Interestingly, across experimental and control groups, there appeared to be patterns in ICCs, with preoccupied anger appearing the easiest to code, probably because it is most overt in transcripts and described concretely in the manual, and idealization particularly hard to code. Coherence, which is often used as a dimensional rating representing total attachment security [2; 8], seemed to benefit most from training and showed perhaps the most linear increase as training increases, possibly because it is an overall rating that requires quite a good and deep understanding of attachment to evaluate. This preliminary evidence of patterns in ICCs across rating groups points to opportunities for expansion and clarification in the CAI manual as well as future research regarding the developmental progression of knowledge specific to internal working models of attachment and how training can scaffold and dovetail with this natural course.

The current study is not without important limitations. First and foremost, the study utilized a relatively small (but adequately powered) number of archival videos, all drawn from the same, relatively homogenous sample of adolescents. Future research should seek to replicate the findings of the current study in diverse samples across developmental stages, given that the CAI is one of few attachment instruments psychometrically vetted for use in middle childhood. Second, we utilized only nine raters (all clinical psychology trainees) from a much larger and diverse global collection of individuals with different backgrounds and training elves who have completed training and reliability certification in the CAI. The small number of homogenous raters precluded examination of individual differences in coding performance, an interesting and important aspect of refining training efforts in the future. Third, we tested moderate agreement utilizing archival videos that had been previously subjected to interrater reliability analyses rather than a training set selected by the developers of the CAI. While our selection of videos was informed by power analyses and captured evidence-based distributions in attachment classifications as measured by the CAI, videos were not screened

or selected for characteristics beyond overall classification. Future research may endeavor to repeat this study utilizing videos that vary in terms of participant characteristics (e.g., clinical vs. community participant; diagnosis; family structure) in order to examine how these characteristics might affect coding accuracy.

Still, important strengths should be noted. First, our raters reflect individuals across three separate research teams who completed face-to-face training in the CAI at different times and in different courses. This variation enhances the generalizability of our findings to the broader population of trained coders. Second, we had access to a large number of archival CAI videos for this study and, thus, could select only those that had already achieved consensus in interrater reliability. This unique data source facilitated the design and execution of the current study, which tested novel questions inaccessible without this data. Third, the current study took place at a time when reliability training was suspended by the CAI's developers and, thus, trained but unreliable coders were available for participation without introducing biases related to trainees who might have self-selected out of reliability training. In the end, our findings support the coding of the CAI by expert (i.e., trained and reliable) coders without prior transcription of the interview, preliminarily support the use of trained coders who have not attempted reliability certification with appropriate caveats, and indicate that development of self-paced training options for the CAI may hold future promise. These contributions erode a number of significant barriers to the current use of the CAI, namely the time- and personnel-cost of transcribing interviews and the inability to train new coders in the absence of available reliability training. Future research can go further by replicating the current findings and testing the feasibility of self-paced training in the CAI.

Summary

Attachment security is associated with a multitude of psychiatric constructs and an insecure working model of attachment is a transdiagnostic risk factor for psychopathology in childhood and across the lifespan. The Child Attachment Interview (CAI) has demonstrated promise for the assessment of internal working models of attachment in youth, yet widespread use in clinical and research settings is thwarted by the need for interview transcription, face-to-face training, and reliability certification. The present study sought to examine the empirical basis for these barriers to the CAI's use. Following power calculations, thirty-five archival CAIs with adequate distribution with regard to attachment classifications were re-coded by: (1) expert coders (i.e., trained and reliable) without access to transcripts, (2) trained coders

who had not completed reliability training, and (3) novice coders who had no formal training. Agreement with consensus classifications was computed with the expectation of moderate agreement. Results supported coding by experts without transcription of the interview. Near-moderate agreement preliminarily supported the use of trained coders who have not attempted reliability certification with appropriate caveats. While moderate agreement was not achieved for novice raters, findings suggest that self-paced training options for the CAI may hold future promise. These contributions erode a number of significant barriers to the current use of the CAI and, moreover, suggest that perhaps the use of attachment interviews in general—including the gold standard Adult Attachment Interview—need not be so cumbersome.

Acknowledgements The authors wish to acknowledge the entire research team responsible for data collection.

Author contribution AV and CS conceived of the research questions. RM, AA, MA, BC, SK, JO, ES, JW, and KW coded the data. RM managed the data set. AV and RM conducted data analyses. AV wrote the first draft; all authors contributed to the manuscript and approved final version.

Funding: This research was funded in part by the University of Houston (no grant number).

Data Availability Data is available by contacting the corresponding author and is not publicly available due to confidentiality concerns.

Tables and Figures.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Informed consent: This work was reviewed by the appropriate IRB and conducted in compliance with all standards for the ethical conduct of research with human participants. Guardians provided signed informed consent and youth provided signed informed assent at the time of the CAI's completion.

References

- Shmueli-Goetz Y, Target M, Fonagy P, Datta A (2008) The Child Attachment Interview: A psychometric study of reliability and discriminant validity. *Dev Psychol* 44(4):939
- Venta A, Shmueli-Goetz Y, Sharp C (2014) Assessing attachment in adolescence: A psychometric study of the Child Attachment Interview. *Psychol Assess* 26(1):238–255 doi:10.1037/a0034712
- Target M, Fonagy P, Shmueli-Goetz Y (2003) Attachment representations in school-age children: The development of the Child Attachment Interview (CAI). *J Child Psychother* 29(2):171–186
- Privizzini A (2017) The Child Attachment Interview: A narrative review. *Front Psychol* 8:384
- Borelli JL, Somers J, West JL, Coffey JK, De Los Reyes A, Shmueli-Goetz Y (2016) Associations between attachment narratives and self-report measures of attachment in middle childhood: Extending evidence for the validity of the Child Attachment Interview. *J Child Fam Stud* 25(4):1235–1246
- Venta A, Sharp C (2015) Mentalizing mediates the relation between attachment and peer problems among inpatient adolescents. *J Infant Child Adolesc Psychother* 14(3):323–340
- Sharp C, Venta A, Schramm A, Vanwoerden S, Ha C, Newlin E, Fonagy P (2015) First evidence for the link between attachment, social cognition, and borderline features in adolescents. *Compr Psychiatr* 64:4–11
- Hart JR, Venta A, Sharp C (2017) Attachment and thought problems in an adolescent inpatient sample: The mediational role of theory of mind. *Comprehensive Psychiatry*, 78, 38–47 doi:10.1016/j.comppsycho.2017.07.002
- Venta A, Hatkevich C, Mellick W, Vanwoerden S, Sharp C (2016) Social cognition mediates the relation between attachment schemas and Post-Traumatic Stress Disorder. *Psychol Trauma: Theory Res Pract Policy* 9(1):88
- Glazebrook K, Townsend E, Sayal K (2015) The role of attachment style in predicting repetition of adolescent self-harm: A longitudinal study. *Suicide and Life-Threatening Behavior* 45(6):664–678
- Venta A, Sharp C, Newlin E (2015) A descriptive study of symptom change as a function of attachment and emotion regulation in a naturalistic adolescent inpatient setting. *European Child & Adolescent Psychiatry*, 24 (1), 95–104 doi:10.1007/s00787-014-0532-0
- Borelli JL, Palmer A, Vanwoerden S, Sharp C (2017) Convergence in reports of adolescents' psychopathology: A focus on disorganized attachment and reflective functioning. *Journal of Clinical Child Adolescent Psychology*
- Shaver PR, Mikulincer M (2004) What do self-report attachment measures assess? In: Rholes WS, Simpson JA (eds) *Adult attachment: Theory, research, and clinical implications*, vol 54. Guilford Press, New York, p 17
- Shmueli-Goetz Y, Target M, Datta A, Fonagy P (2011) *Child attachment Interview (CAI) Coding and Classification Manual, Version VI*. University College London, London, UK
- Cooke JE, Eirich R, Racine N, Lyons-Ruth K, Madigan S (2020) Validation of the AMBIANCE-brief: An observational screening instrument for disrupted caregiving. *Infant mental health journal* 41(3):299–312
- Madigan S, Eirich R, Racine N, Borland-Kerr C, Cooke JE, Devereux C, Lyons-Ruth K (2020) Feasibility of training service providers on the AMBIANCE-Brief measure for use in community settings. *Infant Mental Health Journal* 42(3):438–451
- Vaughn BE, Waters E (1990) Attachment behavior at home and in the laboratory: Q-sort observations and strange situation classifications of one-year-olds. *Child Dev* 61(6):1965–1973
- Van Ijzendoorn MH, Vereijken CM, Bakermans-Kranenburg MJ, Marianne Riksen-Walraven J (2004) Assessing attachment security with the attachment Q sort: Meta-analytic evidence for the validity of the observer AQS. *Child Dev* 75(4):1188–1213
- Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. *Biometrics*, 159–174
- Sharp C, Williams L, Ha C, Baumgardner J, Michonski J, Seals R, Patel A, Bleiberg E, Fonagy P (2009) The development of a mentalization-based outcomes and research protocol for an adolescent in-patient unit. *Bull Menninger Clin* 73(4):311–338
- Borelli JL, Peng X, Hong K, Froidevaux N, Sbarra DA (2019) Accuracy and confidence in perceptions of targets' attachment to former partners: Do judges vary as a function of individual differences in attachment orientation? *J Res Pers* 79:1–12

22. Borchardt V, Krause AL, Li M, van Tol J, Demenescu LR, Buchaim A, Metzger CD, Sweeney-Reed CM, Nolte T, Lord AR, Walter M (2015) Dynamic disconnection of the supplementary motor area after processing of dismissive biographic narratives. *Brain and Behavior*, 5(10), doi: 101002/brb3377
23. Krause AL, Colic L, Borchardt V, Li M, Strauss B, Buchaim A, Wildgruber D, Fonagy P, Nolte T, Walter M (2018) Functional connectivity changes following interpersonal reactivity. *Hum Brain Mapp* 39(2):866–879
24. Rothbaum F, Weisz J, Pott M, Miyake K, Morelli G (2000) Attachment and culture: Security in the United States and Japan. *Am Psychologists* 55(1):1093–1104
25. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15(2):155–163

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.