2020-01-01

The Development Of The Rapport Scales For Investigative Interviews And Interrogations, Observer Version (RS3i-O)

Justin Loren Magee
University of Texas at El Paso

THE DEVELOPMENT OF THE RAPPORT SCALES FOR INVESTIGATIVE INTERVIEWS AND INTERROGATIONS, OBSERVER VERSION (RS3i-O)

JUSTIN L. MAGEE

Doctoral Program in Psychology

APPROVED:

James Wood, Ph.D., Chair

Lawrence Cohn, Ph.D.

Misty Duke, Ph.D.

Jennifer Eno Louden, Ph.D.

Osvaldo Morera, Ph.D.

Stephen Crites, Ph.D.
Dean of the Graduate School

To my mother, Charlotte Ann, the source of all my strength, faith, and compassion.

THE DEVELOPMENT OF THE RAPPORT SCALES FOR INVESTIGATIVE INTERVIEWS AND INTERROGATIONS, OBSERVER VERSION (RS3i-O)

by

JUSTIN L. MAGEE, BA, MA, MS

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

THE UNIVERSITY OF TEXAS AT EL PASO

May 2020

Abstract

Rapport between an interviewer and a source is a vital component of successful investigative interviews and interrogations. Accurate measurement of rapport in these interactions is crucial to understanding its development and maintenance. The Rapport Scales for Investigative Interviews and Interrogations, Source version (RS3i-S; Duke et al., 2018) was developed to measure sources' perception of rapport, but no tool has been developed to measure rapport from a third-party observer's perspective. The primary purpose of this study was to develop the Rapport Scales for Investigative Interviews and Interrogations, Observer version (RS3i-O), a tool that allows observers to rate the rapport generated in investigative contexts. Exploratory and confirmatory factor analyses (EFAs and CFAs) were performed on ratings of 92 simulated investigative interviews made by trained undergraduate raters. The final version of the RS3i-O contained nine items comprising four rapport scales. The internal reliability of these scales was determined to be excellent and their inter-rater reliability was adequate. The convergent, discriminant, and criterion validity of the scales were examined using source ratings of the RS3i-S, a set of previously developed observational rapport scales, measures of source cooperation, and mean differences between interview styles. Substantial evidence of convergent validity was found for all RS3i-O scales, although discriminant validity among the four scales was weak. The secondary aim of the study was to evaluate the convergent and criterion validity of several linguistic variables measured by Linguistic Inquiry and Word Count (LIWC) 2015 and language style matching (LSM), a measure of linguistic convergence. Limited evidence of these linguistic variables' validity as measures of rapport in investigative interactions and interrogations was found.

Table of Contents

List of Tables

List of Figures

Introduction

Obtaining credible information about criminal activity or future threats from potentially uncooperative suspects and witnesses is a chief concern for law enforcement and national security investigators. Rapport, an indicator of relationship quality between two people, has been consistently linked to favorable outcomes (e.g., obtaining credible information) in investigative interviews and interrogations (Alison, Alison, Noone, Elntib, & Christiansen, 2013; Alison et al, 2014; Collins, Lincoln, & Frank, 2002; Vallano & Schreiber Compo, 2011; Walsh & Bull, 2010; Walsh & Bull, 2012).

Compared to the study of rapport in cooperative relationships, relatively little research has been conducted on the role of rapport in investigative interviews and interrogations (Abbe & Brandon, 2012a). However, because rapport's role in successful interviewing is considered vital, research efforts have increased to study the development, maintenance, and expression of rapport across a variety of investigative contexts. A substantial body of research has explored the view that rapport is made up of at least three distinct but highly related constructs, referred to as the "essential components" of rapport: (1) mutual attentiveness, (2) positivity, and (3) coordination (Tickle-Degnen & Rosenthal, 1990, p. 286).

Mutual attentiveness refers to sustained attention by both parties throughout an interaction, and can be an indicator of shared interest or engagement. Positivity can be characterized by feelings of warmth or acceptance experienced by individuals in an interaction. Coordination refers to the balance, flow, and tempo of an interaction – the degree to which interactants are "in sync" with each other. An awkward interaction would thus be described as having little coordination. This theoretical framework can be referred to as the tripartite model of rapport.

The Utility of the Tripartite Model of Rapport

The tripartite model, proposed by Tickle-Degnen & Rosenthal (1990), has become a highly influential psychological theory concerning the building and maintenance of rapport. As described above, this theory proposes that interpersonal rapport is comprised of several lower-order constructs, three of which are most essential to establishing and maintaining rapport over the course of an interaction or relationship: Mutual attentiveness, Positivity, and Coordination.

The model has influenced the study of relationship building across a variety of contexts. This research has consistently supported this model's validity across fields including medicine (e.g., Hall, Roter, Blanch, & Frankel, 2009), education (e.g., Perkins et al., 1995; Jiang-yuan & Wei, 2012), therapy (e.g., Tickle-Degnen, 1998; Tickle-Degnen & Puccinelli, 1999), negotiations (e.g., Drolet & Morris, 2000; Bronstein et al., 2012), and investigative interactions (e.g., Vallano & Schreiber Compo, 2015; Duke, 2018), among others.

Frank Bernieri, an influential researcher in the area of rapport, has applied the tripartite model framework to the development of measurement scales intended to capture the various facets of rapport described by Tickle-Degnen & Rosenthal (1987; 1990). Specifically, Bernieri (1988) originally created a measure of rapport in order to examine the display of relationship quality between individuals in very close relationships (e.g., teachers and their students; mothers and their children. As this program of research progressed, the need to accurately measure rapport led to the development and refinement of several versions of a tool that ultimately became the 18-Item Rapport Questionnaire (RQ; Bernieri, 2005). This self-report measure of rapport was designed to be completed by both individuals after participating in a dyadic interaction and was explicitly influenced by the tripartite model of rapport (Bernieri, Davis, Rosenthal, & Knee, 1994; Bernieri & Gillis, 2001; Bernieri, 2005). Researchers using the RQ

have studied the tripartite model across a variety of interactional contexts including both cooperative and adversarial interactions (Bernieri et al., 1994; Bernieri & Gillis, 2001). Magee (2018) provided a detailed description of these permutations.

Bronstein and his colleagues (2012) adhered closely to the tripartite model when developing scales for measuring rapport during negotiations and conflict resolution. The Negotiators' Rapport Scale (NRS) and Interaction Rapport Scale (IRS) developed by these researchers both contain subscales to measure the essential components of rapport. Bronstein et al. found rapport, as conceived by the tripartite model, was related to successful negotiations. Magee (2018) also recently reported findings which indicated these scales could be used as valid indicators of rapport in investigative interviews.

Vallano & Schrieber Compo (2011; 2015) slightly modified the RQ (Bernieri, 2005) for use in their studies of investigative interviews. These studies provided some support for the validity of this modified scale. Additionally, Duke et al. (2018a), using three scales based on the tripartite model framework, found that rapport was related to several successful investigative interview outcomes.

Though the factor structure of each of the measures used in the studies described above has been inconsistent across studies, the utility of the tripartite theory in accurately and consistently measuring rapport has been generally supported by these findings. The tripartite model is robust and appears to offer a valid approach to conceptualizing and measuring rapport, even within potentially adversarial interactions including investigative interviews and interrogations.

Due to the importance of developing rapport in investigative interviews and interrogations, many researchers have endeavored to develop accurate tools to measure aspects

of rapport in investigative contexts. Some of these instruments are designed for self-report while others are designed for use by observers. Though several observer instruments have been adapted or developed to measure rapport in investigative settings, none has been developed thus far that is specifically designed to measure the constructs of the tripartite model in investigative interviews.

Measuring Rapport

While no measure of rapport has been adopted as a standard in the field (Abbe & Brandon, 2012), several tools have been developed to measure varied theoretical aspects of rapport and rapport-building in investigative interviews and interrogations. For example, researchers have developed tools intended to measure the experience of rapport (Duke, Wood, Bollin, Scullin, & LaBianca, 2018) and interviewer behaviors thought to enhance rapport (Alison et al., 2013; Alison et al., 2014).

The Rapport Scales for Investigative Interviews and Interrogations, Source version (RS3i-S; Appendix A), the best validated self-report measure of rapport in investigative settings, was developed by Duke et al. (2018a). It is a multidimensional self-report questionnaire designed to measure interviewees' experience of rapport during interviews along several dimensions, including each of the three "essential components." For example, interviewees are asked to rate the item "The interviewer really listened to what I had to say." as an indicator of attentiveness. It is important to note here that Duke et al.'s (2018) instrument is published under the name "Rapport Scales for Investigative Interviews and Interrogations, Interviewee version"; however, this dissertation will instead refer to it as the "Source version" (rather than as the "Interviewee version") and use the acronym RS3i-S. This was done to enhance clarity and aid in disambiguation between the acronym referring to the "interviewer" and "interviewee". "Source"

is a commonly used term (e.g., Duke, 2018; Duke, Wood, Magee, & Escobar, 2018) that serves as a logical replacement for the term “interviewee” in this document.

Observational Measures

Although self-report questionnaires like the RS3i-S offer valuable insight into the experience of the interviewee, they may be impractical in some investigative interactions with resistant interviewees who refuse to complete questionnaires or when there is concern about whether responses to a questionnaire are genuine. It may be more practical in applied settings to employ a tool that allows observers to rate rapport behaviors during an interview or interrogation as a proxy measure of the level of rapport experienced by an interviewee.

Along these lines, Alison et al. (2013) have developed the Observing Rapport-Based Interpersonal Techniques (ORBIT), which measures an interviewer’s interpersonal skills and overall adherence to the principals of Motivational Interviewing (MI; Miller & Rollnick, 1992). However, studies using ORBIT have found poor inter-rater reliability for several of its scales, and no published studies have examined these scales' concurrent validity. Additionally, publications on ORBIT do not address the amount of time required to train a rater to accurately use the tool, which is problematic due to many of its scales’ complexity. In fact, the complexity of some ORBIT scales may be prohibitive for use by law-enforcement and national security agents conducting “real world” interviews and interrogations. Further, while a descriptive document has been published for training raters in the use of the ORBIT scales (Alison & Alison, 2017), no procedure to assess the fidelity of or trainees’ comprehension of this training has been provided. Finally, while ORBIT is rooted in theories of interpersonal interaction, its ratings are focused on interviewers’ rapport-building behaviors rather than the interviewee's

experience of rapport. That is, ORBIT measures the behaviors that an interviewer engages in to build rapport, but does measure whether these behaviors actually result in the interviewee experiencing rapport. In fact, no published studies have demonstrated that the behaviors measured by ORBIT correspond to an interviewee's experience of rapport.

Other observational tools have been designed based on the tripartite model framework. Bronstein et al. (2012) developed two observational scales to measure the level of rapport between two participants during a negotiation: (1) the Interaction Rapport Scale (IRS) and (3) the Negotiators' Rapport Scale (NRS). Each scale contains 11 items that are intended to measure Tickle-Degnen and Rosenthal's three essential components of rapport (mutual attention, positivity, and coordination). For instance, the component of attention is reflected in the item "listening", the component of positivity is reflected in the item "pleasant atmosphere", and the component of coordination is reflected in the item "synchrony".), The IRS and NRS are nearly identical; however, the IRS directs the observer to make judgements based on the *interview as a whole* while the NRS requires judges to rate the interviewer and interviewee *separately, as individuals*.

Some research has supported the construct validity of the IRS and NRS., Bronstein et al. (2012) reported that when judges observed participants interacting, observers' ratings of each participant on the NRS were significantly correlated (r = .22) with the self-reported experience of rapport by that participant's partner. However, the authors did not explore the validity of their measures outside of the context of negotiations and conflict management.

Recent research by Magee (2018) suggests that simple, global observer rating scales such as the IRS and NRS can be both reliable and valid tools for measuring rapport in interviews. Magee trained novice observers to rate simulated investigative interviews using several

observational scales designed to measure the three "essential components" of rapport during negotiations (from Bronstein et al., 2012) along with a modified version of an ORBIT scale thought to be related to positivity, which is one of the essential components of rapport. Observers' ratings using these scales were found to correlate significantly with self-report ratings of rapport for the same interviews and with the amount of information shared during the interviews.

Linguistic Measures

In addition to traditional self-report and observer rating scales, recent technological advances have allowed some researchers to examine new approaches to measuring rapport. Specifically, development of linguistic analysis tools such as the Linguistic Inquiry and Word Count 2015 (LIWC2015) program (Pennebaker, Booth, Boyd, & Francis, 2015) has prompted researchers to further study rapport through its potential linguistic indicators.

LIWC analyzes a single text file or group of texts, creating a complete word count and categorizing each word according to the LIWC2015 Dictionary, an internal dictionary containing approximately 6,400 words, word stems, and emoticons compiled by Pennebaker et al. (2015). The Dictionary classifies words across nearly 90 categories, including several standard linguistic dimensions (e.g., verbs and pronouns) and informal language markers (e.g., assents and filler words; see Pennebaker et al., 2015 for a complete list and description of LIWC2015 Dictionary categories). Each category is then output as a distinct linguistic variable expressed as a percent of the total words in that text. For example, a LIWC score of 3.15 for the assent variable indicates that the analyzed text was comprised of 3.15 percent assent words.

Several of these linguistic variables have even been proposed as likely correlates of the "essential components" of rapport in investigative interviews and interrogations (Driskell, Blickensderfer, & Salas, 2013). Specifically, Driskell et al. identified nine LIWC variables that might be hypothesized to correlate with the "essential components" of rapport (see Table 1). For example, as shown in Table 1 below, these researchers conjectured that less use of words related to negative emotion (e.g., "worried", "annoyed", and "sad") may indicate a higher degree of positivity in the interaction.

Table 1

Three-Component Model of Rapport as Measured by LIWC (Driskell et al., 2013, p. 5)

Rapport Component	LIWC categories
Mutual Attentiveness	- use of first-person plural pronouns such as we - use of the present verb tense* - use of words related to social processes
Positivity	- greater positive emotion - less negative emotion
Coordination	- fewer speech errors or non-fluencies - greater expression of assent or agreement - greater expressions of certainty - fewer conjunctions such as but

Note. *This category is not included in LIWC 2015. The new category "focus on present" will replace it.

Additionally, researchers have used LIWC to examine the degree of similarity between dyadic interactants' language style, referred to as Language Style Matching (LSM; Ireland & Pennebaker, 2010; Ireland et al., 2011; Boyd, 2017). This same measure has also been referred to as Linguistic Style Matching (e.g., Niederhoffer & Pennebaker, 2002) with no change in the acronym (LSM). In addition to variables related to psychological processes such as those discussed above, LIWC includes scores for linguistic dimensions such as individuals' use of nine

categories of function words that can be used to calculate LSM scores (see Table 2). LSM scores capture the degree to which two speakers use these categories of function words similarly. Put simply, LSM is thought to be an indicator of linguistic coordination. The specific procedure involved in calculating LSM scores is discussed in the Method section of this proposal.

Table 2

Function Word Categories Provided by LIWC 2015 (adapted from Pennebaker et al., 2015)

Category	Example words	Words in category
Personal pronouns	I, them, her	93
Impersonal pronouns	it, it's, those	59
Articles	a, an, the	3
Prepositions	to, with, above	74
Auxiliary verbs	am, will, have	141
Common adverbs	very, really	140
Conjunctions	and, but, whereas	43
Negations	no, not, never	62
Quantifiers	few, many, much	77

Research has supported a relationship between LSM and successful police crisis negotiations (Rogan, 2011; Taylor & Thomas, 2008) and investigative interviews (Richardson, Taylor, Snook, Conchie, & Bennell, 2014). Further, Driskell et al. (2013, p. 8) examined LSM in investigative interviews and found scores were significantly correlated (r = .48) with LIWC variables they hypothesized to be associated with coordination, one of the so-called "essential components" of rapport (see Table 1). However, recent research has not supported a relationship between LSM and rapport as measured by tools designed to reflect the "essential components". Specifically, Carmody, Mateo, Bowers, & McCloskey (2017) did not find a correlation between LSM and level of rapport in teams participating in a simulated intelligence analysis exercise. It is important to note, though, that these researchers only compared LSM scores to global ratings of

rapport comprised of all three "essential components", rather than to its individual constituent components. This may indicate a limitation in this research, as prior theory holds that "LSM is undoubtedly part of the broader coordination of any human interaction" (Niederhoffer & Pennebaker, 2002, p. 358).

Thus, research has not determined a clear relationship between the LIWC variables identified by Driskell et al. (2013) and favorable investigative interview outcomes. It is also important to note that Driskell et al. only examined sources' (i.e., interviewees') linguistic characteristics and did not include LIWC variables for interviewers in their analyses. Moreover, none of these variables has ever been compared with self-report or observer measures of the "essential components" of rapport. Further, while there seems to be a link between LSM and favorable interview outcomes, no research has examined the relationship of LSM to a validated measure of the coordination component. Thus, it is still unclear to what degree the linguistic variables included in LSM are related to experienced and observed rapport in investigative interviews and interrogations.

If validated as effective indicators of rapport, LIWC variables and LSM may provide an objective means of measuring rapport in research and investigator training. For example, some linguistic variables could offer additional means of establishing the validity of tools developed to measure aspects of relationship quality in a variety of investigative settings (e.g., investigative interviews and probation supervisory meetings). Additionally, during investigator training, LIWC could provide an objective measure of target investigator characteristics (e.g., positivity) during an interview or over the course of several interviews rather than relying on self-report or raters' impressions of that characteristic.

An unpublished pilot study by Quinones (2019), conducted in the same laboratory as the present study, examined the correlation of linguistic variables (LIWC variables and LSM) with scores on self-report rapport scales and observer rapport scales in a sub-sample of the same simulated investigative interviews included in the present study. The pilot study by Quinones indicated that several linguistic variables were strongly and significantly correlated with their theoretically corresponding self-report and observer rapport scales. For example, sources' degree of assent was significantly correlated with RS3i-S Trust/Respect ($r = 0.34$), NRS Positivity ($r = 0.50$), and GMISC Acceptance/Empathy ($r = 0.52$) scales. However, the findings of Quinones are only preliminary results that should be interpreted with caution as analyses were performed on a small number of interviews ($n = 30$). Nonetheless, the results suggest that the linguistic variables' validity is deserving of further examination.

Importance to Investigative Interviewing

The present study seeks to develop a valid and reliable observer measure of rapport for use in investigative interviews and interrogations that requires minimal training and is consistent with theories regarding the three "essential components" (or "Essential Three") of rapport. Specifically, this project will build on the work of Duke et al. (2018) and Magee (2018) to develop the Rapport Scales for Investigative Interviews and Interrogations, Observer version (RS3i-O). A tool like the RS3i-O could be immensely useful across a variety of investigative and criminal justice fields including law enforcement investigative interviews and interrogations, national security HUMINT operations, and potentially to police- or military-civilian interactions. For example, the RS3i-O might allow supervising officers to quickly assess an interviewer's rapport-building skills during training. The RS3i-O may also provide a means for supervising

officers or co-investigators to assess the rapport during an active investigative interview in real time.

The RS3i-O might also benefit future research regarding the development and maintenance of rapport in investigative interactions. Research examining the way rapport is developed over time has suggested that the importance of each components may be variable across the course of an interaction. This requires additional study within the context of investigative interviews and interrogations and a the RS3i-O would be a useful tool in this endeavor. As the RS3i-O is intended to be a quick, simple measure, it could be easily employed at different timepoints throughout an interview.

In addition to the development and validation of the RS3i-O, the present study seeks to examine the validity of the linguistic variables identified by Driskell et al. (2013) as indicators of rapport in investigative interactions. These variables may offer an objective means by which components of rapport may be assessed in an interview. Further, if linguistic variables are significantly correlated with observer ratings of rapport on the newly developed RS3i-O, they may be additionally used in training of the tool to assess raters' accuracy.

Research Questions

In the present study, observers rated simulated investigative interviews using a pool of 30 items intended to measure the three essential components of rapport (Attentiveness, Positivity Coordination) as proposed by Tickle-Degnen and Rosenthal (1990). Exploratory factor analysis (EFA) and structural equation modelling (SEM) were then performed to construct scales from these items. The resulting instrument was titled the Rapport Scales for Investigative Interviews and Interrogations, Observer Version (RS3i-O). Interviews were also transcribed and analyzed using LIWC 2015. The correlation of the resulting linguistic variables with self-report and observer measures of rapport was examined. This study was designed to address the following research questions.

1. Do ratings of the RS3i-O items conform to the theoretical three-factor structure proposed by Tickle-Degnen and Rosenthal (1990)?
2. What are the interrater reliability and internal reliability of the RS3i-O scales?
3. To what extent will scores on the RS3i-O scales be related to (a) interviewees' self-reported experience of rapport as measured by the RS3i-S, (b) scores on other observer scales measuring rapport, and (c) favorable interview outcomes (i.e. the amount of information shared during the interview)?
4. To what extent will linguistic variables thought to be indicative of rapport correlate with observed and experienced rapport, as measured by the RS3i-O, the RS3i-S, and other observer scales measuring rapport?

Method

Overview

The present study used 92 simulated investigative interview videos originally collected and analyzed by Duke et al. (2018). The present project consisted of three parts.

In Part 1, trained undergraduate raters rated each of the 92 interviews using a pool of 30 items (Appendix B) thought to be related to the three "essential components" of rapport as proposed by Tickle-Degnen and Rosenthal (1990). Exploratory factor analysis (EFA) and structural equation modeling (SEM) were used to analyze the ratings of these 30 items and create scales of a new observational measure of rapport, the Rapport Scales for Investigative Interviews and Interrogations, Observer version (RS3i-O).

Part 2 of the study used the same data as Part 1. The interrater reliability and internal reliability of the newly created RS3i-O scales were examined. The relationship between the RS3i-O scales and interview style was examined. Analyses examined the correlation of the RS3i-O scales with the RS3i-S source self-report ratings collected by Duke et al. (2018), the amount of information shared during interviews, self-reported cooperativeness, and observational measures of the essential components of rapport collected by Magee (2018). The purpose of Part 2 of the study was to test the internal and interrater reliability as well as the convergent, discriminant, and criterion validity of the newly developed RS3i-O scales.

In Part 3, trained undergraduate research assistants transcribed each of the 92 interview videos and processed the text for Linguistic Inquiry and Word Count (LIWC) 2015 analyses. The correlation of LIWC variables suggested by Driskell et al. (2013) with RS3i-O and RS3i-S scores was examined. The purpose of Part 3 of the study was to investigate the validity of these linguistic variables as indicators of rapport.

Finally, for Part 4 of the study, a multilevel confirmatory factor analysis (MCFA) was conducted to examine the degree to which the higher-order construct of rapport and its lower-order components are being represented within and between raters.

Participants and Interviews from Duke et al. (2018)

As already stated, the present study used a sample of simulated investigative interviews originally collected by Duke et al. (2018). The present section provides more detail concerning this sample.

Duke et al. (2018) recruited 94 University of Texas at El Paso undergraduate psychology students. Individual participants entered the laboratory, watched a fictionalized "evidence video", and then took part in simulated investigative interviews. Each participant was assigned to play the role of the "source" and was interviewed by a trained interviewer whose goal it was to obtain information about the video. Participants were randomly assigned to be interviewed using one of three distinct interview styles: a style that emphasized rapport-building (Rapport), a style in which the interviewer remained somewhat distant (Neutral), or a style in which the interviewer emphasized a more accusatorial approach (Pressure). Participants were told to withhold some information and to share other information during interviews.

Interviews were video-recorded. After completing the interview, source participants rated the interaction using the RS3i-S. Sources were predominantly young (M = 21.1 years, SD = 5.84), female (62%), and Hispanic (81%). Due to data corruption of the video files, only 92 of the videos are available for rating. The proposed study will use these 92 video-recorded interviews obtained in the Duke et al. (2018) study.

Prior studies have already collected extensive data on this sample of videos. Specifically, Duke et al. (2018) collected source self-report ratings of rapport's essential components using the

RS3i Attentiveness, Trust/Respect, and Connected Flow scales in addition to the total amount of relevant information shared during the interview and sources' self-reported level of cooperativeness. In addition, Magee (2018) collected observer ratings of the essential components using the Attentiveness and Positivity scales of the Negotiators' Rapport Scale (NRS) and the Coordination scale of the Interaction Rapport Scale (IRS) developed by Bronstein et al. (2012), along with the Acceptance/Empathy scale adapted from the Global Motivational Interviewing Skills Code Investigative Interview Adaptation (G-MISC) developed by Alison et al. (2013). The data already collected by Duke et al. and Magee is incorporated into many of the analyses in the present dissertation. Inclusion of this earlier data allows extensive validation analyses in the present study, as explained below in the present study procedures.

In the original study by Duke et al. (2018), the source was informed beforehand, by both the laboratory manager and the informed consent document (see Appendix C), that the interview would be video recorded. In addition, the informed consent form informed participants that their video recorded interview would be saved and used in future studies at the University of Texas at El Paso and that the interviews would be rated by other students in those future studies. The wording on the informed consent form was as follows:

> The entire investigative interview will be recorded. The video will then be shown to students who are participating in this study. The students will be asked to rate your actions, words and emotions during the interview. The video of the interview will probably also be saved and viewed by other students who participate in future studies approved by the UTEP Institutional Review Board. Those students will also be asked to rate your actions, words and emotions during the interview (Appendix C).

Part 1: Exploratory Analyses of the Pool of 30 Observer Rapport Items

Part 1 Raters

Nine undergraduate research assistants (RAs) were recruited from the UTEP Psychology program to rate each of the 92 simulated investigative interview videos using a pool of 30 observer rapport items created by the author of this dissertation. Raters worked independently without knowledge of other raters' ratings. Raters were either volunteers or received course credit for their work in this study. Raters were predominantly young (M = 21.44 years, SD = 2.19), female (66.67%), Hispanic (77.78%), and all were previously entirely unexperienced with behavioral coding.

Each rater was assigned to a "team" of 3 raters, but raters were instructed not to discuss their ratings with other raters, including those assigned to the same team. Raters were assigned to teams in a pseudo-random fashion. That is, the order of rater assignment was randomized and the next available assignment was given to each rater as they were recruited, until each position was filled (no research assistants that applied for participation as a rater were rejected).

A graphical depiction of how teams of raters were assigned to rate the pool of 30 items can be found below in Table 3 on page 18. There were three teams, each including three raters. One team was designated as the "Attentiveness" team (Raters A1, A2, and A3), another team was designated as the "Positivity" team (Raters P1, P2, and P3), and the third team was designated as the "Coordination" team (Raters C1, C2, and C3). Though all raters rated all 92 videos using all 30 items, the order of scoring differed among teams. The "Attentiveness" team first rated the 10 items designed to measure attentiveness, before going on to complete the 10 items designed to measures positivity and the 10 items designed to measure coordination. Similarly, the "Positivity" team first rated the 10 items designed to measure positivity before

completing the 10 coordination and 10 attentiveness items. The Coordination team first rated the 10 items designed to measure "coordination" before completing the attentiveness, then positivity items.

The first ten item ratings made by each team will be referred to as “primary team ratings” and the remaining ratings made after this will be referred to as “secondary team ratings”. Secondary team rating scales were assigned randomly. First, a random value was generated using Microsoft Excel 2016’s random number generation function (“=Rand()”) and assigned to the Attentiveness scale. Next, random values were also generated and assigned to the “Positivity Team” (.33245) and “Coordination Team” (.50031). It was previously determined that the team with the value closest to that generated for the Attentiveness scale would be assigned that scale for their “secondary team ratings”. Thus, the raters on the “Coordination Team” were assigned the Attentiveness scale, the raters of the “Positivity Team” were assigned the Coordination scale, and the raters of “Attentiveness Team” were assigned the Positivity scale for their “secondary team ratings” (Table 3).

Table 3

Overview of Raters' Primary & Secondary Scale Assignments

	Attentiveness Team			Positivity Team			Coordination Team		
	Rater A1	Rater A2	Rater A3	Rater P1	Rater P2	Rater P3	Rater C1	Rater C2	Rater C3
Primary Ratings	**Attentiveness**			**Positivity**			**Coordination**		
Secondary Ratings	Positivity			Coordination			Attentiveness		
Additional Ratings	*Coordination*			*Attentiveness*			*Positivity*		

Part 1 Materials

Simulated interview videos. Ninety-two digitally recorded videos of the interviews conducted by Duke et al. (2018) were rated. Each video showed one interviewer and one source seated across from each other at a square table in a room intended to resemble the style of interrogation booths used in law enforcement and national security investigations. The video camera was placed several feet away from the table, centered directly between the two interactants so that both can clearly be observed throughout the course of the interview. The videos are each 15-20 minutes long.

Rater Information Questionnaire. After recruitment, each rater completed an 11-item self-report questionnaire that asked about raters' demographic information including age, gender, race/ethnicity, education level, language characteristics, and experience with behavioral coding (Appendix D). This questionnaire was not used in the recruitment process. The purpose of the

Rater Information Questionnaire was to gather information necessary to examine possible sources of rater variance due to demographic characteristics.

Pool of 30 Observer Rapport Items' Comprehension Check. Before raters began rating interview videos, raters underwent a training that included an assessment of their familiarity with correct rating procedures, understanding of the target constructs, and their competency in recognizing these constructs in an investigative interview. Thus, as part of this training, raters completed a comprehension check designed to ensure they possess a theoretical understanding of the construct being measured by the items intended to represent each component, practical understanding of how to complete the items, and general competency in employing the rating procedure. Comprehension checks consisted of a series of recognition and free recall items designed to test raters' understanding of the target construct and the rating instructions provided. A unique comprehension check was developed for each set of 10 items intended to represent each of the three "essential components" of rapport. The format of the three comprehension checks was identical. Appendix E presents the comprehension check for the 30 observer rapport items that were evaluated for inclusion in the RS3i-O Attentiveness, Positivity, and Coordination scales. It is important to note that the items in the comprehension checks were presented in a different order to each team during training such that the order of the comprehension check items matched the order of the items' ratings as assigned by Primary and Secondary team ratings.

Comprehension checks included 60 recognition and 63 free recall items for a total of 123 items. There were three types of recall items: (1) Specific questions about the rating procedure (e.g., "Item ratings should be made based on whose behavior?"); (2) synthesizing a unique definition of the target construct (e.g., "Define 'Invested' in your own words."); and (3)

describing observable behaviors related to the construct within the context of interviews (e.g., "Describe, in your own words, behaviors that indicate an interviewer is Focused on the Source."). Recognition items required raters to rate the degree of similarity between target constructs and other characteristics ranging from synonymous to unrelated or antithetical using a Likert-type scale ranging from 1 (Not Similar) to 5 (Very Similar).

Each item was coded by the trainer as either correct (1) or incorrect (0). Free recall questions about the rating procedure had an objectively correct answer. However, free recall items that required raters to define and describe the construct in their own words required some subjectivity in scoring. For these items, the trainer determined whether or not the response satisfactorily indicated rater comprehension of target constructs and coded as either correct or incorrect. For example, when responding to the item "Define 'Trusting' in your own words.", a response of "When you can really rely on what a person says." would be marked as incorrect. This would indicate the rater has confused the characteristic of being trusting with the characteristic of being *trustworthy*.

With regard to Likert items, only responses on the anchor items of the scales (1 and 5) were accepted as correct for completely synonymous, unrelated, or antithetical characteristics. For example, the only accepted answer when rating similarity between "Positive" and "Certain" was 1, as these terms are entirely unrelated in this context. Any answer other than Not Similar would indicate some misunderstanding in the relationship between those characteristics and perhaps a flawed understanding of the target construct. A range of likely responses was accepted as correct for items that required more subjectivity in scoring. For example, rating the similarity between "Attentive" and "Respectful" was likely to produce greater variation in valid answers. Thus, all reasonable responses were accepted as correct. In this case, answers ranging from 3-4

were accepted as each of these responses demonstrating comprehension of some degree of conceptual overlap between the constructs while indicating understanding that the two constructs are distinct in this context. The scoring key for the comprehension checks, is located in Appendix F.

Part 1 Measures

Pool of 30 Observer Rating Items of Rapport Evaluated for Inclusion in the Rapport Scales for Investigative Interviews and Interrogations, Observer version (RS3i-O). Thirty observational rating items were created by the author of this dissertation to measure the three essential components of rapport as proposed by Tickle-Degnan and Rosenthal (1990). Specifically, these items consist of (1) 10 items intended to measure Attentiveness of the interviewer (2) 10 items intended to measure the Positivity of the interviewer, and (3) 10 items intended to measure Coordination between the interviewer and source (see Appendix B). Each of the 30 items was created for the present study using the rational/theoretical approach, based on three criteria: (1) the content of the item appeared to reflect prior theorizing about the way that each component of rapport is manifested in general social interactions (Bernieri, 2005), negotiations (Bronstein et al., 2012), and investigative interviews (e.g., Magee, 2018; Vallano & Schreiber Compo, 2011), (2) the content of the item was closely related to the content of RS3i-S items measuring the same components of rapport, and/or (3) the content of the item was similar to the content of observational items found to be related to the components of rapport in an earlier study by Magee (2018).

The 30 observational rapport items created for the present study did not require observers to engage in frequency-based coding of behaviors. Rather, these items required observers to

globally evaluate whether aspects of behavior related to the "essential components" of rapport were demonstrated during the interview. For example, for the item "Actively Listening" (related to mutual attentiveness), observers rated the extent to which active listening was displayed by the interviewer throughout the interview. For the item "Trusting" (related to positivity), observers rated the extent to which the interviewer appeared to trust the source, and for the item "Awkward" (related to coordination), observers rated how awkward the atmosphere created by the interaction between interviewer and source was. Similar global rating scales have been employed to measure aspects of relationship quality in interrogations (Alison et al., 2013; Alison et al., 2014) and negotiations (Bronstein et al., 2012). Additionally, global rating scales have been used to develop tools designed to measure rapport in broader social interactions (Bernieri & Gillis, 2001; Bernieri, 2005). This measurement approach has also been used by Miller and his colleagues (Miller & Rollnick, 1992) in developing the Motivational Interviewing Skill Code (MISC), a rating instrument designed to measure behaviors linked to rapport-building in therapy sessions that use Motivational Interviewing (MI).

Flesch Reading Ease (RE) Test. Included in Microsoft Word version 16, this test rates text on a 100-point scale where higher scores indicate the text is easier to read (Kincaid, Fishburne, Rogers, & Chissom, 1975). A score of 30 or lower indicates that a text is very difficult to read while a score of 70 or higher indicates a text is very easy to read. Microsoft recommends that most documents should score from 60-70, though other materials designed or adapted for use by criminal justice practitioners have been shown to score somewhat lower (e.g., Ricks & Eno Louden, 2014). The rating instructions and comprehension checks for the pool of 30 items were analyzed for reading ease with the goal of reaching a reading ease score in this range.

Flesch-Kincaid Grade Level (RGL) Test. Included in Microsoft Word version 16, this test rates text by equivalent U.S. grade level necessary to understand the document (Flesch, 1948; Kincaid et al., 1975). A score of 1 indicates that a first-grade student can understand, a score of 2 indicates a second-grade student can understand, and so on. The RGL test has been previously used in the development of measures for use in investigative interactions (Ricks & Eno Louden, 2014).

Rating instructions and comprehension checks were analyzed for the pool of 30 items to ensure a score of 12 or below, meaning that they are easily understood and completed by individuals who have, at most, graduated high school (U.S. grade 12). It is important that materials were designed to require less than 16 years of education while maintaining functional clarity. While a sizeable proportion of law enforcement and national security officers have achieved a two- or four-year degree, there is a great deal of variation in education level between officers, and many employing the tool would likely have terminated their formal education after graduating high school. This is important to ensuring high reliability between officers when employing the tool in training or for investigative support.

Woodcock-Muñoz Language Survey-Revised, Normative Update (WMLS-R NU). The WMLS-R is a standardized objective language assessment that was developed and normed using English and Spanish speakers (Woodcock, Muñoz-Sandoval, Ruef, & Alvarado, 2005). The WMLS-R includes seven subtests in parallel English and Spanish forms including Picture Vocabulary, Verbal Analogies, Letter Word Identification, Passage Comprehension, Understanding Directions, Story Recall, and Dictation designed to evaluate listening, speaking, reading, writing, and comprehension.

In order to assess any potential bias that the training and rating materials may show in varied English vocabulary and comprehension abilities among raters, the WMLS-R NU was employed to examine raters' English language proficiency. Specifically, the WMLS-R NU was employed in the present study to investigate the relationship between raters' English language comprehension, their comprehension of the RS3i-O training materials, and their ratings made using the RS3i-O. In the present study, the Picture Vocabulary subtest was used to calculate raters' age- and U.S. grade-equivalency in English vocabulary. The Verbal Analogies subtest was used to assess raters' general English comprehension age- and grade-equivalency. While the Understanding Directions subtest would have also been highly relevant to the present study, this test was not available to the author of this dissertation. The WMLS-R NU considers scores on both the Picture Vocabulary and Verbal analogies scores, along with the age of the test-taker to provide an Oral Language score. This score "measures listening and speaking in English, including language development and verbal reasoning" (Woodcock et al., 2005). The WMLS-R NU was not used to screen raters for inclusion in the present study; it was only employed retrospectively after each rater had finished rating all 92 interview videos.

Part 1 Procedure

Nine raters were each assigned to a "team" of 3 raters that was instructed to rate the 10 items associated with a single rapport component first, forming the "primary team ratings" for that component. Each team then completed the remaining 20 items associated with the other components forming the "secondary team ratings" for each other component. Team assignments are depicted above in Table 3. That is, one team of three raters produced primary team ratings for the 10 items associated with Attentiveness, one team of three raters produced primary team

ratings for the 10 items associated with Positivity, and the remaining team of three raters produced primary team ratings for the 10 items associated with Coordination. The rating document used by the Attentiveness team is located in Appendix G, but it is important to note that the order of the items varied across ratings documents for the Positivity and Coordination teams to match the order of their primary and secondary rating assignments (see Table 3).

Ratings were all completed in Suite 203 of the UTEP Psychology Building. Videos were viewed using VLC media player on Dell PCs with Behringer HPM1000 noise-cancelling headphones. Raters were limited to three or four 2-hour rating sessions spaced per week to reduce fatigue for a total of six to eight hours of rating per week. Rating sessions were spaced out as much possible (e.g., two hours Monday, two hours Wednesday, and two hours Friday) as well as over varying times of day (morning, midday, or afternoon) where possible; however, as raters were active undergraduate students, it was necessary apply these conditions to varying degrees while working within each rater's individual schedule.

Raters were trained in four steps: (1) the trainer read the scale instructions aloud to the rater and checked verbally for acknowledgement of comprehension; (2) each rater completed a comprehension test designed to ensure that the rater clearly understood each construct being rated and the rating procedure; (3) each rater rated a set of the same five interviews to ensure their mastery of the rating procedure; (4) in order to mitigate rater drift, each rater rated a set of the same 10 interviews whose order was randomly assigned within the unique assignment of each rater.

After reading the rating instructions and items and listening to them read aloud by the trainer, each rater was required to complete a comprehension check document. The contents and form of the comprehension check is described in detail above. There was a total of 123 points

possible on the comprehension check. Raters were only allowed to continue after answering 80% of the items correctly (99 points). If a rater failed the comprehension check, the trainer would lead the rater through the training materials again, highlighting problem areas. The rater would then complete the comprehension check a second time. If the rater did not meet the acceptable threshold this time, they would be excluded from the study. However, this did not occur during the study and no raters were excluded due to failing the comprehension check.

A set of 15 interview videos were randomly selected as test cases used to enhance interrater reliability. These videos were first rated by the author of this dissertation in order to provide "expert ratings" to which raters' ratings was compared. Five of these videos were randomly selected for all raters to complete first, after satisfactory completion of the comprehension check. The rating of these first five videos allowed the trainer to assess the quality of the raters' work and ensure that all of the raters are capable of completing the rating procedure effectively. The remaining ten videos selected as test cases for expert rating were randomly assigned within the unique assignment of each rater. The rating of these test cases allowed the trainer to mitigate rater drift over the course of the several weeks each rater undertook the rating task. The ten test case videos were not identified to the raters.

After each rater completed the first five ratings, the trainer assessed the level of agreement between the rater's and the expert ratings. This was done by calculating the absolute value of the difference between each rater's ratings and the corresponding expert ratings for each item. Across all 30 items for all five interviews, raters were expected to rate 80% of scores within one point, 15% of the scores within two points, and no more than 5% of scores within three points of the corresponding expert ratings. The trainer then met with each rater individually to discuss their performance. If a rater did not meet these thresholds, the trainer would identify

any problematic cases and/or items, viewing the relevant interview video(s) with the rater, and discussing the reason for the difference in ratings. If any item score differed by three points or more, the rater specifically highlighted these cases for discussion. After the meeting to discuss the first five interviews, each rater completed the remaining 87 in a unique, randomized order.

The remaining 10 expert-rated test cases were included in the random assignment of the remaining 87 videos for each rater. That is, these 10 videos were spread out randomly within the unique order of the remaining 87 videos for which each rater was responsible. As each rater completed a test case, the trainer met with them individually to discuss the video. The same criteria were applied to these 10 cases as were to the first five expert rated videos. For all 30 items rated for each test case, raters were expected to rate 80% of scores within one absolute value point, 15% of the scores within two points, and no more than 5% of scores within 3 points of the corresponding expert ratings. If a rater did not meet these thresholds, the trainer would once again identify problematic items, watch that interview video with the rater, and discuss the reason for any large difference in ratings. If any item score differed by 3 points or more, the rater specifically highlighted these cases for discussion.

Part 1 Analyses

The level of agreement between each team's individual raters' primary team ratings was examined and data from the two raters with the highest interrater reliability were used in the study. Specifically, the intraclass correlation coefficient (ICC) for each of the 10 primary team rating items was calculated based on a mean-rating, absolute agreement, two-way mixed effects model. The ICC values were then averaged across items and the data from the two raters with the highest mean ICC were selected for Part 1 analyses. For example, the primary ratings of

"attentiveness" made by the three members of the "Attentiveness" team (Raters A1, A2, and A3) were examined, and it was determined which raters on the team show the highest average interrater reliability. The ratings by these two raters were used in all subsequent analyses, and the ratings by the third rater were not included. This procedure was intended to select the raters most likely to be accurate in their application of the scoring rules. For each of the 10 items included in the primary team ratings, the ratings from the two raters selected as most likely to be accurate were averaged to produce a final score that was used for all Part 1 data analyses.

An exploratory factor analysis (EFA) of all 30 observational items was performed to assess the degree to which the 30 items reflect the hypothesized three-factor structure. The decision to employ EFA rather than approaches such as principle components analysis (PCA) or parallel analysis (PA) was made because the 30 items were hypothesized *a priori* to conform to a 3-factor structure. The EFA was conducted using maximum likelihood (ML) extraction and direct oblimin rotation to allow for correlated factors. The item scores for each item were based on the averaged ratings of each team's primary scores. That is, the EFA was based on (a) the ratings of the 10 "attentiveness" items by two members of the "Attentiveness" team, (b) the ratings of the 10 "positivity" items by two members of the "Positivity" team, and (c) the ratings of the 10 "coordination" items by two members of the "Coordination" team.

MPlus 7 (Muthén & Muthén, 2012) was used to test the fit of each subset of 10 items thought to comprise each factor scale onto a single factor model. The item information of each item was calculated by taking the ratio of each squared standardized item loading to the standardized item residual variance. SPSS 23 was used to compute Cronbach’s alpha for each subset of 10 items thought to comprise each factor scale as well as the alpha that would result from each item’s deletion. The decision to retain or reject items for inclusion in the final RS3i-O

measure were made based on (1) how saliently each loaded onto its respective hypothesized factor, (2) the additional item statistics discussed above, (3) and theoretical concerns related to item content.

After gaining sufficient evidence of the items' factor structure and narrowing the pool of items to a subset of nine items, a series of structural equation models (SEMs) were tested using robust maximum likelihood (MLR) estimation in MPlus 7. MPlus 7 was used for the SEM procedures in part because it allows for MLR estimation, which does not require that data are multivariate normally distributed. Further, MPlus 7 provides several indices of model fit not provided by SPSS 23 that were useful in confirming the new scales' factor structure.

Research regarding the relationship between the construct of rapport and it's lower-order components has produced mixed results. For example, Tickle-Degnen and Rosenthal (1990) originally suggested that rapport is best conceived as being comprised of the three distinct, but intercorrelated lower-order "essential components" (mutual attentiveness, positivity, and coordination). This was supported by the Magee's (2018) analysis of several scales designed to measure the essential components. However, other research examining the factor structure of scales intended to measure rapport has resulted in unidimensional factor solutions best explained by a single rapport factor (e.g., Bernieri et al., 1996). Due to the lack of clarity on this subject, models were tested that examined several possible relationships between the items included in the pool of 30 observer rating items of rapport evaluation for inclusion in the RS3i-O. First, models were tested that conformed to the structure of the correlated traits (CT) model (Reise, Moore, & Haviland, 2010) depicted below in Figure 1. This model assumed that the overall target latent variable being assessed by the measure, rapport, is reflected by the commonality between its constituent components – attentiveness, positivity, and coordination.

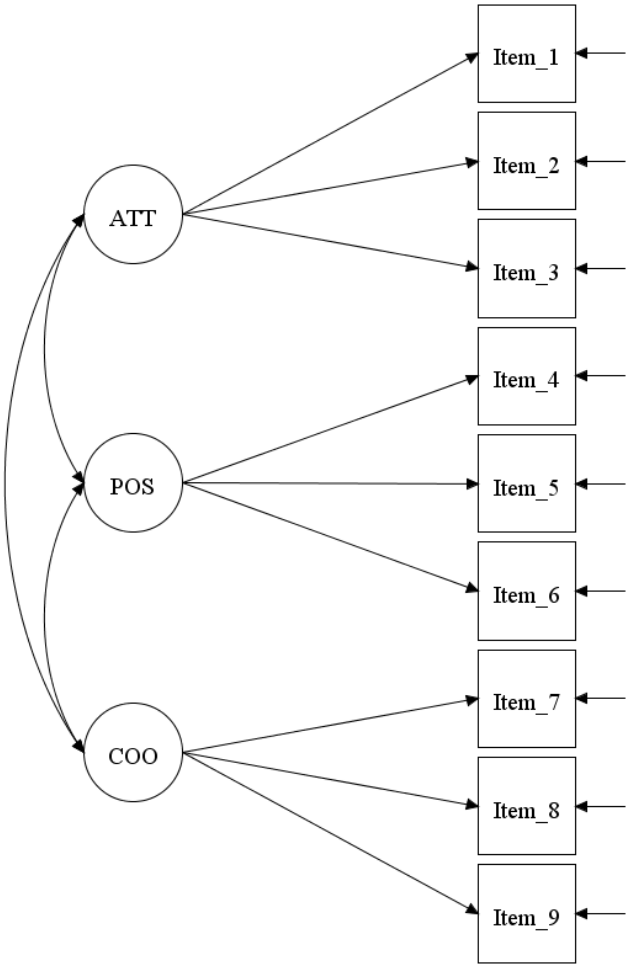


Figure 1. Correlated Traits Model

Note. ATT = Attentiveness, POS = Positivity, COO = Coordination.

In addition to the hypothesized 3-factor CT model, several variations of the items' theoretical factor structure were tested. Each subset of items tested using the CT model were tested using the unidimensional rapport model (Figure 2). This model assumed a unidimensional latent structure among the items representing a single "rapport" factor.

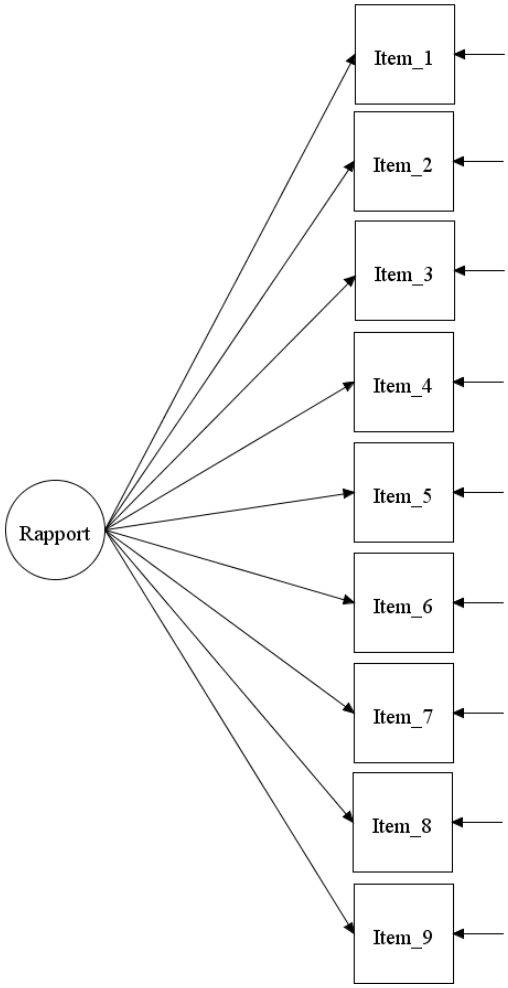


Figure 2. Unidimensional Rapport Model

Variations of an additional multi-dimensional latent factor structure were subsequently tested. The bifactor model (Figure 3) assumed each item loaded to onto a general rapport factor while simultaneously loading onto their respective appropriate orthogonal group rapport component factors.

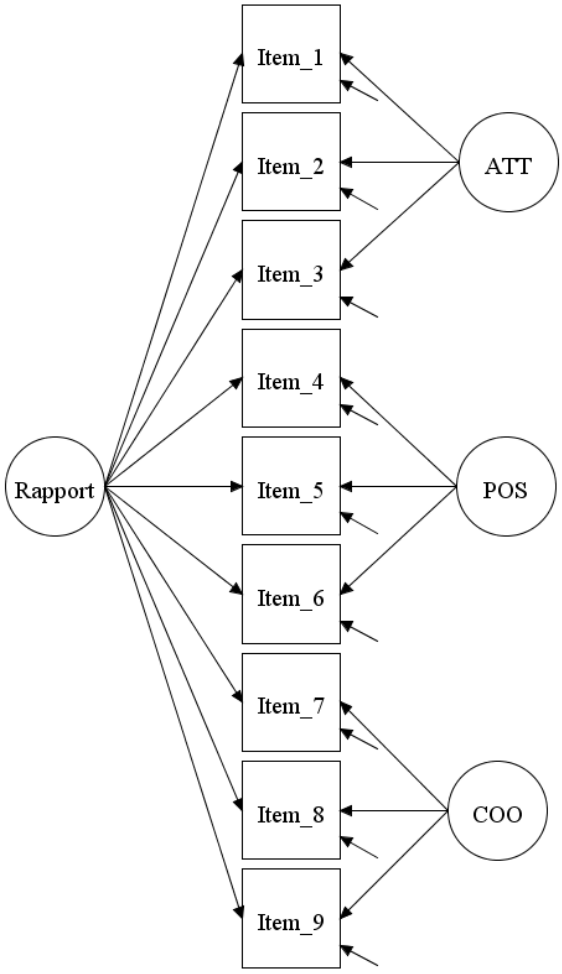


Figure 3. Bifactor Structural Equation Model

Note. ATT = Attentiveness, POS = Positivity, COO = Coordination.

Comparisons of model fit between each SEM were assessed using a variety of fit indices including chi-square, Akaiki Information Criterion (AIC), Root Mean Error of Approximation (RMSEA), and the Comparative Fit Index (CFA). Hu & Bentler's (1999) recommendations for good model fit were initially considered (RMSEA $\leq .06$, CFI ≥ 0.95, and SRMR ≤ 0.08). However, Kenny (2015) notes that, because sampling error is greater for models with small degrees of freedom (df) and low sample size (n), RMSEA in these models can be artificially inflated. For example, using a series of monte carlo simulations, Kenny et al. (2014) illustrated

that models wherein df = 10-20 and n = 100 have a high likelihood of producing RMSEA values greater than 0.08. This has led some researchers to advise against using an RMSEA cutoff of 0.05 or 0.06 (Chen, Curran, Bollen, Kirby, & Paxton, 2008) and others to argue that RMSEA should not even be computed for models with small df and low n (Kenny, Kaniskan, & McCoach, 2014).

However, as RMSEA is one of the most popular fit indices reported in SEM, the present study reported and considered these values. Because the models tested in this present study all have small degrees of freedom (df = 15-25) and relatively low sample size (n = 92), a more liberal interpretation of RMSEA values was taken. MacCallum, Browne, & Sugawara (1996) have recommended using RMSEA values of 0.01, 0.05, and 0.08 to indicate excellent, good, and mediocre fit, respectively, and Kenny (2015) points out that 0.10 may be an appropriate cutoff point to indicate poor fit in these models. Model fit indices was also compared between models such that lower AIC values were considered as an indicator of better model fit (Kenny, 2015). These standards were employed in the present study.

It is also important to note that, when comparing chi-square statistics between models using MLR estimation, it is recommended that a special procedure for computing the value difference between models should be employed (Satorra, 2000; Muthén & Muthén, 2013). A Microsoft Excel Macro developed by Bryant and Satorra (2013) was used to compute these values in the present study.

Based on the results of these analyses, a subset of items from the pool of 30 items was selected for inclusion in the scales of the RS3i-O.

Based on the findings of Magee (2018), it was anticipated that three different factors would emerge from the EFA, each representing one of the three "essential components" of

rapport. Scales were constructed to reflect one of each of these components. Items were selected for each scale if they (a) had relatively high loadings on the factor underlying the scale as shown in the EFA and (b) showed relatively low loadings on the factors underlying the other two scales. Items were also selected on the grounds that they were not highly correlated with other items intended to reflect the same factor in order to reduce redundancy.

Because a goal of the present study was to create a simple, quick measure, as few items as possible were selected for inclusion in each of three RS3i-O scales while maintaining high reliability. Previously findings indicated that a high degree of internal and interrater reliability could be achieved in 3- and 4-item global observer rating scales (Magee, 2018), so it was anticipated that scale development would result in a 9-item, three-scale measure. The degree to which the scales fit to a three-factor model with intercorrelating factors was assessed using Hu & Bentler's (1999) recommendations for good fit (RMSEA $\leq .06$, CFI $\geq .95$, and SRMR $\leq .08$) and was expected to produce a non-significant chi-square statistic.

Part 1 Hypotheses

The analyses outlined above were used to test the following hypotheses:

H1a: The of the pool of 30 observer rapport items will result in three factors interpretable as the essential components of rapport - attentiveness, positivity, and coordination.

H1b: The three factors representing attentiveness, positivity, and coordination will each be significantly, positively, and at least moderately correlated ($r = 0.30$) with the two other factors.

Part 2: Confirmatory Analyses and Scale Validation of the RS3i-O

The purpose of Part 2 of the study was to test the internal and interrater reliability, as well as convergent, discriminant, and criterion validity of the RS3i-O scales. In order to achieve this, RS3i-O scales scores were correlated with data previously collected by Duke et al. (2018) and Magee (2018). A one-way multiple analysis of variance (MANOVA) was performed to assess the relationship between the RS3i-O scales and interviewing styles employed by Duke et al. in order to establish evidence of the scales' criterion validity. Confirmatory factor analysis (CFA) was performed on the final RS3i-O and was expected to confirm the 3-factor structure identified in Part 1. Structural equation models (SEMs) of the final 3-factor RS3i-O and the 4-scale variant including Trust/Respect scale were tested and fit indices were evaluated using Hu & Bentler's (1999) recommendations for good fit (RMSEA $\leq .06$, CFI $\geq .95$, and SRMR $\leq .08$), while considering issues related to inflated RMSEA noted in Part 1, above.

Measures and data from Duke et al. (2018)

In a previous study by Duke et al. (2018), scored the 92 interviews in the present study using several measures including the Rapport Scales for Investigative Interviews and Interrogations (RS3i), the Shared Information Rating Scale (SIRS), and a self-report measure of cooperativeness. Scores for these measures were used in the present study to examine the convergent and criterion validity of the RS3i-O scales. These measures are described in greater detail below.

Rapport Scales for Investigative Interviews and Interrogations (RS3i-S), Source version. The RS3i-S is a 21-item self-report questionnaire on which sources can rate the level of rapport they experienced with an interviewer/interrogator during an investigative interview or

interrogation. Three RS3i-S scales were used to examine convergent validity of the RS3i-O in the current study: Attentiveness, Trust/Respect, and Connected Flow. These three RS3i-S scales demonstrated excellent internal reliability in the study by Duke et al. (2018; Cronbach's alpha = .88, .84, .83, respectively). Further, construct and concurrent validity was established for all three scales. Convergent validity coefficients were large for these scales (average *r*s = .51, .72, and .63 for, respectively). Scale scores were also higher for interviews characterized by rapport behaviors than interviews characterized by pressure tactics and scale scores positively correlated with the amount of information provided during interviews (*r*s = .21, .23, and .20, respectively). The RS3i-S also includes two scales that do not measure the "essential components" of rapport: Expertise and Cultural Similarity. These scales are included in the present study to examine the discriminant validity of RS3i-O scales.

Self-Report Cooperativeness. In the Duke et al. (2018) study, participants rated the degree to which they had been cooperative during the interview using a single self-report item. Participants answered the question "How cooperative were you?" on a 10-point Likert-type scale where 1 represented totally uncooperative and 10 represented totally cooperative.

Shared Information Rating Scale (SIRS). Duke et al. evaluated each interview in terms of the amount of relevant information shared by the source about the evidence video when questioned by the interviewer. The 39-item SIRS was developed with each item representing a relevant fact from the evidence video that could have been mentioned by a source over the course of the interview. SIRS scores were calculated separately for the first half (Phase 1), second half (Phase 2), and for the interview as a whole.

Measures and data from Magee (2018)

A previous study by Magee (2018) scored the 92 interviews in the present study using several observational measures of the essential components of rapport. These measures included the Coordination scale of the Interaction Rapport Scale (IRS) and the Attentiveness and Positivity scales of the Negotiators' Rapport Scale (NRS) developed by Bronstein et al., (2012) along with the Acceptance/Empathy scale adapted from the Global Motivational Interviewing Skills Code Investigative Interview Adaptation (G-MISC) developed by Alison et al. (2013). Scores for these measures were used in the present study to examine the convergent validity of the RS3i-O scales. These measures are described in greater detail below.

Interaction Rapport Scales (IRS). The Interaction Rapport Scale (Bronstein et al., 2012) was designed to rate the three essential components of rapport during negotiations using a transcript of the interaction. The 11-item IRS instructs raters to provide global ratings of Attentiveness, Positivity, and Coordination across eleven items on a 7-point Likert-type scale ranging from 1 (not at all) to 7 (very much). According to Bronstein et al. (2012), scores on these three scales are to be averaged to produce a single score. However, only the 4-item Coordination scale was used by Magee (2018), and this is the only IRS scale that will be used in the present study. Magee (2018) found that scores on the IRS Coordination scale correlated substantially with RS3i Connected Flow (r = .508), other observational measures of rapport (rs > .50), and the amount of information shared during the interview (r = .317).

Negotiators' Rapport Scales (NRS). The NRS (Bronstein et al., 2012) is identical to the Interaction Rapport Scales except that judges are directed to make ratings based on the behavior of one of the individuals during the interaction, rather than the interaction as a whole. Judges rate the interviewer on the eleven 7-point Likert-type items described above and ratings are averaged

to produce a single score including all three scales. In Magee (2018), only the 3-item NRS Attentiveness and 4-item Positivity scales were used by raters, and only these two NRS scales will be used in the present study. These two scales were used to rate the interviewer only and not the source. Magee (2018) found that scores on the NRS Attentiveness scale correlated with RS3i Attentiveness (r = .396), other observational measures of rapport (rs > .50), and the amount of information shared during the interview (r = .216). Similarly, scores on the NRS Positivity scale correlated with RS3i Trust/Respect (r = .314), other observational measures of rapport (rs > .50)

Global Motivational Interviewing Skills Code (G-MISC) Acceptance/Empathy scale. The G-MISC (Alison et al., 2013) is an observer rating instrument that includes five single-item measures allowing raters to rate the degree to which an investigator in an interview adheres to the principles of Motivational Interviewing (MI). Magee (2018) used two G-MISC items, Acceptance and Empathy. Though the G-MISC developers suggested these single-item measures should be treated as independent scales, analyses indicated that they were highly correlated (r *>.90;* Magee, 2018). Further, the Acceptance and Empathy items both appear to reflect aspects of the same "essential three" component: positivity. Thus, Magee (2018) combined these two G-MISC items into a single "G-MISC Acceptance/Empathy" scale, whose scores were calculated by averaging the two items. This G-MISC Acceptance/Empathy scale is the only G-MISC scale used in the present study. Magee (2018) found that scores on the G-MISC Acceptance/Empathy scale correlated substantially with RS3i Trust/Respect (r = .346) and with other observational measures of rapport (rs > .50)

Part 2 Measures

Rapport Scales for Investigative Interviews and Interrogations, Observer version (RS3i-O). The RS3i-O was created based on the results of the EFA described in Part 1. The measure includes three separate 3-item scales, each intended to measure one of the essential components of rapport.

Flesch Reading Ease (RE) Test. RE scores, as described in Part 1, were calculated for the final RS3i-O rating document and comprehension check.

Flesch-Kincaid Grade Level (RGL) Test. RGL scores, as described in Part 1, were also calculated for the final RS3i-O rating document and comprehension check.

Woodcock-Muñoz Language Survey-Revised, Normative Update (WMLS-R NU). Raters' WMLS-R Picture Vocabulary, Verbal Analogies, and Oral Language scores were calculated as described above in Part 1.

Part 2 Procedure

Part 2 used the same interview rating data collected during Part 1, except that RS3i-O secondary team ratings, rather than the primary team ratings, were used to compute item scores used in the confirmatory factor analysis (CFA). Item scores were calculated by taking the average of the two raters' secondary team ratings for each item. Scale scores were calculated by taking the average of the items included in each scale.

Part 2 Analyses

Inter-rater and internal reliability analyses were conducted on each RS3i-O scale using the same analyses and standards used in Part 1. The internal reliability, indexed by Cronbach's

alpha (Cronbach, 1951) and McDonald's omega (McDonald, 1985) was calculated for each scale by SPSS 23 and by hand, respectively. Interrater reliability was assessed for each scale by calculating intraclass correlation coefficient (ICC) between secondary team raters based on a mean-rating, absolute agreement, two-way mixed effects model in SPSS 23. Standards proposed by Hunsley and Mash (2008) for ICCs were used to categorize agreement as adequate (0.70–0.79), good (0.80–0.89), or excellent reliability (>0.90; Hunsley & Mash, 2008).

The convergent validity of each RS3i-O was tested by examining the correlation between each scale and its corresponding RS3i-S scale as well as with the corresponding observer scales collected by Magee (2018). Based on previous findings, it was expected that each RS3i-O scale would correlate at least moderately ($r = 0.30$) with its corresponding RS3i-S scale. Further, it was also expected that each RS3i-O scale would correlate least moderately ($r = 0.30$), but likely higher, with its corresponding observational scales calculated by Magee (2018). The criterion validity of the RS3i-O was tested by examining the correlations between its scales and the information shared during each interview and the sources' self-reported cooperativeness. Criterion validity was also assessed by performing a one-way multiple analysis of variance (MANOVA) comparing mean scores on each RS3i-O scale between the three experimental conditions (Rapport, Neutral, Pressure). Contrasts between each of the conditions were also examined (Rapport vs. Neutral; Rapport vs. Pressure; Neutral vs. Pressure). Each scale was expected to differ by condition and differences between planned contrasts were each expected to be statistically significant.

A confirmatory factor analysis (CFA) was performed to confirm the RS3i-O's hypothesized factor structure established in Part 1. The CFA was performed using maximum likelihood (ML) extraction and direct oblimin rotation, to allow for factor intercorrelation.

Additionally, a structural equation model (SEM) of the hypothesized factor structure established in Part 1 was tested using MPlus 7. The model's fit was assessed using Hu & Bentler's (1999) recommendations for good fit (RMSEA ≤ .06, CFI ≥ .95, and SRMR ≤ .08) and was expected to result in a non-significant scaled chi-square statistic. Given the low degrees of freedom, small n, and high degree of expected intercorrelation between items in this model, an artificially inflated RMSEA were expected. While the model was evaluated using Hu & Bentler's recommendations, considerations outlined in Part 1 analyses were made regarding the potential for an inflated RMSEA.

Part 2 Hypotheses

The analyses outlined above resulted in the following hypotheses:

Convergent Validity.

H2a: RS3i-O Attentiveness will correlate positively and at least moderately (r = .30) with RS3i-S Attentiveness.

H2b: RS3i-O Irritability will correlate Negatively and at least moderately (r = -.30) with RS3i-S Trust/Respect.

H2c: RS3i-O Coordination will correlate positively and at least moderately (r = .30) with RS3i-S Connected Flow.

H2d: RS3i-O Trust/Respect will correlate positively and at least moderately (r = .30) with RS3i-S Trust/Respect.

H3a: RS3i-O Attentiveness will correlate positively and at least moderately (r = .30) with NRS Attentiveness.

H3b: RS3i-O Irritability will correlate Negatively and at least moderately (r = -.30) with NRS Positivity, and GMISC Acceptance/Empathy.

H3c: RS3i-O Coordination will correlate positively and at least moderately (r = .30) with IRS Coordination.

H3d: RS3i-O Trust/Respect will correlate positively and at least moderately (r = .30) with NRS Positivity and GMISC Acceptance/Empathy.

Discriminant Validity.

H4a: Each RS3i-O scale will correlate at least 0.10 less with the RS3i-S and observational scales that measure the other two "essential components" than with the scales intended to measure the same component.

H4b: Each RS3i-O scale will correlate less than r = .20 with the RS3i-S Expertise and Cultural Similarity scales, as they are not "essential components" of rapport.

Criterion Validity.

H5: Each RS3i-O scale correlate at least moderately (r = .30) with SIRS scores for Phase 2 of each interview and the interview as a whole.

H6: Each RS3i-O scale correlate at least moderately (r = .30) with self-report cooperativeness scores.

H7a: Mean ratings for each of the RS3i-O scales will differ significantly by experimental condition (Rapport, Neutral, Pressure).

H7b: Mean ratings for each of the RS3i-O scales will differ significantly by each contrast of the experimental conditions (Rapport vs. Pressure; Rapport vs. Neutral; Neutral vs. Pressure) such that ratings will be higher in the Rapport condition than either the Neutral and Pressure conditions and higher in Neutral condition than in Pressure.

Factorial Validity.

H8a: CFA of the RS3i-O will result in a 3-factor solution that produces a non-significant chi-square statistic.

H8b: The RS3i-O model structure identified in Part 1 will result in indices that meet Hu & Bentler's (1999) recommendations for good fit (RMSEA ≤ .06, CFI ≥ .95, and SRMR ≤ .08).

H9a: CFA of the RS3i-O including the Trust/Respect scale will result in a 3-factor solution that produces a non-significant chi-square statistic.

H9b: The RS3i-O model structure identified in Part 1 including the Trust/Respect scale will result in indices that meet Hu & Bentler's (1999) recommendations for good fit (RMSEA ≤ .06, CFI ≥ .95, and SRMR ≤ .08).

Part 3: Confirmatory Analysis and Validation of LIWC Variables and LSM

Measures and data from Duke et al. (2018)

The data, described above, originally collected by Duke et al. (2018) and used in Part 2 of the present study were used again in Part 3, including the included the RS3i-S, SIRS, self-report cooperativeness scales.

Measures and data from Magee (2018). The data, described above, originally collected by Magee (2018) and used in Part 2 of the present study were used again in Part 3, including the NRS Attentiveness, NRS Positivity, GMISC Acceptance Empathy, IRS Coordination scales.

Measures and data from Part 2 of the Present Study. The RS3i-O scales, including the Attentiveness, Irritability, Coordination, and Trust/Respect scales, were used in Part 3 analyses.

Part 3 Materials

Simulated interview videos. The Duke (2018) simulated investigative interview videos (n = 92) used in Part 1 were used in again in Part 3.

Transcription Instructions. This document describes the general instructions for the transcription process for the present study including detailed instructions on how to address spelling, punctuation and, non-fluencies.

Part 3 Measures

Linguistic Inquiry and Word Count (LIWC) 2015. LIWC software (Pennebaker et al., 2015) was used to analyze the transcribed text of the interviews. Also, LIWC categories previously identified by Driskell et al. (2013) as potential indicators of the essential components of rapport were calculated (e.g., positive language as an indicator of the positivity component). Each of the LIWC categories used in the present study, along with specific examples of their content and their hypothesized components of rapport are listed below in Table 4.

Table 4

Detailed LIWC Categories by Rapport Component (adapted from Driskell et al., 2013)

Rapport component	LIWC categories	Example words	Words in category
Mutual Attentiveness	- First person plural	we, us, our	24
	- Present focus	today, now, is	424
	- Social processes	mate, talk, they	756
Positivity	- Positive emotion	love, nice, sweet	620
	- Negative emotion	hurt, sad, worried	744
Coordination	- Nonfluencies	er, hm, umm	19
	- Assent	agree, OK, yes	36
	- Certainty	always, never	113
	- Conjunctions	and, but, whereas	43

Language style matching (LSM) values were also calculated for each interview. LSM scores were calculated using the difference between two speakers' use of the nine function word categories provided by LIWC (see Table 2). Scores were calculated in three steps. First, the use of each of the nine categories of function words (expressed as a proportion of total words used) was calculated for each speaker using LIWC. Second, LSM scores were calculated across each

category of function words for each dyad using the following formula originally provided by Ireland and Pennebaker (2010):

$$\mathrm{LSM}_{function} = 1 - \frac{|\mathrm{function}_{\mathrm{speaker\ 1}} - \mathrm{function}_{\mathrm{speaker\ 2}}|}{\mathrm{function}_{\mathrm{speaker\ 1}} + \mathrm{function}_{\mathrm{speaker\ 2}} + .0001}$$

Third, LSM scores for each function word category were averaged into a composite LSM score that reflects the degree of similarity in language style. LSM scores range from 0 to 1 with greater values indicating greater linguistic convergence. LSM scores and scores for the nine LIWC variables for both interviewers and sources were used in the present study to examine these scores' convergent validity with the RS3i-O scales and other rapport measures in the present study.

Part 3 Procedure

The purpose of Part 3 of the study was to examine the validity of LIWC variables as indicators of rapport. All 92 simulated interview videos created by Duke et al. (2018) were transcribed and scored using LIWC 2015. LIWC scores for variables thought to be indicators of rapport by Driskell et al. (2013) were correlated with ratings of rapport collected by Duke et al. (2018), Magee (2018), and in Part 1 of the present study to test the linguistic variables' validity. It was expected that LIWC variables thought to reflect each component of rapport would correlate more highly with the self-report and observational ratings scales intended to measure the same component than with scales intended to measure the other two components. For example, LIWC indicators of coordination (e.g., non-fluencies, assent) were expected to correlate with the RS3i-S Connected Flow scale and the RS3i-O Coordination scale. Language

style matching (LSM) scores were calculated using the method described above and correlated with all self-report and observer rapport scales in addition to all non-test criteria used to evaluate the RS3i-O in Part 2.

Transcription procedures took place in three phases: (1) a team of trained research assistants watched all 92 interviews and transcribed each speaker's words verbatim; (2) a second team of research assistants verified the completion and accuracy of the transcript; (3) a third team of RAs then coded each transcript according to the LIWC 2015 manual (Pennebaker et al., 2015), and then split each transcript into separate interviewer and source transcripts and analyzed in LIWC. Split transcript files were then inputted to the LIWC software for analysis to produce scores for linguistic variables for interviewers and sources. Function word variable scores on each split transcript were used to calculate an LSM value for each interview.

In Phase 1, the trainer first read the transcription instructions (Appendix H) aloud to the team of research assistant as they read along on the provided document. After gaining verbal acknowledgement of comprehension of the instructions, each RA was assigned a set of interview videos to transcribe. RAs listened to the audio of each interview while transcribing the interaction verbatim into a spreadsheet using Microsoft Excel 16. Transcribers were instructed to transcribe every word verbatim, including grammatical mistakes, informalities, and contractions. All non-fluencies (e.g., "er", "uh", "hm") were coded as "um". This was done to increase the speed of the transcription process while reducing the potential for mistaken variations that may not be included in the LIWC (e.g., "urm"). As RAs' schedules allowed, transcription shifts were kept under three hours and spaced out along the week in order to reduce fatigue.

In Phase 2, a different team of RAs again listened to the audio of each interview while reviewing the transcriptions made in Phase 1, ensuring accuracy, checking for spelling and punctuation errors, accuracy in non-fluency coding, and consistency of coding.

In Phase 3, a final team of RAs processed the transcription files for analysis in LIWC. Filler words were specifically coded to be read by LIWC (e.g., When the word "like" was used as a meaningless filler, it was coded as "rrlike" so LIWC would read it as filler word instead of considering its preposition, conjunction, noun, or adjective forms.) Other than coding specifically necessary for LIWC, transcripts files remained completely verbatim.

Part 3 Analyses

Linguistic indicators of rapport for both interviewers and sources were correlated with RS3i-S, RS3i-O, NRS, IRS, and G-MISC scale scores. It was expected that LIWC variables and LSM would be weakly correlated ($r = 0.20$) with their corresponding scales on the self-report and observational rapport measures though some variables were expected to produce larger correlations due to their face validity and the direct link to their theoretical constructs (e.g., LIWC use of positive emotion and RS3i-S Trust/Respect). The LIWC variables and LSM were also correlated with non-test criteria including the total information shared in each interview (SIRS) and self-report cooperativeness collected by Duke et al. (2018). It was expected that LIWC variables and LSM would be moderately correlated ($r = 0.30$) with these measures of source cooperativeness.

Part 3 Hypotheses

The analyses outlined above resulted in the following hypotheses:

Convergent Validity.

H10a: Use of first personal plural, present focus and social processes will be at least weakly correlated (r = 0.20) with RS3i-S Attentiveness, RS3i-O Attentiveness, and NRS Attentiveness scales.

H10b: Use of positive emotion and negative emotion will be at least weakly correlated (r = 0.20) with RS3i-S Trust/Respect, RS3i-O Irritability, RS3i-O Trust/Respect, NRS Positivity, and GMISC Acceptance/Empathy scales.

H10c: Use of nonfluencies, assent, certainty, and conjunctions as well as LSM scores will be at least weakly correlated (r = 0.20) with RS3i-S Connected Flow, RS3i-O Coordination, and IRS Coordination scales.

Criterion Validity.

H11: LIWC variables and LSM scores will correlate at least moderately (r = .30) with SIRS scores for Phase 2 of each interview and the interview as a whole.

H12: LIWC variables and LSM scores will correlate at least moderately (r = .30) with self-report cooperativeness scores.

H13a: Mean ratings for each of the LIWC variables and LSM will differ significantly by experimental condition (Rapport, Neutral, Pressure).

H13b: Mean ratings for each of the LIWC variables and LSM will differ significantly by each contrast of the experimental conditions (Rapport vs. Pressure; Rapport vs. Neutral; Neutral vs. Pressure) such that ratings will be higher in the Rapport condition than either the Neutral and Pressure conditions and higher in Neutral condition than in Pressure.

Results

Overview

Analyses in the present study were conducted in three separate parts. In Part 1, exploratory factor analyses (EFA) were performed on the pool of 30 observer rapport items (Appendix B.) in order to test their hypothesized 3-factor structure. Structural equation modeling (SEM) was performed to further explore the items' factor structure. The results of these analyses were used to construct and refine the scales of the Rapport Scales for Investigative Interviews and Interrogations, Observer version (RS3i-O).

In Part 2, the psychometric properties of the RS3i-O were examined. The internal and inter-rater reliabilities of each of the newly developed RS3i-O scales was calculated. Confirmatory factor analysis (CFA) and SEM were performed on the RS3i-O scales in order to confirm the instrument's factor structure identified in Part 1. Analyses were then performed to assess each of the RS3i-O scales' criterion, convergent, and discriminant validity.

In Part 3, the interviews rated in Parts 1 and 2 were transcribed and analyzed using Linguistic Inquiry and Word Count (LIWC) 2015. LIWC variable scores were calculated and language style matching (LSM) scores were calculated using those variable scores. LIWC variable and LSM scores were correlated with the self-report and observer rapport measures used in Part 2 analyses (RS3i-S, NRS, IRS, and GMISC), including the RS3i-O scales.

The analyses present study is summarized below in Table 5.

Table 5

Summary of Study Analyses

Analysis	Sample	Purpose
Part 1		
Exploratory factor analyses Structural equation modeling	9 trained observer ratings of 92 interviews	Explore the factor structure of rapport items Construct RS3i-O scales
Part 2		
Scale analyses	6 trained observer ratings of 92 interviews	Determine internal and interrater reliability of RS3i-O scales
Relationship between RS3i-O and RS3i-S scale scores (correlation)	6 trained observer ratings of 92 interviews 92 source RS3i-S ratings	Examine convergent validity of RS3i-O observer scales with corresponding RS3i-S source scales
Relationship between RS3i-O and other observer rapport scale scores (correlation)	6 RS3i-O ratings and 6 NRS, IRS, & GMISC ratings of 92 interviews	Examine convergent validity of RS3i-O observer scales with corresponding NRS, IRS, and GMISC observer scales
Relationship between RS3i-O scale scores and source cooperation (multiple regression)	6 trained observer ratings of 92 interviews	Predict source cooperation from RS3i-O scale scores
Group comparison: RS3i-O scale means by interview condition (MANOVA)	6 trained observer ratings of 92 interviews	Examine the concurrent validity of the RS3i-O by comparing scale means between interviewing styles
Confirmatory factor analysis	6 trained observer ratings of 92 interviews	Confirm factor structure of the RS3i-O
Structural equation modeling	6 trained observer ratings of 92 interviews	Confirm factor structure of the RS3i-O

Table 5 continued

Summary of Study Analyses

Analysis	Sample	Purpose
Part 3		
Linguistic variable analysis	184 interviewer & source transcripts of 92 interviews	Determine reliability of linguistic variables
Relationship between linguistic variables and rapport scale scores (correlation)	184 interviewer & source transcripts of 92 interviews 6 trained observer RS3i-O ratings 6 observer NRS, IRS, & GMISC ratings 94 source RS3i-S ratings	Examine convergent validity of linguistic variables
Relationship between linguistic variables and source cooperation (multiple regression)	184 interviewer & source transcripts of 92 interviews	Predict source cooperation from linguistic variables
Group comparison: linguistic variable means by interview condition (MANOVA)	184 interviewer & source transcripts of 92 interviews	Examine the concurrent validity of the linguistic variables by comparing means between interviewing styles

Part 1: Exploratory Analyses

The primary goal of Part 1 was to reduce the pool of 30 observer rapport items to a subset of items including the minimum number necessary to produce scales that (1) demonstrate adequate internal and inter-rater reliability, and (2) capture each of the essential components of rapport proposed by Tickle-Degnen and Rosenthal (1990), and (3) are quick and easy to employ in an applied setting. Because Magee (2018) found that scales comprised of as few as three or four items have the potential to accomplish all of these goals, it was expected that the final measure would contain nine items making up three 3-item scales.

First, the agreement between the three raters on each team was assessed and the two raters with the highest level of agreement were chosen as the Final Raters for that team. Next, a series of exploratory factor analyses (EFAs) were conducted to determine the items' factor structure and evaluate individual items' characteristics within the context of (a) each hypothesized scale and (b) the pool of 30 items as a whole. Based on these results, the number of included items was reduced, and structural equation modeling (SEM) was employed to evaluate the remaining items' factor structure.

A detailed description of results of these analyses follows.

Calculation of Item Ratings for Each Team and Item Characteristics

Choosing Final Raters for Rating Teams.

In order to determine which two raters on each team had the highest degree of agreement, and would thus be retained as the Final Raters for that team, the inter-rater reliability of each item included in primary team ratings was assessed between the three raters on each team. The intraclass correlation coefficients (ICCs) based on an absolute-agreement, two-way mixed effects

model, were calculated for each item. The average ICCs for the Attentiveness Team are displayed below in Table 6, for the Positivity Team in Table 7, and for the Coordination Team in Table 8, with the largest average ICC in bold. The rater with the lowest average ICC for the items of their primary team ratings was then disqualified for further analyses in Parts 1, 2, and 3. The final item scores were calculated by taking the average of each item between the remaining two raters ("Final Raters") of each team. This process is described in more detail on page 57 below.

Table 6

Average ICC of Attentiveness Items Between Attentiveness Team Raters (n = 92)

	Rater A1	Rater A2	Rater A3
1. Rater A1	-	0.43	**0.624**
2. Rater A2	0.43	-	0.339
4. Rater A3	**0.624**	0.339	-

Note. Largest Average ICC appears in bold

Raters A1 and A3 clearly had a much higher average ICC (0.624) than that between raters A1 and A2 (0.430) or between raters A2 and A3 (0.339). Based on these results, ratings made by rater A2 were eliminated from all further analyses, and Raters A1 and A3 were selected as the two Final Raters for the Attentiveness Team.

Table 7

Average ICC of Positivity Items Between Positivity Team Raters (n = 92)

	Rater P1	Rater P2	Rater P3
1. Rater P1	-	0.746	0.725
2. Rater P2	0.746	-	**0.777**
4. Rater P3	0.725	**0.777**	-

Note. Largest average ICC appears in bold.

Raters P2 and P3 had a slightly higher average ICC (0.777) than that between raters P1 and P2 (0.746) or between raters P1 and P2 (0.725). Based on these results, ratings made by rater P1 were eliminated from all further analyses, and Raters P2 and A3 were selected as the two Final Raters for the Positivity Team.

Table 8

Average ICC of Coordination Items Between Coordination Team Raters (n = 92)

	Rater C1	Rater C2	Rater C3
1. Rater C1	-	0.323	**0.664**
2. Rater C2	0.323	-	0.267
4. Rater C3	**0.664**	0.267	-

Note. Largest average ICC appears in bold.

Raters C1 and C3 had a higher average ICC (0.664) than that between raters C1 and C2 (0.323) or between raters C2 and C3 (0.276). Based on these results, ratings made by rater C2

were eliminated from all further analyses, and Raters D1 and D3 were selected as the two Final Raters for the Coordination Team.

Calculation of Item Ratings.

The item ratings used in subsequent analyses were calculated using ratings made by the two Final Raters of each rating team. Specifically, the average between the two Final Raters on each team was calculated for each item. These averaged item ratings were either labeled as "Primary Item Ratings," which were used for all Part 1 analyses, or "Secondary Item Ratings," which were used for all Part 2 and Part 3 analyses.

Primary Item Ratings were calculated by averaging between the Final Raters for each of the items on their assigned Primary Rating scale (see Table 3, page 18). To calculate Primary Item Ratings for each of the Attentiveness items (1-10), the average rating for that item between Final Raters on the Attentiveness Team (Raters A1 and A3) was calculated. Next, to calculate Primary Item Ratings for each of the Positivity items (11-20), the average rating for that item between Positivity Team Final Raters (Raters P2 and P3) was calculated. Last, to calculate Primary Item Ratings for each of the Coordination items (21-30), the average rating for that item between Final Raters on the Positivity Team (Raters P2 and P3) was calculated. Primary Item Ratings were used for all Part 1 analyses.

Secondary Item Ratings were calculated using the same method described above, except ratings between the two Final Raters for each of the items on their assigned Secondary Rating scale were used (Table 3). To calculate Secondary Item Ratings for each of the Attentiveness items, the average rating for that item between Final Raters on the Coordination Team was calculated. Similarly, to calculate Secondary Item Ratings for each of the Positivity items, the

average rating for that item between Attentiveness Team Final Raters was calculated. Last, to calculate Secondary Item Ratings for each of the Coordination items, the average rating for that item between Final Raters on the Positivity Team was calculated. Secondary Item Ratings were used in all Part 2 and 3 analyses.

It is also important to note that reverse-coded item (e.g., 5. Distracted) scores were calculated (0=6, 1=5, 2=4, 3=3,4=2, 5=1, 6=0) and used in some analyses to aid in the interpretation of some indices (e.g., internal reliability analyses).

Item Characteristics.

While some items' distributions approximated normality (e.g., 1. Actively Listening and 23. Productive), most items were not normally distributed. For example. some items' distributions exhibited strong negative (e.g., 7. Focused on Source = -1.18) or positive (e.g., 13. Frustrated = 1.21) skew. Other item characteristics were generally as expected in that no outliers were identified and no data were unexpectedly missing. Item means, standard deviations, skewness statistics, and intraclass correlation coefficients (ICCs) are listed below in Table 9.

The intraclass correlation coefficient (ICC) was calculated for every item of the primary team ratings based on an absolute-agreement, two-way mixed effects model. Two ICC values were calculated for each item. In Table 9, "Primary ICC" show the agreement between the two Final Raters from each team for their "primary items." For example, the Primary ICC for the Attentiveness scale (items 1-10) indicates the agreement between the Final Raters on the Attentiveness Team (Raters A1 and A3; see Table 6). The "Secondary ICC" listed for each item refers to the absolute agreement among the four Final Raters not included in the Primary ICC. For example, the Secondary ICCs for Attentiveness scale (items 1-10) indicate the agreement

among the Final Raters from the Positivity (Raters P2 and P3) and Coordination (Raters C1 and C3) teams.

Table 9

30 Observer Rapport Items Means, Standard Deviations, Skewness Statistics, and ICCs (n = 92)

Hypothesized Rapport Component	Item	Mean *(SD)*	Skew	Primary ICC	Secondary ICC
Attentiveness	1. Actively Listening	4.38 *(1.05)*	-0.75	.414	.456
	2. Alert	2.90 *(1.50)*	-0.04	.765	.579
	3. Interested	3.66 *(1.20)*	-0.23	.353	.548
	4. Involved	3.84 *(1.34)*	-0.48	.707	.539
	5. Distracted*	0.57 *(0.82)*	1.64	.617	.310
	6. Invested	3.92 *(1.17)*	-0.79	.496	.401
	7. Focused on Source	4.73 *(0.86)*	-1.18	.735	.469
	8. Thoughtful	3.32 *(1.66)*	-0.16	.713	.297
	9. Bored*	1.39 *(1.44)*	0.94	.749	.602
	10. Attentive	4.63 *(0.89)*	-0.82	.689	.469
Positivity	11. Trusting	3.25 *(1.68)*	-0.35	.810	.888
	12. Understanding	2.89 *(1.67)*	0.34	.421	.900
	13. Frustrated*	1.03 *(1.45)*	1.21	.813	.840
	14. Approachable	2.43 *(2.04)*	0.33	.871	.884
	15. Respectful	2.99 *(2.11)*	-0.26	.881	.938
	16. Honest	3.41 *(1.56)*	-0.36	.655	.338
	17. Aggressive*	1.43 *(1.95)*	0.91	.863	.941
	18. Pleasant	2.52 *(2.05)*	0.19	.843	.900
	19. Relaxed *(Interviewer)*	3.51 *(1.60)*	-0.61	.750	.851
	20. Positive	2.23 *(2.17)*	0.52	.818	.889
Coordination	21. Peaceful	3.95 *(1.22)*	-0.72	.853	.826
	22. Awkward*	2.28 *(1.11)*	0.39	.448	.629
	23. Productive	3.41 *(1.25)*	-0.25	.726	.703
	24. Tense*	2.04 *(1.38)*	0.68	.761	.846
	25. Smooth	3.77 *(0.98)*	-0.47	.513	.652
	26. Cooperative	3.50 *(1.19)*	-0.48	.748	.751
	27. Relaxed *(Atmosphere)*	3.79 *(1.32)*	-0.71	.894	.859
	28. Communicative	3.49 *(1.13)*	-0.28	.706	.761
	29. Shared Expectations	3.78 *(1.01)*	-0.46	.477	.750
	30. Coordinated	3.85 *(0.89)*	-0.48	.513	.712

Note. *Reverse-coded item. Primary ICCs calculated between primary team raters for each scale.

Secondary ICCs calculated between remaining four raters not included in primary ICCs.

It is important to note that two items (19 and 27) share identical content ("Relaxed"), though they refer to two distinct characteristics of the interaction. Item 19 refers to the degree to which the interviewer seems relaxed while item 27 refers to the degree the atmosphere of the interaction is relaxed. There was no need to disambiguate between the two for raters, because the target of each item was made clear in the rating instructions. However, in order to disambiguate between the two in the following pages, both items appear with an italicized, parenthetical descriptor indicating the target of the item as described to raters in their respective training materials. Item 19 appears as Relaxed *(Interviewer)*, and item 27 appears as Relaxed *(Atmosphere)* in tables and text below. These descriptors are for purposes of disambiguation in this text only and did not appear in the actual item content in the materials provided to raters.

Correlations Among Item Ratings.

Correlations were calculated among the ratings of the 30 Observer Rapport items listed in Table 9. The resulting correlation matrix was too large to be effectively displayed here and is located in Appendix I. It is important to note that, as expected, the items all displayed a high degree of intercorrelation, even between items intended to measure different components of rapport.

Exploratory Factor Analyses (EFA) of Pool of Observer Rapport Items

Exploratory factor analyses (EFAs) were performed to determine the factor structure of all 30 items and which combinations of items were most likely to produce scales with the desired qualities enumerated above on page 54. Items were initially identified for inclusion in subsequent

analyses based on the degree to which they saliently loaded onto each factor. Items that loaded most strongly onto their hypothesized factor were retained for inclusion in the final measure.

EFA of Individual Hypothesized Scales.

First, MPlus 7 was used to test the fit of each subset of 10 items thought to comprise each factor scale onto a single factor model. Standardized item loadings for each single factor model are located in Appendix J. The item information of each item was calculated by taking the ratio of each standardized item loading to the standardized item residual variance. SPSS 23 was used to compute Cronbach's alpha for each subset of 10 items thought to comprise each factor scale, the alpha that would result from each item's deletion, and each item's corrected item-total correlation value. These scale and item statistics are also located in Appendix J. These additional item statistics, as well as theoretical concerns related to each items' content, were also considered in decision to retain or reject items for inclusion in the final RS3i-O.

Each 10-item, single factor scale had high reliability ($\alpha s > .90$). Standardized item loadings were generally high for items on their respective scales ($\lambda s > .70$) with the exception of items 22. Awkward ($\lambda = .478$) and 25. Smooth ($\lambda = .681$) on the Coordination scale. Each subset of items was highly reliable, so the Cronbach's alpha if item deleted value was not useful in determining problematic items.

EFA Including all 30 Observer Rapport Items.

Tests were first performed by SPSS to assess the degree of interrelationship between items and, thus, the appropriateness of employing factor analysis on the data. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy indicates the proportion of the items' variance that

may be shared among items. Kaiser and Rice (1974) suggested the following values for interpreting the KMO test: .90s marvelous, .80s meritorious, .70s middling, .60s mediocre, .50s miserable, <.50 unacceptable. The KMO test thus indicates whether there is a high enough degree of inter-item correlation for the data to be appropriate for factor analysis.

The KMO value for the 30 Observer Rapport items, KMO = .931, indicated a strong interrelationship among the pool of 30 rapport items. Bartlett's test of sphericity, which tests the null hypothesis that the data's correlation matrix is an identity matrix, indicated a high degree of inter-item correlation among the rapport items (χ^2 (435) = 4333.91, $p < .001$). These results supported the use of factor analysis on these data.

Using SPSS 23, an EFA using maximum likelihood (ML) extraction based on eigenvalues above one was performed on all 30 rapport items. The direct oblimin rotation method was employed, as the hypothesized factors were expected to be significantly intercorrelated. Small factor loadings (below .40) were suppressed from appearing in the resulting matrices and were thus not considered in these analyses. While a factor loading cutoff value of .32 has been recommended (Tabachnick & Fidell, 2013), the extracted factors were expected to be highly intercorrelated, and it was expected that some items would cross-load onto multiple factors. Table 10 displays the standardized factor loadings for the resulting 3-factor solution, with loadings below .40 suppressed.

Table 10

Standardized Factor Loadings for EFA of Pool of 30 Observer Rapport Items (n = 92)

Rapport Item	Factor 1	Factor 2	Factor 3
1. Actively Listening		.658	
2. Alert		.819	
3. Interested		.839	
4. Involved		.862	
5. Distracted		-.821	
6. Invested		.916	
7. Focused on Source		.863	
8. Thoughtful	.524	.605	
9. Bored		-.933	
10. Attentive		.685	
11. Trusting	.851		
12. Understanding	.690		
13. Frustrated	-.933		
14. Approachable	.714		
15. Respectful	.924		
16. Honest	.825		
17. Aggressive	-1.013		
18. Pleasant	.816		
19. Relaxed *(Interviewer)*	.938		
20. Positive	.679	.438	
21. Peaceful	.851		
22. Awkward			
23. Productive			.960
24. Tense	-.788		
25. Smooth			
26. Cooperative			.470
27. Relaxed *(Atmosphere)*	.867		
28. Communicative			.925
29. Shared Expectations			.696
30. Coordinated			.572

The first EFA factor included items intended to measure aspects of positivity and was thus tentatively titled Positivity. The second factor included items that measured aspects of attentiveness and was tentatively titled Attentiveness. The third factor included items intended to measure aspects of interpersonal coordination and so was tentatively titled Coordination. Initial eigenvalues indicated that the first three factors explained 57.38%, 15.46%, and 8.90% of the variance among items, respectively. The 3-factor solution explained 82.19% of the variance. However, a Chi-square test of goodness of fit indicated overall poor model fit (χ^2 (348) = 866.75, *p <.001*). The resulting scree plot (Appendix K) clearly leveled off at 3 factors.

It was expected that the EFA of the pool of 30 observer rapport items would result in three factors interpretable as the essential components of rapport: Attentiveness, Positivity, and Coordination (Hypothesis 1_a) and that these factors would each be at least moderately correlated (r = 0.30) with the two other factors (Hypothesis 1_b). Based on the results above, Hypothesis 1_a was supported and a 3-factor solution was retained. Each of the RS3i-O model factors was expected to be at least moderately (*r* = 0.30) with the two other scales (Hypothesis 1b). Table 11 shows the intercorrelations between the three factors.

Table 11

Pool of 30 Observer Rapport Items Factor Intercorrelations (n = 92)

Factor	1	2	3
1. Attentiveness	-	.723*	.492*
2. Positivity	.723*	-	.845*
3. Coordination	.492*	.845*	-

Note. *Correlation is significant at the .001 level.

Hypothesis 1b was supported in that there were moderate-to-strong correlations between the Positivity and Attentiveness factors (r = .723), the Positivity and Coordination factors (r = .845), and the Attentiveness and Coordination factors (r = .492).

A primary goal of Part 1 analyses was to reduce the number of items included in each factor scale. The final measure was expected to include three 3-item scales to facilitate quickness and ease of use; however, ensuring the scales' reliability was paramount. For this reason, the pool of rapport items was reduced in steps to explore several variations of scales that may demonstrate both ease of use and reliability. The high degree of intercorrelation among items and between factors suggested that the first step in removing items from consideration was to identify items that strongly cross-loaded onto multiple factors. Items that did not load saliently onto any factor were also identified at this time. This was done with the intention of choosing items that would ensure the most orthogonal scales possible while effectively capturing the target constructs.

Two items (8. Thoughtful and 20. Positive) loaded onto the Attentiveness and Positivity factors simultaneously, while three Coordination items (21. Peaceful, 23. Tense, and 27. Relaxed *(Atmosphere)*) loaded strongly (>.70) onto the Positivity factor. One Coordination item (Awkward) did not produce a factor loading that met the .40 threshold for presentation in the pattern matrix while another had a low loading close to the limit (Smooth; .470).

Based on these results, these seven were removed from the pool of potential items. An EFA was then performed on the remaining 23 observer rapport items, as described in the next subsection. Statistics related to the appropriateness of EFA and the general factor structure of the items were not expected to change meaningfully with each subset of items being analyzed, but were calculated out of due diligence and reported here.

EFA Including Remaining 23 Observer Rapport Items.

As described in the previous subsection, seven items were removed from the pool of items. An EFA with maximum likelihood (ML) extraction and direct oblimin rotation was then performed on the remaining 23 rapport items. KMO and Bartlett's tests supported the appropriateness of factor analysis on the remaining items (KMO = .927; χ^2 (253) = 3134.57, $p <$ *.001*). Table 12 displays the standardized factor loadings for the resulting three-factors with those below .40 suppressed.

Table 12

Standardized Factor Loadings for EFA of 23 Observer Rapport Items (n=92)

Rapport Item	Factor 1	Factor 2	Factor 3
1. Actively Listening		.656	
2. Alert		.809	
3. Interested		.829	
4. Involved		.857	
5. Distracted		-.834	
6. Invested		.914	
7. Focused on Source		.856	
9. Bored		-.950	
10. Attentive		.683	
11. Trusting	.852		
12. Understanding	.699		
13. Frustrated	-.905		
14. Approachable	.714		
15. Respectful	.915		
16. Honest	.838		
17. Aggressive	-.983		
18. Pleasant	.819		
19. Relaxed (Interviewer)	.929		
23. Productive			.963
26. Cooperative			.923
28. Communicative			.935
29. Shared Expectations			.703
30. Coordinated			.571

Initial eigenvalues indicated that the first three factors (interpretable as Positivity, Attentiveness, and Coordination) explained 57.04%, 17.36%, and 10.30% of the variance among items, respectively. The resulting 3-factor solution explained 84.69% of the variance among items, though the chi-square goodness of fit test statistic remained significant (χ^2 (187) = 434.76, $p < .001$), indicating poor fit.

The remaining items with the lowest loadings on the factors attributed to Attentiveness and Positivity were identified and removed. No additional Coordination factor items were removed during this step because several had been removed in the previous step. Among the Attentiveness items, 1. Actively Listening (.656), 2. Alert (.809), 3. Interested (.829) and 10. Attentive (.683) were identified as the lowest loadings. Among the Positivity items, 12. Understanding (.699), 14. Approachable (.714), 16. Honest (.838) .and 18. Pleasant (.819) were identified as the lowest loadings. These eight items were removed and an EFA was performed on the remaining 15 items, as described in the next subsection.

EFA Including Remaining 15 Observer Rapport Items.

As explained in the previous subsection, so the final RS3i-O measure was expected to contain three 3-item scales. However, it was initially unclear if the scales formed from only three items would maintain reliability, so the pool of items was first reduced to a subset of 15 items to provide alternative scales should the 3-item scales' reliability prove inadequate. Thus, eight items were removed from the pool of remaining items in order to form three potential 5-item scales. An EFA with maximum likelihood (ML) extraction and direct oblimin rotation was performed on the remaining 15 rapport items. KMO and Bartlett's tests still indicated the appropriateness of factor analysis (KMO = .788; χ^2 (36) = 882.41, $p < .001$). Table 13 displays the standardized factor loadings for the resulting three factors (loadings below .40 suppressed) which explained 85.01% of the variance among these items, though a test of goodness of fit still indicated poor fit (χ^2 (63) = 146.12, $p < .001$).

Table 13

Standardized Factor Loadings for EFA of 15 Observer Rapport Items (n=92)

Item	Factor 1	Factor 2	Factor 3
4. Involved			.827
5. Distracted			-.850
6. Invested			.872
7. Focused on Source			.859
9. Bored			-.935
11. Trusting	.919		
13. Frustrated	-.877		
15. Respectful	.757		
17. Aggressive	-.985		
19. Relaxed (Interviewer)	.905		
23. Productive		.944	
26. Cooperative		.918	
28. Communicative		.941	
29. Shared Expectations		.708	
30. Coordinated		.583	

The original intention of the Part 1 analyses was to create scales with as few items as possible in order to promote practical ease of use while maintaining psychometric soundness. It was anticipated that each factor scale would contain three items, creating a 3-factor, 9-item instrument. So, the two items with the lowest loadings were then removed from each scale. Two items (4. Involved and 5. Distracted) were removed because they had the lowest loadings (.832 and -.850, respectively) of the remaining Attentiveness items. One of these items (4. Involved) was also highly correlated with the Positivity factor (.438). Similarly, item 11. Trusting (.825) and 15. Respectful (.881) were removed because they had the lowest loadings remaining of the Positivity factor items. Last, two items were removed because they had the lowest loadings among the remaining Coordination factor items (29. Shared Expectations, .708; 30. Coordinated,

.583). Based on these results, these six items were removed. An EFA was then performed on the remaining 9 observer rapport items.

EFA Including Remaining 9 Observer Rapport Items.

The six items listed above were removed from the pool of items and an EFA with maximum likelihood (ML) extraction and direct oblimin rotation was performed on the remaining 9 rapport items. KMO and Bartlett's tests still indicated the appropriateness of factor analysis (KMO = .791; χ^2 (36) = 847.56, $p < .001$) on the remaining items. The resulting 3-factor model explained 90.84% of the variance among items, and a Chi-square test of goodness of fit indicated a good fit for the 9-item, 3-factor model (χ^2 (12) = 7.61, $p = .815$). However, the resulting model structure matrix indicated that some remaining items were highly correlated with multiple factors, so the model was further examined through structural equation modelling (SEM).

Structural Equation Modeling (SEM) Including 9-item Subsets

After confirming the 9-item subset's 3-factor structure, a series of SEMs were tested using robust maximum likelihood (MLR) estimation in MPlus 7. First models were tested that conformed to the structure of the "correlated traits (CT) model" depicted in Figure 1 in the Method section above. This model assumes that the target latent variable being assessed by the measure ("Rapport") is reflected by the commonality between its constituent components

The first 9-item correlated traits model included the nine items retained during the last EFA conducted (6. Invested, 7., Focused on Source, 9. Bored, 13. Frustrated, 17. Aggressive, 19. Relaxed *(Interviewer)*, 23. Productive, 26. Cooperative, and 28. Communicative). Variations of

the correlated traits model were tested based on the results of each preceding test and the modification indices provided by MPlus 7. Two models were also tested that explored alternative theoretical relationships between the essential components of the rapport scales (Attentiveness, Positivity, & Coordination) and rapport as a higher-order factor. The unidimensional rapport model (Figure 2) assumed a unidimensional latent structure among the items representing a single rapport factor. The bifactor model (Figure 3) assumed each item loaded to onto a general rapport factor while simultaneously loading onto their respective appropriate orthogonal group rapport component factors. The model fit of each SEM was assessed using Hu & Bentler's (1999) recommendations for good model fit (RMSEA ≤ 0.06, CFI ≥ 0.95, and SRMR ≤ 0.08).

Initial 9-Item Correlated Traits (CT) Model.

An initial SEM was tested that included the subset of items identified in the EFAs of Part 1 analyses. These items are displayed below in Table 14 along with their standardized factor loadings into each of the three factors

Table 14

Standardized Factor Loadings for EFA of Initial Correlated Traits Model (n=92)

Item	Factor 1	Factor 2	Factor 3
6. Invested			.888
7. Focused on Source			.865
9. Bored			-.885
13. Frustrated		.940	
17. Aggressive		.977	
19. Relaxed *(Interviewer)*		-.927	
23. Productive	.963		
28. Communicative	.971		
29. Shared Expectations	.843		

The chi-square goodness of fit test value was statistically significant (χ^2 (24) 40.88, p = *.017*) indicating poor model fit of this 9-item, 3-factor model. Other fit indices produced mixed results (AIC = 1989.98, RMSEA = 0.087, RMSEA 90% CI = (0.037, 0.132), CFI = 0.960, and SRMR = 0.051). While, the CFI and SRMR values were respectively above and below the thresholds recommended by Hu & Bentler (1999; RMSEA ≤ .06, CFI ≥ .95, and SRMR ≤ .08), the RMSEA value of this model was much higher than the criterion value for accepting the model.

Model modification indices provided by MPlus 7 indicated that, given this subset of items, the Coordination item 23. Productive may be loading significantly onto the Positivity factor, a relationship that would need to be accounted for in the model. Due to this cross-loading, a different item was chosen to replace this item in the model. Rather than make structural changes to the model, an attempt was made to first determine if there was a subset of items that fit the hypothesized relationships between items. Item 29. Shared Expectations was the last Coordination factor item to be removed on the basis of factor loading magnitude, and was thus chosen to replace the item 23. Productive in the Coordination Scale.

Alternative CT Model 1 (Final RS3i-O Model).

A new 3-factor correlated traits (CT) SEM was tested using the subset of items listed in Table 15. This model was identical to the previous model CT except that the item 23. Productive was replaced with the item 29. Shared Expectations. At the time the model was tested, it was tentatively labeled “Alternative CT Model 1.” This model was ultimately chosen as the final version of the RS3i-O, and is labeled as such here.

Table 15

Standardized Factor Loadings for the Final RS3i-O Scales (n=92)

Item	Factor 1	Factor 2	Factor 3
6. Invested			.888
7. Focused on Source			.865
9. Bored			-.885
13. Frustrated		.940	
17. Aggressive		.977	
19. Relaxed (Interviewer)		-.927	
26. Cooperative	.963		
28. Communicative	.971		
29. Shared Expectations	.843		

The chi-square goodness of fit test value was not statistically significant, which indicated good model fit of this version of the 9-item, 3-factor model (χ^2 (24) 32.817, $p = .108$). Additional fit indices also suggested good model fit (AIC = 1947.67, RMSEA = 0.063, RMSEA 90% CI = (0.000, 0.113), CFI = 0.986, and SRMR = 0.045). CFI and SRMR values both met the thresholds recommended for good fit, but the RMSEA value was 0.003 higher than the RMSEA value of .06 recommended for acceptable fit by Hu and Bentler (1999), but well below the cutoff value of 0.10 suggested for poor fit by Kenny (2015).

This model showed significantly improved fit over the Initial 9-Item CT Model, met all of the necessary requirements outlined in the purpose of this study, and was ultimately chosen as the final RS3i-O model. However, the modification indices provided by MPlus 7 for this model indicated that fit may be improved if the residual variance of (1) items 26. Cooperative and 28. Communicative and (2) of items 13. Frustrated and 19. Relaxed were allowed to correlate.

When the residual variances of these items were allowed to correlate in the model, the fit significantly improved. The chi-square goodness of fit test statistic was not statistically

significant (χ^2 (23) 25.10, $p = .292$). Additional fit indices also suggested good model fit (AIC = 1942.57, RMSEA = 0.039, RMSEA 90% CI = (0.000, 0.99), CFI = 0.995, and SRMR = 0.033).

Model Variants

Alternative Correlated Traits Models.

Additional 3-factor correlated traits (CT) SEMs were tested that only differed slightly from the final model. For example, one of these models was titled the "Alternative CT Model 2" in which the item 26. Cooperative was replaced with the item 25. Smooth for reasons discussed in the following paragraph. This model's and its fit indices offered marked improvement over the final RS3i-O. The chi-square goodness of fit test value was not statistically significant, which indicated good model fit of this version of the 9-item, 3-factor model (χ^2 (24) 27.225, $p = .294$). Additional fit indices also suggested excellent model fit (AIC = 2044.90, RMSEA = 0.038, RMSEA 90% CI = (0.000, 0.096), CFI = 0.994, and SRMR = 0.037).

Despite its good fit to the data, the Alternative CT Model 2 was problematic because it included the item Smooth. This item is one of the more challenging items to rate and train others to rate. The item relies on impressions of timing between interactant responses, which may mitigate inter-rater agreement. In the present study, the inter-rater agreement for this item (ICC = .413) was very poor. In contrast, the reliability of the item Cooperative (ICC = .748) was higher and in the ranged considered acceptable. Furthermore, the item information for item 26. Cooperative in the final RS3i-O model (5.98) was considerably higher than the information for item 25. Smooth (0.99) in the "Alternative CT Model 2." For these reasons, this model was ultimately rejected.

Unidimensional Rapport Models.

A unidimensional model was tested for each item configuration in the correlated traits models. In every instance, this model performed severely worse than its correlated traits counterpart. For example, a test of the unidimensional rapport model using the nine items included in the Final RS3i-O model (Table 16) shows significantly worse fit across every index (AIC = 2388.78, χ^2 (36) 371.977, $p < .001$, RMSEA = 0.373, RMSEA 90% CI = (0.340, 0.407), CFI = 0.466, and SRMR = 0.222). Similar model fit indices were obtained when testing the unidimensional model of each previously-tested subset of items.

Bifactor Models.

Variations of an additional multi-dimensional latent factor structure were also considered. The bifactor model (Figure 3) assumed each item loaded to onto a general rapport factor while simultaneously loading onto their respective appropriate orthogonal group rapport component factors. However, this model was not identifiable using only 9 RS3i-O items, as the number of predicted parameters exceeded the number of observed variables. Bifactor model versions were tested for each of the correlated traits models. Earlier versions of the pool of observer rapport items were then tested using the bifactor framework (30-item, 23-item, 15-item). Each time this model was tested, no items loaded significantly onto the general Rapport factor. Thus, this model was rejected as a plausible theoretical alternative to the 3-factor, correlated traits model.

Criteria for Decision on Final RS3i-O Model.

The decision for the final model was based on (1) item characteristics, such as how saliently items loaded onto each hypothesized factor, (2) model fit indices for each subset of items, and (3) the content of the included items.

Each model was considered independently using the recommendations for good model fit (Hu & Bentler, 1999; RMSEA ≤ .06, CFI ≥ .95, and SRMR ≤ .08). Fit indices were also compared between models to assess if one model showed significantly improved fit over another. Specifically, the Akaike Information Criterion (AIC) for each model was compared such that lower AIC values were considered to be indicators of better comparative fit. The AIC indices for each of the 9-item SEMs described above are displayed in Table 16 below.

Table 16

Comparison of Akaike Information Criteria (AIC) Across Select Models

Model	AIC
Initial CT Model	1989.98
Unidimensional Variant	2399.45
Final RS3i-O	1947.67
Unidimensional Variant	2388.78
With Item Residual Correlations	1942.57
Final RS3i-O with Trust/Respect Scale	2536.04
Unidimensional Variant	3045.24
With Item Residual Correlations	2530.47

Final formation of the RS3i-O Scales

Formation of the Attentiveness, Coordination and Irritability Scales of the RS3i-O.

Based on the results of these analyses, the nine items listed above in Table 14 were chosen for inclusion in final version of the RS3i-O. Three scales were formed from these items: One scale, which was assigned the name Attentiveness, consisted of the following three items: Invested, Focused on Source, and Bored. A second scale, which was assigned the name Coordination, consisted of the following three items: Cooperative, Communicative, and Shared Expectations.

A third scale, which was originally intended to be the "Positivity" scale, consisted of the following three items: Frustrated, Aggressive, and Relaxed. As may be seen, two of the items of this scale were actually descriptors of a *lack* of positivity: Frustrated and Aggressive. Therefore, this scale was assigned the name Irritability, as the commonality in item content was best described by this label.

Figure 4 below illustrates all statistically significant standardized parameter estimates for the three RS3i-O scales just described. The model fit indices for the Final RS3i-O model containing these three scales are listed above (page 74) and summarized in Table 16 (page 78) under the label “Final RS3i-O”.

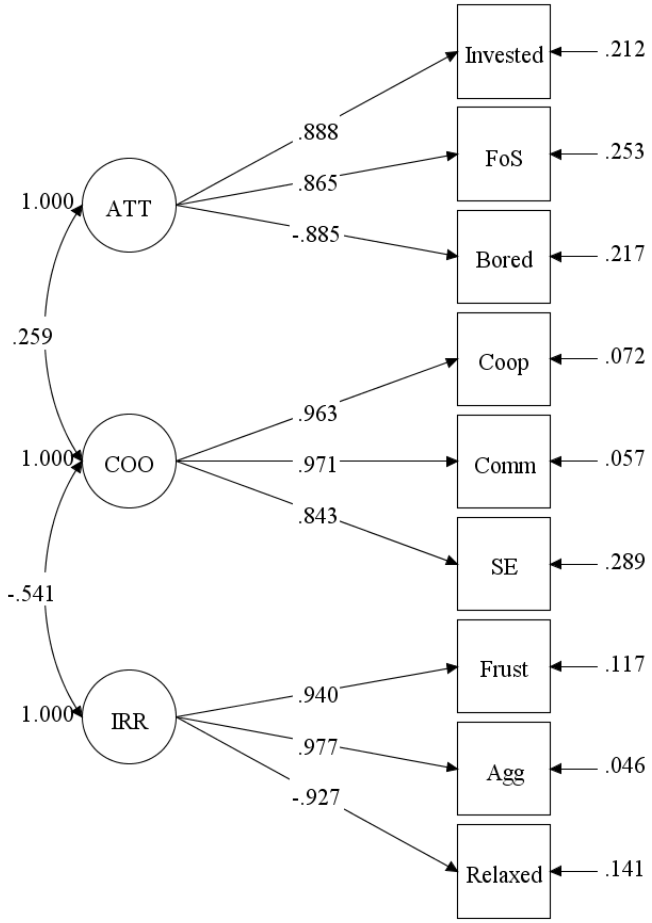


Figure 4. Model of the RS3i-O Attentiveness, Cooperation, and Irritability Scales and Their Items

Note. ATT = Attentiveness, IRR = Irritability, COO = Coordination, FoS = Focused on Source, Coop = Cooperative, Comm = Communicative, SE = Shared Expectations, Frust = Frustrated, Aggress = Aggressive

Addition of the Trust/Respect Scale to the RS3i-O.

As reported in the preceding subsection, the present efforts to develop an observational measure of Positivity ultimately resulted in a three-item scale, Irritability, whose items predominantly measure the *negative* pole of the Positivity dimension. Repeated analyses of the data indicated that the three items of the Irritability scale provided better fit to the Positivity

dimension than any other items included in this study. Furthermore, as will be reported later in this dissertation, the Irritability scale has very good reliability and consistent evidence of validity.

Although the findings of this dissertation clearly support the Irritability scale as an indicator of the negative pole of the Positivity dimension, future researchers may also want an observational measure whose item content represents the *positive* pole of this same dimension. Therefore, a fourth scale, Trust/Respect, has been added to the RS3i-O. This scale is composed of three items – 1. Trusting, 5. Respectful, and 6., Honest – which demonstrated good fit to the Positivity dimension in the analyses reported here. A structural equation model (SEM) that includes the items of the Trust/Respect scale and of the other three RS3i-O scales is shown in Figure 5, with all significant, standardized parameter estimates. The fit indices produced by this model are discussed below on page 82.

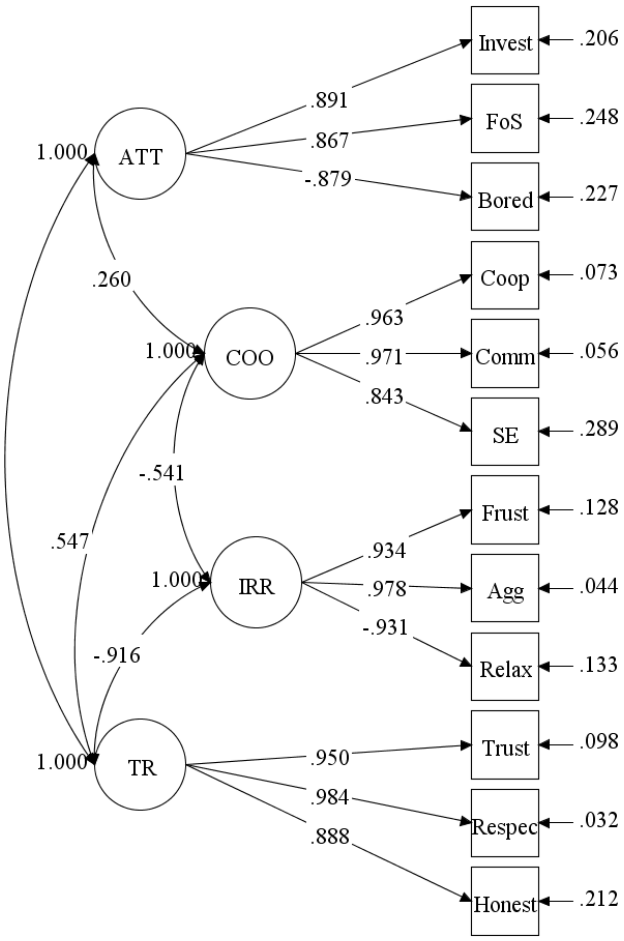


Figure 5. Model of the RS3i-O Attentiveness, Cooperation, Irritability, and Trust/Respect Scales and Their Items

Note: ATT = Attentiveness, COO = Coordination, IRR = Irratibility, TR = Trust/Respect, Invest = Invested, FoS = Focused on Source, Coop = Cooperative, Comm = Communicative, SE = Shared Expectations, Frust = Frustrated, Aggress = Aggressive, Relax = Relaxing, Trust = Trusting, Respec = Respectful.

This model produced mixed fit indices (χ^2 (48) 70.80, $p = .018$, RMSEA = 0.072, RMSEA 90% CI = (0.031, 0.106), CFI = 0.979, and SRMR = 0.049). While CFI and SRMR values were acceptable, this model produced chi-square value that indicated significantly worse fit than the RS3i-O model without the Trust/Respect Scale included (Figure 4). This information

is summarized above in Table 16 (page 79). The modification indices indicated that fit may be improved if the items Cooperative and Communicative were allowed to correlate in the model.

As with the final 3-factor RS3i-O model (though not to the same degree), fit improved when the residual variances of some items within scales were allowed to correlate in the 4-factor model. When the disturbances of items 26. Cooperative and 28. Communicative correlated, the chi-square goodness of fit test statistic was marginally not significant (χ^2 (47) 63.40, $p = .056$). Additional fit indices also suggested improved model fit (AIC = 2528.68, RMSEA = 0.062, RMSEA 90% CI = (0.000, 0.98), CFI = 0.985, and SRMR = 0.037. Furthermore, when the disturbances of items 23. Frustrated and 29. Relaxed *(Interviewer)* were allowed to correlate, some fit indices improved while others worsened fit (AIC = 2530.63, χ^2 (47) 63.47, $p = .045$). RMSEA = 0.064, RMSEA 90% CI = (0.011, 0.100), CFI = 0.984, and SRMR = 0.037). The difference in chi-square values between this model and the model without residual item correlations remained statistically significant. The comparative fit of each model, based on AIC, is summarized above in Table 16 (page 77).

Numbering of the Items of the RS3i-O.

Table 18 below lists the four scales and 12 items of the RS3i-O. The table lists (a) each item's number in the original pool of 30 observer rapport items and (b) each item's number in the 12-item RS3i-O.

Table 18

Rapport Scales for Investigative Interviews and Interrogations, Observer version (RS3i-O)

Items by Factor Scale

Scale	Name and Item Number in RS3i-O	Item Number in Original Item Pool
Attentiveness		
	1. Invested	6
	2. Thoughtful	7
	3. Bored	9
Irritability		
	4. Frustrated	13
	5. Aggressive	17
	6. Relaxed *(Interviewer)*	19
Coordination		
	7. Cooperative	26
	8. Communicative	28
	9. Shared Expectations	29
Trust/Respect		
	10. Trusting	21
	11. Honest	26
	12. Respectful	25

Summary of Part 1 Analyses

The primary goals of Part 1 analyses were to (1) explore the factor structure of the pool of rapport items used to rate investigative interviews videos and (2) identify a subset of items to develop a quick, reliable set of factors scales that adhere to the Tripartite Theory of rapport

(Tickle-Degnen, & Rosenthal, 1990). Exploratory factor analyses (EFAs) confirmed the hypothesized 3-factor structure of the observer rapport items and indicated the items that loaded most saliently onto these factors. Structural equation modeling (SEM) was then employed to refine the scales of the final RS3i-O and confirm its model fit.

It was hypothesized that the items comprising the pool of 30 observer rapport items would load onto three distinct factors representing Tickle-Degnen & Rosenthal's (1990) "essential components" of rapport (Hypothesis 1a). Specifically, items 1 – 10 were expected to load onto a single factor representing the construct Attentiveness, items 21 – 20 were expected to load onto a single factor representing Positivity, and items 21 – 30 were expected to load onto a single factor representing Coordination (Hypothesis 1b). This hypothesis was generally supported, as EFA of the 30 items produced a 3-factor structure and nearly all of the items loaded primarily onto one of the hypothesized factors. Additionally, each of the RS3i-O model factors was expected to be at least moderately (r = 0.30) with the two other scales (Hypothesis 1b). This hypothesis was supported. Each 10-item factor was significantly correlated with the other two factors at larger magnitudes than this threshold.

Hypotheses 1a and 1b are also partially supported when their criteria are applied to the final, 9-item RS3i-O model. Model fit indices clearly supported the 3-factor structure of the final measure (Hypothesis 1a). Significant factor intercorrelation was only partially supported in that, while correlations between Attentiveness and Coordination factors and between Coordination and Positivity were statistically significant, only the latter reached the hypothesized threshold (r = .30). The correlation between the Attentiveness and Positivity scales was not significant and did not meet the hypothesized threshold.

Based on the results described above, the final version of the Rapport Scales for Investigative Interviews and Interrogations, Observer version (RS3i-O) was chosen (Table 18) and all analyses in future study parts were conducted using this measure.

Part 2: Psychometric Properties of the RS3i-O Scales

The purpose of Part 2 of the study was to examine the psychometric properties of the RS3i-O scales, including their factor structure, intercorrelations, internal and interrater reliability, and convergent, discriminant, and criterion validity.

RS3i-O Factor Structure

Confirmatory factor analysis (CFA) was performed on the final RS3i-O and was expected to confirm the 3-factor structure identified in Part 1 (Hypothesis 8a). CFA was also performed on the RS3i-O measure including the Trust/Respect scale, despite the addition of this new scale, the CFA was expected to result in a 3-factor structure (Hypothesis 9a). Structural equation models (SEMs) of the final 3-factor RS3i-O and the 4-scale variant including Trust/Respect scale were tested and fit indices were evaluated according to Hu & Bentler's (1999) recommendations for good fit (RMSEA $\leq$ .06, CFI $\geq$ .95, and SRMR $\leq$ .08; Hypotheses 8b and 9b).

Structural Equation Models (SEMs).

Structural equation models (SEMs) of the final 9-item, 3-factor RS3i-O and the 12-item, 4-scale variant including were tested using MPlus7. Fit indices were evaluated according to Hu & Bentler's (1999) recommendations for good fit (RMSEA $\leq$ .06, CFI $\geq$ .95, and SRMR $\leq$ .08).

This 9-item, 3-scale RS3i-O was expected to demonstrate good fit, according to these criteria (Hypothesis 8b). The 12-item, 4-scale RS3i-O with Trust/Respect version was also expected to meet these fit criteria (Hypothesis 9b).

9-Item, 3-Scale RS3i-O.

The RS3i-O model including nine items across Attentiveness, Irritability, and Coordination scales was tested using MPlus 7. Based on the results of Part 1 and previous Part 2 analyses, the Irritability scale was expected not to correlate with the Attentiveness scale. Further, based on the final SEM in Part 1, it was anticipated that modification indices produced by MPlus for this model may suggest allowing the residual variances of items 4. Frustrated and 6. Relaxed as well as those of items 7. Cooperative and 8. Communicative to correlate.

The chi-square goodness of fit test value was statistically significant (χ^2 (25) 52.53, p = *.001*) indicating poor model fit of this 9-item, 3-factor model. Other fit indices produced mixed results (AIC = 1693.63, RMSEA = 0.110, RMSEA 90% CI = (0.068, 0.151), CFI = 0.960, and SRMR = 0.063). While, the CFI and SRMR values were respectively above and below the thresholds recommended by Hu & Bentler (1999; RMSEA ≤ .06, CFI ≥ .95, and SRMR ≤ .08), the RMSEA value of this model was much higher than the criterion value for accepting the model. Given that this model had small degrees of freedom (df = 25) and sample size (n = 92), an inflated RMSEA was expected; however, the value for this model exceeded even the more liberal standards proposed (Kenny, 2015).

As expected, modification indices produced by MPlus 7 for this model indicated several modifications to the model that would improve fit. MPlus suggested both of the expected modifications to the model: (2) allowing the residual variance of items 4. Frustrated and 6.

Relaxed to correlate (2) allowing the residual variance of items 7. Cooperative and 8. Communicative to correlate, and (3) allowing item 5. Aggressive to cross-load onto the Coordination factor. Models incorporating these modifications were then examined. Rather than making major structural modifications (e.g., cross-loading items) to the model, only the within-factor item correlations were allowed.

When only the residual variances of items 7. Cooperative and 8. Communicative were allowed to correlate in the model, fit indices did not significantly improve (AIC = 1693.33, χ^2 (24) 50.51, $p = .001$, RMSEA = .109, CFI = .961, SRMR =.061). When the residual variances of items 4. Frustrated and 6. Relaxed were also allowed to correlate in this model, the resulting model produced an improved, but statistically significant chi-square goodness of fit value (χ^2 (23) 36.171, $p = .040$), as well as improved fit indices (RMSEA = .079, CFI = .981, SRMR =.058). Based on these results, the 3-factor structure of the RS3i-O (Hypothesis 9a) had mixed support. Figure 6 depicts the SEM for the unmodified, 9-Item RS3i with significant, standardized parameter estimates.

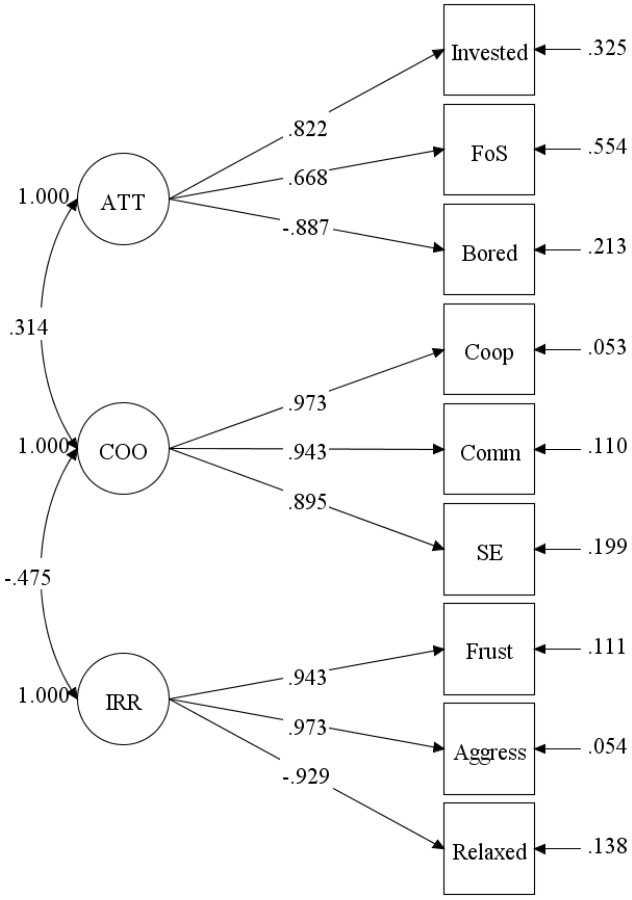


Figure 6. Confirmatory RS3i-O Structural Equation Model

Note. ATT = Attentiveness, COO = Coordination, IRR = Irritability, FoS = Focused on Source, Coop = Cooperative, Comm = Communicative, SE = Shared Expectations, Frust = Frustrated, Aggress = Aggressive.

12-Item, 4-Scale RS3i-O Including Trust/Respect.

An RS3i-O model including 12 items across Attentiveness, Irritability, Coordination, and Trust/Respect scales was tested using MPlus 7. As in the previous model, the Irritability scale was expected not to correlate with the Attentiveness scale. It was also anticipated that MPlus would produce modification indices similar to with the 9-item RS3i.

The chi-square goodness of fit test value was statistically significant (χ^2 (49) 70.80, $p = .012$) indicating poor model fit of this 12-item, 4-factor model. Other fit indices produced mixed results (AIC = 2536.04, RMSEA = 0.072, RMSEA 90% CI = (0.031, 0.106), CFI = 0.979, and SRMR = 0.049). While, the CFI and SRMR values were respectively above and below the thresholds recommended by Hu & Bentler (1999; RMSEA ≤ .06, CFI ≥ .95, and SRMR ≤ .08), the RMSEA value of this model was higher than the criterion value for good model fit. Even using a more liberal criterion to account for degrees of freedom and sample size (Kenny, 2015), this RMSEA value indicated between mediocre and poor fit. Figure 7 depicts the SEM for the unmodified, 12-item, 4-factor RS3i with significant, standardized parameter estimates.

When the residual variances of items 7. Cooperative and 8. Communicative as well as items 4. Frustrated and 6. Relaxed were allowed to correlate in this model, the resulting model still produced a statistically significant chi-square goodness of fit value (χ^2 (46) 63.47, $p = .044$) indicating poor model fit, though other fit indices were somewhat improved, though mixed (AIC = 2530.63, RMSEA = .064, RMSEA 90% CI = (0.011, 0.111) CFI = .984, SRMR =.037). Based on these results, the 4-factor structure of the RS3i-O with the included Trust/Respect scale (Hypothesis 9b) also had mixed support.

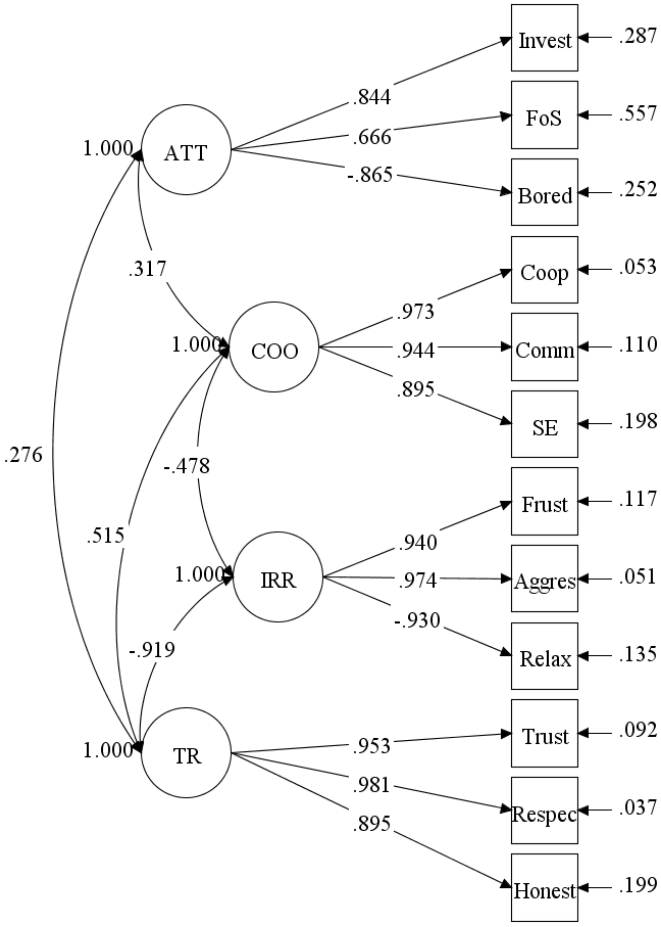


Figure 7. Confirmatory RS3i-O Model Including Trust/Respect Scale

Note. ATT = Attentiveness, COO = Coordination, IRR = Irritability, TR = Trust/Respect, Invest = Invested, FoS = Focused on Source, Coop = Cooperative, Comm = Communicative, SE = Shared Expectations, Frust = Frustrated, Aggres = Aggressive, Relax = Relaxed, Respec = Respectful.

RS3i-O Scale Reliability

Means, standard deviations, internal consistency, and interrater reliability of the three RS3i-O rating scales are displayed in below in Table 19. Intraclass correlation coefficients (ICC) were computed using ratings made for each interview by the raters on each secondary rating

team. Rater 1 and Rater 2 scores for the items of each scale were averaged to produce final scores for each item. These final item scores were used to compute the internal consistency of each observer scale. Cronbach's alpha (α) was computed using SPSS 23. McDonald's omega (ω) was computed by hand.

Table 19

RS3i-O Means, Standard Deviations, Internal Consistency, and Interrater Reliability (n=92)

RS3i-O Scale	M *(SD)*	α	ω	Secondary Team ICC	All Raters ICC
1. Attentiveness	4.34 *(0.66)*	.889	.938	.678	.799
2. Irritability	1.65 *(1.61)*	.956	.988	.830	.956
3. Coordination	3.72 *(0.93)*	.945	.977	.733	.871
4. Trust/Respect	3.22 *(1.72)*	.953	.987	.730	.925

Note. Secondary Team ICC calculated between two Final Raters on Secondary Rating Team for that scale. All raters ICC calculated across all six raters.

The internal reliability of each of the scales was high (αs > .88; ωs > .93). The ICC values between Secondary Team Final Raters indicating inter-rater reliability of the Irritability, Coordination, and Trust/Respect scales met the standards proposed by Hunsley and Mash (2008) as either adequate (>0.70) or excellent (>0.80) inter-rater reliability. However, the ICC between raters for the Attentiveness scale (.678) fell somewhat below (0.022) the threshold for adequate interrater reliability. It is important to note that, when ICC values were calculated across all six raters included in the study, each scale met the criteria for excellent inter-rater reliability (>.80).

RS3i-O Scale Validity

Analyses were performed to examine the inter-correlations among the RS3i-O scales to assess their convergent validity. As all scales were intended to measure dimensions of rapport, the scales were all expected to correlate with each other at least moderately. The intercorrelations between the RS3i-O are presented in Table 20 below.

Table 20
RS3i-O Scale Intercorrelations (n = 92)

RS3i-O Scale	1	2	3	4
1. Attentiveness	-	-.118	.260*	.419**
2. Irritability	-.118	-	-.557**	-.872**
3. Coordination	.260*	-.557**	-	.553**
4. Trust/Respect	.419**	-.872**	.553**	-

Note. *Correlation is significant at the .05 level.

** Correlation is significant at the .01 level.

As all scales were intended to measure dimensions of rapport, the scales were all initially expected to correlated with each other at least moderately (producing Pearson product-moment correlation coefficients greater than .30).

As can be seen in Table 19, there was a high degree of intercorrelation between the RS3i-O scales, including the Trust/Respect scale, and nearly all scales were significantly correlated with the other three. However, the Attentiveness scale was not significantly correlated with the Irritability scale (r = -.118, p = .264). though it was significantly correlated with the Coordination scale (r = .260, p = .012), the relationship did not meet the expected threshold.

Convergent Validity: Relationship Between RS3i- O and RS3i-S Scales.

A principal purpose of Part 2 of the present study was to determine the degree to which the observational RS3i-O scales are related to sources' self-report experience of rapport. To further examine the RS3i-O scales' convergent validity, RS3i-O observer scales correlations with the self-report RS3i-S source scales intended to measure "essential components" of rapport (mutual attentiveness, positivity, and coordination) were calculated. Each RS3i-O scale was expected to correlate moderately ($r \geq .30$) and in the theoretically consistent direction with the corresponding RS3i-S scale intended to measure the same dimension of rapport (Hypotheses 2a-d). Specifically, RS3i-O Attentiveness was expected to correlate with RS3i-S Attentiveness (Hypothesis 2a), RS3i-O Irritability was expected to correlate with RS3i-S Trust/Respect (Hypothesis 2b), and RS3i-O Coordination was expected to correlate with RS3i-S Connected/Flow (Hypothesis 2c). Additionally, RS3i-O Trust/Respect was expected to correlate with RS3i-S Trust/Respect (Hypothesis 2d). Table 21 contains the zero-order Pearson product-moment correlation coefficients for RS3i-O and RS3i-S scales.

Table 21

Correlations Between RS3i-O Observer Scales and RS3i-S Self-Report Scales (n = 92)

	RS3i-S Scale				
RS3i-O Scale	Attentiveness	Trust/Respect	Connected Flow	Expertise	Cultural Similarity
1. Attentiveness	**.199**	.115	.128	.316*	.134
2. Irritability	-.358**	**-.339****	-.482**	-.004	-.043
3. Coordination	.354**	.361**	**.421****	.149	.065
4. Trust/Respect	.293*	**.335****	.417**	.039	.073

Note. Hypothesized correlations appear in bold.

* Correlation is significant at the .01 level.

** Correlation is significant at the .001 level.

The correlations reported in Table 21 did not support Hypothesis 2a, but did support Hypotheses 2b-d. With the exception of the Attentiveness scale, each RS3i-O scale was significantly correlated moderately ($r > .30$) and in the theoretically consistent direction with the corresponding RS3i-S scale intended to measure the same dimension of rapport. The RS3i-O Attentiveness scale exhibited a small correlation with the RS3i-S Attentiveness scale ($r = .199$, $p = .055$), but this relationship did not reach statistical significance. These results generally support the convergent validity of the RS3i-O Irritability, Coordination, and Trust/Respect scales.

Convergent Validity: Relationship Between RS3i- O and Magee (2018) Observational Scales.

The convergent validity of the RS3i-O scales was also assessed by correlating the RS3i-O scales with the observer scales included in the study by Magee (2018) that examined the same interviews included in the present study. Specifically, the Magee study included the Negotiators' Rapport Scale (NRS), Interaction Rapport Scale (IRS), & Global Motivational Interviewing Skills Code (GMISC) Acceptance/Empathy observational scales. Hypotheses 3a-d predicted that each RS3i-O scale would correlate least moderately ($r = 0.30$) with its corresponding observational scale calculated by Magee (2018). Specifically, RS3i-O Attentiveness was expected to correlate with NRS Attentiveness (Hypothesis 3a), RS3i-O Irritability was expected to correlate negatively with NRS Positivity and GMISC Acceptance/Empathy (Hypothesis 3b), RS3i-O Coordination was expected to correlate with IRS Coordination (Hypothesis 3c), and RS3i-O Trust/Respect was expected to correlate positively with NRS Positivity and GMISC Acceptance/Empathy (Hypothesis 3d). Table 22 contains the zero-order Pearson product-moment correlation coefficients for RS3i-O scales and observer scales from Magee (2018).

Table 22

Correlations Between RS3i-O Observer Scales and Magee (2018) Observer Scales (n = 92)

RS3i-O Scale	Magee (2018) Observer Scales			
	NRS Attentiveness	NRS Positivity	G-MISC Acceptance / Empathy	IRS Coordination
1. Attentiveness	**.494****	.213*	.180	.265*
2. Irritability	-.376**	**-.850***	**-.878***	-.614*
3. Coordination	.456**	.452**	.473*	**.632***
4. Trust/Respect	.561*	**.932***	**.473***	.651*

Note. Hypothesized correlations appear in bold.

* Correlation is significant at the .05 level.

** Correlation is significant at the .01 level.

Hypotheses 3a-d were clearly supported by the resulting correlation matrix. Each of the RS3i-O scales was correlated moderately to very strongly with its analogous observer scale from Magee (2018). These results generally support the convergent validity of all the RS3i-O scales.

Discriminant Validity: Relationship Between RS3i- O, Non-Corresponding RS3i-S Scales of the Essential Components, and Non-Essential Component RS3i-S Scales.

Analyses examined the discriminant validity of the RS3i-O scales for distinguishing between the "essential three" constructs of the tripartite model (Tickle-Degnen & Rosenthal, 1990). First, the scales' ability to distinguish between the "essential components" as measured by the Rapport Scales for Investigative Interviews and Interrogations, Source version (RS3i-S) Attentiveness, Trust/Respect, and Connected Flow scales was tested.

It was expected that each RS3i-O observer scale would correlate not only with its corresponding RS3i-S source scale but also to some degree with the two other "non-corresponding" RS3i scales meant to measure one of the other two components of the “essential three." Correlations between RS3i-O scales and non-corresponding RS3i-S scales are located above in Table 20. Each RS3i-O scale was expected to correlate at least 0.10 less with the RS3i-S and observational scales that measure the other two “essential components” than with the scales intended to measure the same component (Hypothesis 4a). Further, each RS3i-O scale was expected to correlate less than $r = .20$ with the RS3i-S Expertise and Cultural Similarity scales, as they do not measure “essential components” of rapport (Hypothesis 4b).

Hypothesis 4a was not supported. None of the RS3i-O scales were correlated at least 0.10 less with non-corresponding RS3i-S “essential component” scales than with their analogous RS3i-S scale. Hypothesis 4b was generally supported. Each RS3i-O scale correlated at least 0.20 less with the RS3i-S Expertise and Cultural Similarity scales than with its analogous RS3i-S scale, with the exception of RS3i-O Attentiveness. This scale was correlated significantly ($r = .316$, $p = .002$) with the RS3i-S Expertise scale. These results support the RS3i-O scales’ ability to discriminate between “essential components” and other aspects of rapport, as perceived by the interview source, but not their ability to discriminate between the “essential components” themselves.

Discriminant Validity: Relationship between RS3i-O and Non-Corresponding Magee (2018) Observer Scales.

Next, the RS3i-O scales’ ability to distinguish between the “essential components” as measured by the observational rapport scales employed by Magee (2018). These included

Bronstein et al.'s (2012) Negotiators' Rapport Scale (NRS) Attentiveness and Positivity scales, the Interaction Rapport Scale (IRS) Coordination scale, and the Global Motivational Interviewing Skills Code (GMISC) Acceptance/Empathy scale developed by Alison et al. (2013) and computed by Magee (2018). It was expected that each RS3i-O observer scale would correlate not only with its analogous observer scale but also to some degree with the other observer scales meant to measure one of the other two rapport components (Hypothesis 4a). This hypothesis was supported, as each RS3i-O scale was correlated at least .10 less with non-corresponding observer rapport scales. Correlations between observer rating scales and non-corresponding observer scales are located above in Table 22.

Criterion Validity: Relationship Between RS3i-O Scales and Amount of Information Shared.

The RS3i-O scales' criterion validity was first assessed by examining the relationship of each scale with the amount of relevant information shared during the interview as measured by the Shared Information Rating Scale (SIRS). Separate analyses were conducted using SIRS scores for the first half of the interviews (Phase 1), the second half of the interviews (Phase 2), and the total information shared for the entire interviews. It was expected that each RS3i-O scale would correlate at least moderately (r = .30) with SIRS scores for Phase 2 of each interview and the interview as a whole (Hypothesis 5). Table 23 lists the Pearson product-moment correlation between each RS3i-O scale and the amount of information shared by the source during Phase 1, Phase 2, and total over the course of the interview.

Table 23

Correlations Between Observational Rating Scales and SIRS (n = 92)

RS3i-O Scale	Information Shared in Phase 1	Information Shared in Phase 2	Total Information Shared
1. Attentiveness	.115	-.004	.055
2. Irritability	-.100	.163	.081
3. Coordination	-.097	.564**	.491**
4. Trust/Respect	.071	.200	.032

Note. SIRS = Shared Information Rating Scale.

** Correlation is significant at the .001 level.

Hypothesis 5 was unsupported for most RS3i-O scales. However, the RS3i-O Coordination scale was significantly and substantially correlated with the amount of information shared in the interview during Phase 2 ($r = .564$, $p < .001$) and the interview as a whole ($r = .491$, $p < .001$).

Criterion Validity: Relationship Between RS3i-O Scales and Sources' Self-Reported Cooperativeness.

Criterion validity was also tested by examining the correlation of each RS3i-O scale with sources' self-report cooperativeness, reported on a 10-point Likert-type scale where 1 represented totally uncooperative and 10 represented totally cooperative. Hypothesis 6 predicted that each RS3i-O scale correlate at least moderately ($r = .30$) with self-report cooperativeness scores. Table 24 lists the Pearson product-moment correlations between each RS3i-O scale and self-report cooperativeness.

Table 24

Correlations Between RS3i-O Scales and Self-report Cooperativeness (n = 92)

RS3i-O Scale	Cooperativeness
1. Attentiveness	.080
2. Irritability	-.222*
3. Coordination	.410**
4. Trust/Respect	.180

Note. Cooperativeness = *"How cooperative were you?".* * Correlation is significant at the .05 level. ** Correlation is significant at the .001 level.

Hypothesis 6 was only partially supported. The Irritability scale was correlated with self-report cooperativeness (r = -.222, p = *.039*), but not to the specified threshold (r = .30). The Coordination scale was significantly, and moderately correlated with sources' self-reports of their own cooperativeness during interviews (r = .410, p < *.001*).

Criterion Validity: Correlations between RS3i-O and Interview Style.

Criterion validity of the RS3i-O scales was further examined by exploring the relationship between the scores on each scale and the interviewing styles employed by Duke et al. (2018). A one-way multiple analysis of variance (MANOVA) test was performed comparing mean scores on each observer scale between the three experimental conditions (Rapport, Neutral, Pressure). Contrasts between each of the conditions were also examined (Rapport vs. Pressure; Rapport vs. Neutral; Neutral vs. Pressure). Hypothesis 7a predicted a main effect of interview style on each of the RS3i-O scales, and differences between each interview condition were each

expected to be statistically significant (Hypothesis 7b). RS3i-O scale means and standard deviations for each interview condition are displayed below in Table 25. Statistically significant differences in means between condition are denoted in Table 25 with subscript.

Table 25

Means for RS3i-O Scales by Interview Condition (n=92)

	Interview Condition		
	Rapport *(n=30)*	Neutral *(n=29)*	Pressure *(n=33)*
RS3i-O Scale	M	M	M
1. Attentiveness	4.80_a	3.85_b	4.36_c
2. Irritability	0.47_a	0.65_{ab}	3.60_b
3. Coordination	4.24_a	3.73_{ab}	3.23_b
4. Trust/Respect	5.05_a	3.64_b	1.18_c

Note. For each scale, means sharing a common subscript are not significantly different at $p < .05$ according to the Tukey HSD procedure.

A main effect of experimental condition (Rapport vs. Neutral vs. Pressure) was found for the RS3i-O Attentiveness Scale, ($F [2,89] = 22.73$, $p < .001$, partial $\eta^2 = .338$), supporting Hypothesis 7a. Post-hoc Tukey HSD comparisons indicated that, as predicted by Hypothesis 7b, Attentiveness was rated significantly higher for interviews in the Rapport condition than in the Neutral and Pressure conditions.

A main effect of experimental condition (Rapport vs. Neutral vs. Pressure) was found for the RS3i-O Irritability Scale, ($F [2,89] = 217.78$, $p < .001$, partial $\eta^2 = .830$), supporting Hypothesis 7a. Post-hoc comparisons indicated that, as predicted, Irritability was rated significantly lower for interviews in the Rapport condition than in the Pressure condition, and

significantly lower in Neutral than in Pressure interviews. Contrary to prediction, there was no statistically significant difference in Attentiveness scores between the Rapport and Neutral conditions (Hypothesis 7b).

A main effect of experimental condition (Rapport vs. Neutral vs. Pressure) was found for RS3i-O Coordination Scale, (F [2,89] = 11.54, $p < .001$, partial $\eta^2 = .206$). Post-hoc comparisons indicated that, as predicted, Coordination was rated significantly higher for interviews in the Rapport condition than in the Pressure condition, and significantly higher in Neutral than in Pressure interviews. Contrary to prediction (Hypothesis 7b), there was no statistically significant difference in Attentiveness scores between the Rapport and Neutral conditions.

A main effect of experimental condition (Rapport vs. Neutral vs. Pressure) was found for RS3i-O Trust/Respect Scale, (F [2,89] = 414.40, $p < .001$, partial $\eta^2 = .903$), adding further support for Hypothesis 7a. Post-hoc comparisons indicated that, as predicted, Trust/Respect was rated significantly higher for interviews in the Rapport condition than in the Neutral and Pressure condition and significantly higher in Neutral than in Pressure interviews, supporting Hypothesis 7b.

Post-Hoc Analyses

Effect of Raters' English Language Ability.

Raters' English language abilities, as measured by the WMLS-R Picture Vocabulary (M = 17.11, SD = 5.82), Verbal Analogies (M = 28.11, SD = 5.67), and Oral Language (M = 23.89, SD = 6.23) age-equivalency scores were correlated with Comprehension Check and RS3i-O scale scores. No significant correlations were found between any of these scales. However, given

the very small sample size (n = 9), these analyses did not reach acceptable power to accurately assess even very large correlations.

Comprehension of Training Materials.

Raters' Comprehension Check scores (M = 107.33, SD = 7.35) were correlated with their RS3i-O scale scores. No significant correlations were found.

Gender.

A one-way multiple analysis of variance (MANOVA) was performed to examine the effect of rater gender on Comprehension Check and RS3i-O scale scores. No significant effects were found for gender on any of these variables.

Summary of Part 2 Analyses

The purpose of Part 2 of the present study was to assess the psychometric properties of the RS3i-O scales developed in Part 1.

The factor structure of the RS3i-O was tested by performing SEM on the final 9-item, 3-scale version. Analysis of the hypothesized factor structure produced mixed fit indices. If some slight modifications were made to the model, though, model fit improved significantly. These results provide limited support for the 3-factor structure of the RS3i-O.

Part 2 results supported the reliability and validity of the RS3i-O scales. Each RS3i-O scale demonstrated excellent internal reliability. Further, each scale demonstrated at least adequate inter-rater reliability with the exception of the Attentiveness scale. This scale's ICC fell just below the standards proposed by Hunsley and Mash (2008) to indicate adequate reliability (>.70).

Based on the tripartite theory framework (Tickle-Degnen & Rosenthal, 1990), each RS3i-O scale was expected to correlate with the others. However, considering the results of Part 1 of the present study, the Irritability scale was not expected to correlate with the Attentiveness scale. These two scales were not significantly correlated, but the other scale intercorrelations were demonstrated as expected. Further, the addition of the Trust/Respect scale and its significant correlation with Attentiveness supported the theorized relationships between latent Attentiveness, Positivity, and Coordination factors.

To evaluate their convergent and discriminant validity, the RS3i-O scales were correlated with other self-report (RS3i-S) and observer (NRS, IRS, and GMISC) rapport scales. It was hypothesized that each scale would correlate positively and at least moderately (r = .30) with its corresponding RS3i-S scale, respectively (Hypotheses 2a-d). This hypothesis was supported for all scales other than Attentiveness. This scale was only weakly correlated with its self-report analog (r = .20, p = *.055*), and this correlation did not reach statistical significance. It was also hypothesized that each RS3i-O scale would correlate positively and at least moderately (r = .30) with its corresponding observer scale measured by Magee (2018; Hypotheses 3a-d). These hypotheses were all strongly supported, with each scale demonstrating large convergent validity coefficients (see Table 22).

Each RS3i-O scale was also expected to correlate less with RS3i-S scales not intended to measure the "essential components" of rapport (Hypothesis 4a) as well as with non-analogous observer scales measured by Magee (Hypothesis 4b). These hypotheses were largely supported, indicating acceptable discriminate validity. The one exception to this finding was once more the Attentiveness scale. This scale did not demonstrate an ability to discriminate between self-report rapport scales and was, in fact, most highly correlated to the RS3i-S Expertise scale.

To assess the RS3i-O scales' criterion validity, they were correlated with SIRS and self-report cooperativeness scores. Each scale was expected to correlate at least moderately with the amount of information shared in the interview (Hypothesis 5) and sources' perception of their own cooperativeness (Hypothesis 6). These hypotheses were largely unsupported, except with regard to the Coordination scale. This scale's criterion validity was clearly demonstrated. The scales' criterion validity was further examined by comparing scale means across interview conditions. It was predicted that each scale would differ significantly between conditions, and this was largely supported. In particular, every scale differed between rapport and pressure interview conditions.

Part 3: Linguistic Variable Validation

LIWC Variable Validity

Convergent Validity: Relationship Between LIWC Variables and RS3i-S and RS3i-O Scales.

In order to assess the convergent validity of the LIWC variables identified by Driskell et al. (2013), each of the hypothesized variables for both interviewers and sources were correlated with scores on the RS3i-S and RS3i-O. Table 26 below contains the zero-order correlations of the interviewer LIWC variables with these rapport scales, and Table 27 contains those for the source LIWC variables.

Both interviewer and source LIWC variables were expected to correlate at least weakly (r = .20) with all corresponding RS3i-S and RS3i-O scales thought to measure the same "essential component" (Hypotheses 10a-c).

Table 26

Correlations of Interviewer LIWC Variables with RS3i-S and RS3i-O Scales

Rapport Scale		Interviewer LIWC Variables								
		First Person Plural	Present Focus	Social Processes	Positive Emotion	Negative Emotion	Assent	Nonfluencies	Conjunctions	Certainty
RS3i-S Scales	Attentiveness	**.093**	**-.134**	**-.039**	.317**	-.088	.209*	.155	.235*	-.249*
	Trust/Respect	.244*	.101	.116	**.236***	**-.272****	.099	-.039	.023	-.214*
	Connected Flow	.103	-.094	-.043	.352**	-.176	**.240***	**.068**	**.219***	**-.301****
RS3i-O Scales	Attentiveness	**-.064**	**-.056**	**-.201**	.235*	.113	.126	.395**	.173	.026
	Irritability	-.035	.260*	.196	**-.453****	**.445***	-.297*	-.182	-.157	.600**
	Coordination	-.097	.005	-.362**	.445*	-.275**	**.345****	**.117**	**.193**	**-.226***
	Trust/Respect	-.002	.192	-.255*	**.502****	**-.362***	.295**	.258*	.202*	-.530**

Note. Hypothesized correlations appear in bold.

*Correlation is significant at the 0.05 level (2-tailed).

Hypotheses 10a-c were largely unsupported by these results, as most of the Interviewer LIWC variables did not correlate to the hypothesized degree or in the hypothesized direction with their corresponding RS3i-S and RS3i-O scales. However, interviewer positive emotion, negative emotion, and assent each met the criteria for establishing convergent validity with these scales. Specifically, interviewer use of positive emotion (e.g., "happy") was positively correlated with RS3i-S Trust/Respect (r = .236, $p = .023$) and RS3i-O Trust/Respect (r = .502, $p < .001$), and negatively correlated with RS3i-O Irritability (r = -.453, $p < .001$). Conversely, interviewer use of negative emotion (e.g., "angry") was negatively correlated with RS3i-S Trust/Respect (r = -.272, $p = .009$) and RS3i-O Trust/Respect (r = -362, $p < .001$), and positively correlated with RS3i-O Irritability (r = .445, $p < .001$). These results supported the convergent validity of these two LIWC variables with both sources' and observers' perceptions of interviewers' Positivity. Additionally, degree of interviewer assent (e.g., "yes") was positively correlated with both RS3i-S Connected Flow (r = .240, $p = .021$) and RS3i-O Coordination (r = .345, $p = .001$), demonstrating this LIWC variable's convergent validity.

Table 27

Correlations of Source LIWC Variables with RS3i-S and RS3i-O Scales

Rapport Scale		Source LIWC Variables								
		First Person Plural	Present Focus	Social Processes	Positive Emotion	Negative Emotion	Assent	Nonfluencies	Conjunctions	Certainty
RS3i-S Scales	Attentiveness	**-.097**	**.036**	**.091**	.042	-.138	.272*	-.132	-.037	-.131
	Trust/Respect	-.036	.008	.089	**-.150**	**-.039**	.156	-.102	-.020	-.217*
	Connected Flow	-.006	.140	.054	-.046	-.091	**.249***	**-.168**	**-.031**	**-.165**
RS3i-O Scales	Attentiveness	**.250***	**.180**	**-.108**	.208*	-.189	.334*	.292**	.063	.001
	Irritability	.056	-.140	.212*	**-.057**	**.228***	-.255*	.011	.056	.170
	Coordination	-.169	.005	-.193	-.220*	-.313**	**.175**	**.012**	**.028**	**-.272***
	Trust/Respect	.049	.192	-.268*	.153	-.212*	.405**	.118	-.050	-.217*

Note. Hypothesized correlations appear in bold.

*Correlation is significant at the 0.05 level (2-tailed).

Again, Hypotheses 10a-c were not supported by the results above, as most of the source LIWC variables did not correlate to the hypothesized degree or in the hypothesized direction with their corresponding RS3i-S and RS3i-O scales. Further, those variables that did correlate with either their corresponding self-report or observer scales did not correlate with both. It is important to note though that source use of negative emotion were significantly correlated with most RS3i-S and RS3i-O scales.

Convergent Validity: Relationship Between LIWC Variables and Observer Rapport Scales from Magee (2018).

The convergent validity of the LIWC variables identified by Driskell et al. (2013) was assessed by examining their correlations with scores on the observer rapport scales collected by Magee (2018). LIWC variables were measured separately for the interviewer and source in each interviewer. Table 26 below contains the zero-order correlations of the interviewer LIWC variables with these rapport scales, and Table 27 contains those for the source LIWC variables.

Both interviewer and source LIWC variables were expected to correlate at least weakly (r = .20) with all corresponding RS3i-S and RS3i-O scales thought measure the same "essential component" (Hypotheses 10a-c).

Table 28

Correlations of Interviewer LIWC Variables with Observer Rapport Scales from Magee (2018)

	Interviewer LIWC Variables								
Rapport Scale	First Person Plural	Present Focus	Social Processes	Positive Emotion	Negative Emotion	Assent	Nonfluencies	Conjunctions	Certainty
NRS Attentiveness	**.020**	**-.136**	**-.290**	.614**	-.217*	.432**	.357**	.392**	-.247*
NRS Positivity	.062	-.243*	-.304**	**.500****	**-.322****	.261*	.262*	.218*	-.089
GMISC Acceptance/Empathy	.050	-.245*	-.295**	**.532****	**-.337****	.290**	.276**	.212*	-.139
IRS Coordination	.023	-.164	-.239*	.419**	-.260*	**.211***	**.026**	**.222***	**-.143**

Note. Hypothesized correlations appear in bold.

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

Similar to previous results, most of the interviewer LIWC variables did not correlate to the hypothesized degree or in the hypothesized direction with their corresponding observer scales. However, interviewer positive emotion, negative emotion, and assent were identified as particularly robust indicators of rapport whose convergent validity was further supported. Specifically, interviewer use of positive emotion was positively correlated with NRS Positivity (r = .500, $p < .001$) and GMISC Acceptance/Empathy (r = .532, $p < .001$), while use of negative emotion was negatively correlated with these variables (r = -.322, $p = .002$; *and* r = -.337, $p = .001$, *respectively*). These results further supported the convergent validity of these two LIWC variables as indicators of Positivity. Additionally, the degree of interviewer assent was positively correlated with IRS Coordination (r = .211, $p = .044$), demonstrating evidence of its validity as an indicator of Coordination

Table 29

Correlations of Source LIWC Variables with Other Observer Scales

Rapport Scale	Source LIWC Variables								
	First Person Plural	Present Focus	Social Processes	Positive Emotion	Negative Emotion	Assent	Nonfluencies	Conjunctions	Certainty
NRS Attentiveness	**-.004**	**.184**	**-.265***	.083	-.132	.498**	.155	.006	-.247*
NRS Positivity	.086	.245*	-.346**	**.190**	**-.206***	.397*	.079	-.031	-.089
GMISC Acceptance/Empathy	.036	.242*	-.323**	**.134**	**-.203**	.402**	.068	-.070	-.139
IRS Coordination	.127	.199	-.247*	.041	-.193	**.217***	**-.044**	**.060**	**-.143**

Note. Hypothesized correlations appear in bold.

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

The previously demonstrated pattern of results persisted in that most source LIWC variables were not correlated with observer rapport scales as predicted. However, source use of negative emotion was negatively correlated with NRS Positivity (r = -.206, *p* = *.049*) and degree of assent was positively correlated with IRS Coordination (r = .217*p* = *.038),* providing additional evidence of their convergent validity.

Convergent Validity: Relationship Between LSM and RS3i-S, RS3i-O, and Other Observer Rapport Scales.

Language Style Matching (LSM) scores were correlated with RS3i-S (Table 30), RS3i-O (Table 31), and other observer rapport scales reported by Magee (2018; Table 32). LSM scores were expected to be at least weakly correlated (r = .20) with RS3i-S Connected Flow, RS3i-O Coordination, and IRS Coordination (Hypothesis 10c).

Table 30

Correlations of LSM with RS3i-S Scales

	RS3i-S Attentiveness	RS3i-S Trust/ Respect	RS3i-S Connected Flow
LSM	.052	-.064	.067

Note. *Correlation is significant at the 0.05 level (2-tailed).

Table 31

Correlations of LSM with RS3i-O Scales

	RS3i-O Attentiveness	RS3i-O Irritability	RS3i-O Coordination	RS3i-OTrust/ Respect
LSM	.211*	.182	-.018	-.163

Note. *Correlation is significant at the 0.05 level (2-tailed).

Table 32

Correlations of Source LIWC Variables with Magee (2018) Observational Scales

	NRS Attentiveness	NRS Positivity	GMISC Acceptance/Empathy	IRS Coordination
LSM	.030	-.176	-.176	-.064

Hypothesis 10c was not supported by these analyses. LSM scores were not significantly correlated with any self-report or observer measures of coordination. These results do not provide evidence of LSM's validity as an indicator of Coordination in investigative interviews.

Criterion Validity: Relationship Between LIWC Variables and Information Shared.

To evaluate both interviewer and source LIWC variables' criterion validity, they were each correlated with the amount of information the source shared during the interview as measured by the Shared Information Rating Scale (SIRS). LIWC variables were each predicted to correlate at least moderately (r = .30) with SIRS scores for Phase 2 of each interview and the interview as a whole (Hypothesis 11). The results of these analyses are presented below in Table 33.

Table 33

Correlations of Interviewer and Source LIWC Variables with SIRS Scores

	Interviewer LIWC Variables								
	First Person Plural	Present Focus	Social Processes	Positive Emotion	Negative Emotion	Assent	Nonfluencies	Conjunctions	Certainty
Information Shared Phase 2	-.229*	-.176	-.246*	.076	-0.253*	.035	-.083	-.077	-.060
Total Information Shared	-.188	-.210*	-.199	.066	-0.268*	.022	-.073	-.128	-.100

	Source LIWC Variables								
	First Person Plural	Present Focus	Social Processes	Positive Emotion	Negative Emotion	Assent	Nonfluencies	Conjunctions	Certainty
Information Shared Phase 2	-.331**	-.239*	-.204	-.277**	-.420**	-.049	-.040	.185	-.293**
Total Information Shared	-.295**	-.269*	-.084	-.308**	-.328**	-.098	-.072	.253	-.313**

Note. *Significant at the .05 level. **Significant at the .01 level.

While many of the LIWC variables for both interviewer and source were significantly correlated with the amount of information shared during the interview, they were not correlated in the predicted direction. Most LIWC variables predicted to positively correlate with information sharing were in fact negatively correlated with SIRS scores. This is likely attributable to the fact that, as sources used more present focus, for example, they were less likely to be discussing the past events counted as relevant details in SIRS scoring. One variable did show consistent evidence of criterion validity, however. Both interviewer and source use of negative emotion was negatively correlated to information shared in both Phase 2 and the interview as a whole.

Criterion Validity: Relationship Between LIWC Variables and Self-Report Cooperativeness.

To further assess interviewer and source LIWC variables' criterion validity, they were also correlated with sources' self-report of their perceptions of their own cooperativeness. LIWC variables were predicted to correlate at least moderately (r = .30) with self-report cooperativeness scores (Hypothesis 12). The results of these analyses are presented below in Table 34.

Table 34

Correlations of LIWC Variables and Self-Report Cooperativeness

	Interviewer LIWC Variables								
	First Person Plural	Present Focus	Social Processes	Positive Emotion	Negative Emotion	Assent	Nonfluencies	Conjunctions	Certainty
Self-Report Cooperativeness	.030	.014	-.153	.328**	-.103	.250*	-.002	.107	-.217*

	Source LIWC Variables								
	First Person Plural	Present Focus	Social Processes	Positive Emotion	Negative Emotion	Assent	Nonfluencies	Conjunctions	Certainty
Self-Report Cooperativeness	-.191	.102	-.114	-.121	-.103	.041	-.069	-.062	-.108

Only two LIWC variables were correlated as expected with self-report cooperativeness. Interviewer use of positive emotion (r = .328, p = .002) and degree of assent (r = .250, p = .020) were weakly-to-moderately correlated with self-report cooperativeness, providing additional evidence of their criterion validity.

Criterion Validity: Relationship Between LSM and Source Cooperation.

Language style matching (LSM) scores were also correlated with the amount of information shared during Phase 2 and the interview as a whole, as well as self-report cooperativeness. LSM was expected to correlated at least moderately (r = .30) with both measures of source cooperation (Hypotheses 11-12). The results of these analyses are located in Table 35 below.

Table 35

Correlations LSM with SIRS and Self-Report Cooperativeness

	Information Shared Phase 2	Total Information Shared	Self-Report Cooperativeness
LSM	.026	-.043	.128

Note. *Correlation is significant at the 0.05 level (2-tailed).

LSM was not significantly correlated with SIRS scores nor self-report cooperativeness. These results did not support LSM's validity.

Criterion Validity: Relationship Between LIWC Variables and Interview Style.

A one-way multiple analysis of variance (MANOVA) was performed to assess the relationship between interviewer and LIWC variables and interviewing styles employed by Duke

et al. (2018) in order to further establish evidence of the variables' criterion validity. Mean ratings for each of the LIWC variables and LSM were expected to differ significantly by experimental condition (Rapport, Neutral, Pressure), such that each of the variables (except Negative Emotion, Conjunctions, and Nonefluencies) was higher in the rapport condition than the pressured condition (Hypotheses 13a-b). Negative Emotion, Conjunctions, and Nonefluencies were expected to be lowest in the pressure interview condition and highest in the rapport condition. The results of the MANOVA F-tests are displayed below in Table 36 and the post-hoc Tukey's Honestly Significant Difference (HSD) contrasts are displayed in Table 37.

Table 36

Between Subjects Effects of Interview Condition on Interview & Source LIWC Variables

		Type III SS	df	Mean Square	F	Sig.	η^2
Interviewer	First Person Plural	0.11	2	0.06	0.66	.519	.015
	Present Focus	43.47	2	17.84	5.37	.006	.108
	Social Processes	28.64	2	14.32	7.39	.001	.142
	Positive Emotion	92.43	2	46.22	16.38	.000	.269
	Negative Emotion	3.96	2	1.98	2.51	.000	.220
	Assent	25.88	2	12.94	3.99	.022	.082
	Nonfluencies	13.19	2	6.59	5.72	.005	.114
	Conjunctions	20.84	2	10.421	3.33	.040	.070
	Certainty	6.70	2	3.35	23.56	.000	.346

		Type III SS	df	Mean Square	F	Sig.	η^2
Source	First Person Plural	0.50	2	0.03	0.72	.490	.016
	Present Focus	43.47	2	21.74	5.35	.006	.107
	Social Processes	63.73	2	31.87	4.96	.009	.100
	Positive Emotion	4.44	2	2.22	3.11	.050	.065
	Negative Emotion	2.83	2	1.42	2.51	.087	.053
	Assent	19.20	2	9.60	21.94	.000	.330
	Nonfluencies	7.92	2	3.96	1.76	.178	.038
	Conjunctions	0.76	2	0.38	0.15	.859	.003
	Certainty	0.59	2	0.29	0.80	.454	.018

Table 37

LIWC Variable Means by Interview Condition

	Interviewer								
	First Person Plural	Present Focus	Social Processes	Positive Emotion	Negative Emotion	Assent	Nonfluencies	Conjunctions	Certainty
Rapport	1.12a	14.05ab	18.00a	6.41a	0.98a	4.33a	1.55a	7.59a	1.38a
Neutral	1.02a	13.47a	19.26b	5.25b	0.77a	4.33a	0.68a	6.58a	1.40a
Pressured	1.06a	14.97b	19.12b	3.99c	1.27b	3.22b	0.98b	6.58a	1.96b
	Source								
	First Person Plural	Present Focus	Social Processes	Positive Emotion	Negative Emotion	Assent	Nonfluencies	Conjunctions	Certainty
Rapport	0.17a	9.54a	16.63a	2.81a	2.16a	1.88a	2.28a	7.42a	0.97a
Neutral	0.11a	7.97b	17.91ab	2.30b	2.53a	0.84a	1.55a	7.45a	1.09a
Pressured	0.14a	8.20b	18.63b	3.99c	2.54a	0.99b	1.97a	7.62a	1.17a

Note. For each scale, means sharing a common subscript are not significantly different at $p < .05$ according to the Tukey HSD procedure.

A significant main effect of interview style was observed for multiple interviewer LIWC variables, but for relatively few source LIWC variables (see Table 36). Further, post hoc analyses (Table 37) indicated that the majority of LIWC variables did not differ as predicted by interview style contrast. Once more, the strongest effects here were found for interviewer positive emotion, negative emotion, and assent, as well as source negative emotion and assent. These results provide limited support for LIWC variable criterion validity with the exception of those variables.

Criterion Validity: Relationship Between LSM and Interview Style.

An additional one-way MANOVA was performed to examine the differences in LSM by interview condition. A main effect of experimental condition (Rapport vs. Neutral vs. Pressure) was found for LSM scores, ($F [2,89] = 7.54$, $p < .001$, partial $\eta^2 = .145$), supporting Hypothesis 13a. Post-hoc Tukey HSD comparisons indicated that LSM was significantly higher for interviews in the Rapport and Pressure conditions than in the Neutral condition, and higher in the Pressure than the Rapport condition. This finding did not support Hypothesis 13b.

Summary of Part 3 Analyses

The purpose of Part 3 was to evaluate the convergent and criterion validity of the LIWC variables identified by Driskell et al. (2013) as likely correlates of the "essential components" of rapport as well as language style matching (LSM) between interviewers and sources. By and large, the results reported above do not consistently support the validity of the majority of these variables. However, Positive Emotion, Negative Emotion, and Assent, as measured by LIWC, repeatedly demonstrated evidence of validity as measures of rapport in investigative interviews and interrogations. These variables were most consistently related to components of rapport

judged by sources and observers, as well as measures of source cooperation. It is also important to note here the LIWC variable Certainty was consistently strongly correlated with convergent and criterion variables in the opposite direction as predicted. This was initially thought to be function of the semi-scripted experimental manipulation employed by Duke et al. (2018) when conducting the simulated interviews. However, ANOVA did not support this interpretation. Further investigation into the relationship between Certainty and investigative interview characteristics is required.

Discussion

The primary purpose of the present study was to build off of the work of Tickle-Degnen and Rosenthal (1990) as well as Duke et al. (2018) to develop a new measure that would allow observers to rate the rapport developed between an investigator and their source during an interview. The Rapport Scales for Investigative Interviews and Interrogations, Observer version (RS3i-O) was developed based on the results of exploratory factor analysis (EFA) of observer ratings of interviews of simulated interviews across 30 items. Analyses of the RS3i-O scales demonstrated their reliability and provided some evidence of their validity. Further, a confirmatory factor analysis (CFA) provided limited support of the instrument's 3-factor structure.

The secondary purpose of this study was to evaluate the validity of several linguistic variables measured by Linguistic Inquiry and Word Count (LIWC) thought to be indicators of rapport (Driskell et al., 2013) as well as language style matching (LSM), a variable calculated using LIWC thought to be related to interpersonal coordination. Analyses generally did not support these variables' validity as indicators of rapport in investigative interviews and interrogations.

RS3i-O Factor Structure

The RS3i-O's 3-factor structure identified in the EFA phase of this study was replicated later in a CFA, with mixed results. Fit indices including the comparative fit index (CFI) and standardized root mean square residual (SRMR) indicated good model fit. However, the chi-square test and, subsequently, the root mean square error of approximation (RMSEA), did not meet general standards of good fit, without making minor modifications to the model, When the residual variances of two items on the Irritability scale (Frustrated and Relaxed) and two on the

Coordination scale (Cooperative and Communicative) were allowed to correlate, model fit significantly improved. As discussed earlier, it is possible that different response patterns for these items will be found in actual investigative interviews and interrogations. Thus, the factor structure of the RS3i-O should be examined in an applied setting.

RS3i-O Items and Scales

The RS3i-O is comprised of twelve items forming four separate 3-item scales designed to be completed by an observer watching an investigative interview or interrogation (see Table 18, p. 85). The measure, and its accompanying material, was developed so that it is easily understood by those with English language abilities equivalent to a U.S. high-school student.

Each of the RS3i-O's scales is intended to measures one of the "essential components" of rapport proposed by the tripartite theory (Tickle-Degnen & Rosenthal, 1990). The Attentiveness scale is made up of items intended to reflect the degree to which an interviewer is exhibiting attentiveness towards a source and is intended to reflect an aspect of the Mutual Attentiveness component. The Irritability scale is made up of items that reflect an interviewer's temperament and is intended to be a negative indicator the Positivity component. The Coordination scale was designed to measure the interpersonal coordination exhibited between the interviewer and source in an interview. This scale is intended to reflect the "essential component" of coordination. The Trust/Respect scale is comprised of items that measure the degree to which an interviewer seems to be trustful and treating the source respect. This scale, like the Irritability scale, is also intended to measure the "essential component" of Positivity. The Trust/Respect scale is composed of items at the positive pole of the Positivity dimension, whereas the Irritability scale is composed of items at the negative pole.

RS3i-O Scale Reliability

The internal reliability of each RS3i-O scale was evaluated by computing Cronbach's alpha (α) and McDonald's omega (ω). All RS3i-O scales, including the supplementary Trust/Respect scale demonstrated a high degree of internal reliability as indicated by both of these statistics (see Table 19, p. 88). Each scale's interrater reliability was assessed by calculating the intraclass correlation coefficient (ICC) between raters on each team. All RS3i-O scales demonstrated at least adequate reliability, with the exception of the Attentiveness scale. This scale's interrater reliability fell slightly below the standards proposed by Hunsley and Mash (2008) to indicate adequate reliability (ICC ≥ .70).

However, when ICC values were calculated across all six raters included in the study, each scale met the criteria for excellent inter-rater reliability (>.80). This improvement of the "All Raters ICCs" is likely a result of the fact that they include primary raters. Final raters were specifically chosen on the basis of demonstrating the highest ICC values, and including these raters resulted in somewhat inflated ICC values. It is also important to note that, while scale scores averaged between raters were used in the validity analyses for the present study, inter-rater reliability (i.e., ICC) values were calculated between the individual raters rather than using these averaged scores. Thus, the All Raters ICCs for each RS3i-O scale were not further inflated because of including primary raters.

RS3i-O Scale Convergent Validity

The convergent validity of the RS3i-O was assessed by calculating the correlations of its scales with other self-report (RS3i-S) and observer (NRS, IRS, GMISC) scales designed to measure aspects of rapport. The RS3i-O Irritability, Coordination, and Trust/Respect scales, but not the Attentiveness scale, were all significantly correlated with their RS3i-S counterparts. All

of the RS3i-O scales demonstrated excellent convergent validity with other observer scales measuring the “essential components” of rapport.

RS3i-O Scale Discriminant Validity

The discriminant validity of the RS3i-O scales was assessed by examining their correlations with (1) RS3i-S scales designed to measure other “essential” components, (2) RS3i-S scales designed to measures aspects of rapport not included in the “essential three”, and (3) additional observer rapport scales designed to measure non-corresponding components of rapport. All of the RS3i-O scales showed excellent ability to discriminate between other observer scales of rapport. Further, the Irritability, Coordination, and Trust/Respect scales also demonstrated excellent ability to discriminate between self-report rapport scales designed to measure corresponding components of rapport and those not designed to measure other, non “essential” components of rapport (i.e., Expertise and Cultural Similarity). However, each of the RS3i-O scales demonstrated much less ability to discriminate between the RS3i-S scales designed to measure the “essential three.”

The Attentiveness scale demonstrated the least ability to discriminate between “essential three” components of rapport as perceived by the source. In fact, RS3i-O Attentiveness was only significantly correlated with the RS3i-S Expertise scale. It is possible that the items that make up the Attentiveness scale (Invested, Focused on Source, and Bored) were perceived by observers as expertise, given that interviewers were students rather than professional investigators.

RS3i-O Criterion Validity

The criterion validity of the RS3i-O scales was tested by comparing scale means between interview styles and by examining the scales' correlations with measures of source cooperation.

All of the scales were rated significantly higher (or lower, with respect to the Irritability scale) when interviewers used a rapport-based approach compared to an accusatorial, pressuring approach, providing evidence of their concurrent validity.

Further evidence of the Irritability and Coordination scales' criterion validity was demonstrated in that they were both significantly correlated with sources' perception of their own cooperativeness. However, only the Coordination was significantly correlated with the actual amount of information shared by sources during interview. This mirrors findings from Magee (2018), wherein it was found that only the IRS Coordination scale was correlated significantly with the amount of information shared in Phase 2 ($r = .323$, $p = .002$) as well as with total amount of information shared during the interview ($r = .317$, $p = .002$), but not in Phase 1 ($r = .111$).

RS3i-O Attentiveness Scale

It must be noted that the Attentiveness scale demonstrated the weakest psychometric properties of all the RS3i-O scales. As described in the preceding sections, the Attentiveness scale demonstrated insufficient inter-rater reliability and produced limited evidence of convergent and discriminant validity. It is possible that this scale's target construct, Attentiveness, is in some way more difficult for raters to capture than those measured by other RS3i-O scales. Abbe and Brandon (2013) have argued that the mutual attentiveness "essential component" may less applicable to interviews than in other types of interpersonal interactions.

However, previous research has illustrated that a similarly constructed observer scale performed much better than the RS3i-O Attentiveness scale. Specifically, the 3-item NRS Attentiveness scale developed by Bronstein et al. (2012) has shown good psychometric properties. In fact, NRS Attentiveness demonstrated excellent interrater reliability as well as

strong evidence of convergent and discriminant validity when employed on the same simulated interview videos rated in the present study (Magee, 2018).

This discrepancy in results is likely due to the content of the items on the three scales. That is, the item content on the RS3i-S and NRS Attentiveness scales are far more similar than either of those scales and the RS3i-O Attentiveness scale. For example, the RS3i-S Attentiveness scale items 'The Interviewer was attentive to me.' and 'The interviewer really listened to what I had to say.' contain nearly identical content to the NRS Attentiveness scale items Listening and Attentive, respectively. The corresponding RS3i-O scale does not include items as similar in content. The original pool of 10 attentiveness-related observer rapport items evaluated for inclusion in the RS3i-O did include items with this content; however, these items were ultimately not selected for inclusion in the final measure. The author of present study only had access to computed RS3i-S scale scores rather that individual item ratings. Future development of the RS3i-O would benefit from examining the relationships between individual RS3i-O and RS3i-S items.

RS3i-O Coordination Scale

It is also important to note an issue that arises from the observed relationship between the RS3i-O Coordination scale and the measures of source cooperation used to establish evidence of the scale's criterion validity. It is possible that these validity coefficients are inflated because raters simply saw a source being cooperative and relied solely on that as an indicator of coordination.

It is important to note here that primary goal of investigative interviews and interrogations is to gain information relevant to the current investigation (i.e., cooperation). In addition to cooperativeness, the RS3i-O Coordination scale also includes items intended to

measure shared expectations and communication. It is entirely possible to for an investigator and source to carry on a well-coordinated, highly communicative interaction that never reveal any relevant details. In fact, this is approach is sometimes taken by sources who seek to avoid sharing information by using counter-interrogation tactics (CITs; Alison et al., 2014). Thus, it does not seem likely that the relationships between this scale and measures of source cooperation is overly inflated.

However, this issue still ultimately brings into question the utility of the Coordination scale for applied settings. Does the scale have any predictive, rather than postdictive, validity? Would the scale be useful to investigators if it can only tell them how cooperative a source was *after* the interview was completed?

This issue actually highlights one of the RS3i-O's greatest strengths. Given the tool's very quick format and ease of use, it is possible to employ the RS3i-O and obtain scale ratings in a matter of minutes. This means that the RS3i-O scales could feasibly be employed at several points of time across the course of an interview. In addition to providing a novel means by which researcher can study the development of rapport over the course of an interview, this feature of the RS3i-O allows for a greater ability to test the validity of its scales. Future study should be performed in which *a priori* hypotheses are tested regarding the relationship between aspects of rapport displayed during the early stages of an interview and favorable outcomes that occur later during the interaction.

Linguistic variable validity

Most linguistic variables examined in the present study, including the nine LIWC variables identified by Driskell et al. (2013) and LSM, failed to demonstrate convergent validity with self-report and observer measures of rapport or source cooperation. However, three LIWC

variables -- Positive Emotion, Negative Emotion, and Assent -- were found to have consistent correlations with rapport in investigative interactions. These three variables demonstrated limited evidence of convergent validity with self-report and observer measures of rapport. Validity coefficients between these interviewer linguistic variables and observer measures designed to measure interviewer behavior were higher than were those between source linguistic variables and source ratings of rapport.

Implications of the Present Findings for Applied Settings, Theory, and Future Research

Applied Use

RS3i-O.

The RS3i-O provides a quick, simple means of accurately rating the rapport developed during an investigative interview or interrogation. The measure has clearly demonstrated accessibility, reliability, and validity. A final version of the RS3i-O rating sheet (Appendix L) and comprehension check (Appendix M) were created using only the items included in the final RS3i-O measure. In order to ensure high reliability between officers when employing the tool in applied settings such as during training or investigative support, it is important that the RS3i-O and its accompanying documents are easily understood and employed. While a sizeable proportion of law enforcement and national security officers have achieved a two- or four-year degree, there is a great deal of variation in education level between officers, and many employing the tool would likely have terminated their formal education after graduating high school. The Reading Ease (RE) score for the RS3i-O document itself was 58.9 with a Reading Grade Level (RGL) of 86., indicating the document is appropriate for a reader below U.S. ninth grade. The Comprehension Check resulted in a RE score of 46.4 with an RGL of 10.1, indicating

the document is appropriate for a reader in the U.S. tenth grade. Thus, it is a tool that should be easily accessible to any member of the law enforcement and national security communities.

Given the ease of training and use, excellent internal reliability, adequate inter-rater reliability, and strong evidence of validity demonstrated for its scales, the RS3i-O is recommended for use in applied investigative settings. Supervising or training officers can use the RS3i-O scales to assess how well an investigator is developing rapport with their source during training. The RS3i-O scales' demonstrated relationship with source perceptions of rapport also support the instrument's use in live interviews. Of particular value is the measure's potential utility as a proxy measure of sources' perceptions of rapport. Two important practical benefits of employing the RS3i-O's observer scales in addition to the RS3i-S self-report scales are that (1) the RS3i-O can be used during an ongoing interview and (2) RS3i-O does not rely on source cooperation (e.g., accurately completing a questionnaire).

Further, its's scales offer improvements over existing observational measures of rapport (e.g., Alison et al., 2013; Bronstein et al., 2012). The RS3i-O scales generally demonstrated strong evidence of convergent validity, which has not been demonstrated for other measures. An additional useful aspect of the RS3i-O is the ease with which raters can be trained to accurately employ the tool. This study clearly demonstrated that the RS3i-O scales can achieve a high degree of internal and interrater reliability with minimal training provided to novice raters. The training process is also uncomplicated and clearly outline in this document. Other published observational rating scales have demonstrated very little evidence of validity in investigative contexts, are either extremely complicated to employ, or are vague in their training processes.

Last, the instrument's short format requires few resources to train and employ and will allow the tool to be employed at any point during (or multiple points throughout) an interview.

This will allow for expanded study of how rapport is developed and maintained during investigative interviews and will allow for ratings of officers' rapport building over the course of an interview or across several interviews with the same source over time.

While the RS3i-O is recommended for use by law enforcement and national security practitioners in applied settings, some cautionary considerations are recommended. Specifically, the Attentiveness scale showed limited evidence of validity with regard to its relationship to sources' perceptions of interviewers' attentiveness and its scores should thus be interpreted with some degree of caution. The relationship between RS3i-O Attentiveness and sources' perceptions of interviewer Expertise requires further analysis. It is also recommended that practitioners employ the ad-hoc Trust/Respect scale along with the Attentiveness, Irritability, and Coordination scales.

Linguistic Variables.

The results of the present study did not support the validity of most of the LIWC variables examined or LSM as measures of the "essential components" of rapport. They would thus be of limited use to practitioners who wish to use them to evaluate the rapport developed during interviews.

The majority of these variables demonstrated no significant correlations with any ratings of rapport components, while several were significantly correlated in the direction opposite to what was predicted. No clear patterns emerged among these unpredicted relationships in the present study, though, with the exception of the LIWC variable, certainty. Interviewer expression of certainly exhibited consistently negative weak-to-moderate correlations (rs > .20) with self-report rapport scales and some large correlations (*r*s > .50) with observer rapport scales. In other

words, contrary to prediction, sources and raters perceived less rapport as interviewers expressed more certainty.

This pattern of relationships was not observed between interview outcome variables and source certainty, which suggests that these results are likely due to the partially-scripted nature of the experimental design employed by Duke et al. (2018). Recall that there were three distinct interview styles employed in that study: (1) an approach that emphasized behaviors thought to increase rapport, (2) a completely neutral approach, and (3) a pressuring, accusatorial approach. It is possible that interviewers employing the accusatorial approach were more likely to use words associated with certainty (e.g., "I am sure you know something you're not telling me."). Future analysis should be conducted to further examine if the expression of certainty is a meaningful indicator of rapport in investigative interviews or if these results are a byproduct of experimental design.

Importantly, some linguistic variables did demonstrate significant relationships as predicted by *a priori* hypotheses. Specifically, LIWC measurements of negative emotion, positive emotion, and assent were consistently correlated with observer and source ratings of rapport, demonstrating potential utility in measuring components of rapport in investigative interviewing. Interviewer's positive and negative emotion, and assent were correlated as predicted with both sources' RS3i-S ratings and observers' RS3i-O, NRS, GMISC, and IRS ratings related to positivity and coordination, respectively.

Considering their relationship with observer ratings, these three variables may be used in training or evaluation to objectively assess rater' accuracy in rating, though doing so is likely more practical in a research context rather than an applied setting, given the lengthy and effortful

process by which scores are obtained. To calculate each variable score, interviews must first be concluded and subsequently transcribed.

Further study of these variables is recommended prior to use in applied settings. Future research may focus on LIWC variables more relevant to the particular scenario being studied. For example, focus on past may be a better indicator of source engagement (related to mutual attentiveness) than focus on present.

Theoretical Implications

The results of this study add further support for the utility of the tripartite theory for conceptualizing and measuring rapport in investigative contexts. Tickle-Degnen & Rosenthal's (1990) proposed model was ultimately supported by factor analyses in the present study. An additional scale designed to measure the "essential component" Positivity was developed and this scale resulted in the inter-factor correlations proposed by the tripartite theory. It is important to note, though, that each "essential component" may be further separated into lower order constructs whose relationships have yet to be fully explored. This was illustrated in the present study by the Irritability and Trust/Respect scales. Though these scales were designed to measure the "essential component" of Positivity, they produced unique validity coefficients and provided different information in structural equation models (SEMs). Factor analyses indicated that the six items comprising both scales all load onto a single factor best interpreted as Positivity.

It is important to note that results indicated that RS3i-O Coordination scale may be the most useful scale included in the measure. Not only was the Coordination scale significantly correlated with every self-report and observer rapport scale, but it was the only RS3i-O scale to demonstrate criterion validity with regard to post-diction of source cooperation. This may seem

tautological, as Cooperative is one of the items on the RS3i-O Coordination scale. However, there are clearly other aspects of coordination that also play an important role in the valid measurement of the construct (e.g., effective communication and shared expectations between interviewer and source).

Though the Coordination scale demonstrated relatively greater utility, retaining all of the RS3i-O scales is recommended for future use, because scale provides unique information regarding source perceptions of rapport. The scales seem to represent related, but unique components and analyses did not support the proposition that the scales or their items all load onto a single underlying rapport factor. Three factor models demonstrated superior fit to unifactorial and bifactor model variants. Further, post-hoc SEM was conducted in which each of the RS3i-O scales loaded onto a higher-order factor, and all indices indicated poor model fit ($\chi 2$ (49) 114.61, $p < .001$, RMSEA = 0.121, RMSEA 90% CI = (0.092, 0.150), CFI = 0.939, and SRMR = 0.191). Rapport in investigative interviews may therefore best be conceived as a set of intercorrelated constructs, particularly for the purposes of measurement.

Future Research

Future researchers are highly recommended to use the RS3i-O. The items and rating instructions for final RS3i-O can be obtained in the appendices of this dissertation. The final RS3i-O rating form is located in in Appendix L and its accompanying training materials are located in Appendix M. Training and scoring procedures are described in detail in the Part 1 procedure section of the present study (page 24). Additionally, a manual is forthcoming that will compile all relevant information regarding the RS3i-O's training, rating, and scoring procedures and will be available from the author of this dissertation. Future versions of these materials can

be obtained by contacting the author of this dissertation (see Vita on page 220 for contact information).

As previously mentioned, training raters in use of the RS3i-O requires very little time and resources. The measure is quick to employ, and is the only such observer measure of rapport that has been specifically developed and validated for use in investigative interviews and interrogations. Future research may also validate the RS3i-O in other investigative contexts such as supervisory meetings between probation officers and their clients. The primary goal of these meetings is similar to that of investigative interviews – to gain relevant, credible information from the source. The instrument and accompanying training materials would require little-to-no modifications for this purpose.

Limitations and Future Directions

There were two main limitations of this study related to (1) the experimental design of the simulated interviews rated, and (2) the small sample size of both interviews and raters.

First, the RS3i-O was developed based on observations made using partially-scripted, simulated investigative interviews conducted in an experimental psychology laboratory. A clear next step in this line of research is to evaluate the tool's performance using non-simulated law enforcement interviews. It is possible that response patterns for the rapport items included in this study would differ during actual law enforcement and national security interviews. For example, in the present sample, the item Bored was a better indicator of attentiveness than the item Distracted. However, this may be a function of the experimental design employed by Duke et al. (2018) in conducting the simulated interviews rated in this study. Interviewers were student research assistants who were specifically tasked with conducting interviews. Because they were explicitly instructed to follow certain procedures, it is more likely that they would appear bored

than distracted while going about their task. For example, student interviewers did not have access to any other materials (other than a script) while interviews were conducted. In contrast, it may be more likely that a law enforcement or national security investigator conducting an interview would appear distracted than bored. As expert interviewers, they may endeavor to not appear bored; however, as investigators have access to paper files and electronic devices during interviews and are sometimes required to leave the room or talk with other investigators, it is likely that an interviewer could appear distracted to observers or their source.

For this reason, is recommended that three previously rejected items may be included in future observer ratings of rapport. The items Distracted, Approachable, and Productive may perform differently when applied to non-simulated interviews. Furthermore, the partially-scripted nature of the interactions may have resulted in a restriction of range of the language used by both the interviewer and the source. This may have affected the relationships between linguistic variables, ratings of rapport, and interview outcomes.

The second limitation of the present study was the small sample size employed. Because of the small sample size, it is possible that results may have been inflated due to the initial procedure of selecting items based on the strength of item characteristics and factor loadings. However, this does not represent a serious problem in the present study, as these procedures were only employed during the exploratory factor analyses (EFAs) and final formation of the scales. All confirmatory and validity analyses in Part 2 were performed using secondary ratings that were unrelated to the ratings used to formulate the scales. They therefore did not benefit from the procedures employed in Part 1. RS3i-O item factor loadings are not expected to differ significantly in future analyses, as they did not differ dramatically between exploratory and confirmatory factor analyses.

Further, the limited sample also resulted in insufficient statistical power to detect effects of rater characteristics (e.g., gender, ethnicity) if there were present in the data. The limited number of interviews (n = 92) and raters (n = 6) was chosen for practical purposes of completing this study; however, future analysis of the RS3i-O should seek to validate its scales across a broader range of raters in order to explore possible effects. Now that the RS3i-O has been developed and its items pared down from 30 to nine, a follow up to this study should employ a much larger sample of raters to observe recorded non-simulated interviews. Further, the RS3i-O is ultimately intended to be employed by trained law enforcement and national security agents. Thus, another potential future avenue of research involves developing a manual for training law enforcement officers and evaluating their use of the RS3i-O.

The small number of raters may have also contributed to rater fatigue. Each rater was responsible for rating 30 items across 92 interviews for a total of 2760 individual ratings. While this is undoubtedly a large number of ratings to be performed accurately, rater fatigue was mitigated by spreading out ratings as much as practically possible. Raters were limited to 2-hour rating sessions each day and no more than 3 days of rating per week. Where possible, rating sessions were spread out across the week (e.g., two hours Monday, two hours Wednesday, two hours Friday). Taking frequent breaks in between rating videos was encouraged. Ratings were spread out over a 5-month period, such that raters only rated an average of 4.6 interviews per week for a weekly total of 138 individual item ratings. For these reasons, it is not likely that rater fatigue caused a serious issue in the present study. Regardless, collecting RS3i-O ratings from a much larger sample of raters in future studies would greatly reduce the burden on individual raters while achieving statistical power necessary for analyses.

One final limitation of this study is that it relied on observer rapport measures that did not control for order effects of items. That is, this study used data from observational rapport ratings previously collected without controlling for the potential effect of item order. Magee (2018) did not randomize item order on the rating scales as they were simply employed as published by their developers (e.g., Bronstein et al., 2012). Likewise, no attempt was made to control for order effects during the development of the RS3i-O scales. This study sought to validate the training materials and procedures accompanying the RS3i-O in addition to the tool itself. The items were not randomized during the development because their order would not be randomized when used in applied settings. In order to evaluate the scales' utility in applied settings, rating sheets, training materials, and procedures were presented to raters exactly as they would be to law enforcement investigators. Future study of the RS3i-O may benefit from controlling for effects of item order, though, to gain further insight into the validity and effectiveness of each scale as well as the items included therein.

Conclusion

The primary purpose of this study was the development of a practical observer measure of rapport for use in investigative interviews and interrogations. This measure, the Rapport Scales for Investigative Interviews and Interrogations, Observer version (RS3i-O), was constructed and the reliability and validity of each of its scales was demonstrated. While future work needs to be done to explore the psychometric properties of the RS3i-O in "real world" interviews, this tool presents an exciting new means of measuring rapport across a variety of investigative interactions.

References

Alison, L. J., Alison, E., Noone, G., Elntib, S., & Christiansen, P. (2013). Why tough tactics fail and rapport gets results: Observing rapport-based interpersonal techniques (ORBIT) to generate useful information from terrorists. *Psychology, Public Policy, & Law*, *19*, 411-431. https://doi.org/10.1037/a0034564

Abbe, A., & Brandon, S. E. (2012). The role of rapport in investigative interviewing: A review. *Journal of Investigative Psychology and Offender Profiling*, *10*(3), https://doi.org/10.1002/jip.1386

Abbe, A., & Brandon, S. E. (2013). Building and maintaining rapport in investigative interviews. *Police Practice and Research, 15*(3), 207-220. https://doi.org/10.1080/15614263.2013.827835

Akaike, H. (1987) *Factor Analysis and AIC.* In: Parzen E., Tanabe K., Kitagawa G. (eds) Selected Papers of Hirotugu Akaike. Springer Series in Statistics (Perspectives in Statistics). Springer, New York, NY. https://doi.org/10.1007/978-1-4612-1694-0_29

Alison, L. J., Alison, E., Noone, G., Elntib, S., Waring, S., & Christiansen, P. (2014). The efficacy of rapport-based techniques for minimizing counter-interrogation tactics among a field sample of terrorists. *Psychology, Public Policy, & Law, 20*(4), 421-430. https://doi.org/10.1037/law0000021

Bernieri, F. J., Davis, J. M., Rosenthal, R., Knee, C. R. (1994). Interactional synchrony and rapport: measuring synchrony in displays devoid of sound and facial affect. *Personality and Social Psychology Bulletin, 20,* 303-311. https://doi.org/10.1177/0146167294203008

Bernieri, F. J., & Gillis, J. S. (2001). Judging rapport employing Brunswik's lens model to study interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.*), Interpersonal sensitivity* (pp. 67-88). Mahwah, NJ: Lawrence Erlbaum Associates.

Bernieri, F. (2005). The expression of rapport. In V. Manusov (Ed.), *The Sourcebook of Nonverbal Measures*: *Going Beyond Words*. Lawrence Erlbaum Associates.

Boyd, R.L. (2017). Psychological text analysis in the digital humanities. In: S. Hai-Jew (Ed.) *Data Analytics in Digital Humanities: Multimedia Systems and Applications.* Springer, Cham. https://doi.org/10.1007/978-3-319-54499-1_7

Bronstein, I., Nelson, N., Livnat, Z., & Ben-Ari, R. (2012). Rapport in negotiation: The contribution of the verbal channel. *Journal of Conflict Resolution, 56*(6), 1089-1115. https://doi.org/10.1177/0022002712448913

Bryant, F. B., & Satorra, A. (2013). EXCEL macro file for conducting scaled difference chi-square tests via LISREL 8, LISREL 9, EQS, and MPlus. Available from the authors.

Carmody, P. C., Mateo, J. C., Bowers, D., & McCloskey, M. J. (2017). Linguistic coordination as an unobtrusive, dynamic indicator of rapport, prosocial team processes, and performance in team communication. *Proceedings of the Human Factors and Ergonomics Annual Meeting, 61,* 140-144. https://doi.org/10.1177/1541931213601518

Chen F, Curran PJ, Bollen KA, Kirby J, Paxton P. (2008). An empirical evaluation of the use of fixed cutoff points in RMSEA test statistic in structural equation models. *Sociological Methods & Research, 36,* 462–494. https://doi.org/10.1177/0049124108314720.

Collins, R., Lincoln, R., & Frank, M. G. (2002). The effect of rapport in forensic interviewing. *Psychiatry, Psychology and Law, 9*(1), 69-78. https://doi.org/10.1375/pplt.2002.9.1.69

Cronbach, L. J., (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. https://doi.org/10.1007/BF02310555

Driskell, J. E., & Driskell, T. (2011). Rapport and the investigative interview. Report provided to Federal Bureau of Investigation.

Driskell, T., Blickensderfer, E. L., & Salas, E. (2013). Is three a crowd? Examining rapport in investigative interviews. *Group Dynamics: Theory, Research, and Practice, 17*(1), 1-13. https://doi.org/10.1037/a0029686

Drolet, A. L., & Morris, M. W. (2000). Rapport in conflict resolutions: Accounting for how nonverbal exchange fosters cooperation on mutually beneficial settlements to mixed motive conflicts. *Journal of Experimental Social Psychology, 36*, 26-50. https://doi.org/10.1006/jesp.1999.1395

Duke, M. C., Wood, J. M., Bollin, B., Scullin, M., & LaBianca, J. (2018a). Development of the rapport scales for investigative interviews and interrogations (RS3i), interviewee version. *Psychology, Public Policy, and Law, 24*(1), 64-79. https://doi.org/10.1037/law0000147

Duke, M. C., Wood, J. M., Magee, J., & Escobar, H. (2018b). The effectiveness of army field manual interrogation approaches for educing information and building rapport. *Law and Human Behavior, 42*(5), 442-457. https://doi.org/10.1037/lhb0000299

Flesch, R. (1948). A new readability yardstick. *Journal of Applied Psychology, 32*, 221–233. https://doi.org/10.1037/h0057532

Hall, J. A., Roter, D. L., Blanch, D. C., & Frankel, R. M. (2009). Observer-rated rapport in interactions between medical students and standardized patients. *Patient Education and Counseling, 76*, 323-327. https://doi.org/10.1016/j.pec.2009.05.009

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6,* 1–55.

Hunsley, J., & Mash, E. J. (2008). Developing criteria for evidence-based assessment: An introduction to assessments that work. In J. Hunsley & E. J. Mash (Eds.), *A guide to assessments that work* (pp. 3-14). New York: Oxford University Press.

Ireland, M. E., & Pennebaker, James, W. (2010). Language style matching in writing: Synchrony in essays, correspondence, and poetry. *Journal of Personality and Social Psychology, 99*(3), 349-571. https://doi.org/10.1037/a0020386

Ireland, M. E., Slatcher, R. B., Eastwick, P. W., Scissors, L. E., Finkel, E. J. & Pennebaker, J. W. (2011). Language Style Matching Predicts Relationship initiation and stability. *Psychological Sciences, 22*(1), 39-44. https://doi.org/10.1177/0956797610392928

Jiang-yuan, Z., & Wei, G. (2012). Who is controlling the interaction? The effect of nonverbal mirroring on teacher-student rapport. *US-China Education Review A7,* 662-669.

Kaiser, H. F., & Rice, J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement, 34,* 111-117. https://doi.org/10.1177/001316447403400115

Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research, 44*(3), 486–507. https://doi.org/10.1177/0049124114543236

Kenny, D. A. (2015*). Measuring model fit*. Retrieved from http://www.davidakenny.net/cm/fit.htm

Kincaid, J. P., Fishburne, R. P., Rogers, R. L., & Chissom, B. S. (1975). *Derivation of new readability formulas (Automated Readability Index, Fog Count, and Flesch Reading Ease Formula) for Navy enlisted personnel* (Research Branch Report 8–75). Millington, TN:

Naval Air Station Memphis. Retrieved from http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA006655

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Magee, J. M. (2018). Convergent validity of five observational rating scales as measures of rapport in investigative interviews and interrogations. (Unpublished master's thesis). University of Texas at El Paso, El Paso, TX.

McDonald, R.P. (1985). *Factor analysis and related methods.* Hillsdale, NJ: Lawrence Erlbaum.

Miller, W. R., & Rollnick, S. (1992). *Motivational interviewing: Preparing people for change*. New York, NY: Guilford Press. https://doi.org/10.1002/casp.2450020410

Muthén, L. K., & Muthén, B.O. (2012). Mplus user's guide. Seventh edition. Los Angeles, CA: Muthén & Muthén.

Muthén, L. K., & Muthén, B.O. (2013). Chi-square difference testing using the Satorra-Bentler scaled chi-square. Retrieved from http://www.statmodel.com/chidiff.shtml

Niederhoffer, K. G. & Pennebaker, J. W. (2002). Linguistic style matching in social interaction. *Journal of Language and Social Psychology, 21*(4), 337-360. https://doi.org/10.1177/026192702237953

Pennebaker, J. W., Boyd, R. L., Jordan, K., & Blackburn, K. (2015). *The development and psychometric properties of LIWC2015*. Austin, TX: University of Texas at Austin. Retrieved from http://hdl.handle.net/2152/31333

Perkins, D., Schenk, T. A., Stephan, L., Vrungos, S., & Wynants, S. (1995). Effects of rapport, intellectual excitement, and learning on students' perceived ratings of college instructors. *Psychological Reports, 76*, 627-635. https://doi.org/10.2466/pr0.1995.76.2.627

Quinones, A. (2019). Linguistic variables as components of rapport in investigative interactions. (Unpublished undergraduate honors thesis). University of Texas at El Paso, El Paso, TX.

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544-559. https://doi.org/10.1080/00223891.2010.496477

Richardson, B. H., Taylor, P. J., Snook, B., Conchie, S. M., & Bennell, C. (2014). Language style matching and police interrogation outcomes. *Law and Human Behavior, 38*(4), 357-366. https://doi.org/10.1037/lhb0000077

Ricks, E. P., & Eno Louden, J. (2014). The relationship between officer orientation and supervision strategies in community corrections. *Law and Human Behavior, 39*(2), 130-141. https://doi.org/10.1037/lhb0000098

Rogan, R. G. (2011) Linguistic style matching in crisis negotiations: A comparative analysis of suicidal and surrender outcomes. *Journal of Police Crisis Negotiations, 11*(1), 20-39. https://doi.org/10.1080/15332586.2011.523302

Satorra, A. (2000). Scaled and adjusted restricted tests in multi-sample analysis of moment structures. In Heijmans, R.D.H., Pollock, D.S.G. & Satorra, A. (eds.), *Innovations in multivariate statistical analysis*. London: Kluwer Academic Publishers.

Tabachnick, B. G., & Fidell, L. S., (2013). *Using multivariate statistics.* Boston, MA: Pearson.

Taylor, P. J. & Thomas, S. (2008). Linguistic style matching and negotiation outcome. *Negotiation and Conflict Management Research, 1*(3), 263-281. https://doi.org/10.1111/j.1750-4716.2008.00016.x

Tickle-Degnen, L., & Rosenthal, R. (1987). *Group rapport and nonverbal behavior*. In C. Hendrick (Ed.), Review of personality and social psychology, Vol. 9. Group processes and intergroup relations (p. 113–136). Sage Publications, Inc.

Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry, 1*, 285-293. https://doi.org/10.1207/s15327965pli0104_1

Tickle-Degnen, L. (1998). Working well with others: The Prediction of Students' Clinical Performance. *The American Journal of Occupational Therapy, 52*(2), 133-142. https://doi.org/10.5014/ajot.52.2.133

Tickle-Degnen, L., & Puccinelli, N. M. (1999). The nonverbal expression of negative emotions: Peer and supervisor responses to occupational therapy student's emotional states. *The Occupational Therapy Journal of Research, 19*(1), 18-37. https://doi.org/10.1177/153944929901900102

Vallano, J. P., & Schreiber Compo, N. (2011). A Comfortable Witness Is a Good Witness: Rapport-Building and Susceptibility to Misinformation in an Investigative Mock-Crime Interview. *Applied Cognitive Psychology, 25*, 960-970. https://doi.org/10.1002/acp.1789

Vallano, J. P., & Schreiber Compo, N. (2015). Rapport-building with cooperative witnesses and criminal suspects: a theoretical and empirical review. *Psychology, Public Policy, and Law, 21*, 85-99. https://doi.og/10.1037/law0000035

Walsh, D., & Bull, R. (2010). What really is effective interviews with suspects? A study comparing interviewing skills against interviewing outcomes. *Legal and Criminological Psychology*, *15*, 305-321. https://doi.org/10.1348/135532509X463356

Walsh, D., & Bull, R. (2012). Examining rapport in investigative interviews with suspects: Does its building and maintenance work? Journal of Police and Criminal Psychology, *27*, 73-84. https://doi.org/10.1007/s11896-011-9087-x

Woodcock, R. W., Muñoz-Sandoval, A. F., Ruef, M. L., & Alvarado, C. G. (2005). *Woodcock Muñoz Language Survey-Revised.* Rolling Meadows, IL: Riverside Publishing.

Appendix A.

Rapport Scales for Interrogations and Investigative Interviews, Source version (RS3i-S)

1. I think the Interviewer is generally honest with me.

SD D N A SA

2. The Interviewer did his/her job with skill during the interview.

SD D N A SA

3. The Interviewer respects my knowledge.

SD D N A SA

4. The Interviewer and I have our culture in common.

SD D N A SA

5. The Interviewer performed expertly during the interview.

SD D N A SA

6. I think that the Interviewer can generally be trusted to keep his/her word.

SD D N A SA

7. The Interviewer and I probably share the same ethnicity.

SD D N A SA

8. The Interviewer really listened to what I had to say.

SD D N A SA

9. I was motivated to perform well during the interview.

SD D N A SA

10. I feel I can trust the Interviewer to keep his/her word to me.

SD D N A SA

11. The Interviewer made an effort to do a good job.

SD D N A SA

12. The Interviewer acted like a professional.

SD D N A SA

13. The Interviewer paid careful attention to my opinion.

SD D N A SA

14. The Interviewer and I got along well during the interview.

SD D N A SA

15. The Interviewer and I worked well together as a team.

SD D N A SA

16. The Interviewer probably shares my culture.

SD D N A SA

17. I wanted to do a good job during the interview.

SD D N A SA

18. The Interviewer was attentive to me.

SD D N A SA

19. Communication went smoothly between the Interviewer and me.

SD D N A SA

20. The Interviewer was interested in my point of view.

SD D N A SA

21. I felt committed to accomplishing the goals of the interview.

SD D N A SA

Select **SD** if the statement is definitely false or if you **strongly disagree**.

Select **D** if the statement is mostly false or if you **disagree**.

Select **N** if the statement is about equally true or false, if you cannot decide, or if you are **neutral** about the statement.

Select **A** if the statement is mostly true or if you **agree**.

Select **SA** if the statement is definitely true or if you **strongly agree**.

Appendix B.

Pool of 30 Observer Rating Items of Rapport Evaluated for Inclusion in the Rapport Scales for Investigative Interviews and Interrogations, Observer version (RS3i-O)

1. Actively
2. Alert
3. Interested
4. Involved
5. Distracted
6. Invested
7. Focused on Source
8. Thoughtful
9. Bored
10. Attentive

11. Trusting
12. Understanding
13. Frustrated
14. Approachable
15. Respectful
16. Honest
17. Aggressive
18. Pleasant
19. Relaxed
20. Positive

21. Peaceful
22. Awkward
23. Productive
24. Tense
25. Smooth
26. Cooperative
27. Relaxed
28. Communicative
29. Shared Expectations
30. Coordinated

Note. Attentiveness items (1-10), positivity items (11-20), and coordination items (21-30). Items in bold are those included in the final RS3i-O.

Appendix C.

Informed Consent Document (Duke et al., 2018)

University of Texas at El Paso (UTEP) Institutional Review Board
Informed Consent Form for Research Involving Human Subjects

Protocol Title: Rapport in investigative interviews - Source Version
Principal Investigator: Brock Bollin; Misty Duke, PhD
UTEP: Psychology
Proposal 609970: Concurrent Validity of the Rapport Scales for Investigative Interviews and Interrogations (RS3i)
Approved by UTEP IRB: 6-12-2014
Closed: 7-30-2015

---

1. Introduction

You are being asked to take part voluntarily in the research project described below. Please take your time making a decision and feel free to discuss it with your friends and family. Before agreeing to take part in this research study, it is important that you read the consent form that describes the study. Please ask the study researcher or the study staff to explain any words or information that you do not clearly understand.

2. Why is this study being done?

You have been asked to take part in a research study of to develop a measure of rapport between a Source and an Investigator in the context of an investigative interview. Approximately, 90, will be enrolling in this study at UTEP. You are being asked to be in the study because you are a student at UTEP If you decide to enroll in this study, your involvement will last about 1 ½ hours.

3. What is involved in the study?
If you agree to take part in this study, you will participate in an investigative interview as a Source. You will be required to view a video and respond to questions posed by interviewer. You will complete questionnaires after the interview.

The entire investigative interview will be recorded. The video will then be shown to students who are participating in this study. The students will be asked to rate your actions, words and emotions during the interview. The video of the interview will probably also be saved and viewed by other students who participate in future studies approved by the UTEP Institutional Review Board. Those students will also be asked to rate your actions, words and emotions during the interview. It is also possible that the video of the interview will be shown during scientific presentations or course presentations at UTEP or other universities.

4. What are the risks and discomforts of the study?

There are no known risks associated with this research. The interview in which you will participate will be about the video that you watched and will not deal with embarrassing or highly personal matters.

5. What will happen if I am injured in this study?

The University of Texas at El Paso and its affiliates do not offer to pay for or cover the cost of medical treatment for research related illness or injury. No funds have been set aside to pay or reimburse you in the event of such injury or illness. You will not give up any of your legal rights by signing this consent form. You should report any such injury to Brock Bollin at bcbollin@miners.utep.edu and to the UTEP Institutional Review Board (IRB) at (915-747-8841) or irb.orsp@utep.edu.

6. Are there benefits to taking part in this study?

There will be no direct benefits to you for taking part in this study. (You will receive class credit or payment for participation in this study. You will receive 1 ½ credits. As a student, you will benefit through learning about how psychological research is conducted. This research may help us to understand how to improve Source cooperation through the development of rapport in investigative interviews.

7. What other options are there?

You have the option not to take part in this study. There will be no penalties involved if you choose not to take part in this study.

8. Who is paying for this study?

Internal Funding:

Funding for this study is provided by UTEP Department of UTEP Department of Psychology.

External funding:

UTEP and *list the names of the investigators* are receiving funding from *list the name of the sponsor or organization* to conduct this study.

9. What are my costs?

There are no direct costs. You will be responsible for travel to and from the research site and any other incidental expenses.

10. Will I be paid to participate in this study?

You will not be paid for taking part in this research study.

Appendix D.

Rater Information Questionnaire

1. What is your age? _____

2. What is your gender? □ Male □ Female □ Other

3. What is your race/ethnicity?

   □ Hispanic
   □ Non-Hispanic White
   □ African-American
   □ Asian-American
   □ Native American
   □ Other (please specify): _____

4. What is your student status (freshman, sophomore, etc.)? _____

5. What is your major? _____

6. What is your minor? _____

7. What is your GPA? _____

8. What was the first language you learned?

   □ English □ Spanish □ Other

9. What is the language you speak the most now?

   □ English □ Spanish □ Other

10. Do you have prior experience rating interactions?

    □ Yes □ No

11. If 'Yes', please explain your experience.

    _____

    _____

Appendix E.

Pool of 30 Observer Rapport Items' Comprehension Check

A1. Attentiveness item ratings should be made based on whose behavior?

_____

A2. Provide a definition of **Actively Listening** in your own words.

_____

_____

A3. Rate how similar the following characteristics are to **Actively Listening** on a scale of 0 (Not similar) to 4 (Very similar).

A3.1 Distracted

0	1	2	3	4
Not similar				Very similar

A3.2 Accepting

0	1	2	3	4
Not similar				Very similar

A4. Describe, in your own words, behaviors that indicate an individual is **Actively Listening**.

_____

_____

_____

A5. Provide a definition of **Alert** in your own words.

_____

_____

_____

A6. Rate how similar the following characteristics are to being **Alert** on a scale of 0 (Not similar) to 4 (Very similar).

A6.1 Active

0	1	2	3	4
Not similar				Very similar

A6.2 Sharp

0	1	2	3	4
Not similar				Very similar

A7. Describe, in your own words, behaviors that indicate an individual is **Alert**.

_____
_____
_____

A8. Provide a definition of **Interested** in your own words.

_____
_____

A9. Rate how similar the following characteristics are to being **Interested** on a scale of 0 (Not similar) to 4 (Very similar).

A9.1 Excited

0	1	2	3	4
Not similar				Very similar

A9.2 Uncaring

0	1	2	3	4
Not similar				Very similar

A10. Describe, in your own words, behaviors that indicate an individual is **Interested**.

_____
_____
_____

A11. Provide a definition of **Involved** in your own words.

_____
_____
_____

A12. Rate how similar the following characteristics are to being **Involved** on a scale of 0 (Not similar) to 4 (Very similar).

A12.1 Complicated

0	1	2	3	4
Not similar				Very similar

A12.2 Participating

0	1	2	3	4
Not similar				Very similar

A13. Describe, in your own words, behaviors that indicate an individual is **Involved**.

_____
_____
_____

A14. Provide a definition of **Distracted** in your own words.

_____
_____
_____

A15. Rate how similar the following characteristics are to being **Distracted** on a scale of 0 (Not similar) to 4 (Very similar).

A15.1 Upset

0	1	2	3	4
Not similar				Very similar

A15.2 Distant

0	1	2	3	4
Not similar				Very similar

A16. Describe, in your own words, behaviors that indicate an individual is **Distracted**.

_____
_____
_____

A17. Provide a definition of **Invested** in your own words.

_____
_____
_____

A18. Rate how similar the following characteristics are to being **Invested** on a scale of 0 (Not similar) to 4 (Very similar).

A18.1 Happy

0	1	2	3	4
Not similar				Very similar

A18.2 Caring

0	1	2	3	4
Not similar				Very similar

A19. Describe, in your own words, behaviors that indicate an individual is **Invested**.

_____
_____
_____

A20. Provide a definition of **Focused on Source** in your own words.

_____
_____
_____

A21. Rate how similar the following characteristics are to being **Focused on Source** on a scale of 0 (Not similar) to 4 (Very similar).

A21.1 Attracted

0	1	2	3	4
Not similar				Very similar

A21.2 Concentrating

0	1	2	3	4
Not similar				Very similar

A22. Describe, in your own words, behaviors that indicate an individual is **Focused on Source**.

_____
_____
_____

A23. Provide a definition of **Thoughtful** in your own words.

_____
_____
_____

A24. Rate how similar the following characteristics are to being **Thoughtful** on a scale of 0 (Not similar) to 4 (Very similar).

A24.1 Helpful

0	1	2	3	4
Not similar				Very similar

A24.2 Polite

0	1	2	3	4
Not similar				Very similar

A25. Describe, in your own words, behaviors that indicate an individual is **Thoughtful**.

_____
_____
_____

A26. Provide a definition of **Bored** in your own words.

_____
_____
_____

A27. Rate how similar the following characteristics are to being **Bored** on a scale of 0 (Not similar) to 4 (Very similar).

A27.1 Uninteresting

0	1	2	3	4
Not similar				Very similar

A27.2 Tired

0	1	2	3	4
Not similar				Very similar

A28. Describe, in your own words, behaviors that indicate an individual is **Bored**.

_____
_____
_____

A29. Provide a definition of **Attentive** in your own words.

_____
_____
_____

A30. Rate how similar the following characteristics are to being **Attentive** on a scale of 0 (Not similar) to 4 (Very similar).

A30.1 Watchful

0	1	2	3	4
Not similar				Very similar

A30.2 Respectful

0	1	2	3	4
Not similar				Very similar

A31. Describe, in your own words, behaviors that indicate an individual is **Attentive**.

_____
_____
_____

A32. Provide a definition of **Peaceful** in your own words.

_____
_____

P1. Positivity item ratings should be made based on whose behavior?

_____

P2. Provide a definition of **Trusting** in your own words.

_____

_____

P3. Rate how similar the following characteristics are to being **Trusting** on a scale of 0 (Not similar) to 4 (Very similar).

P3.1Trustworthy

0	1	2	3	4
Not similar				Very similar

P3.2 Accepting

0	1	2	3	4
Not similar				Very similar

P4. Describe, in your own words, behaviors that indicate an individual is **Trusting.**

_____

_____

_____

P5. Provide a definition of **Understanding** in your own words.

_____

_____

_____

P6. Rate how similar the following characteristics are to being **Understanding** on a scale of 0 (Not similar) to 4 (Very similar).

P6.1 Kind

0	1	2	3	4
Not similar				Very similar

P6.2 Accepting

0	1	2	3	4
Not similar				Very similar

P7. Describe, in your own words, behaviors that indicate an individual is **Understanding**.

_____
_____
_____

P8. Provide a definition of **Frustrated** in your own words.

_____
_____

P9. Rate how similar the following characteristics are to being **Frustrated** on a scale of 0 (Not similar) to 4 (Very similar).

P9.1 Upset

0	1	2	3	4
Not similar				Very similar

P9.2 Angry

0	1	2	3	4
Not similar				Very similar

P10. Describe, in your own words, behaviors that indicate an individual is **Frustrated**.

_____
_____
_____

P11. Provide a definition of **Approachable** in your own words.

_____
_____
_____

P12. Rate how similar the following characteristics are to being **Approachable** on a scale of 0 (Not similar) to 4 (Very similar).

P12.1 Distant

0	1	2	3	4
Not similar				Very similar

P12.2 Friendly

0	1	2	3	4
Not similar				Very similar

P13. Describe, in your own words, behaviors that indicate an individual is **Approachable**.

_____
_____
_____

P14. Provide a definition of **Respectful** in your own words.

_____
_____
_____

P15. Rate how similar the following characteristics are to being **Respectful** on a scale of 0 (Not similar) to 4 (Very similar).

P15.1 Polite

0	1	2	3	4
Not similar				Very similar

P15.2 Kind

0	1	2	3	4
Not similar				Very similar

P16. Describe, in your own words, behaviors that indicate an individual is **Respectful**.

_____
_____
_____

P17. Provide a definition of **Honest** in your own words.

_____
_____
_____

P.18 Rate how similar the following characteristics are to being **Honest** on a scale of 0 (Not similar) to 4 (Very similar).

P18.1 Reliable

0	1	2	3	4
Not similar				Very similar

P18.2 Direct

0	1	2	3	4
Not similar				Very similar

P19. Describe, in your own words, behaviors that indicate an individual is **Honest**.

_____
_____
_____

P20. Provide a definition of **Aggressive** in your own words.

_____
_____
_____

P21. Rate how similar the following characteristics are to being **Aggressive** on a scale of 0 (Not similar) to 4 (Very similar).

P21.1 Intimidating

0	1	2	3	4
Not similar				Very similar

P21.2 Cocky

0	1	2	3	4
Not similar				Very similar

P22. Describe, in your own words, behaviors that indicate an individual is **Aggressive**.

_____
_____
_____

P23. Provide a definition of **Pleasant** in your own words.

_____
_____
_____

P24. Rate how similar the following characteristics are to being **Pleasant** on a scale of 0 (Not similar) to 4 (Very similar).

P24.1 Happy

0	1	2	3	4
Not similar				Very similar

P24.2 Agreeable

0	1	2	3	4
Not similar				Very similar

P25. Describe, in your own words, behaviors that indicate an individual is **Pleasant**.

_____
_____
_____

P26. Provide a definition of **Relaxed** in your own words.

_____
_____
_____

P27. Rate how similar the following characteristics are to being **Relaxed** on a scale of 0 (Not similar) to 4 (Very similar).

P27.1 Casual

0	1	2	3	4
Not similar				Very similar

P27.2 Calm

0	1	2	3	4
Not similar				Very similar

P28. Describe, in your own words, behaviors that indicate an individual is **Relaxed**.

_____
_____
_____

P29. Provide a definition of **Positive** in your own words.

_____
_____
_____

P30. Rate how similar the following characteristics are to being **Positive** on a scale of 0 (Not similar) to 4 (Very similar).

P30.1 Certain

0	1	2	3	4
Not similar				Very similar

P30.2 Happy

0	1	2	3	4
Not similar				Very similar

P31. Describe, in your own words, behaviors that indicate an individual is **Positive**.

_____
_____
_____

C1. Coordination item ratings should be made based on whose behavior?

_____

C2. Rate how similar the following characteristics are to being **Peaceful** on a scale of 0 (Not similar) to 4 (Very similar).

C2.1 Quiet

0	1	2	3	4
Not similar				Very similar

C2.2 Unfriendly

0	1	2	3	4
Not similar				Very similar

C3. Describe, in your own words, behaviors that indicate an interview is **Peaceful.**

_____
_____
_____

C4. Provide a definition of **Awkward** in your own words.

_____
_____
_____

C5. Rate how similar the following characteristics are to being **Awkward** on a scale of 0 (Not similar) to 4 (Very similar).

C5.1 Rude

0	1	2	3	4
Not similar				Very similar

C5.2 Unpleasant

0	1	2	3	4
Not similar				Very similar

C6. Describe, in your own words, behaviors that indicate an interview is **Awkward**.

_____
_____
_____

C7. Provide a definition of **Productive** in your own words.

_____
_____

C8. Rate how similar the following characteristics are to being **Productive** on a scale of 0 (Not similar) to 4 (Very similar).

C8.1 Useless

0	1	2	3	4
Not similar				Very similar

C8.2 Constructive

0	1	2	3	4
Not similar				Very similar

C9. Describe, in your own words, behaviors that indicate an interview is **Productive**.

_____
_____
_____

C10. Provide a definition of **Tense** in your own words.

_____
_____
_____

C11. Rate how similar the following characteristics are to being **Tense** on a scale of 0 (Not similar) to 4 (Very similar).

C11.1 Dramatic

0	1	2	3	4
Not similar				Very similar

C11.2 Casual

0	1	2	3	4
Not similar				Very similar

C12. Describe, in your own words, behaviors that indicate an interview is **Tense**.

_____
_____
_____

C13. Provide a definition of **Smooth** in your own words.

_____
_____
_____

C14. Rate how similar the following characteristics are to being **Smooth** on a scale of 0 (Not similar) to 4 (Very similar).

C14.1 Soft

0	1	2	3	4
Not similar				Very similar

C14.2 Civilized

0	1	2	3	4
Not similar				Very similar

C15. Describe, in your own words, behaviors that indicate an interview is **Smooth**.

_____
_____
_____

C16. Provide a definition of **Cooperative** in your own words.

_____
_____
_____

C17. Rate how similar the following characteristics are to being **Cooperative** on a scale of 0 (Not similar) to 4 (Very similar).

C17.1 Supportive

0	1	2	3	4
Not similar				Very similar

C17.2 Harmonious

0	1	2	3	4
Not similar				Very similar

C18. Describe, in your own words, behaviors that indicate an interview is **Cooperative**.

_____
_____
_____

C19. Provide a definition of **Relaxed** in your own words.

_____
_____
_____

C20. Rate how similar the following characteristics are to being **Relaxed** on a scale of 0 (Not similar) to 4 (Very similar).

C20.1 Casual

0	1	2	3	4
Not similar				Very similar

C20.2 Tolerant

0	1	2	3	4
Not similar				Very similar

C21. Describe, in your own words, behaviors that indicate an interview is **Relaxed**.

_____
_____
_____

C22. Provide a definition of **Communicative** in your own words.

_____
_____
_____

C23. Rate how similar the following characteristics are to being **Communicative** on a scale of 0 (Not similar) to 4 (Very similar).

C23.1 Talkative

0	1	2	3	4
Not similar				Very similar

C23.2 Informative

0	1	2	3	4
Not similar				Very similar

C24. Describe, in your own words, behaviors that indicate an interview is **Communicative**.

_____
_____
_____

C25. Provide a definition of **Shared Expectations** in your own words.

_____
_____
_____

C26. Rate how similar the following characteristics are to being **Shared Expectations** on a scale of 0 (Not similar) to 4 (Very similar).

C26.1 Like-minded

0	1	2	3	4
Not similar				Very similar

C26.2 Agreeing

0	1	2	3	4
Not similar				Very similar

C27. Describe, in your own words, behaviors that indicate an interaction shows **Shared Expectations**.

_____
_____
_____

C28. Provide a definition of **Coordinated** in your own words.

_____
_____
_____

C29. Rate how similar the following characteristics are to being **Coordinated** on a scale of 0 (Not similar) to 4 (Very similar).

C29.1 Harmonized

0	1	2	3	4
Not similar				Very similar

C29.2 Disorganized

0	1	2	3	4
Not similar				Very similar

C30. Describe, in your own words, behaviors that indicate an interview is **Coordinated**.

_____
_____
_____

Appendix F.

Pool of 30 Observer Rapport Items' Comprehension Check Scoring Key

Instructions: Acceptable answers to Likert-type questions are highlighted. All definitions and descriptions should be theoretically consistent. Score each item as either correct (1) or incorrect (O).

C31. Positivity item ratings should be made based on whose behavior?
INTERVIEWER

C32. Provide a definition of **Trusting** in your own words.

C33. Rate how similar the following characteristics are to being **Trusting** on a scale of 0 (Not similar) to 4 (Very similar).

3.1 Trustworthy

1	2	3	4	5
Not similar				Very similar

3.2 Accepting

1	2	3	4	5
Not similar				Very similar

C34. Describe, in your own words, behaviors that indicate an individual is **Trusting.**

C35. Provide a definition of **Understanding** in your own words.

_____
_____
_____

C36. Rate how similar the following characteristics are to being **Understanding** on a scale of 0 (Not similar) to 4 (Very similar).

6.1 Kind

1	2	3	4	5
Not similar				Very similar

6.2 Accepting

1	2	3	4	5
Not similar				Very similar

C37. Describe, in your own words, behaviors that indicate an individual is **Understanding**.

_____
_____
_____

C38. Provide a definition of **Frustrated** in your own words.

_____
_____

C39. Rate how similar the following characteristics are to being **Frustrated** on a scale of 0 (Not similar) to 4 (Very similar).

9.1 Upset

1	2	3	4	5
Not similar				Very similar

9.2 Angry

1	2	3	4	5
Not similar				Very similar

C40. Describe, in your own words, behaviors that indicate an individual is **Frustrated**.

_____
_____
_____

C41. Provide a definition of **Approachable** in your own words.

_____
_____
_____

C42. Rate how similar the following characteristics are to being **Approachable** on a scale of 0 (Not similar) to 4 (Very similar).

12.1 Distant

1	2	3	4	5
Not similar				Very similar

12.2 Friendly

1	2	3	4	5
Not similar				Very similar

C43. Describe, in your own words, behaviors that indicate an individual is **Approachable**.

_____
_____
_____

C44. Provide a definition of **Respectful** in your own words.

_____
_____
_____

C45. Rate how similar the following characteristics are to being **Respectful** on a scale of 0 (Not similar) to 4 (Very similar).

15.1 Polite

1	2	3	4	5
Not similar				Very similar

15.2 Kind

1	2	3	4	5
Not similar				Very similar

C46. Describe, in your own words, behaviors that indicate an individual is **Respectful**.

_____
_____
_____

C47. Provide a definition of **Honest** in your own words.

_____
_____
_____

C48. Rate how similar the following characteristics are to being **Honest** on a scale of 0 (Not similar) to 4 (Very similar).

18.1 Reliable

1	2	3	4	5
Not similar				Very similar

18.2 Direct

1	2	3	4	5
Not similar				Very similar

C49. Describe, in your own words, behaviors that indicate an individual is **Honest**.

_____
_____
_____

C50. Provide a definition of **Aggressive** in your own words.

_____
_____
_____

C51. Rate how similar the following characteristics are to being **Aggressive** on a scale of 0 (Not similar) to 4 (Very similar).

21.1 Intimidating

1	2	3	4	5
Not similar				Very similar

21.2 Cocky

1	2	3	4	5
Not similar				Very similar

C52. Describe, in your own words, behaviors that indicate an individual is **Aggressive**.

_____
_____
_____

C53. Provide a definition of **Pleasant** in your own words.

_____
_____
_____

C54. Rate how similar the following characteristics are to being **Pleasant** on a scale of 0 (Not similar) to 4 (Very similar).

24.1 Happy

1	2	3	4	5
Not similar				Very similar

24.2 Agreeable

1	2	3	4	5
Not similar				Very similar

C55. Describe, in your own words, behaviors that indicate an individual is **Pleasant**.

_____
_____
_____

C56. Provide a definition of **Relaxed** in your own words.

_____
_____
_____

C57. Rate how similar the following characteristics are to being **Relaxed** on a scale of 0 (Not similar) to 4 (Very similar).

27.1 Casual

1	2	3	4	5
Not similar				Very similar

27.2 Calm

1	2	3	4	5
Not similar				Very similar

C58. Describe, in your own words, behaviors that indicate an individual is **Relaxed**.

_____
_____
_____

C59. Provide a definition of **Positive** in your own words.

_____
_____
_____

C60. Rate how similar the following characteristics are to being **Positive** on a scale of 0 (Not similar) to 4 (Very similar).

30.1 Certain

1	2	3	4	5
Not similar				Very similar

30.2 Happy

1	2	3	4	5
Not similar				Very similar

C61. Describe, in your own words, behaviors that indicate an individual is **Positive**.

_____
_____
_____

C62. Attentiveness item ratings should be made based on whose behavior?

_____INTERVIEWER_____

C63. Provide a definition of **Actively Listening** in your own words.

_____
_____

C64. Rate how similar the following characteristics are to **Actively Listening** on a scale of 0 (Not similar) to 4 (Very similar).

34.1 Distracted

1	2	3	4	5
Not similar				Very similar

34.2 Accepting

1	2	3	4	5
Not similar				Very similar

C65. Describe, in your own words, behaviors that indicate an individual is **Actively Listening**.

_____
_____
_____

C66. Provide a definition of **Alert** in your own words.

_____
_____
_____

C67. Rate how similar the following characteristics are to being **Alert** on a scale of 0 (Not similar) to 4 (Very similar).

37.1 Active

1	2	3	4	5
Not similar				Very similar

37.2 Sharp

1	2	3	4	5
Not similar				Very similar

C68. Describe, in your own words, behaviors that indicate an individual is **Alert**.

_____
_____
_____

C69. Provide a definition of **Interested** in your own words.

_____
_____

C70. Rate how similar the following characteristics are to being **Interested** on a scale of 0 (Not similar) to 4 (Very similar).

40.1 Excited

1	2	3	4	5
Not similar				Very similar

40.2 Uncaring

1	2	3	4	5
Not similar				Very similar

C71. Describe, in your own words, behaviors that indicate an individual is **Interested**.

_____
_____
_____

C72. Provide a definition of **Involved** in your own words.

_____
_____
_____

C73. Rate how similar the following characteristics are to being **Involved** on a scale of 0 (Not similar) to 4 (Very similar).

43.1 Complicated

1	2	3	4	5
Not similar				Very similar

43.2 Participating

1	2	3	4	5
Not similar				Very similar

C74. Describe, in your own words, behaviors that indicate an individual is **Involved**.

_____
_____
_____

C75. Provide a definition of **Distracted** in your own words.

_____
_____
_____

C76. Rate how similar the following characteristics are to being **Distracted** on a scale of 0 (Not similar) to 4 (Very similar).

45.1 Upset

1	2	3	4	5
Not similar				Very similar

45.2 Distant

1	2	3	4	5
Not similar				Very similar

C77. Describe, in your own words, behaviors that indicate an individual is **Distracted**.

_____
_____
_____

C78. Provide a definition of **Invested** in your own words.

_____
_____
_____

C79. Rate how similar the following characteristics are to being **Invested** on a scale of 0 (Not similar) to 4 (Very similar).

49.1 Happy

1	2	3	4	5
Not similar				Very similar

49.2 Caring

1	2	3	4	5
Not similar				Very similar

C80. Describe, in your own words, behaviors that indicate an individual is **Invested**.

_____
_____
_____

C81. Provide a definition of **Focused on Source** in your own words.

_____
_____
_____

C82. Rate how similar the following characteristics are to being **Focused on Source** on a scale of 0 (Not similar) to 4 (Very similar).

52.1 Attracted

1	2	3	4	5
Not similar				Very similar

52.2 Concentrating

1	2	3	4	5
Not similar				Very similar

C83. Describe, in your own words, behaviors that indicate an individual is **Focused on Source**.

_____
_____
_____

C84. Provide a definition of **Thoughtful** in your own words.

_____
_____
_____

C85. Rate how similar the following characteristics are to being **Thoughtful** on a scale of 0 (Not similar) to 4 (Very similar).

55.1 Helpful

1	2	3	4	5
Not similar				Very similar

55.2 Polite

1	2	3	4	5
Not similar				Very similar

C86. Describe, in your own words, behaviors that indicate an individual is **Thoughtful**.

_____
_____
_____

C87. Provide a definition of **Bored** in your own words.

_____
_____
_____

C88. Rate how similar the following characteristics are to being **Bored** on a scale of 0 (Not similar) to 4 (Very similar).

58.1 Uninteresting

1	2	3	4	5
Not similar				Very similar

58.2 Tired

1	2	3	4	5
Not similar				Very similar

C89. Describe, in your own words, behaviors that indicate an individual is **Bored**.

_____
_____
_____

C90. Provide a definition of **Attentive** in your own words.

_____
_____
_____

C91. Rate how similar the following characteristics are to being **Attentive** on a scale of 0 (Not similar) to 4 (Very similar).

61.1 Watchful

1	2	3	4	5
Not similar				Very similar

61.2 Respectful

1	2	3	4	5
Not similar				Very similar

C92. Describe, in your own words, behaviors that indicate an individual is **Attentive**.

_____
_____
_____

C93. Provide a definition of **Peaceful** in your own words.

_____
_____

C94. Coordination item ratings should be made based on whose behavior?

_____INTERVIEWER & SOURCE (INTERVIEW AS A WHOLE)_____

C95. Rate how similar the following characteristics are to being **Peaceful** on a scale of 0 (Not similar) to 4 (Very similar).

65.1 Quiet

1	2	3	4	5
Not similar				Very similar

65.2 Unfriendly

1	2	3	4	5
Not similar				Very similar

C96. Describe, in your own words, behaviors that indicate an interview is **Peaceful.**

_____
_____
_____

C97. Provide a definition of **Awkward** in your own words.

_____
_____
_____

C98. Rate how similar the following characteristics are to being **Awkward** on a scale of 0 (Not similar) to 4 (Very similar).

68.1 Rude

1	2	3	4	5
Not similar				Very similar

68.2 Unpleasant

1	2	3	4	5
Not similar				Very similar

C99. Describe, in your own words, behaviors that indicate an interview is **Awkward**.

_____
_____
_____

C100. Provide a definition of **Productive** in your own words.

_____
_____

C101. Rate how similar the following characteristics are to being **Productive** on a scale of 0 (Not similar) to 4 (Very similar).

71.1 Useless

1	2	3	4	5
Not similar				Very similar

71.2 Constructive

1	2	3	4	5
Not similar				Very similar

C102. Describe, in your own words, behaviors that indicate an interview is **Productive**.

_____
_____
_____

C103. Provide a definition of **Tense** in your own words.

_____
_____
_____

C104. Rate how similar the following characteristics are to being **Tense** on a scale of 0 (Not similar) to 4 (Very similar).

74.1 Dramatic

1	2	3	4	5
Not similar				Very similar

74.2 Casual

1	2	3	4	5
Not similar				Very similar

C105. Describe, in your own words, behaviors that indicate an interview is **Tense**.

_____
_____
_____

C106. Provide a definition of **Smooth** in your own words.

_____
_____
_____

C107. Rate how similar the following characteristics are to being **Smooth** on a scale of 0 (Not similar) to 4 (Very similar).

77.1 Soft

1	2	3	4	5
Not similar				Very similar

77.2 Civilized

1	2	3	4	5
Not similar				Very similar

C108. Describe, in your own words, behaviors that indicate an interview is **Smooth**.

_____
_____
_____

C109. Provide a definition of **Cooperative** in your own words.

_____
_____
_____

C110. Rate how similar the following characteristics are to being **Cooperative** on a scale of 0 (Not similar) to 4 (Very similar).

80.1 Supportive

1	2	3	4	5
Not similar				Very similar

80.2 Harmonious

1	2	3	4	5
Not similar				Very similar

C111. Describe, in your own words, behaviors that indicate an interview is **Cooperative**.

_____
_____
_____

C112. Provide a definition of **Relaxed** in your own words.

_____
_____
_____

C113. Rate how similar the following characteristics are to being **Relaxed** on a scale of 0 (Not similar) to 4 (Very similar).

83.1 Casual

1	2	3	4	5
Not similar				Very similar

83.2 Tolerant

1	2	3	4	5
Not similar				Very similar

C114. Describe, in your own words, behaviors that indicate an interview is **Relaxed**.

_____
_____
_____

C115. Provide a definition of **Communicative** in your own words.

_____
_____
_____

C116. Rate how similar the following characteristics are to being **Communicative** on a scale of 0 (Not similar) to 4 (Very similar).

86.1 Talkative

1	2	3	4	5
Not similar				Very similar

86.2 Informative

1	2	3	4	5
Not similar				Very similar

C117. Describe, in your own words, behaviors that indicate an interview is **Communicative**.

_____
_____
_____

C118. Provide a definition of **Shared Expectations** in your own words.

_____
_____
_____

C119. Rate how similar the following characteristics are to being **Shared Expectations** on a scale of 0 (Not similar) to 4 (Very similar).

89.1 Like-minded

1	2	3	4	5
Not similar				Very similar

89.2 Agreeing

1	2	3	4	5
Not similar				Very similar

C120. Describe, in your own words, behaviors that indicate an interaction shows **Shared Expectations**.

_____
_____
_____

C121. Provide a definition of **Coordinated** in your own words.

_____
_____
_____

C122. Rate how similar the following characteristics are to being **Coordinated** on a scale of 0 (Not similar) to 4 (Very similar).

92.1 Harmonized

1	2	3	4	5
Not similar				Very similar

92.2 Disorganized

1	2	3	4	5
Not similar				Very similar

C123. Describe, in your own words, behaviors that indicate an interview is **Coordinated**.

_____
_____
_____

TOTAL CORRECT: _____

SCORE (TOTAL/123): _____

Appendix G.

Rating Document for Pool of 30 Observer Rapport Items

Think about the interview you just watched. Read each characteristic below and rate how well it describes **INTERVIEWER** from Not at all (0) to Very well (6).

1. Actively Listening – Did the interviewer respond appropriately to what the source said? Did the interviewer correctly summarize what the source previously said?

0 1 2 3 4 5 6

Not at all Very well

2. Alert – Did the interviewer show they were lively or energetic?

0 1 2 3 4 5 6

Not at all Very well

3. Interested – Did the interviewer seem interested in what the source had to say?

0 1 2 3 4 5 6

Not at all Very well

4. Involved – Did interviewer actively participate in the interview?

0 1 2 3 4 5 6

Not at all Very well

5. Distracted – Did the interviewer seem distracted during the interview?

0 1 2 3 4 5 6

Not at all Very well

6. Invested – Did the interviewer seem to care about the result of the interview?

0 1 2 3 4 5 6

Not at all Very well

7. Focused on Source – Did the interviewer concentrate on the source during the whole interview?

0 1 2 3 4 5 6

Not at all Very well

8. Thoughtful – Did the interviewer seem to think carefully about what they said and how they acted?

0 1 2 3 4 5 6

Not at all Very well

9. Bored – Did the interviewer seem uninterested during the interview?

0 1 2 3 4 5 6

Not at all Very well

10. Attentive – How much did the interviewer pay attention to the source? Did they listen without interrupting while the source was speaking?

Attentiveness

0 1 2 3 4 5 6

Not at all Very well

11. Trusting –Did the interviewer seem to trust the source? Did the interviewer seem to believe what the source said?

0 1 2 3 4 5 6

Not at all Very well

12. Understanding – Did the interviewer accept the source and what the source said? Were they tolerant of the source's point of view? (Remember, the interviewer can disagree while remaining understanding.)

0 1 2 3 4 5 6

Not at all Very well

13. Frustrated – Did the interviewer seem frustrated with the source or what the source said?

0 1 2 3 4 5 6

Not at all Very well

14. Approachable – Did the interviewer seem agreeable or friendly?

0 1 2 3 4 5 6
Not at all Very well

15. Respectful – Did the interviewer treat the source with respect? Did the interviewer express their point of view without insulting, intimidating, or threatening the source?

0 1 2 3 4 5 6

Not at all Very well

16. Honest – Did the interviewer seem honest?

0 1 2 3 4 5 6

Not at all Very well

17. Aggressive – Did the interviewer seem aggressive, hostile, or intimidating?

0 1 2 3 4 5 6

Not at all Very well

18. Pleasant – Did the interviewer contribute to creating a peaceful and calm (rather than hostile or argumentative) interview?

0 1 2 3 4 5 6
Not at all Very well

19. Relaxed – Did the interviewer seem calm and laid-back?

0 1 2 3 4 5 6

Not at all Very well

20. Positive – Did the interviewer show positivity toward the source during the interaction? Did they encourage the source to participate or try to make the source feel accepted?

0 1 2 3 4 5 6

Not at all Very well

Think about the interview you just watched. Read each characteristic below and rate how well it describes **INTERACTION AS A WHOLE** from Not at all (0) to Very well (6).

21. Peaceful – Was the interview calm and peaceful?

0 1 2 3 4 5 6

Not at all Very well

22. Awkward – Was the interview awkward? Were there frequent interruptions or awkward silent periods?

0 1 2 3 4 5 6

Not at all Very well

23. Productive – Was the interview productive?

0 1 2 3 4 5 6

Not at all Very well

24. Tense – Was the interview tense?

0 1 2 3 4 5 6

Not at all Very well

25. Smooth – Was the interaction smooth? Did the conversation flow smoothly from one topic to the next?

0 1 2 3 4 5 6

Not at all Very well

26. Cooperative – How much did the interviewer and source cooperate with each other? Did they work together toward the same goal?

0 1 2 3 4 5 6

Not at all Very well

27. Relaxed – Was the interaction calm and relaxed?

0 1 2 3 4 5 6

Not at all Very well

28. Communicative – Did the interviewer and source communicate effectively? Were they open and candid during their conversation?

0 1 2 3 4 5 6

Not at all Very well

29. Shared Expectations – Were the interviewer and source "on the same page"? Did they know what to expect from each other and the interview?

0 1 2 3 4 5 6

Not at all Very well

30. Coordinated – How much were the interactants "in sync"? Did the conversation switch from one speaker to the next without interruptions? Did the interviewer and source's postures and physical expressions seem appropriate compared to their partner's?

0 1 2 3 4 5 6

Not at all Very well

NOTES

Use these pages to take notes while watching the interview. You may find taking notes helpful in making ratings. You may pause the video at any time. Notes should include relevant behaviors or observations to help you make ratings after you have finished watching the interview.

1. Actively Listening

_____
_____
_____

2. Alert

_____
_____
_____

3. Interested

_____
_____
_____

4. Involved

_____
_____
_____

5. Distracted

_____
_____
_____

6. Invested

_____
_____
_____

7. Focused on Source

8. Thoughtful

9. Bored

10.Attentive

11. Trusting

12. Understanding

13. Frustrated

14. Approachable

15. Respectful

16. Honest

17. Aggressive

18. Pleasant

19. Relaxed

20. Positive

21. Peaceful

22. Awkward

23. Productive

24. Tense

25. Smooth

26. Cooperative

27. Relaxed

28. Communicative

29. Shared Expectations

30. Coordinated

Appendix H.

Research Assistant Transcription Instructions

1. Labeling system for files
    - **Interview#_Initials_TranscriptionDate**
        - 32_JM_090317 is *Interview #32 by JM on September 3, 2017*

2. Make a copy of the blank transcription file & label accordingly

3. Every time someone speaks (until the next person speaks), this is a **turn**
    - Interviewer turns will be labeled I1, I2, I3, etc
    - Source turns will be labeled S1, S2, S3, etc.
    - Example:

        I1: Hi! How are you doing today? Are your classes going well?
        S1: I'm okay.
        I2: What about your classes?
        S2: They are okay.
    - Continue numbering through break to Phase 2

4. If you cannot understand something that is said, transcribe as much as possible and fill blank spaces with a bolded, bracketed, **[inaudible]** statement

5. **Interruptions:** note with a dash at the point of interruption " – "
   (Be sure dash has **spaces** on both sides)
    - Example: I2: How are you -
      S2: Fine!

6. **Nonfluency spelling:**
    - Um (**General Nonfluency**)
    - Mhmm (**Yes**)
    - Huh? (**Questioning**)
    - Mm-mm (**No**)

7. **Mark Timestamp:**
    - Beginning of each Phase
        - Mark end of Phase if there is a major time break between phases
    - If a phrase is **[inaudible]**
    - If something notable/strange occurs
    - If you make a judgement call as to what was said / how it was said
    - Last Turn of the interview

Appendix I.

Correlations Between Items in the Pool of 30 Observer Rapport Items

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.	28.	29.	30.
1. Actively Listening	1																													
2. Alert	.77	1																												
3. Interested	.82	.88	1																											
4. Involved	.79	.86	.91	1																										
5. Distracted	-.57	-.63	-.67	-.69	1																									
6. Invested	.64	.78	.75	.76	-.67	1																								
7. FoS	.64	.70	.78	.73	-.76	.76	1																							
8. Thoughtful	.77	.82	.88	.82	-.60	.62	.65	1																						
9. Bored	-.73	-.80	-.82	-.85	.82	-.78	-.77	-.72	1																					
10. Attentive	.82	.77	.80	.76	-.69	.67	.73	.79	-.76	1																				
11. Trusting	.64	.64	.64	.57	-.30	.33	.38	.81	-.41	.65	1																			
12. Understanding	.65	.73	.70	.66	-.34	.47	.46	.78	-.48	.64	.89	1																		
13. Frustrated	-.48	-.32	-.40	-.36	.15*	-.01*	-.13*	-.62	.14	-.45	-.82	-.66	1																	
14. Approachable	.65	.75	.74	.68	-.38	.49	.51	.85	-.50	.63	.88	.91	-.68	1																
15. Respectful	.63	.60	.63	.56	-.31	.31	.40	.80	-.39	.60	.93	.86	-.83	.89	1															
16. Honest	.59	.64	.66	.56	-.31	.33	.43	.78	-.41	.57	.87	.83	-.67	.89	.88	1														
17. Aggressive	-.45	-.32	-.37	-.32	.09*	-.02	-.12	-.62	.14	-.43	-.85	-.68	.92	-.72	-.89	-.72	1													
18. Pleasant	.64	.71	.73	.65	-.38	.43	.49	.85	-.47	.66	.91	.91	-.73	.96	.93	.91	-.79	1												
19. Relaxed	.48	.38	.44	.38	-.11*	.07*	.18	.65	-.15*	.45	.84	.71	-.87	.76	.86	.75	-.91	.82	1											
20. Positive	.66	.79	.77	.71	-.44	.54	.56	.84	-.55	.66	.87	.93	-.65	.96	.87	.87	-.68	.95	.72	1										
21. Peaceful	.45	.44	.47	.39	-.16	.14	.21	.67	-.26	.47	.82	.70	-.83	.77	.86	.72	-.87	.80	.81	.72	1									
22. Awkward	-.43	-.39	-.47	-.42	.31	-.42	-.45	-.40	.39	-.41	-.28	-.35	.23	-.38	-.33	-.30	.22	-.35	-.26	-.40	-.32	1								
23. Productive	.29	.28	.16*	.14*	-.02*	.23	.14*	.21	-.10*	.34	.27	.23	-.26	.25	.22	.08	-.23	.21	.27	.20	.33	-.28	1							
24. Tense	-.48	-.48	-.49	-.40	.19	-.20	-.21	-.66	.26	-.43	-.78	-.72	.78	-.77	-.83	-.69	.82	-.78	-.79	-.73	-.92	.39	-.32	1						
25. Smooth	.52	.51	.49	.50	-.09*	.33	.19	.49	-.33	.43	.51	.59	-.43	.57	.52	.46	-.43	.52	.42	.54	.54	-.57	.36	-.53	1					
26. Cooperative	.41	.43	.31	.27	-.04*	.24	.15*	.43	-.16*	.42	.54	.49	-.52	.53	.50	.39	-.51	.50	.55	.49	.63	-.35	.81	-.62	.61	1				
27. Relaxed	.46	.43	.44	.39	-.16	.13*	.17*	.65	-.26	.43	.81	.72	-.85	.78	.87	.72	-.89	.79	.82	.73	.95	-.31	.32	-.92	.54	.62	1			
28. Communicative	.46	.46	.35	.30	-.09*	.30	.21	.44	-.19	.44	.52	.51	-.47	.56	.51	.41	-.47	.50	.50	.50	.59	-.43	.81	-.61	.63	.94	.60	1		
29. SE	.51	.48	.45	.37	-.12*	.31	.24	.52	-.28	.49	.59	.56	-.53	.60	.60	.50	-.54	.56	.54	.56	.67	-.52	.64	-.69	.64	.80	.69	.82	1	
30. Coordinated	.50	.45	.45	.40	-.22	.33	.29	.51	-.33	.50	.52	.54	-.42	.59	.53	.49	-.48	.56	.51	.55	.58	-.69	.53	-.59	.67	.67	.58	.71	.77	1

Note. FoS = Focused on Source. SE = Shared Expectations.

*Correlation **not** significant at the .05 level.

Appendix J.

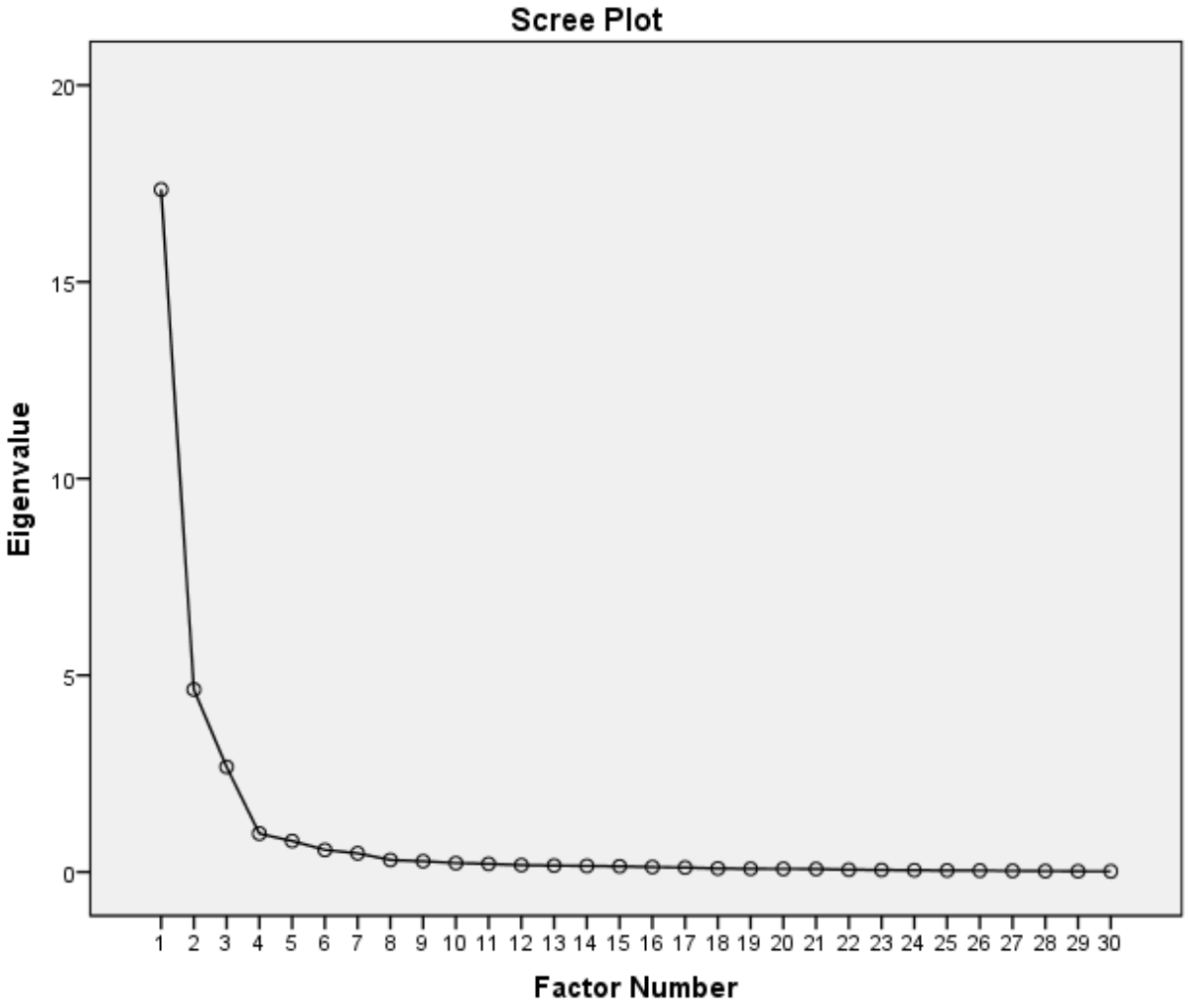
Reliability and Item Statistics by Hypothesized Scale

Hypothesized Scale	Cronbach's Alpha for Scale	Item	Standardized Item Loading	Item Information	Corrected Item-Total Correlation	Cronbach's Alpha if Item Deleted
Attentiveness	0.963					
		1. Actively Listening	0.848	3.02	0.827	0.959
		2. Alert	0.915	5.61	0.897	0.956
		3. Interested	0.956	11.12	0.935	0.955
		4. Involved	0.939	7.96	0.917	0.955
		5. Distracted*	0.75	1.71	0.755	0.962
		6. Invested	**0.811**	**2.37**	**0.803**	**0.96**
		7. Focused on Source	**0.811**	**2.36**	**0.809**	**0.961**
		8. Thoughtful	0.878	3.82	0.843	0.961
		9. Bored*	**0.887**	**4.18**	**0.884**	**0.957**
		10. Attentive	0.851	3.08	0.851	0.959
Positivity	0.979					
		11. Trusting	**0.937**	**7.14**	**0.948**	**0.975**
		12. Understanding	0.926	6.00	0.895	0.976
		13. Frustrated*	**0.771**	**13.82**	**0.813**	**0.979**
		14. Approachable	0.963	12.7	0.929	0.975
		15. Respectful	**0.951**	**9.42**	**0.963**	**0.974**
		16. Honest	**0.919**	**5.41**	**0.895**	**0.977**
		17. Aggressive*	**0.814**	**1.96**	**0.848**	**0.978**
		18. Pleasant	0.984	31.23	0.962	0.974
		19.Relaxed *(Interviewer)*	**0.838**	**2.36**	**0.866**	**0.977**
		20. Positive	0.953	9.98	0.907	0.976
Coordination	0.935					
		21. Peaceful	0.704	0.98	0.784	0.926
		22. Awkward*	0.478	0.30	0.500	0.94
		23. Productive	0.773	1.48	0.586	0.937
		24. Tense*	0.709	1.01	0.785	0.927
		25. Smooth	0.681	0.87	0.690	0.931
		26. Cooperative	**0.945**	**8.42**	**0.846**	**0.923**
		27. Relaxed *(Atmosphere)*	0.708	1.00	0.779	0.927
		28. Communicative	**0.953**	**9.98**	**0.859**	**0.922**
		29. Shared Expectations	**0.88**	**3.44**	**0.872**	**0.923**
		30. Coordinated	0.771	1.46	0.797	0.927

Note. Items included in final RS3i-O appear in bold. All scale and item statistics are based on single factor models including only items in each hypothesized factor. *Item is reverse-coded.

Appendix K.

Scree Plot Resulting from Exploratory Factor Analysis (EFA) of Pool of 30 Observer Rapport Items



Appendix L.

Rapport Scales for Investigative Interviews and Interrogations,

Observer version (RS3i-O)

Think about the interview you just watched. Read each characteristic below and rate how well it describes **INTERVIEWER** from Not at all (0) to Very well (6).

1. Invested – Did the interviewer seem to care about the result of the interview?

0 1 2 3 4 5 6

Not at all Very well

2. Focused on Source – Did the interviewer concentrate on the source during the whole interview?

0 1 2 3 4 5 6

Not at all Very well

3. Bored – Did the interviewer seem uninterested during the interview?

0 1 2 3 4 5 6

Not at all Very well

4. Frustrated – Did the interviewer seem frustrated with the source or what the source said?

0 1 2 3 4 5 6

Not at all Very well

5. Aggressive – Did the interviewer seem aggressive, hostile, or intimidating?

0 1 2 3 4 5 6

Not at all Very well

6. Relaxed – Did the interviewer seem calm and laid-back?

0 1 2 3 4 5 6

Not at all Very well

Now, think about the interview you just watched. Read each characteristic below and rate how well it describes **INTERACTION AS A WHOLE** from Not at all (0) to Very well (6).

7. Cooperative – How much did the interviewer and source cooperate with each other? Did they work together toward the same goal?

0 1 2 3 4 5 6

Not at all Very well

8. Communicative – Did the interviewer and source communicate effectively? Were they open and candid during their conversation?

0 1 2 3 4 5 6

Not at all Very well

9. Shared Expectations – Were the interviewer and source "on the same page"? Did they know what to expect from each other and the interview?

0 1 2 3 4 5 6

Not at all Very well

NOTES

Use these pages to take notes while watching the interview. You may find taking notes helpful in making ratings. You may pause the video at any time. Notes should include related behaviors or observations to help you make ratings.

1. Invested

_____
_____
_____

2. Focused on Source

_____
_____
_____

3. Bored

_____
_____
_____

4. Frustrated

_____
_____
_____

5. Aggressive

_____
_____
_____

6. Relaxed

_____
_____
_____

7. Cooperative

_____
_____
_____

8. Communicative

_____
_____
_____

9. Shared Expectations

_____

_____

_____

Appendix M.

RS3i-O Comprehension Check

1. Ratings should be made based on whose behavior?

_____
_____
_____

2. Provide a definition of **Invested** in your own words.

_____
_____
_____

3. Rate how similar the following characteristics are to being **Invested** on a scale of 0 (Not similar) to 4 (Very similar).

3.1 Happy

1	2	3	4	5
Not similar				Very similar

3.2 Caring

1	2	3	4	5
Not similar				Very similar

4. Describe, in your own words, behaviors that indicate an individual is **Invested**.

_____
_____
_____

5. Provide a definition of **Focused on Source** in your own words.

_____
_____
_____

6. Rate how similar the following characteristics are to being **Focused on Source** on a scale of 0 (Not similar) to 4 (Very similar).

6.1 Attracted

1	2	3	4	5
Not similar				Very similar

6.2 Concentrating

1	2	3	4	5
Not similar				Very similar

7. Describe, in your own words, behaviors that indicate an individual is **Focused on Source**.

_____
_____
_____

8. Provide a definition of **Bored** in your own words.

_____
_____
_____

9. Rate how similar the following characteristics are to being **Bored** on a scale of 0 (Not similar) to 4 (Very similar).

9.1 Uninteresting

1	2	3	4	5
Not similar				Very similar

9.2 Tired

1	2	3	4	5
Not similar				Very similar

10. Describe, in your own words, behaviors that indicate an individual is **Bored**.

_____
_____
_____

11. Provide a definition of **Frustrated** in your own words.

_____
_____
_____

12. Rate how similar the following characteristics are to being **Frustrated** on a scale of 0 (Not similar) to 4 (Very similar).

12.1 Upset

1	2	3	4	5
Not similar				Very similar

12.2 Angry

1	2	3	4	5
Not similar				Very similar

13. Describe, in your own words, behaviors that indicate an individual is **Frustrated**.

_____
_____
_____

14. Provide a definition of **Respectful** in your own words.

_____
_____
_____

15. Provide a definition of **Aggressive** in your own words.

_____
_____
_____

16. Rate how similar the following characteristics are to being **Aggressive** on a scale of 0 (Not similar) to 4 (Very similar).

16.1 Intimidating

1	2	3	4	5
Not similar				Very similar

16.2 Cocky

1	2	3	4	5
Not similar				Very similar

17. Describe, in your own words, behaviors that indicate an individual is **Aggressive**.

_____
_____
_____

18. Provide a definition of **Relaxed** in your own words.

_____
_____
_____

19. Rate how similar the following characteristics are to being **Relaxed** on a scale of 0 (Not similar) to 4 (Very similar).

19.1 Casual

1	2	3	4	5
Not similar				Very similar

19.2 Calm

1	2	3	4	5
Not similar				Very similar

20. Describe, in your own words, behaviors that indicate an individual is **Relaxed**.

_____
_____
_____

21. Provide a definition of **Cooperative** in your own words.

_____
_____
_____

22. Rate how similar the following characteristics are to being **Cooperative** on a scale of 0 (Not similar) to 4 (Very similar).

22.1 Supportive

1	2	3	4	5
Not similar				Very similar

22.2 Harmonious

1	2	3	4	5
Not similar				Very similar

23. Describe, in your own words, behaviors that indicate an interview is **Cooperative**.

_____
_____
_____

23. Provide a definition of **Communicative** in your own words.

_____
_____
_____

24. Rate how similar the following characteristics are to being **Communicative** on a scale of 0 (Not similar) to 4 (Very similar).

24.1 Talkative

1	2	3	4	5
Not similar				Very similar

24.2 Informative

1	2	3	4	5
Not similar				Very similar

25. Describe, in your own words, behaviors that indicate an interview is **Communicative**.

_____
_____
_____

26. Provide a definition of **Shared Expectations** in your own words.

_____
_____
_____

27. Rate how similar the following characteristics are to being **Shared Expectations** on a scale of 0 (Not similar) to 4 (Very similar).

27.1 Like-minded

1	2	3	4	5
Not similar				Very similar

27.2 Agreeing

1	2	3	4	5
Not similar				Very similar

28. Describe, in your own words, behaviors that indicate an interaction shows **Shared Expectations**.

_____
_____
_____

Vita

Justin Loren Magee was born in Little Rock, Arkansas to Charlotte and Henry Magee. He graduated from Russellville High School in Russellville, Arkansas in 2006. He has an older brother, Jared Magee, whose brilliance has consistently inspired Justin to strive toward academic excellence. He graduated from Arkansas Tech University in Russellville in 2011, earning a Bachelor of Arts degree in Sociology with a minor in French. After completing his undergraduate degree, Justin earned an international Teaching English as a Foreign Language (TEFL) certification in Prague, Czech Republic in 2012. He then moved to El Paso, Texas in 2013, later earning a Master of Science degree in Intelligence and National Security Studies from the University of Texas at El Paso in 2015. Justin has continued his graduate studies at UTEP and earned his Master of Arts degree in Experimental Psychology in 2018. He is currently working toward earning his Ph.D. in the UTEP Legal Psychology program. Justin's research interests focus on improving the quality of interactions between criminal justice practitioners and civilians, evaluating criminal justice policy and procedure, and developing practical tools for law enforcement and national security practitioners.

Email: jlmagee@utep.edu