

CONVERGENT VALIDITY OF FIVE OBSERVATIONAL RATING SCALES  
AS MEASURES OF RAPPORT IN INVESTIGATIVE INTERVIEWS AND  
INTERROGATIONS

JUSTIN L. MAGEE

Master's Program in Psychology

APPROVED:

---

James M. Wood, Ph.D., Chair

---

Lawrence D. Cohn, Ph.D.

---

Michael A. Zárate, Ph.D.

---

Misty C. Duke, Ph.D.

---

Charles Ambler, Ph.D.  
Dean of the Graduate School

CONVERGENT VALIDITY OF FIVE OBSERVATIONAL RATING SCALES  
AS MEASURES OF RAPPORT IN INVESTIGATIVE INTERVIEWS AND  
INTERROGATIONS

by

JUSTIN L. MAGEE, BA, MS

THESIS

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

MASTER OF ARTS

Department of Psychology

THE UNIVERSITY OF TEXAS AT EL PASO

December 2018

## **Abstract**

Research suggests that rapport can play an integral role in the success of investigative interviews. Accurate measurement of rapport is therefore critical. However, published scales designed to measure rapport in investigative interviews require interviewee self-report or complex behavioral coding. Global observer ratings of rapport may be more practical for use in law enforcement and national security interview training. In the present study, novice observers rated 92 simulated investigative interviews on three dimensions of rapport using quick, global rating scales. All observer ratings scales were significantly related to interviewee self-report ratings of analogous dimensions of rapport. Further, scores on all observer scales were found to be higher when interviewers used rapport-based tactics. Two scales were significantly related to amount of information shared by sources during interviews. These results provide evidence for the validity of observational global ratings in investigative interviews. Global observer scales that are quick to complete and require little training can provide useful ratings of rapport in both research and practical applications.

## Table of Contents

Abstract .....	iii
Table of Contents .....	iv
List of Tables .....	ix
List of Figures .....	x
Background .....	1
The “Essential Components” of Rapport .....	2
Rapport in Investigative Interviews and Interrogations .....	3
Measures of Rapport .....	5
Rapport Scales for Investigative Interviews and Interrogations (RS3i). ....	6
Observing Rapport-Based Interpersonal Techniques (ORBIT).....	7
Global Motivational Interviewing Skills Code (G-MISC). ....	7
Bronstein Suite.....	8
Interaction Rapport Scale (IRS).....	9
Negotiators’ Rapport Scale (NRS).....	9
Present Study .....	10
Method .....	14
Overview .....	14
Participants in Duke et al. (2018) .....	15
Materials in Duke et al. (2018) .....	15

Measures in Duke et al. (2018) .....	16
Procedure in Duke et al. (2018) .....	17
Participants.....	20
Materials .....	20
Measures .....	21
Procedure .....	25
Analyses & Hypotheses .....	30
Convergent validity.....	31
Discriminant validity. ....	32
Concurrent validity. ....	33
Predictive validity. ....	34
Exploratory analyses.....	34
Construct validity.....	35
Incremental validity. ....	35
Factor structure. ....	35
Construct validity through SEM. ....	36
Results.....	37
Overview.....	37
Confirmatory analyses .....	37
Descriptive statistics, interrater reliability and internal consistency. ....	37

Convergent validity: Correlations between observational rating scales. ....	38
Convergent validity: Observational rating scale correlations with RS3i self-report ratings on Essential 3 Components of rapport. ....	39
Discriminant validity. Observational rating scale correlations with RS3i self-report ratings on non-corresponding “Essential 3” components of rapport. ....	41
Discriminant validity. Observational rating scale correlations with RS3i self-report ratings on non- “Essential 3” components. ....	41
Concurrent validity: Relationship between observational rating scales and interviewing styles. ....	42
Negotiators’ Rapport Scale (NRS) Attentiveness. ....	43
Negotiators’ Rapport Scale (NRS) Positivity. ....	43
Global Motivational Interviewing Skills Code (G-MISC) Acceptance/Empathy. ....	44
Interaction Rapport Scale (IRS) Coordination. ....	44
Predictive validity: Relationship between observational rating scales and amount of information shared. ....	44
Exploratory Analyses .....	45
Convergent validity: Relationship between observational rating scales and self-report cooperativeness. ....	45
Incremental validity: Semi-partial correlations between observational rating scales and amount of information shared. ....	47
Factor structure. ....	48

Predictive validity. ....	49
Construct validity through SEM. ....	51
Discussion .....	58
Convergent and Concurrent Validity .....	58
Discriminant Validity.....	61
Interaction Rapport Scale (IRS) Coordination Scale .....	62
Limitations, Practical Implications, and Future Directions .....	63
Conclusion .....	66
References:.....	68
Appendix A. Review of Measures of Rapport.....	74
Appendix B. Rapport Scales for Interrogations and Investigative Interviews (RS3i) .....	151
Appendix C. Negotiator Rapport Scale (NRS) Attentiveness .....	154
Appendix D. Negotiator Rapport Scale (NRS) Positivity .....	156
Appendix E. Modified Global Motivational Interviewing Skills Code – Investigative Interview Adaptation (G-MISC – IIA) Instructions.....	158
Appendix F. Modified G-MISC Rating Form .....	159
Appendix G. Interaction Rapport Scale (IRS) Coordination .....	161
Appendix H. Informed Consent Document (Duke et al., 2018) .....	163
Appendix I. Negotiator Rapport Scale (NRS) Attentiveness Comprehension Check .....	165
Appendix J. Negotiator Rapport Scale (NRS) Positivity Comprehension Check .....	168

Appendix K. Modified G-MISC Comprehension Check .....	172
Appendix L. Interaction Rapport Scales (IRS)Coordination Comprehension Check .....	175
Vita.....	179



## List of Tables

Table 1: Descriptions of rapport measures in present study .....	24
Table 2. Observational Rating Scale Means, Standard Deviations, Internal Consistency, and Interrater Reliability .....	38
Table 3. Observational Rating Scale Intercorrelations (n = 92) .....	39
Table 4. Correlations Between Observational Rating Scales and RS3i Self-Report Scales (n = 92) .....	40
Table 5. Means and Standard Deviations for Observational Rating Scales by Interview Condition (n=92).....	43
Table 6. Correlations Between Observational Rating Scales and SIRS (n = 92) .....	45
Table 7. Correlations Between Observational Rating Scales and Self-report Cooperativeness (n = 92) .....	46
Table 8. Semi-partial Correlations Between Observational Rating Scales and SIRS (n = 92) ....	47
Table 9. Observational Rating Scale Factor Loadings for Unifactorial Solutions .....	49
Table 10. Correlations Between Observational Rating Scale Items and SIRS (n = 92) .....	50

## List of Figures

Figure 1. Tripartite model SEM.....	52
Figure 2. Higher-order rapport model SEM .....	53
Figure 3. Single rapport factor model SEM.....	54
Figure 4. Scale type model SEM .....	55
Figure 5. Observer scale only model SEM .....	56
Figure 6. Cross-modal scale correlation model SEM .....	57

## Background

“Clicking”, “jiving”, and “getting along well” - these are all phrases commonly used to describe positive feelings one may have about interactions with another person. When two people meet for the first time and mutually enjoy the interaction, they are said to have “hit it off” or “established rapport.” Though concisely defining rapport can be difficult, the construct reflects an important part of many interpersonal relationships. For instance, Capella (1990) argues that “rapport is arguably one of the central, if not the central, construct necessary to understanding successful helping relationships” (p. 303).

Empirical research has supported Capella’s (1990) assertion, indicating that rapport contributes to the quality of interactions in a wide variety of contexts: in the medical field between doctors and patients, in education between students and teachers, in psychotherapeutic environments between therapists and clients, and in business between service providers and their customers (see Driskell & Driskell, 2011 for review). Rapport has also been shown to influence the outcomes of negotiations (Drolet & Morris, 2000; Morris, Nadler, Kurtzberg, & Thompson, 2002). Though the literature overwhelmingly indicates that rapport is important to relationships across a variety of contexts, some evidence suggests that rapport may be expressed differently depending on the context of an interaction (Bernieri, Gillis, Davis, & Grahe, 1996). In other words, “rapport must be considered a function of the interactants *and* the situation” (Bernieri, 2005, p. 352). There are, however, some aspects that have been recognized as foundational to the development of rapport over a variety of contexts.

## The “Essential Components” of Rapport

Tickle-Degnen and Rosenthal (1990) proposed what has become a highly influential psychological theory concerning rapport. They posited that rapport may be best operationally defined as a composite construct comprised of three lower-order constructs that are conceptually discrete but likely to correlate across varying interactive contexts. These three “essential components” of rapport are (1) mutual attentiveness, (2) positivity, and (3) coordination (Tickle-Degnen and Rosenthal, 1990, p. 286). Mutual attentiveness refers to sustained attention by both parties throughout a dyadic interaction, and can be an indicator of shared interest or engagement with the interaction. Positivity can be characterized by feelings of warmth or friendliness experienced by individuals taking part in an interaction (hereafter referred to as *interactants*). Coordination is perhaps the most difficult component to conceptualize, though also certainly the most thoroughly researched (Bernieri 1988a; Bernieri, 1988b; Bernieri, Davis, Rosenthal, & Knee, 1994; Bernieri et al., 1996; Bernieri & Gillis, 2001; Bernieri & Rosenthal, 1991; Dabbs, 1969; Grahe & Bernieri, 1999). It refers to the balance and flow of an interaction – the degree to which interactants are “in sync”.

Researchers after Tickle-Degnen and Rosenthal (1990) have more or less come to agree with them on the nature of these three components of rapport across varying situations (Bernieri, 2005). However, contextual characteristics of interactions may affect the expression of rapport or the behaviors that lead to its development. For example, it is possible that rapport may be expressed differently in adversarial interactions than in cooperative contexts. It is not hard to imagine that rapport is expressed differently between two fellow investigators than between an investigator and a detained suspect. Duke (2013, p. 7) notes that the model implied by Tickle-Degnen and Rosenthal’s theory “may be only partially applicable to rapport as it develops

within... interrogations and investigative interviews” because such interrogations and interviews differ from socially cooperative interactions in their goals, purpose, and overall emotional climate. Empirical research has also supported the idea that there are key differences in the way rapport is expressed in cooperative versus adversarial interactions (e.g., Bernieri et al., 1994; Bernieri et al., 1996).

### **Rapport in Investigative Interviews and Interrogations**

Most of the scholarly literature concerning rapport has focused on its role in cooperative or friendly relationships. However, researchers have begun to examine the role of rapport in some of the most potentially adversarial situations: investigative interviews and interrogations. While rapport has been frequently linked to favorable interview outcomes (Collins, Lincoln, & Frank, 2002; Holmberg & Christianson, 2002; Kelly, Miller, & Redlich, 2016; Walsh & Bull, 2012), relatively little research has examined the process of rapport-building or measurement of rapport in investigative interactions (Abbe & Brandon, 2013). This has led several researchers to develop instruments to measure rapport and rapport-building in investigative interviews and interrogations (Alison, Alison, Noone, Elntib, & Christiansen, 2013; Alison et al., 2014; Duke, Wood, Bollin, Scullin & LaBianca, 2018; Vallano & Schreiber-Compo, 2011; Vallano et al., 2015).

For example, Duke et al. (2018) published the Rapport Scales for Investigative Interviews and Interrogations (RS3i), a multidimensional self-report questionnaire designed to measure interviewees’ experience of rapport in investigative interactions. While the RS3i is a valuable tool for rapport research and interviewer training, its nature as a self-report questionnaire that must be completed after an interaction has concluded means that it offers limited practical utility to investigators who would like to assess the rapport between interviewer and interviewee during

an ongoing interaction. On the other hand, Alison and colleagues (Alison et al, 2013; Alison et al., 2014; Alison & Alison, 2017) developed the Observing Rapport-Based Interpersonal Techniques (ORBIT) instrument, which is comprised of a set of scales intended to allow individuals observing investigative interviews and interrogations to rate the interviewer's rapport-building skills. While ORBIT has demonstrated some validity in its relationship to favorable interview outcomes (Alison et al., 2013; Alison et al., 2014), there are potential issues impeding its practical utility among practitioners. For example, this instrument is somewhat cumbersome, taking up to 70 minutes to code an interview (Alison et al., 2013, p. 418). ORBIT's complexity may also hinder investigators with little psychological training from using it effectively. Further, it is important to note that ORBIT is intended to measure interviewer rapport-building skills rather than the rapport generated during an interaction and thus offers its users limited insight into the psychological experience of a source.

There currently exists no observational measure of rapport that is psychometrically sound, practical, and has demonstrated a relationship with the experience of rapport across the three essential components of rapport. The Rapport Scales for Investigative Interviews and Interrogations and Observing Rapport-Based Interpersonal Techniques instruments are currently the strongest tools available to assess rapport and rapport building in investigative interactions, but research into both rapport in general and in its role in investigative interviews and interrogations has in fact produced a variety of measures intended to assess aspects of rapport. However, the psychometric soundness, validity, and practical utility of these measures is often limited or unknown. The following section provides a brief description of several of these instruments.

## Measures of Rapport

In the past, researchers have used a variety of methods to measure rapport in investigative interviews. Some researchers have relied upon a single, self-report Likert-type item (e.g., Evans et al., 2014), while others have used multi-item self-report scales designed specifically to measure the construct of rapport (e.g., Bernieri & Gillis, 2001). In a literature review preparatory to the present thesis, six sets of rapport measures were identified that have either been used in prior studies of investigative interviews or have potential usefulness in such studies: (1) The 18-Item Rapport Questionnaire (RQ) developed by Bernieri, 2005), (2) The Interaction Questionnaire (IQ) developed by Vallano and Schreiber Compo (2015) based on Bernieri's RQ, (3) The Working Alliance Inventory (WAI) developed by Hovarth and Greenburg (1986) for use in psychotherapy research, (4) the Rapport Scales for Investigative Interviews and Interrogations (RS3i) developed by Duke et al. (2018), (5) the Observing Rapport-Based Interpersonal Techniques (ORBIT) system developed by Alison and Alison (2017), and (6) a suite of rapport measures developed by Bronstein, Nelson, Livnat, and Ben-Ari (2012) for use in negotiation research.

The following sections of this literature review provides a brief description of several of these measures. For the purpose of brevity, only the three measures most relevant to the present study are described in these sections – the Rapport Scales for Investigative Interviews and Interrogations, the Observing Rapport-Based Interpersonal Techniques instrument, and the Bronstein suite of measures. A more detailed description of these three measures and of the other rapport measures listed in the preceding paragraph is provided in Appendix A.

### **Rapport Scales for Investigative Interviews and Interrogations (RS3i).**

The self-report Rapport Scales for Investigative Interviews and Interrogations (RS3i; Duke, Wood, Bollin, Scullin, & LaBianca, 2018) were developed with the intent of measuring Tickle-Degnen and Rosenthal's (1990) three essential components of rapport (mutual attentiveness, positivity, and coordination) while also accounting for other aspects of a source's perception of rapport that are particularly relevant to an investigative interaction (e.g., the interviewer's expertise). The RS3i asks sources (i.e., interviewees) to rate their perceptions of an interviewer on 21 items using a 5-point Likert-type scale (Table 4). Eighteen of the 21 items inquire about rapport experienced as a result of interviewer behavior, and the remaining three items inquire about the source's engagement in the interaction, or Commitment to Communication (CtC).

Two confirmatory factor analyses (CFAs) of different samples indicate a five-factor solution for the 18 items in the RS3i rapport subscales (Attentiveness, Trust/Respect, Expertise, Cultural Similarity, and Connected Flow), as reported by Duke et al., (2018, p. 4). The internal consistency of the RS3i scales is generally good: Duke et al. (2018) reported Chronbach  $\alpha$ 's above .70 (.71-.88) for all scales except Connected Flow ( $\alpha = .69$ ).

Validity of the RS3i scales has been demonstrated in two ways. First, the construct validity of the RS3i scales was demonstrated by showing that they significantly correlated with conceptually related scales developed by prior researchers (see Duke et al., 2018, p. 6 for a detailed review of each scale used), with good convergent and discriminant validity for most of the scales. Second, concurrent validity of the RS3i scales was demonstrated by showing that in an experiment involving simulated investigative interviews, interviewee's ratings on the RS3i scales were significantly higher when interviewers used rapport-building techniques than when



they did not (Duke et al., 2018, p. 70). In addition, in the same study, it was shown that the amount of information shared by interviewees was significantly correlated with scores on the RS3i Attentiveness scale ( $r = .23$ ,  $p = .03$ , 95% CI: .03, .41) and the RS3i Trust/Respect scale ( $r = .21$ ,  $p = .04$ , 95% CI: .01, .40).

### **Observing Rapport-Based Interpersonal Techniques (ORBIT).**

The Observing Rapport-Based Interpersonal Techniques (ORBIT) system, is a multi-scale observational coding framework developed by Alison et al. (2013, Alison et al. 2014, Alison & Alison, 2017). Unlike self-report measures such as the RS3i (Duke et al., 2018) that are completed by interview participants at the conclusion of their interaction, ORBIT is a rating tool that can be completed by observers after they have watched the interaction in progress.

The ORBIT includes numerous rating scales designed to measure various aspects of investigative interviews. Only one set of these scales, the Global Motivational Interviewing Skills Code – Investigative Interview Adaptation (abbreviated here as G-MISC), was included in the present study. The psychometric properties of this set of scales are summarized in the section that follows. The remaining ORBIT scales will not be discussed in the following section, although they are discussed in detail in Appendix A

### ***Global Motivational Interviewing Skills Code (G-MISC).***

ORBIT includes a set of five observer rating scales that are intended to measure the degree to which an investigative interviewer's communication strategies are consistent with the counseling skills of Motivational Interviewing. These five rating scales, which are collectively called the Global Motivational Interviewing Skills Code: Investigative Interview Adaptation (G-MISC), are largely modeled on the Motivational Interviewing Skills Code 1.1 (Glynn & Moyers,

2009), a set of rating scales designed to evaluate counselors' Motivational Interviewing skills in therapy sessions. The five scales of the G-MISC are intended to measure the following interviewer skills: (1) Acceptance, (2) Empathy, (3) Adaptation, (4) Evocation, and (5) Autonomy (see Alison & Alison, 2017 for full description of each skill's operational definition).

Raters using the Global Motivational Interviewing Skills Code make global ratings of an interviewer on each of these five skills at the end of an interview or a lengthy interview sample. Each skill is rated on a 7-point Likert-type scale ranging from 1 (poor skill) to 7 (positive skill).

Each of the five G-MISC scales consists of a single item. These five scales behave similarly to a single scale composed of five closely related items. The scales are highly intercorrelated and load on a single factor, which the test developers have called the "MI latent variable" (Alison et al., 2013, p. 423; Alison et al., 2014). The unifactorial nature of these five scales indicates that there is substantial redundancy among them, and that most of the reliable variance of these scales reflects the same underlying construct (the MI latent variable). No interrater reliability figures have been reported for the five scales of the G-MISC. However, research supports their validity: The five G-MISC scales have been shown to significantly correlate with the amount of information shared by interviewees during real (not simulated) police interrogations ( $r = .32-.36$ ; Alison et al., 2013, p. 425). In addition, there is some evidence that higher ratings on G-MISC scales are related to decreased use of some types of counterinterrogation tactics by interviewees in real interrogations (Alison et al., 2014).

### **Bronstein Suite.**

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 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 IRS item "listening", the component of positivity is reflected in the item "pleasant atmosphere", and the component of coordination is reflected in the item "synchrony".

### ***Interaction Rapport Scale (IRS).***

The first scale, the Interaction Rapport Scale (IRS) is an 11-item observational tool intended to measure how rapport is expressed in the *interaction* between the interviewer and interviewee. Thus, each IRS item refers to the interaction of the two interview participants and is rated on a 7-point Likert-type scale ranging from 1 (not at all) to 7 (very much). Item ratings are averaged to produce a single IRS score with high scores indicating that the interaction of the interviewer and interviewee expresses a high level of rapport. Bronstein et al. (2012) reported that they factor analyzed the items of the IRS and that the resulting factor pattern was unifactorial. However, Bronstein et al did not actually publish this factor pattern. Bronstein et al. (2012) reported that the IRS has shown high internal consistency (Cronbach's  $\alpha = .96$ ). Relevant to construct validity, Bronstein et al. (2012) reported that when judges observed participants interacting, the judges' ratings of these interactions on the IRS were significantly correlated ( $r = .36$ ) with the participants' own self-reported experience of rapport.

### ***Negotiators' Rapport Scale (NRS).***

The second observer scale developed by Bronstein et al. (2012), the Negotiators' Rapport Scale (NRS) is nearly identical to the Interaction Rapport Scale and consists of the same eleven items. However, the NRS requires judges to rate the interviewer and interviewee separately, as individuals, on the eleven items, in contrast to the Interaction Rapport Scale, which asks judges

to rate *the interaction* between the interviewer and interviewee. Put another way, the NRS asks judges to rate each of the participants in the interview separately, whereas the Interaction Rapport Scale (IRS) asks judges to rate the interaction of these participants. Ratings on the eleven NRS items are averaged to produce a single NRS score for each interactant, with higher scores indicating higher levels of rapport-related behavior. Bronstein et al. (2012) reported that they factor analyzed the items of the NRS and that the resulting factor pattern was unifactorial. However, Bronstein et al. did not actually publish this factor pattern. Bronstein et al. reported that the NRS has high internal reliability (Cronbach's  $\alpha = .93$ ). Relevant to construct validity, Bronstein et al. (2012) reported that when judges observed participants interacting, the judges' ratings of each participant on the NRS were significantly correlated ( $r = .22$ ) with the self-reported experience of rapport by that participant's partner.

## **Present Study**

The current study seeks to identify reliable and valid scales for the observational measurement of rapport in investigative interactions. Studying rapport using self-report measures is useful in helping to better understand what behaviors actually lead to the development and maintenance of rapport across a variety of contexts. However, self-report measures are often impractical in real-world investigative settings for two reasons. First, self-report rapport scales such as the Rapport Scales for Investigative Interviews and Interrogations (RS3i) require that a source complete them only after the conclusion of an interaction. Thus, these tools cannot aid an investigator seeking insight while an interview is actually being conducted. Second, in many investigative interviews (i.e., criminal interrogations, terrorism interrogations), it is impractical to ask the interviewee to complete a self-report measure. For these two reasons, observational

measure of rapport for investigative interactions are probably more practical than self-report scales for many real-life investigations.

Simply put, the purpose of the present research is to identify a valid, reliable set of observational rating scales for assessing sources' experience of rapport in investigative interactions. Observational rating scales are sought that (1) correlate significantly with sources' self-reported perceptions of rapport as measured by the Rapport Scales for Investigative Interviews and Interrogations (RS3i), (2) do not require extensive training, (3) are as short as possible and relatively quick to employ, and (4) are useful to employ at any point in an investigative interaction. The first step in the present study is to have observers watch simulated investigative interviews and rate these interviews on dimensions of rapport, using the Global Motivational Interviewing Skills Code, Interaction Rapport Scale, and Negotiators' Rapport Scale. These observer ratings will then be compared with sources' self-reported experience of rapport during the interviews, as measured by the RS3i. If the observer ratings of rapport are found to correlate with sources' self-reported experience of rapport, these findings will support the validity of using observer measures of rapport to measure rapport in investigative interactions.

Observer scales for the present study were chosen based on 3 criteria: (1) practical utility, (2) theoretical relevance, and (3) validity. First, regarding practical utility, the observer scales in the present study were selected because they have the potential to be practically useful within the context of real-life investigative interviews and interrogations. That is, the present study only included observer scales that can be scored quickly and relatively easily by observers without disrupting an interview. Second, regarding theoretical relevance, the observer scales in the present study were selected because they are designed to measure the three essential components

of rapport identified by Tickle-Degnen and Rosenthal (1990) that are also measured by the self-report scales of the Rapport Scales for Investigative Interviews and Interrogations (Mutual Attentiveness, Positivity, and Coordination). Third, regarding validity, the observer scales in the present study were selected because they have been used in prior studies and are supported by at least some evidence of construct validity.

Based on these criteria, five observer scales were selected for inclusion in the present study: (1) the Global Motivational Interviewing Skills Code (G-MISC) Acceptance scale, (2) the G-MISC Empathy scale, (3) the Negotiators' Rapport Scale (NRS) Attentiveness scale, (4) the NRS Positivity scale, and (5) the Interaction Rapport Scale (IRS) Coordination scale. The G-MISC Acceptance and Empathy scales are both global observer scales included in the Observing Rapport-Based Interpersonal Techniques (ORBIT) system, as already described. Both of these scales are intended to measure constructs related to the same rapport construct, Positivity, that is measured by the Trust/Respect scale of the Rapport Scales for Investigative Interviews and Interrogations (RS3i). Research outlined above has also supported the construct validity of the G-MISC Acceptance and Empathy scales by demonstrating significant relationships with desirable interview outcomes (Alison et al., 2013; Alison et al., 2014).

Similarly, the Negotiators' Rapport Scale (NRS) Attentiveness and Positivity scales are global observer scales intended to measure rapport constructs (Attentiveness; Positivity) that are measured by the Attentiveness and Trust / Respect scales, respectively, of the RS3i. As already discussed, the validity of the NRS scales has been supported in a study that showed they were both correlated with a commonly used self-report measure of rapport (Bronstein et al., 2012).

Lastly, the Interaction Rapport Scale (IRS) Coordination scale is an observational measure of the rapport construct of Coordination, which is measured by the Connected Flow

scale of the RS3i. The validity of the IRS Coordination scale has been supported in a study which found this scale was significantly correlated with participants' experience of rapport during an interaction (Bronstein et al., 2012).

All the observer scales included in the present study can be rated by observers who watch an investigative interaction. All the scales can be completed relatively quickly. Each of the scales is intended to measure the three essential components of rapport proposed by Tickle-Degnen & Rosenthal (1990), and each scale corresponds to one of the self-report scales of the RS3i. Furthermore, all of these scales have shown some evidence of validity in previous research.

## **Method**

### **Overview**

The primary aim of the present study is to examine the validity of observational rating scales as measures of rapport in investigative interviews. To achieve this aim, the study used a pool of 92 videos of simulated investigative interviews that were originally collected as part of an earlier study by Duke et al. (2018). In that study, undergraduate participants viewed a fictional portrayal of a domestic terrorism incident and then were questioned by trained experimenters who conducted simulated-investigative interviews using a prepared script. These simulated investigative interviews were video-recorded. After questioning, participants completed the Rapport Scales for Investigative Interviews and Interrogations (RS3i; Appendix B), a questionnaire that measured how much participants experienced rapport with the interviewer during the simulated investigative interview.

In the present study, trained undergraduate research assistants observed the video interviews from the Duke et al. (2018) study and then, based on their observations, rated specific dimensions of rapport displayed between the interviewer and interviewee. Several dimensions of rapport were rated, using the following observational scales: (1) the Attentiveness and Positivity scales of the Negotiators' Rapport Scale (NRS; Bronstein et al., 2012), (2) the Acceptance and Empathy scales of the Global Motivational Interviewing Skills Code Investigative Interview Adaptation (G-MISC; Alison & Alison, 2017), (3) and the Coordination scale of the Interaction Rapport Scale (IRS; Bronstein et al., 2012). The rating sheets and instructions for these scales are provided in Appendices C – G).

The central hypothesis of the study is that the level of rapport in these interviews, as rated by the research assistants using the observer rating scales just listed, will correlate significantly



with the self-report rapport ratings for the same interviews that were made by the participants in the Duke et al. (2018) study on the RS3i immediately after they were questioned. If observer ratings of rapport correlate with interviewee's self-reports of rapport on the RS3i, this provides evidence for the construct validity of the observer ratings as measures of rapport.

The following section provides a detailed description of the procedures used by Duke et al. (2018) to develop the interview videos and gather participants' self-reported perceptions of rapport. The section after that provides a detailed description of the present study's procedures for training undergraduate raters and collecting their observer ratings of rapport.

*Creation of simulated interview videos: Procedures in the original study by Duke et al. (2018)*

### **Participants in Duke et al. (2018)**

In the original study by Duke et al. (2018), participants were 94 undergraduate psychology students recruited from the University of Texas at El Paso's online SONA undergraduate research pool to play the role of interviewee in a simulated-crime interview. Participants were awarded course credit for participation and provided a small monetary incentive to motivate secrecy during interviews. The members of the sample were predominantly young ( $M = 20.1$  years,  $SD = 4.3$ ), female (62%), and Hispanic (81%; Duke et al., 2018). A full description of the sample obtained has been previously reported by Duke et al. (2018).

### **Materials in Duke et al. (2018)**

***Evidence video.*** As already described, a 5-minute "evidence video" created by the researchers was watched by all participants in the Duke et al. (2018) study. This video depicted

recent activities of a fictional young man, Eric, who had been implicated in an ongoing domestic terrorism investigation (see Duke et al., 2018). For example, some scenes in the video indicated that Eric had recently displayed paranoid and threatening behavior online and at work. Other scenes showed Eric in a gun shop inquiring about concealable weapons and later unloading heavy bags on a university campus. Although the video did not explicitly depict Eric engaging in any criminal activity, it included many indications that he could be a culprit in the domestic terrorism incident in which he had been implicated.

### **Measures in Duke et al. (2018)**

*Rapport Scales for Investigative Interviews and Interrogations (RS3i).* As indicated in earlier, the Rapport Scale for Investigative Interviews and Interrogations (RS3i) is a 21-item self-report questionnaire on which interviewees (sources) can rate the level of rapport they have felt with an interviewer/interrogator during an investigative interview or interrogation. The full 21-item measure is presented in Appendix B, though only three scales were analyzed in the current study: Attentiveness, Trust/Respect, and Connected Flow.

An older, 33-item version of the RS3i was administered to the 94 participants in the original study conducted by Duke et al. (2018), from which the participants of the present study were taken. The 33 items of this older version included all 21 items of the current version of the RS3i, plus 12 additional items that are not included in the current version. Thus, accurate scores for the current RS3i scales can be calculated using the older version of the RS3i administered in the original study. For the present study, scores for the 21 items of the current RS3i were extracted and used to calculate participants' scores on the five RS3i rapport scales. Thus, all analyses reported here using the RS3i are based on the 21-item version of the RS3i and its scales as described below.

***Shared Information Rating Scale (SIRS).*** In the Duke et al. (2018) study, each interview was evaluated in terms of the amount of relevant information shared by the source about the evidence video when questioned by the interviewer. The Shared Information Rating Scale (SIRS) was developed by Duke and her colleagues, with each item on the SIRS representing a relevant fact from the evidence video that could have been mentioned by a source over the course of the interview. The SIRS included 39 details from across the 5 scenes of the evidence video (7 from the first scene, 10 from the second, 4 from the third, 10 from the fourth, and 8 from the fifth). For example, items included “Eric [the young man in the video] had brown hair” and “Eric was carrying a duffel bag” (Duke et al., 2018, p. 68). Each fact was scored as a 1 if it was mentioned by the source and as a 0 if it was not and scores were summed. SIRS scores were calculated for the first half (Phase 1) and second half (Phase 2) of each interview, as well as for the interview as a whole.

***Self-Report Cooperativeness.*** Participants were asked to rate the degree to which they had been cooperative during the interview using a single ad-hoc, self-report item. Participants answered the question “How cooperative were you?”, rating their cooperativeness on a 10-point Likert-type scale where 1 represented totally uncooperative and 10 represented totally cooperative.

### **Procedure in Duke et al. (2018)**

The procedure employed by Duke et al. (2018) was approved by the UTEP IRB. In this study, participants arrived at the laboratory, were greeted by an undergraduate laboratory manager, and were then asked to complete an informed consent document (Appendix H) describing the purpose of and risks involved with the study. Interviewee participants then watched a fictional 5-minute "evidence video" created by Duke et al. for the purposes of their

study. This evidence video portrayed a series of recent scenes in the life of a fictional university student named Eric. These scenes provided substantial evidence that Eric had recently become erratic and paranoid, had formed a strong antipathy toward one of his professors, and may have joined with another student in a domestic terrorism incident to harm the professor. Participants were told that they were about to participate in a role play. Their task was to play the role of Eric's close friend, who was about to be interviewed by a police investigator. They were told that they should feel free to share much of the information in the video with the interviewer. However, they were also told that a few details depicted in the video were secret and that they should be kept from the interviewer. Interviewee participants were given a brief knowledge test to ensure they had understood the video and instructions. They were then led to an interview room where they waited briefly for the interviewer to arrive.

Next, an interviewer (actually a trained undergraduate or graduate research assistant) entered the interview room and introduced himself or herself as the detective investigating a recent domestic terrorism incident that had occurred on campus. The interviewer described some details of the fictional incident and explained that the interviewee/participant's close friend Eric, was the primary suspect. The interviewer further explained that the interviewee was thought to have knowledge of Eric's recent behaviors and thus needed to share everything they knew to protect their friend and, potentially, themselves. The interviewer then conducted a simulated investigative interview using scripted questions to gain information about Eric's activities from the interviewee.

Each interview was conducted using one of 3 distinct styles that were randomly assigned across participants (Rapport, Neutral, Pressure). Each interview style followed essentially identical scripts with regard to the information being presented and requested; however, rapport-

building behaviors differed between conditions. In the Rapport condition, interviewers engaged in several behaviors previously identified as likely to build rapport (e.g. introducing themselves by first name to the source, sharing a few personal details with the source). Conversely, interviewers in the Pressure condition did not engage in these behaviors and instead employed behaviors thought to increase pressure on the source (e.g., introducing themselves as Detective [Last Name] to the source; emphasizing the consequences of lying during the interview). Interviewers employing the Neutral style simply moved through the scripted set of questions without engaging in rapport-building or pressuring behaviors.

All interviews in the Duke et al. (2018) study were recorded with a digital video camcorder. The interviewee was informed beforehand, by both the laboratory manager and the informed consent document (see Appendix H), 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 H).

The digital video camera used to record the interviews was visible to participants in the Duke et al. (2018) study while they were interviewed. The camera was placed on a nearby tabletop to capture the entire figures of both participants. After the interview was completed and the interviewer had left, the interviewee was taken by an experimenter to a different room and asked to rate the interviewer across several dimensions of rapport using the Rapport Scales for Investigative Interviews and Interrogations (RS3i), as described later in this Method section. Laboratory managers then debriefed and thanked participants, and then provided them with course credit.

#### *Rating of simulated interview videos: Procedures of the present study*

### **Participants**

**Sources.** The sample of participants from the original Duke et al. (2018) were used in the present study. Sample demographics are described above.

**Raters.** Eight raters were recruited as members of the research team of the present study. Their role was to rate the videos from the Duke et al. (2018) study using the observational scales described in the following sections. The raters were all University of Texas at El Paso undergraduate students with no prior experience in studying rapport, rating interactions, or using observational measures. The recruited raters were predominantly young ( $M = 21.63$  years,  $SD = 1.87$ ) and Hispanic (62.5%). Five of the eight raters were female (62.5%).

### **Materials**

**Interview Videos.** Digitally recorded videos of the interviews conducted by Duke et al. (2018) were rated. Each video shows one source and one interviewer seated across from each

other at a square table in a bare room intended to resemble the style of interrogation booths used in law enforcement and national security investigations. The digital 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. Each video begins with the interviewer introducing themselves to the source and ends with the interviewer concluding the interview and leaving the room. Interview videos had an average length of 15 to 20 minutes. Two of the original 94 interview video files were corrupted and were not rated. Thus, ratings made for a total of 92 interview videos were included in analyses.

## **Measures**

***Comprehension Checks.*** Before raters rated the videos, they were administered a comprehension check to assess their understanding of the target constructs they were rating in the study, and to assess their competency in recognizing these constructs in a dyadic interaction. 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 observer scale. The appendices present the comprehension checks for the Negotiators' Rapport Scale (NRS) Attentiveness scale (Appendix I), the NRS Positivity scale (Appendix J), the Global Motivational Interviewing Skills Code Acceptance/Empathy scale (Appendix K), and the Interaction Rapport Scale (IRS) Coordination scale (Appendix L).

***Modified Global Motivational Interview Skills Code Investigative Interview Adaptation (G-MISC).*** The G-MISC (Alison and Alison, 2017) is an observer rating instrument that allows raters to rate the degree to which an investigator in an interview uses the techniques of Motivational Interviewing. The G-MISC includes five single-item measures intended to measure

the degree to which an interviewer in an investigative interview or interrogation adheres to the principles of Motivational Interviewing. The present study used two G-MISC items, Acceptance and Empathy (see Appendix F). Though the G-MISC developers suggested these single-item measures should be treated as independent scales, analyses indicated that they were highly related (Alison et al., 2013; Alison et al., 2014). Further, the Acceptance and Empathy items both appear to reflect aspects of the same “essential three” component, Positivity. Thus, the present study combined these two G-MISC items into a single “G-MISC Acceptance/Empathy” scale, whose scores were calculated by averaging the two items. Analyses reported here were performed using this scale rather than the single Acceptance and Empathy items.

***Interaction Rapport Scales (IRS).*** The 11-item Interaction Rapport Scale (Bronstein et al., 2012) instructs raters to provide global ratings of rapport for an interaction across eleven items on a 7-point Likert-type scale ranging from 1 (not at all) to 7 (very much). The IRS includes three scales that are to be averaged to produce a single score, but only the 4-item IRS Coordination scale (Appendix G) was used in the present study.

***Negotiators’ Rapport Scales (NRS).*** The NRS (Bronstein et al., 2012) allows for separate rating of each dyadic interactant (interviewer and source, in this case); however, a version of the NRS focusing solely on the interviewer was used in the present study. The NRS is identical to the Interaction Rapport Scales except that judges are directed to make ratings based on the behavior of the interviewer 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 for each the three scales. In the present study, only the 3-item NRS Attentiveness (Appendix C) and 4-item Positivity (Appendix D) scales were used by raters. These scales were used to rate the interviewer only and not the source.



Table 1 provides a brief description of (a) the RS3i scales that measure the Tickle-Degnen and Rosenthal (1990) “essential three” rapport components and (b) the observer rating scales from the present study that are hypothesized to measure these same three components.

Table 1: Descriptions of rapport measures in present study

Tool	Scales	Rating Source	Description
RS3i	Attentiveness	Self-report by Source	Reflects the degree to which a source feels an interviewer has been attentive
	Trust / Respect	Self-report by Source	Reflects the degree to which a source feels an interviewer trusts and respects them
	Connected Flow	Self-report by Source	Reflects the degree to which a source feels their interaction with an interviewer was smooth and well-coordinated
G-MISC	Acceptance	Observer rating of Interviewer	Reflects the degree to which an observer feels the interviewer displayed skill in creating an open, accepting atmosphere
	Empathy	Observer rating of Interviewer	Reflects the degree to which an observer feels the interviewer displayed skill in understanding the source's point of view
IRS	Coordination	Observer rating of interaction of Interviewer and Source	Reflects the degree to which an observer feels an interview was interpersonally smooth and well-coordinated
NRS	Attentiveness	Observer rating of Interviewer	Reflects the degree to which an observer feels the interviewer was attentive to the source
	Positivity	Observer rating of Interviewer	Reflects the degree to which an observer feels the interviewer's behavior contributed to an open, accepting atmosphere

## Procedure

***Rater Training.*** In the present study, 8 undergraduate research assistants from the University of Texas at El Paso, acting as raters, rated characteristics related to rapport for each of the 92 videotaped interviews originally collected by Duke et al (2018). Specifically, two raters each rated the 92 interview videos using the NRS Attentiveness scale; another two raters used the NRS Positivity scale; another two raters rated the interviews using two of the G-MISC scales, Acceptance and Empathy; and the remaining two raters rated the interviews using the IRS Coordination scale. A brief description of each scale, including the constructs measured and sources of ratings, is provided above (Table 1).

Each rater received training in using the scale or scales they were assigned to score. The entire training procedure lasted 45-65 minutes. The three steps of the training procedure were as follows:

*Step 1:* Raters were presented with the scoring instructions for the scale or scales they had been assigned to use when rating interviews. These scoring instructions were delivered both verbally by the trainer and in written form. The scoring instructions presented (a) concise descriptions of the purpose and makeup of each scale, (b) concepts that raters should keep in mind while making ratings, as well as (c) the rating procedures and scoring rules for each scale. The scoring instructions for each rating scale are provided in the appendices: the instructions for the Global Motivational Interviewing Skills Code (G-MISC) Acceptance/Empathy scale are provided in Appendix E, the instructions for the Interaction Rapport Scale (IRS) Coordination scale are provided in Appendix G, the instructions for the Negotiators' Rapport Scale (NRS) Attentiveness scale are provided in Appendix C, and the instructions for the NRS Positivity

Scale are presented in Appendix D. The average amount of time spent conducting Step 1 was between 15 – 20 minutes.

*Step 2:* After reading the scoring instructions and hearing them explained by the trainer, raters completed comprehension check for their assigned scale or scales. The purpose of the comprehension check was to ensure that raters understand the theoretical meaning of the construct, how these constructs may be expressed in dyadic behavior, and how to correctly complete the rating procedure. The comprehension check for the Global Motivational Interviewing Skills Code (G-MISC) Acceptance/Empathy scale is presented in Appendix K, the comprehension check of the Interaction Rapport Scale (IRS) Coordination scale is presented in Appendix L, the comprehension check for the Negotiators' Rapport Scale (NRS) Attentiveness scale is presented Appendix I, and the comprehension check for the NRS Positivity scale is presented in Appendix J.

The comprehension check for each scale included three tasks. First, raters were asked to articulate a definition of the target construct in their own words. Second, raters completed a series of items requiring them to correctly classify characteristics in relation to the target construct. For example, raters rated characteristics such as "judgmental" and "respectful" in relationship to their similarity to the construct Acceptance as measured by the Global Motivational Interviewing Skills Code (G-MISC) Acceptance/Empathy scale. Third, raters were required to provide a brief description of an interaction that would reflect a strong presence of the target construct. For example, as part of the Negotiators' Rapport Scale (NRS) Positivity comprehension check (Appendix J), raters were required to respond to the following: "Describe, in your own words, behaviors that indicate to you that an individual is contributing to a Pleasant

Atmosphere.” The average amount of time spent conducting Step 2 for each rater was between 20 – 30 minutes.

*Step 3:* The trainer evaluated each rater’s performance on the comprehension check. First, the trainer ensured that the rater had answered at least 80% of recognition items in the comprehension check correctly. Second, the trainer examined raters' performance on the free recall items to ensure that raters had adequately demonstrated that they understood the meaning of the target construct and the procedures to rate each scale. If the rater demonstrated good performance on the recognition and recall items, the rater was allowed to begin rating interviews for the main part of the study. However, if the rater did not demonstrate acceptable performance on the recognition and recall items, the Trainer guided the rater through the training materials a second time, beginning again at Step 1. If the rater did not demonstrate adequate performance on the recognition and recall items after completing the training materials a second time, the rater would have been dismissed from this assignment and a new rater would have been trained to take their place; however, no raters were dismissed in this way during the present study. The average amount of time spent conducting Step 3 was between 10 – 15 minutes.

***Observation and Rating of Video-recorded Interviews.*** After the trainer determined that a rater understood the scoring rules, the rater began rating the 92 videotaped interviews from the Duke et al. (2018) study using their assigned rating scale or scales. Raters worked in teams of two, according to the following procedures:

(a) Each team was named after the rating scale the raters used. For instance, one team was called the "NRS Attentiveness team" whereas another was called the "G-MISC Acceptance/Empathy team". Each team consisted of two members, who were designated as Rater A and Rater B.

(b) The two raters on each team independently rated all 92 interviews using the rating scale or scales assigned to their team.

(c) Before rating began, each rater was assigned a unique order in which to view and rate the videos. The randomization procedure of the videos is described in the following section.

(d) Before rating each interview, the rater reviewed the scale's scoring sheet in order to remind them of the ratings they would be making and the rating procedure, as ratings were made over several weeks, and raters went days at a time in between viewing some videos. The raters then viewed the interview in full before rating the relevant characteristics on the assigned scale. The raters were encouraged to take notes while viewing the interview and had unlimited ability to pause or replay any section of the video as often as needed. After viewing the interview and reviewing his or her notes, the raters rated the interview on the assigned scale. Each videoed interview lasted between 15 – 20 minutes took approximately 20-25 minutes to view and rate. Because there were 92 interviews, the total amount of time necessary for a single rater to rate all interviews on a particular scale was approximately 31-39 hours. Raters viewed and rated the videos in 2 - 3 hours shifts. When possible, raters were limited to 2 hour shifts to minimize fatigue, though scheduling required some shifts to last 3 hours.

(f) After each of the 92 interviews was rated on a particular scale by both Rater A and Rater B, the two sets of ratings were averaged for each interview. This average produced the ratings used as that interview's final rating on each scale, and these final ratings were used in all subsequent analyses in this study.

***Randomization of Video-recorded Interviews.*** Video order was randomized using the sequence generator tool provided by random.org. Each scale was assigned to 2 raters. That is, 2 raters were assigned the Negotiators' Rapport Scale (NRS) Attentiveness scale, 2 were assigned

the NRS Positivity scale, 2 were assigned the Interaction Rapport Scale (IRS) Coordination scale, and 2 were assigned the G-MISC Acceptance/Empathy scale. The pair of raters assigned to each scale are referred to as a team; however, it is important to note that each rater was blind to the ratings of their counterpart. Further, raters completed ratings of interviews in a randomized order that was unique to each individual rater. The randomization procedure is described below.

The first 10 interview videos from the original data collected by Duke et al. (2018) were used to familiarize raters with rating procedures and calibrate ratings between raters on each scale. All raters across all scales rated the first 10 videos in the same order. This was done in order to allow the first author to check that there was reasonable agreement between raters while keeping their exposure to the data at a minimum.

At this point, it is important to note that randomization procedures employed to obtain pilot data differed from those used for the bulk of the data collection after the rating of the first 10 videos. As only the Interaction Rapport Scale (IRS) Coordination and Negotiators' Rapport Scale (NRS) Positivity scales were included in the pilot observations, the randomization procedure for these scales were different from those used for the other three scales.

After completing the first 10 videos in the same order, raters assigned to the IRS Coordination and NRS Positivity scales were presented the next 40 videos (about half of the total number of interviews) in two distinct randomized orders, Order A and Order B. Each of these presented the next 40 videos in a uniquely randomized order. Raters on each team were randomly assigned either Order A or Order B, such that within each team raters viewed the interview videos in completely different orders. For example, on the IRS Coordination Team, Rater A was assigned to randomization Order A while Rater B was assigned randomization Order B. On the other hand, Rater A on the NRS Positivity Team was assigned to randomization

Order B while Rater B was assigned randomization Order A. This was done in order to obtain pilot data in order to determine if the training and rating procedures had a high enough likelihood of success to warrant the investigation proposed by the current thesis. Further, it was important to obtain a sufficient number of comparisons in order to reach reliable estimates in a limited amount of time. The order of the remaining 42 videos was randomized uniquely for each rater from this point on.

The randomization procedure for the remaining scales (Negotiators' Rapport Scale Attentiveness, Global Motivational Interviewing Skills Code Acceptance/Empathy) differed from that described above. After completing ratings for the first 10 videos, each rater for the remaining scales completed the remaining 82 videos in a distinct randomized order. That is, each individual rater received a personalized random order for the remaining videos that was different from all other raters, including those rating other scales.

## **Analyses & Hypotheses**

First, the interrater reliability of each rating scale was examined. Specifically, the level of agreement between Rater A and Rater B for each of the observer rating scales in the study, as indexed by the intraclass correlation coefficient (ICC), was calculated based on a mean-rating, absolute-agreement, two-way mixed-effects model. 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).

All further analyses were conducted using the average rating between raters. A series of hypotheses related to these analyses is presented below. Experimental hypotheses regarding the convergent, discriminant, concurrent, and predictive validity of the rating scales will be



presented followed by a discussion of purely exploratory analyses testing incremental validity, factor structure, and construct validity through structural equation modelling (SEM).

### **Convergent validity.**

The primary analyses of the present study examined the zero-order correlations of the observer rating scales (Global Motivational Interviewing Skills Code [G-MISC] Acceptance/Empathy Scale, Interaction Rapport Scale [IRS] Coordination Scale, Negotiators' Rapport Scale [NRS] Attentiveness and Positivity Scales) with the self-report Rapport Scales for Investigative Interviews and Interrogations (RS3i) scales that measure corresponding constructs. These analyses were performed in order to assess whether the observer rating scales in the present study have convergent validity as measures of rapport. These primary analyses tested four hypotheses regarding convergent validity of the observer rating scales with the corresponding RS3i self-report scales:

H1: NRS Attentiveness will correlate positively and at least moderately with RS3i Attentiveness

H2: NRS Positivity will correlate positively and at least moderately with RS3i Trust/Respect

H3: G-MISC Acceptance/Empathy will correlate positively and at least moderately with RS3i Trust/Respect

H4: IRS Coordination will correlate positively and at least moderately with RS3i Connected Flow

Next, analyses were performed to examine the correlations of the observer rating scales with each other, to further assess these scales' convergent validity. These analyses tested the following hypothesis regarding convergent validity:

H5: G-MISC Acceptance/Empathy will correlate positively and at least moderately with NRS Positivity

### **Discriminant validity.**

Analyses examined the discriminant validity of the five observer rating rater scales for distinguishing among the "essential three" constructs measured by the Rapport Scales for Investigative Interviews and Interrogations (RS3i) Attentiveness, Trust/Respect, and Connected Flow scales. It was expected that each observer scale would correlate not only with its corresponding RS3i scale (see Hypotheses 1 to 5), but also to some degree with the two other "non-corresponding" RS3i scales which measure the "essential three." However, for purposes of the present study, an observer rating scale would be considered to have discriminant validity only if (a) its correlation with its corresponding RS3i scale is .30 or higher, and (b) that correlation is at least .10 larger than the correlations of the observer rating scales with the two other non-corresponding RS3i scales that measure the "essential three." For instance, if the correlation of the Global Motivational Interviewing Skills Code (G-MISC) Acceptance/Empathy scale with RS3i Trust/Respect scale is .35, the G-MISC Acceptance/Empathy scale will be considered to have discriminant validity only if its correlation with the RS3i Attentiveness and Connected Flow scales is .25 or lower. Thus, the following hypothesis was proposed in regards to discriminant validity:

H6: Each of the observer rating scales listed in H1-H4 (a) will correlate  $r = 0.30$  or higher with its corresponding RS3i scale, and (b) will correlate at least 0.10 less with the two other "non-corresponding" RS3i scales that measure the "essential three" components.

To further test discriminant validity, the following hypothesis was proposed:

H7: Each of the observer rating scales will correlate less than 0.20 with the RS3i Expertise and Cultural Similarity scales, which do not measure the "essential three" components of rapport.

#### **Concurrent validity.**

Concurrent validity of the observational scales was examined by exploring the relationship between the scores on each scale and the interviewing styles employed by Duke et al. (2018). A series of one-way analysis of variance (ANOVA) tests were 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). Each scale was expected to differ by condition and differences between planned contrasts were each expected to be statistically significant.

H8: Mean ratings for each of the observer rating scales listed in H1-H4 will differ significantly by experimental condition (Rapport, Neutral, Pressure).

H9: Mean ratings for each of the observer rating scales listed in H1-H4 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.

### **Predictive validity.**

The observer rating scales' validity was further tested 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). The zero-order correlations between each observer rating scale and SIRS scores was examined. Because Duke et al. (2018), found differences between self-report scales' scores relationship with SIRS scores for the first and second halves of the interviews, separate analyses will be conducted using SIRS scores for the first half of the interviews, the second half of the interviews, and the total information shared for the entire interviews, resulting in the following hypotheses:

H10: Each observer rating scale will correlate positively and at least moderately with the total amount of information shared (SIRS) during interview Phase 2.

H11: Each observer rating scale will correlate positively and at least moderately with the total amount of information shared over the entire interview.

### **Exploratory analyses.**

A series of exploratory analyses were conducted in order to evaluate the psychometric properties, concurrent, and predictive validity of the observer rating scales used in the present study. Analyses were conducted to further examine the relationship of the observer rating scales and self-report scales with the theoretical constructs they are intended to measure. Other analyses assessed the relationship between the observational scales and relevant criteria (e.g., information

yield). These analyses were all exploratory in nature, and no specific hypotheses were tested. These analyses are briefly described below.

### **Construct validity.**

To further assess the validity of the observational scales, their relationship with sources' self-report cooperativeness was examined. Sources answered the question "How cooperative were you?", rating their cooperativeness on a 10-point Likert-type scale where 1 represented totally uncooperative and 10 represented totally cooperative. All observational scales were expected to correlate significantly with scores on the self-report cooperativeness measure. A MANOVA was also performed to assess the concurrent validity of the observational scales, including the rating scales as dependent variables and experimental condition as an independent variable.

### **Incremental validity.**

The incremental validity of scales was evaluated by examining their semi-partial correlations with the amount of information shared during the interviews. This denotes the amount of variance in information shared is explained by each rating scale, above and beyond that explained by the other scales. Separate analyses were conducted using the information shared during the first half of the interview, the second half of the interview, and the total information shared for the entire interview.

### **Factor structure.**

Exploratory factor analyses (EFA) was conducted on each of the rating scales to assess the degree to which they conform to the factor structures theorized by their developers. Items on the Negotiators' Rapport Scale (NRS) Attentiveness, NRS Positivity, and Interaction Rapport

Scale Coordination scales were expected to load onto a single factor each based on the results of Bronstein et al. (2012). The Global Motivational Interviewing Skills Code Acceptance/Empathy scale items were expected to load onto a single factor based on the structural equation models (SEMs) presented by Alison et al. (2013) and Alison et al. (2014).

### **Construct validity through SEM.**

In order to further test the construct validity of both the rating scales and the RS3i self-report scales, structural equation models (SEMs) were tested to examine how each of the scales relates to the theoretical component of rapport they are intended to measure. The primary models tested reflect the tripartite model of rapport proposed by Tickle-Degnen & Rosenthal (1990). G-MISC Attentiveness, G-MISC Empathy, NRS Positivity, and RS3i Trust/Respect were expected to load onto a single Positivity factor, NRS Attentiveness and RS3i Attentiveness were expected to load onto a single Attentiveness factor, and IRS coordination and RS3i Connected Flow were expected to load onto a single Coordination factor.

In one of these models, each of three factors was expected to correlate with the other two (Figure 1), and a second model replaced the factor intercorrelation with a higher order factor representing rapport (Figure 2). A third model was tested that replaced the lower-order factors intended to represent the Essential 3 components a single factor intended to represent the construct of rapport (Figure 3). Models were assessed using Hu & Bentler 's (1999) recommendations for good fit ( $RMSEA \leq .06$ ,  $CFI \geq .95$ , and  $SRMR \leq .08$ ) and were expected to produce non-significant chi-square statistics. Different imputations of thses models were tested as the results of previous models were examined. That is, a series of ad hoc models were also tested based on the results of the previous models rather than on the tripartite theory of rapport.

## **Results**

### **Overview**

The present study employed a series of analyses to examine the validity of five observational rating scales as measures of the “essential three” components of rapport in investigative interviews and interrogations. First, a series of confirmatory analyses were performed that tested hypotheses related to their convergent, discriminant, concurrent, and predictive validity (see Hypotheses 1 – 11). Second, a series of exploratory analyses to evaluate the psychometric properties, convergent validity, concurrent validity, predictive validity, and incremental validity of the observer rating scale. Analyses were conducted to further examine the relationship of the observer rating scales and self-report scales with the theoretical constructs they are intended to measure. These analyses were all exploratory in nature, and no specific hypotheses were tested. The following sections report the results of the confirmatory and exploratory analyses.

### **Confirmatory analyses**

#### **Descriptive statistics, interrater reliability and internal consistency.**

Means, standard deviations, internal consistency, and interrater reliability of the five observer rating scales are displayed in below in Table 2. Intraclass correlation coefficients (ICC) were computed using ratings made for each interview by Rater A and Rater B on each scale. ICCs were .70 or higher for all observational rating scales, indicating adequate to excellent interrater reliability. Rater A and Rater B 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. Internal consistency indexed by coefficient alpha was .70 or higher for all the scales, indicating adequate to excellent internal reliability.

Table 2. Observational Rating Scale Means, Standard Deviations, Internal Consistency, and Interrater Reliability

Scale	M ( <i>SD</i> )	$\alpha$	ICC
NRS Attentiveness	5.29 (.988)	.721	.811
NRS Positivity	4.11 (1.18)	.977	.986
G-MISC Acceptance/Empathy	4.22 (2.44)	.995	.957
IRS Coordination	4.09 (.965)	.813	.742

**Convergent validity: Correlations between observational rating scales.**

Analyses were performed to examine the inter-correlations among the observer rating 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. Further, since the Negotiators' Rapport Scale (NRS) Positivity Scale and Global Motivational Interviewing Skills Code (G-MISC) Acceptance/Empathy Scale were intended to measure dimensions of the same "essential three" component (positivity), specific hypotheses predicted these scales would inter-correlate positively and at least moderately (see Hypothesis 5). The intercorrelations between the five observational rating scales are presented in Table 3 below.



Table 3. Observational Rating Scale Intercorrelations (n = 92)

	1	2	3	4
1. NRS Attentiveness	-	.535**	.553**	.513**
2. NRS Positivity	.535**	-	.980**	.710**
3. G-MISC Acceptance/Empathy	.553**	.980**	-	.715**
4. IRS Coordination	.513**	.710**	.715**	-

*Note.* \*\* Correlation is significant at the .01 level

As can be seen, all observational rating scales were strongly correlated with Pearson product-moment correlation coefficients of .51 or higher (average  $r = .723$ ). As predicted by Hypothesis 5, Negotiators' Rapport Scale Positivity was very strongly correlated with Global Motivational Interviewing Skills Code (G- MISC) Acceptance/Empathy. Thus, the convergent validity of these scales, both intended to measure the underlying construct of "Positivity", was strongly supported. Further, though not reported in Table 3, the two G-MISC ("Acceptance" and "Empathy") included in the G-MISC Acceptance/Empathy Scale were almost perfectly correlated with each other ( $r = .990, p < .001$ ), supporting the decision to include them in a single scale. Overall, the pattern of scale intercorrelations supports the validity of each of these scales as measures of rapport as conceptualized by the tripartite theory.

**Convergent validity: Observational rating scale correlations with RS3i self-report ratings on Essential 3 Components of rapport.**

The primary analyses of the present study examined the correlations of the observational rating scales with the self-report Rapport Scales for Investigative Interviews and Interrogations (RS3i) scales intended to measure "essential three" components of rapport (mutual attentiveness, positivity, and coordination). Each observational scale was expected to correlate moderately ( $r > .30$ ) and positively with the self-report scale intended to measure an analogous dimension of

rapport (see Hypotheses 1-4). Specifically, Negotiators' Rapport Scale (NRS) Attentiveness was expected to correlate with RS3i Attentiveness, NRS Positivity and Global Motivational Interviewing Skills Code Acceptance/Empathy were expected to correlate with RS3i Trust / Respect, and Interaction Rapport Scale Coordination was expected to correlate with RS3i Connected Flow. The results supported these hypotheses. The correlations between the observer and self-report scales are presented in Table 4 below.

Table 4. Correlations Between Observational Rating Scales and RS3i Self-Report Scales

(n = 92)

RS3i Scale	NRS Attentiveness Scale	NRS Positivity Scale	G-MISC Acceptance / Empathy Scale	IRS Coordination Scale
1. Attentiveness	<b>.396**</b>	.313**	.305**	.320**
2. Trust / Respect	.259*	<b>.314**</b>	<b>.346**</b>	.422**
3. Connected Flow	.369**	.409**	.437**	<b>.508**</b>
4. Expertise	.197	.055	.037	.238**
5. Cultural Similarity	.135	.070	.069	.089

*Note.* Hypothesized correlations appear in bold. \* Correlation is significant at the .05 level, \*\* Correlation is significant at the .01 level

Each observer scale was significantly correlated with the corresponding RS3i self-report scale intended to measure an analogous dimension of rapport. The correlations predicted in Hypotheses 1 – 4 are presented in bold in Table 4. These results strongly support the convergent validity of the observational rating scales.

**Discriminant validity. Observational rating scale correlations with RS3i self-report ratings on non-corresponding “Essential 3” components of rapport.**

Analyses examined the discriminant validity of the observer rating rater scales for distinguishing among the "essential three" constructs measured by the Rapport Scales for Investigative Interviews and Interrogations (RS3i) Attentiveness, Trust/Respect, and Connected Flow scales. It was expected that each observer scale would correlate not only with its corresponding RS3i scale but also to some degree with the two other "non-corresponding" RS3i scales which measure the “essential three.” An observer rating scale would be considered to have discriminant validity only if (a) its correlation with its corresponding RS3i scale is .30 or higher, and (b) that correlation is at least .10 larger than the correlations of the observer rating scales with the two other non-corresponding RS3i rapport scales (Hypothesis 6). Correlations between observer rating scales and non-corresponding RS3i rapport scales are located above in Table 4. As can clearly be seen in Table 4, none of the observational scales met the criteria for adequate discriminant validity.

**Discriminant validity. Observational rating scale correlations with RS3i self-report ratings on non- “Essential 3” components.**

To further test discriminant validity of the five observer scales, analyses were conducted to examine their correlations with Rapport Scales for Investigative Interviews and Interrogations (RS3i) self-report scales (Expertise and Cultural Similarity) that are not intended to measure the three dimensions of rapport in Tickle-Degnen and Rosenthal’s (1990) tripartite model. It was expected that all observer ratings scales would correlate weakly ( $r < .20$ ) if at all with the RS3i Expertise and Cultural Similarity scales (Hypothesis 7). The results supported this hypothesis.

Correlations between observer rating scales and non-corresponding RS3i rapport scales are presented in Table 4. Only one correlation fell above the hypothesized criterion: IRS Coordination was significantly correlated with the RS3i Expertise scale ( $r = .238, p = .022$ ). These findings add further support for the validity of the observational rating scales as measures of rapport as conceptualized by the tripartite model (Tickle-Degnen & Rosenthal, 1990).

**Concurrent validity: Relationship between observational rating scales and interviewing styles.**

Concurrent validity of the observational scales was examined by exploring the relationship between the scores on each scale and the interviewing styles employed by Duke et al. (2018). A series of one-way analysis of variance (ANOVA) tests were 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). Each scale was expected to differ by condition, and differences between planned contrasts were each expected to be statistically significant (Hypotheses 8 & 9). The results supported these hypotheses.

Observational rating scale means and standard deviations for each interview condition are displayed below in Table 5. Statistically significant differences in means between condition are denoted in Table 5 with subscript.

Table 5. Means and Standard Deviations for Observational Rating Scales by Interview Condition

(n=92)

Observational Rating Scale	Interview Condition		
	Rapport (n=30)	Neutral (n=29)	Pressure (n=33)
	M (SD)	M (SD)	M (SD)
1. NRS Attentiveness	6.25 (.49) <sub>a</sub>	4.83 (.89) <sub>b</sub>	4.81 (.75) <sub>b</sub>
2. NRS Positivity	5.45 (.09) <sub>a</sub>	4.36 (.22) <sub>b</sub>	2.67 (.24) <sub>c</sub>
3. GMISC Acceptance/Empathy	6.91 (.18) <sub>a</sub>	4.89 (.47) <sub>b</sub>	1.19 (.19) <sub>c</sub>
4. IRS Coordination	4.96 (.68) <sub>a</sub>	4.11 (.57) <sub>b</sub>	3.27 (.74) <sub>c</sub>

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

#### ***Negotiators' Rapport Scale (NRS) Attentiveness.***

A main effect of experimental condition (Rapport vs. Neutral vs. Pressure) was found for NRS Attentiveness, ( $F [2,89] = 38.84, p < .001$ , partial  $\eta^2 = .466$ ), Post-hoc Tukey HSD comparisons indicated that, as predicted by Hypothesis 9, Attentiveness was rated significantly higher for interviews in the Rapport condition than in the Neutral ( $d = .1.97$ ) and Pressure conditions ( $d = 2.27$ ). Contrary to prediction, there was no statistically significant difference in Attentiveness scores between the Neutral and Pressure conditions.

#### ***Negotiators' Rapport Scale (NRS) Positivity.***

A main effect of experimental condition (Rapport vs. Neutral vs. Pressure) was found for NRS Positivity, ( $F [2,89] = 255.84, p < .001$ , partial  $\eta^2 = .977$ ). Post-hoc comparisons indicated that, as predicted, Positivity was rated significantly higher for interviews in the Rapport condition than in the Neutral ( $d = 6.44$ ) and Pressure condition ( $d = 15.35$ ), and significantly higher in Neutral than in Pressure interviews ( $d = 7.26$ ).

### ***Global Motivational Interviewing Skills Code (G-MISC) Acceptance/Empathy.***

A main effect of experimental condition (Rapport vs. Neutral vs. Pressure) was found for G-MISC Acceptance/Empathy, ( $F [2,89] = 2932.91, p < .001$ , partial  $\eta^2 = .985$ ). Post-hoc comparisons indicated that, as predicted, Acceptance/Empathy was rated significantly higher for interviews in the Rapport condition than in the Neutral ( $d = 5.24$ ) and Pressure condition ( $d = 24.17$ ), and significantly higher in Neutral than in Pressure interviews ( $d = 9.22$ ).

### ***Interaction Rapport Scale (IRS) Coordination.***

A main effect of experimental condition (Rapport vs. Neutral vs. Pressure) was found for IRS Coordination, ( $F [2,89] = 22.34, p < .001$ , partial  $\eta^2 = .527$ ). Post-hoc comparisons indicated that, as predicted, Coordination was rated significantly higher for interviews in the Rapport condition than in the Neutral ( $d = 1.35$ ) or Pressure condition ( $d = 2.38$ ), and significantly higher in Neutral than in Pressure interviews ( $d = 1.27$ ).

### **Predictive validity: Relationship between observational rating scales and amount of information shared.**

The observer rating scales' validity was further tested 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. Each observer rating scale was expected to correlate positively and at least moderately with the total amount of information shared during interview Phase 2 and with the total amount of information shared over the entire interview (Hypotheses 10 & 11). The results only partially supported these hypotheses. The

correlations between each observer rating scale and information shared are presented in Table 6 below.

Table 6. Correlations Between Observational Rating Scales and SIRS (n = 92)

RS3i Scale	Information Shared in Phase 1	Information Shared in Phase 2	Total Information Shared
NRS Attentiveness	.107	.200	.216*
NRS Positivity	-.051	.085	.043
G-MISC Acceptance/Empathy	-.064	.102	.049
IRS Coordination	.111	.323**	.317**

*Note.* SIRS = Shared Information Rating Scale. \* Correlation is significant at the .05 level.

\*\* Correlation is significant at the .01 level.

As can be seen above, none of the observational rating scales intended to measure the dimension of positivity were significantly correlated with information shared at any point in the interview. The Interaction Rapport Scale (IRS) Coordination was most strongly correlated with the amount of information shared during Phase 2 ( $r = .323, p = .002$ ) and the total amount of information shared over the course of the interview ( $r = .317, p = .002$ ).

## Exploratory Analyses

### **Convergent validity: Relationship between observational rating scales and self-report cooperativeness.**

To further assess the construct validity of the observational scales, their relationship with sources' self-report cooperativeness was examined. The self-report cooperativeness measure had sources' rate their cooperativeness on a 10-point Likert-type scale where 1 represented totally uncooperative and 10 represented totally cooperative. All observational scales were expected to

correlate significantly with scores on the self-report cooperativeness measure. The zero-order correlations between observational scale ratings and self-report cooperativeness are presented in Table 7 below.

Table 7. Correlations Between Observational Rating Scales and Self-report Cooperativeness (n = 92)

RS3i Scale	Cooperativeness
NRS Attentiveness	.253*
NRS Positivity	.203
G-MISC Acceptance/Empathy	.252*
IRS Coordination	.407**

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

All observational scales, with the exception of the Negotiators' Rapport Scale (NRS) Positivity, were significantly correlated with sources' self-report cooperativeness. NRS Positivity was also weakly correlated with cooperativeness did not reach statistical significance ( $r = .20$ ,  $p = .059$ ). These results add further support for the construct validity of all the observational scales as measures of rapport.

A MANOVA was also performed to assess the concurrent validity of the observational scales. This MANOVA included all the observational scales as dependent variables and experimental condition (interview style: Rapport vs. Neutral vs. Pressure) as an independent variable. This test showed a significant multivariate effect of interview style on ratings of rapport made using the observational rating scales ( $F [8,174] = 41.30$ ,  $p < .001$ , partial  $\eta^2 = .655$ ). Univariate analyses for the effect of interview style were identical to those reported earlier in this document.



**Incremental validity: Semi-partial correlations between observational rating scales and amount of information shared.**

The incremental validity of each observer rating scale was assessed by examining the semi-partial correlation of each observational scale with the amount of information shared during the interview. The semi-partial correlation denotes the amount of variance in information shared that is explained by each rating scale, above and beyond that explained by the other scales. Since no scales were significantly correlated with information shared for the first half of the interview, no analyses were conducted using these scores. Separate analyses were conducted using information shared during the second half of the interview and the total information shared for the entire interview (including information from the first half). The semi-partial correlations between observational rating scales and information shared are presented below in Table 8.

Table 8. Semi-partial Correlations Between Observational Rating Scales and SIRS (n = 92)

RS3i Scale	Information Shared in Phase 2	Total Information Shared
1. NRS Attentiveness	.097	.145
2. NRS Positivity	-.098	-.052
3. GMISC Acceptance/Empathy	.032	-.035
4. IRS Coordination	.340***	.370***

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

When controlling for the effects of the other observational rating scales, only the Interaction Rapport Scale (IRS) Coordination Scale was significantly correlated with both information shared in Phase 2 and total information above and beyond the other scales.

### **Factor structure.**

Exploratory factor analyses (EFA) using unweighted least squares (UWLS) extraction were conducted on each of the rating scales to assess the degree to which they conform to the factor structures theorized by their developers. Items included in each of the scales were expected to load onto a single factor each.

A unifactorial solution for the Negotiators' Rapport Scale (NRS) Attentiveness scale explained 67.58% of the variance in scores. A unifactorial solution for the NRS Positivity scale explained 93.7% of the variance, and a single factor solution for the Global Motivational Interviewing Skills Code Acceptance/Empathy scale explained 95.91% of the variance in scores on these scales. Last. A unifactorial solution for the Interaction Rapport Scale Coordination scale explained 64.95% of the variance in scores. Item factor loadings for each scale are presented in Table 9 below.

Table 9. Observational Rating Scale Factor Loadings for Unifactorial Solutions

Scale	Item	Factor 1
1. NRS Attentiveness	Listening	.891
	Tolerance	.393
	Attentiveness	.887
2. NRS Positivity	Pleasant Atmosphere	.999
	Absence of Aggression	.950
	Positivity	.966
	Negativity ( <i>reversed</i> )	.896
3. G-MISC Acceptance/Empathy	G-MISC Acceptance	.979
	G-MISC Empathy	.979
4. IRS Coordination	Cooperation	.778
	Like-mindedness	.910
	Synchrony	.685
	Flexibility	.573

### **Predictive validity.**

In order to further explore the predictive validity of the observational rating scales, the correlations between the individual items for each scale and the amount of information shared during the interviews were examined. The zero-order correlations between each scale's items and the amount of information shared in Phase 1, Phase 2, and over the course of the whole interview are presented below in Table 10.

Table 10. Correlations Between Observational Rating Scale Items and SIRS (n = 92)

Scale	Item	Information Shared in Phase 1	Information Shared in Phase 2	Total Information Shared
NRS Attentiveness	Listening	.137	.195	.226**
	Tolerance	-.032	.189	.136
	Attentiveness	.190	.086	.165
NRS Positivity	Pleasant Atmosphere	-.047	.093	.052
	Absence of Aggression	-.052	.075	.034
	Positivity	-.047	.091	.050
	Negativity <sup>†</sup>	-.040	.067	.034
G-MISC Acceptance / Empathy	Acceptance	-.071	.105	.049
	Empathy	-.058	.098	.050
IRS Coordination	Cooperation	.061	.587***	.504***
	Like-mindedness	.172	.267**	.302***
	Synchrony	.092	.027	.068
	Flexibility	.040	.146	.139

*Note.* SIRS = Shared Information Rating Scale. <sup>†</sup>Item score reversed. \*\* Correlation is significant at the .01 level. \*\*\*Correlation is significant at the .001 level.

Few of the observational scale items were independently correlated with the amount of information shared during either interview phase or the interview as a whole. While the Negotiators' Rapport Scale Attentiveness scale was correlated significantly with the total amount of information shared, this was accounted for by a single item: "Listening". Similarly, though the Interaction Rapport Scale Coordination scale was significantly correlated with information shared during the interviews, this relationship seems to be mostly accounted for by 2 items: "Cooperation" and "Like-mindedness among participants".

### **Construct validity through SEM.**

In order to further test the construct validity of both the rating scales and the Rapport Scales for Investigative Interviews and Interrogations (RS3i) self-report scales, structural equation models (SEMs) were tested to examine how each of the scales relates to the theoretical component of rapport they are intended to measure. The primary models tested reflect the tripartite model of rapport proposed by Tickle-Degnen & Rosenthal (1990). Global Motivational Interviewing Skills Code (G-MISC) Attentiveness/Empathy, Negotiators' Rapport Scale (NRS) Positivity, and RS3i Trust/Respect were expected to load onto a single Positivity factor, NRS Attentiveness and RS3i Attentiveness were expected to load onto a single Attentiveness factor, and Interaction Rapport Scale (IRS) Coordination and RS3i Connected Flow were expected to load onto a single Coordination factor.

In one of these models, each of three factors was expected to correlate with the other two (Figure 1), and a second model replaced the factor intercorrelation with a higher order factor representing rapport (Figure 2). A third model was tested that removed the lower-order factors intended to represent the “essential three” components and assumed all observational and self-report items loaded onto a single rapport factor (Figure 3). Models were assessed using Hu & Bentler 's (1999) recommendations for good fit ( $RMSEA \leq .06$ ,  $CFI \geq .95$ , and  $SRMR \leq .08$ ) and were expected to produce non-significant chi-square statistics. A series of ad hoc models were also tested, based on the results of the previous models rather than on the tripartite theory of rapport.

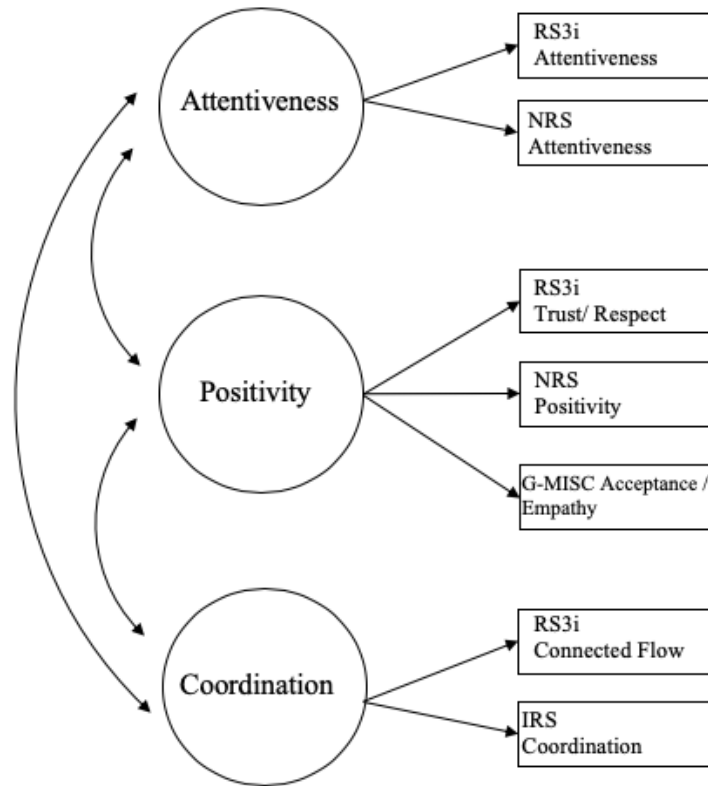


Figure 1. Tripartite model SEM

The SEM analysis of the tripartite model was performed with Mplus 7, using maximum likelihood (ML) estimation to fit the data to the hypothesized model. Model fit was assessed using the Hu & Bentler (1999) recommendations for good fit ( $RMSEA \leq .06$ ,  $CFI \geq .95$ , and  $SRMR \leq .08$ ). Results demonstrate indices reflecting poor model fit:  $RMSEA = 0.312$ ;  $CFI = 0.818$ ; and  $SRMR = 0.141$ . Furthermore, the chi square statistic rejected the null hypothesis which suggests that the model does a poor job at fitting the data,  $\chi^2 (11, N = 94) = 111.78, p < .001$ . Overall, the SEM indices indicate poor model fit.

A second model was tested that included a higher order rapport factor. This higher order rapport model was identical to the previously tested tripartite model, but included a single higher

order factor intended to represent the construct of rapport that replaced the intercorrelation between lower order factors (Figure 2).

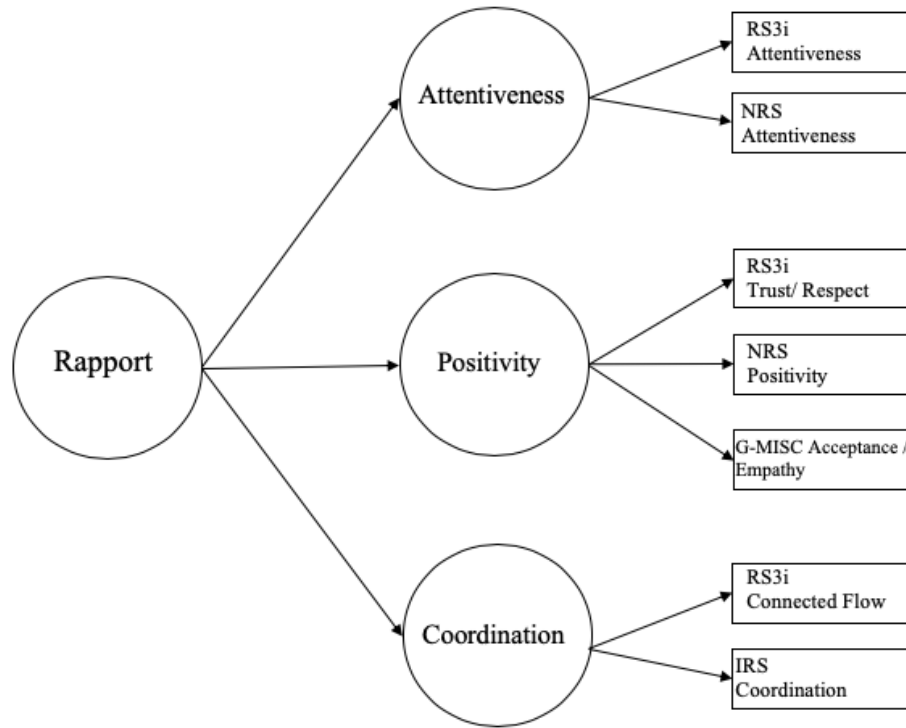


Figure 2. Higher-order rapport model SEM

Results for this model were identical to the previously tested tripartite model: ((RMSEA  $\leq .06$ , CFI  $\geq .95$ , and SRMR  $\leq .08$ ,  $\chi^2 [11] = 111.78$ ,  $p < .001$ ). Again, these indices generally indicate that the higher-order rapport model does not fit the data.

A third model was tested that to examine whether each of these scales all load onto a single rapport factor. This single rapport factor model included all three RS3i scales and all observational scales loading onto a single Rapport factor (Figure 3).

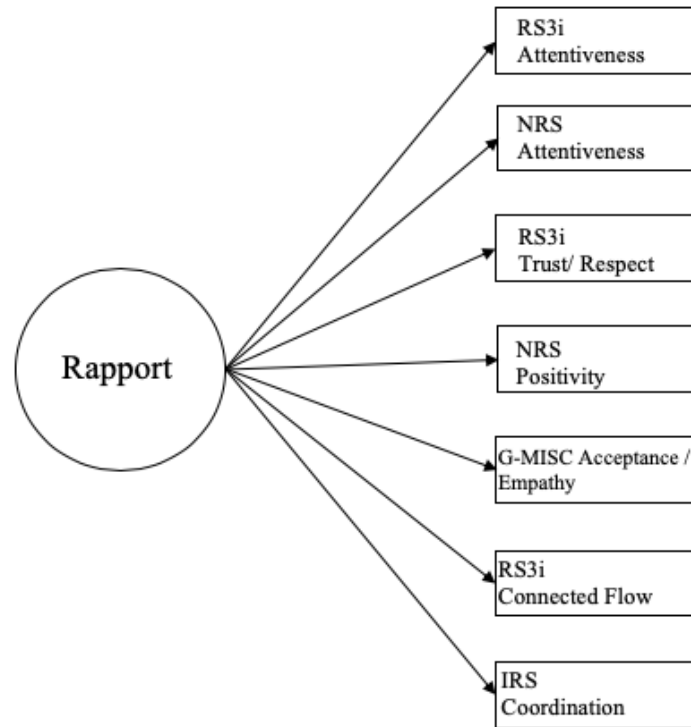


Figure 3. Single rapport factor model SEM

Results once more demonstrate reflected poor model fit: RMSEA = 0.304; CFI = 0.780; and SRMR = 0.163. Further, the chi square statistic once more rejected the null hypothesis suggesting that single rapport factor model does a poor job at fitting the data,  $\chi^2$  (14, N = 94) = 135.93,  $p < .001$ . These results indicate poor model fit for the single rapport factor model.

Given the failure of all three *a priori* models, three additional exploratory ad hoc SEMs were tested, adjusting paths to try and identify a better fitting model. Based on the results of an EFA including all rapport scales (both self-report & observational), it was thought that the self-report scales may load onto a single factor while the observational scales would load onto a second factor while the self-report factor and observational factor are correlated. This model will be referred to as the scale type model (Figure 4).



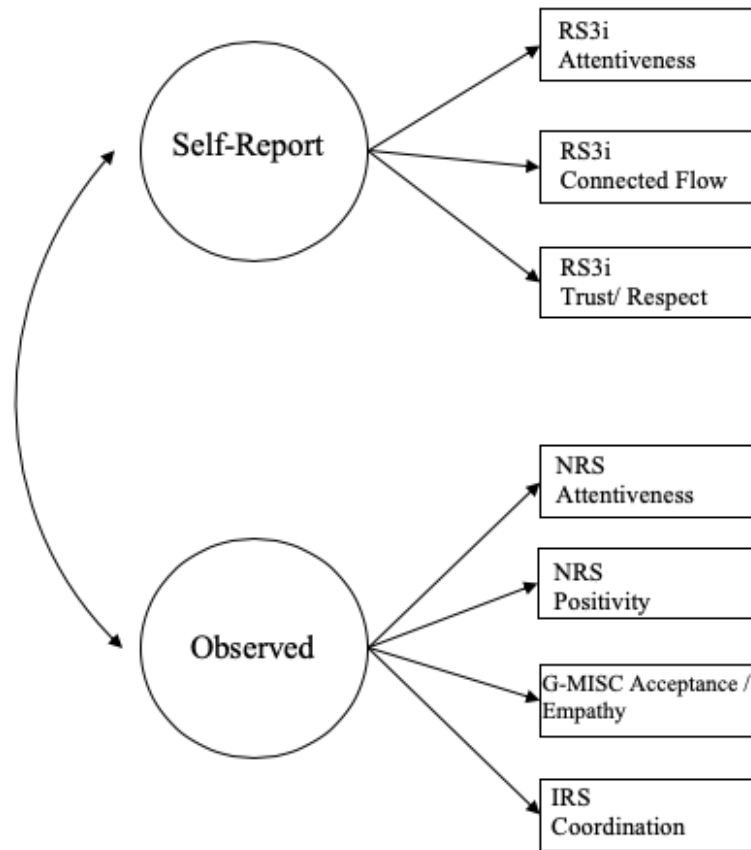


Figure 4. Scale type model SEM

The resulting indices were mixed. Some indices met the set criteria (CFI = 0.972 and SRMR = 0.068); however, some did not (RMSEA = 0.113). The chi square statistic rejected the null hypothesis which suggests that the model does a poor job at fitting the data,  $\chi^2 (19, N = 94) = 28.55, p = .008$ . Overall, the SEM indices indicate poor model fit; however, the model fit was improved over the previously tested models based on *a priori* theory.

A fifth model tested the theoretical relationship between only the observational scales proposed by the tripartite theory. Negotiators' Rapport Scale (NRS) Positivity and Global Motivational Interviewing Skills Code (G-MISC) Acceptance/Empathy were expected to load onto a single Positivity factor, while NRS Attentiveness and Interaction Rapport Scale (IRS)

Coordination were simply treated as observed scores in place of Attentiveness and Coordination factors, respectively. The Positivity factor and Attentiveness and Coordination scores were each expected to correlate with the other two (Figure 5).

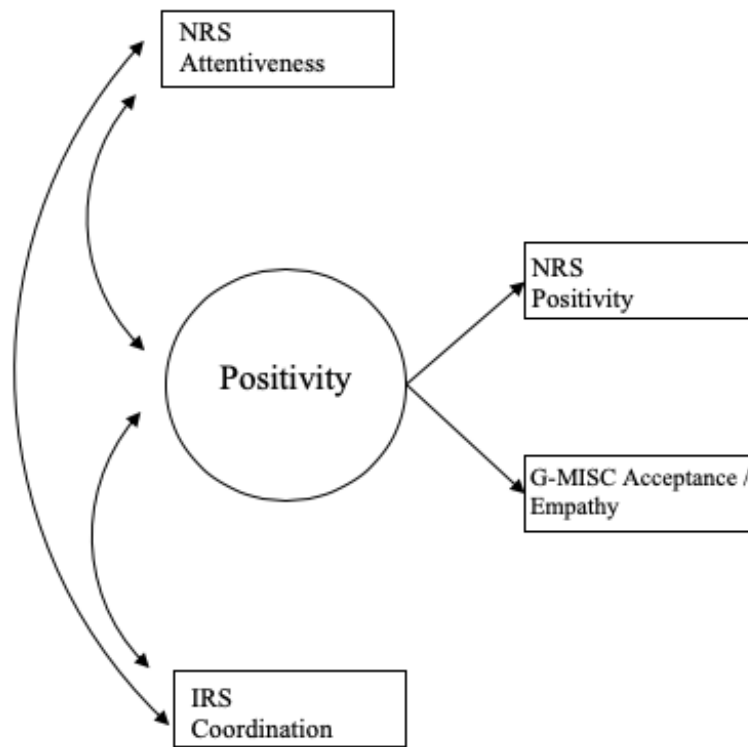


Figure 5. Observer scale only model SEM

The resulting SEM indices for the observer scales only model did not meet all of the set criteria ( $RMSEA < 0.001$ ;  $CFI = .98$ ; and  $SRMR = 0.109$ ). Further, the chi square statistic was statistically significant, which also suggests that the model does not fit the data well,  $\chi^2 (2, N = 92) = 6.64, p = .036$ . These results indicate poor model fit.

Last, a model was tested that combined the scale type model with the observer only model. This cross-modal scale correlation model assumed that NRS Positivity and G-MISC Acceptance/Empathy would load onto a single Positivity factor which, along with observed

Negotiators' Rapport Scale (NRS) Attentiveness and Interaction Rapport Scale (IRS)

Coordination scores, would load onto a single “observed rapport factor.” Additionally, the three Rapport Scales for Investigative Interviews and Interrogations (RS3i) scales were expected to load onto a single “self-report rapport” factor. Both of these factors were expected to load onto a higher order factor intended to represent the construct of rapport. Lastly, each observational scale was expected to correlate with its corresponding self-report scale. This model is presented below (Figure 6).

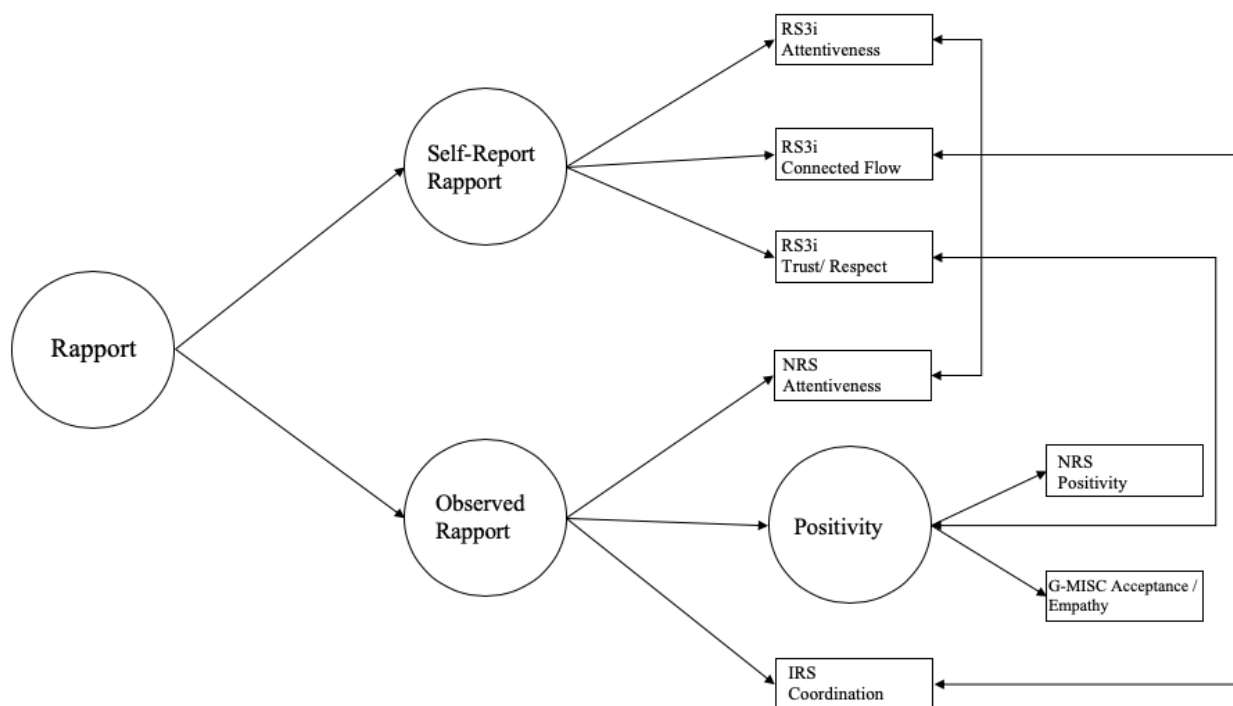


Figure 6. Cross-modal scale correlation model SEM

The resulting fit indices for this model met the set criteria: RMSEA = 0.011; CFI = 1.00; and SRMR = 0.017. Further, the chi square statistic suggested that this model fits the data well,  $\chi^2 (9, N = 94) = 9.09, p = .429$ . These results indicate excellent model fit, adding additional support for the construct validity of the observational rating scales and RS3i scales included.

## **Discussion**

Three findings from the present study are particularly notable. First, valid observer rating scales were identified for the Tickle-Degnen & Rosenthal (1990) “essential three” components of rapport. The Negotiators’ Rapport Scale (NRS) Attentiveness scale was found to be a valid measure of Attentiveness. NRS Positivity and Global Motivational Interviewing Skills Code (G-MISC) Acceptance/Empathy were found to be valid measures of Positivity. Interaction Rapport Scale (IRS) Coordination was found to be a valid measure of Coordination. Second, although the observer scales were found to have good convergent validity for identifying rapport within the context of investigative interviews and interrogations, they did not discriminate well among the “essential three” components. Third, the strongest evidence of validity was found for the IRS Coordination scale. IRS Coordination demonstrated stronger convergent, discriminant, and predictive validity. Each of these findings will be discussed in detail in the sections that follow.

### **Convergent and Concurrent Validity**

The results of the present study clearly demonstrated the convergent and concurrent validity of each of the observational ratings scales examined. Further, two of the scales demonstrated predictive validity.

The convergent validity of all the observational scales was demonstrated by their zero-order correlations with self-report scales intended to measure theoretically analogous constructs. That is, each of the scales was significantly correlated with the Rapport Scales for Investigative Interviews and Interrogations (RS3i) self-report scale intended to measure the same component of rapport. Specifically, (a) Negotiators’ Rapport Scale (NRS) Attentiveness was significantly correlated with RS3i Attentiveness, (b) NRS Positivity and Global Motivational Interviewing Skills Code (G-MISC) Acceptance/Empathy were significantly correlated with RS3i

Trust/Respect, and (c) IRS Coordination was significantly correlated with RS3i Connected Flow. Further, each of these relationships met the criterion set by the present study to demonstrate acceptable convergent validity ( $r > .30$ ; see Hypotheses 1 -5), indicating that these scales are meaningfully related to the constructs they are intended to measure. In addition to this, 3 scales intended to measure aspects of the Positivity component (NRS Positivity, G-MISC Acceptance/Empathy) were all highly intercorrelated ( $r$ 's  $> .975$ ) further supporting these scales' convergent validity. These correlations are especially impressive considering that the NRS Positivity ratings were made independently by a different team than the team that made the ratings for the two G-MISC scales.

The concurrent validity of the observational scales was demonstrated on the basis of hypothesized differences between the experimental conditions used in the simulated interviews employed by Duke et al. (2018). A series of one-way ANOVAs were performed comparing mean scores on each observer scale between the three experimental conditions (Rapport, Neutral, Pressure). All scales differed significantly between these interviewing styles. Planned contrasts between each of the conditions (Rapport vs. Pressure; Rapport vs. Neutral; Neutral vs. Pressure) indicated that each of the observational rating scales differed significantly between each of the interviewing styles. With the exception of one contrast (NRS Attentiveness, Neutral vs. Pressure), all of the planned contrasts showed significances between conditions such that ratings were highest in the Rapport condition and lowest in the Pressure condition, with Neutral condition ratings falling in between. When interviewers engaged in rapport-building behaviors, ratings on the observational rating scales were higher. Conversely, ratings using the observational scales were lower when interviewers did not engage in these behaviors and lowest when interviewers instead employed a more accusatorial interviewing style. These findings add

further support for the construct validity of all five observational rating scales as measures of rapport.

The observational rating scales' predictive validity was tested by examining the relationship of each scale with the amount of relevant information shared during the second half of the interviews and over the course of the entire interview. Each observer rating scale was expected to correlate with information shared during interview Phase 2 and with the total amount of information shared over the entire interview (Hypotheses 12 & 13). However, the results only partially supported these hypotheses. NRS Attentiveness was not significantly correlated with Phase 2 information but was significantly related to the total amount of information shared over the course of the interview. IRS Coordination was significantly related to both Phase 2 and total information, demonstrating the strongest predictive validity of the 5 scales examined. None of the observational rating scales intended to measure aspects of the Positivity component (NRS Positivity, G-MISC Acceptance/Empathy) were significantly correlated with information shared at any point in the interview. While this finding did not support the study's hypotheses, it is in line with the results of previous studies. Specifically, self-report ratings on the RS3i Trust/Respect scale were not significantly correlated with the total amount of information shared during these same interviews (Duke et al., 2018). These results support the predictive validity of the NRS Attentiveness and IRS Coordination scales and indicate that observational rating scales used to measure rapport during interviews may be useful in predicting favorable interview outcomes.

## **Discriminant Validity**

The results of the present study clearly support the convergent validity of the five observational scales examined; however, equally important to determining the scales' construct validity is their ability to discriminate between (a) non-“essential three” aspects and (b) non-corresponding “essential three” components. To address this issue, the correlations between each of the five observational rating scales and (a) RS3i Expertise and Cultural Similarity and (b) RS3i scales intended to measure non-corresponding components were examined (see Hypotheses 8 & 9).

First, the correlations between observational ratings scales and the RS3i scales intended to measure aspects of rapport distinct from the “essential three” were examined. None of the observational scales were significantly related to the RS3i Cultural Similarity scale. Further, only the IRS Coordination scale was significantly related to the RS3i Expertise scale. It is possible that sources' perceptions of the interviewer's expertise were based at least in part by the interviewer's ability to facilitate a well-coordinated interaction. This would explain why observers' ratings of the coordination displayed during interaction was related to sources' ratings of interviewer expertise. It is also possible that sources were generally more cooperative when they perceived the interviewer's expertise to be higher, resulting in higher ratings on the IRS Coordination scale, as one of its items reflects the degree of cooperation observed. Overall, these results demonstrate the ability of the observational scales to adequately discriminate between the “essential three” and RS3i self-report scales intended to measure other dimensions of rapport.

Second, the correlations between observational ratings scales and the RS3i scales intended to measure non-corresponding “essential three” scales were examined. Though the

results of the present study indicate that the observational scales are valid measures of rapport, they do not perform well in discriminating between the three dimensions of rapport proposed by Tickle-Degnen and Rosenthal's (1990) tripartite model. It was hypothesized that each observational rating scale would correlate at least .10 more with the RS3i scales intended to measure the same component of rapport than with the non-corresponding RS3i "essential three" scales. While the differences between a few correlations were at least .10, the results did not meet this criterion overall. The observational rating scales' correlations with non-corresponding RS3i scales were generally equivalent to their correlations with corresponding RS3i scales. In fact, the NRS Positivity and G-MISC Acceptance/Empathy scales were all most highly correlated with the RS3i Connected Flow scale, though they were expected to correlate strongest with the RS3i Trust / Respect scale. While the correlations between the observational rating scales and their corresponding RS3i scales demonstrates their validity as measures of rapport in investigative interviews and interrogations, the present findings suggest that these scales are primarily measuring a single underlying construct of "rapport," rather than the individual components they are intended to capture.

### **Interaction Rapport Scale (IRS) Coordination Scale**

The Interaction Rapport Scale (IRS) Coordination scale was the observational scale that most consistently supported *a priori* hypotheses and generally demonstrated the strongest evidence of validity. IRS Coordination had the largest correlations with all RS3i self-report scales, and was more strongly related to its corresponding RS3i scale than the other 4 observational scales were with their corresponding self-report scales. Furthermore, while the magnitude of its correlations did not meet the hypothesized criterion, this scale was the most effective at discriminating among the "essential three" components of rapport.



The IRS Coordination scale was also the strongest predictor of interview outcome (i.e., interviewee cooperativeness). That is, scores on the IRS Coordination scale were significantly correlated with the amount of information shared by sources during Phase 2 of the interview and also with the amount of details shared over the course of the entire interview. Exploratory examination of semi-partial correlations illustrate that, when included in a model with the other 4 observational scales, IRS Coordination was the only significant predictor of information shared during the interview.

The direction of the relationship between observed coordination and source cooperation is not fully explained by these analyses, however. That is, the degree to which the source was cooperating with the interviewer over the course of the interaction may have affected observer ratings of the coordination displayed. This potential is illustrated by the fact that exploratory analyses indicated the IRS Coordination item “Cooperation” was most strongly correlated with information shared (measured by SIRS scores). However, only certain pieces of relevant information shared by sources were included in the SIRS scores for each interview. Raters were not privy to which pieces of information shared by sources was included in the SIRS scores, so it is unlikely that source cooperativeness affected observer ratings in this way. That is, just because a source was providing a great deal of information (or simply speaking a great deal), this did not necessarily increase their SIRS scores. Thus, the IRS Coordination scale likely captured aspects of coordination other than simply source participation.

### **Limitations, Practical Implications, and Future Directions**

The present study presented a limitation related to the nature of using experimental data produced by Duke et al. (2018). The simulated investigative interviews rated in this study relied on semi-scripted interactions between interviewers and sources. Some sections of the interviews

within each condition (Rapport, Neutral, Pressure) featured similar or sometimes identical interviewer behavior. While these interactions were ecologically valid, this may have represented an attenuation of the range of behaviors an investigator may display during a single interaction. It is also possible that the experimental manipulation resulted in inflated reliability and validity coefficients. For example, raters clearly displayed excellent reliability at discriminating between experimental conditions using Negotiators' Rapport Scale (NRS) Positivity and Global Motivational Interviewing Skills Code (G-MISC) Acceptance/Empathy; however, there was little variation in ratings on these scales *within* each condition, indicating that raters may have been better at identifying the experimental condition than at accurately rating rapport using these scales. It remains unclear if the same level of interrater reliability would be found in a sample of "real world" investigative interactions that potentially display a more varied range of interviewer behavior and expression of rapport.

Future study should replicate the effects of the present study using videos of interactions from law enforcement or national security interview training, or true investigative interviews in which some measure of rapport was taken in order to address this issue. Unfortunately, such training materials are not commonly made publicly available and measuring sources' perceived rapport after investigative interviews is not common practice. This represents a challenge in studying rapport in this manner.

The present study identified several individual scales that can be employed as valid measures of rapport in investigative interviews. However, there still currently exists no cohesive observational measure of rapport that is psychometrically sound, practical, and has demonstrated a relationship with the experience of rapport across the three essential components of rapport. The content of the scales used in the present study, their psychometric descriptions, and

instructions are all spread across multiple publications. Further, their scoring and training materials are unclear or embedded with other complicated scales unrelated to the measurement of rapport. The use of these scales by practitioners (e.g., law enforcement and national security investigators) may represent a laborious process. Future work should be done to develop a single observational tool that is psychometrically sound, well-validated, practical, and is accompanied by clear and uncomplicated scoring and training materials. The development of a such a measure would likely increase practitioners' likelihood of use while potentially decreasing the likelihood of error. Given the importance of rapport across several types of interactions, a tool should be useful in a variety of investigative settings. Further, care should be taken to ensure that such a measure requires little training and is easily completed by novice raters with little-to-no psychological experience. In developing a single observational measure of rapport, results of the present study indicate that researchers may find focusing on the Coordination dimension particularly fruitful.

## **Conclusion**

The present study provides important evidence that observational scales can be reliable and valid tools for measuring rapport in investigative interviews and interrogations. The central hypothesis of the study was that the level of rapport in simulated interviews, as rated by the research assistants using the observational rating scales discussed herein, would correlate significantly with the self-report rapport ratings for the same interviews made by the participants in the Duke et al. (2018) study using the RS3i immediately after they were questioned. This hypothesis was supported by the results of the present study. Quick, global scales can be employed by novice raters with limited training to rate components of rapport rather than relying on cumbersome, multi-instrument measures that require lengthy training and intense behavioral coding.

These results are particularly meaningful for the measurement of rapport in real-world investigative settings like law enforcement and national security interview training as well as the rating of real police and intelligence agency interviews and interrogations. While studying rapport using self-report measures is useful in helping to better understand what behaviors actually lead to the development and maintenance of rapport, they are often impractical in these applied settings. Self-report measures require that a source complete them only after the conclusion of an interaction so they cannot aid an investigator seeking insight into the quality of the interaction while an interview is actually being conducted. It is also important to note that many investigative interviews are conducted with uncooperative sources, so it is unlikely that the source will reliably complete a self-report measure such as the RS3i. For these reasons, observational measures of rapport are probably more practical than self-report scales. Observational measures, such as those examined here, can provide investigators with valuable

information about the quality of in interview or interrogation while it is being conducted without relying on the source to provide direct feedback. Further, the five observational measures employed in the present study require an hour or less of training and only minutes to complete. Thus, they offer a significant improvement over existing observational tools that require lengthy and complex training and coding procedures.

Finally, while the scales identified by the present study may be useful in research and training, it is recommended that a single observational measure of rapport be developed in the future that maintains the validity these scales demonstrated while further increasing the ease of use for practitioners. This will further facilitate and simplify the accurate measurement of rapport in investigative interviews and interrogations.

## 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.
- 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.
- Alison, L., & Alison, E. (2017). ORBIT: *Observation of rapport based interview techniques: Coding manual*. Liverpool: University of Liverpool.
- Archer, D., & Akert, R. M. (1977). Words and everything else: Verbal and nonverbal cues in social interpretation. *Journal of Personality and Social Psychology*, 35, 443-449.
- Bernieri, F. (1988a). Coordinated movement and rapport in teacher-student interactions. *Journal of Nonverbal Behavior*, 12, 120-138.
- Bernieri, F. (1988b). Coordinated movement in human interaction: Synchrony, posture similarity, and rapport. *Journal of Nonverbal Behavior*, 12(2), 120.
- Bernieri, F. (2005). The expression of rapport. In V. Manusov (Ed.), *The Sourcebook of Nonverbal Measures: Going Beyond Words*. Lawrence Erlbaum Associates.
- 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.

- Bernieri, F. J., Gillis, J. S., Davis, J. M., & Grahe, J. E. (1996). Dyad rapport and the accuracy of its judgment across situations: A lens model analysis. *Journal of Personality and Social Psychology*, 71(1), 10-129.
- 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. & Rosenthal, R. (1991). Interpersonal coordination: Behavior matching and interactional synchrony. In R. S. Feldman & B. Rime (Eds.), *Fundamentals of Nonverbal Behavior*. (401-432.) Cambridge, MA: Cambridge University Press.
- Berry, D. S., Pennebaker, J. W., Mueller, J. S., & Hiller, W. S. (1997). Linguistic bases of social perception. *Personality and Social Psychology Bulletin*, 23, 226-528.
- Birtchnell, J. (2002). Psychotherapy and the interpersonal octagon. *Psychology and Psychotherapy: Theory, Research and Practice*, 75, 349– 363.
- 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.
- Brown, P., & Levinson, S. C. (1987). *Politeness: Some universals in language usage*. Cambridge: Cambridge University Press.
- Capella, J. N. (1990). On defining conversational coordination and rapport. *Psychological Inquiry*, 1, 303-305.
- Dabbs, J. M. (1969). Similarity of gestures and interpersonal influence. Proceedings, 77th Annual Convention, 4, 337-338. (summary)
- Driskell, J. E., & Driskell, T. (2011). Rapport and the investigative interview. *Report provided to Federal Bureau of Investigation*.

- Drolet, A. L., & Morris, M. W. (1995). Communication media and interpersonal trust in conflicts: The role of rapport and synchrony on nonverbal behavior. Paper presented at the Academy of Management meeting, Vancouver, Canada.
- Drolet, A. L., & Morris, M. W. (2000). Rapport in conflict resolution: accounting for how nonverbal exchange fosters cooperation on mutually beneficial settlements to mixed motive conflicts. *Journal of Experimental Social Psychology*, 36, 26-50.
- Duke, M. C. (2013) The development of the rapport scales for investigative interviews and interrogations. (Unpublished doctoral dissertation). University of Texas at El Paso, El Paso, TX.
- Duke, M. C., Wood, J. M., Bollin, B., Scullin, M., & LaBianca, J. (2018). Development of the rapport scales for investigative interviews and interrogations (RS3i), interviewee version. *Psychology, Public Policy, and Law*, 24(1), 64-79.
- Evans, J. R., Houston, K. A., Meissner, C. A., Ross, A. B., LaBianca, J. R., Woestehoff, S. A., & Kleinman, S. M. (2014). An empirical evaluation of intelligence-gathering interrogation techniques from the united states army field manual. *Applied Cognitive Psychology*, 28, 867-875. doi: 10.1002/acp.3065
- Gfeller, J. D., Lynn, S. J., & Pribble, W. E. (1987). Enhancing hypnotic susceptibility: Interpersonal and rapport factors. *Journal of Personality and Social Psychology*, 52, 586-595.
- Glynn, L. H., & Moyers, T. B. (2009). *Manual for the motivational interviewing skill code (MISC), Version 1.1: Addendum to MISC 1.0*. Retrieved from [https://casaa.unm.edu/download/misc\\_1.1\\_manual.pdf](https://casaa.unm.edu/download/misc_1.1_manual.pdf)



Grahe, J. E., & Bernieri, F. J. (1999). The importance of nonverbal cues in judging rapport.  
*Journal of Nonverbal Behavior*, 23(4), 253-269.

Grahe, J. E., & Bernieri, F. J. (2002). Self-awareness of judgement policies of rapport.  
*Personality and Social Psychology Bulletin*, 28(10), 1407-1418.

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.

Kieckhaefer, J. M., Vallano, J. P., Schreiber Compo, N. (2014). Examining the positive effects of  
rapport building: When and why does rapport-building benefit adult eyewitness memory?  
*Memory*, 22(8), 1010-1023.

Kleinman. (2006). KUBARK counterintelligence interrogation review: Observations of an  
interrogator – Lessons learned and avenues for further research. In Intelligence Science  
Board (Ed.) *Educing Information. Interrogation: Science and art. Foundations for the  
future*. Washington, D.C.: National Defense Intelligence College Press. ISBN: 1-932946-  
17-9

Leary, T. F., & Coffey, H. S. (1954). The prediction of interpersonal behavior in group  
psychotherapy. *Group Psychotherapy*, 7, 7-51.

Miller, W. R., & Rollnick, S. (1992). *Motivational interviewing: Preparing people for change*.  
New York, NY: Guilford Press.

- Miller, W. R., Moyers, T. B., Ernst, D., & Amrhein, P. (2008). *Manual for the motivational interviewing skill code (MISC): Version 2.1*. Retrieved from <https://casaa.unm.edu/download/misc.pdf>
- Morris, M. W., Nadler, J., Kurtzberg, T. R., & Thompson, L. (2002). Schmooze or lose: Social friction and lubrication in e-mail negotiations. *Group Dynamics: Theory, Research, and Practice*, 6, 89-100.
- Rapport. Entry 2. (2008). In Oxford English dictionary online (3rd ed.), Retrieved from: <http://www.oed.com/view/Entry/158209?rskey=PR6qBs&result=1&isAdvanced=false#eid>
- Rollnick, S., & Miller, W. R. (1995). What is motivational interviewing? *Behavioral and Cognitive Psychotherapy*, 23, 325-334
- Schumacker, R. E., & Lomax, R. G. (2004). *A beginner's guide to structural equation modeling* (2nd ed.). Mahwah, NJ: Erlbaum.
- Sharpley, C. F. (1997). The influence of silence upon client perceived rapport. *Counselling Psychology Quarterly*, 10, 237-246.
- Sharpley, C. F., Fairnie, E., Tabaray-Collins, E., Bates, R., & Lee, P. (2000). The use of counselor verbal response modes and client-perceived rapport. *Counselling Psychology Quarterly*, 13, 99-116.
- Tickle-Degnen, L., & Rosenthal, R. (1990). The nature of rapport and its nonverbal correlates. *Psychological Inquiry*, 1, 285-293.
- 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.

- Vallano, J. P., & Schreiber Compo, N. (2015). Rapport-building with cooperative witnesses and criminal suspects: A theoretical and empirical review. *Psychology, Public Policy, & Law*, 21, 85-99.
- Vallano, J. P., Evans, J. R., Schreiber Compo, N., & Kieckhafer, J. M. (2015). Rapport-building during witness and suspect interviews: A survey of law enforcement. *Applied Cognitive Psychology*, 29, 369-380.
- 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.
- 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.
- Wang, N., & Gratch, J. (2009). Can virtual human build rapport and promote learning? *International Conference on Artificial Intelligence in Education* Brighton, United Kingdom.
- Whichman, H. (1970). Effect of isolation and communication on cooperation in a two person game. *Journal of Personality and Social Psychology*, 16, 114-120.

## **Appendix A. Review of Measures of Rapport**

The following sections of this review of the extant literature provides a painstakingly detailed analysis of several measures that have been used in research relevant to the investigation of rapport in investigative interviews. Each instrument is described in terms of its development, composition, and psychometric properties. As several versions of each measure have been published, information relevant to the disambiguation of alternate versions of each tool is also presented. These sections may go into more detail than the average reader finds necessary to sufficiently understand the purpose of the present study, so they have been included here as an appendix.

### **Bernieri's "18-Item Rapport Questionnaire" (RQ-2001)**

The measure of dyadic rapport most widely used by researchers today was first presented in a book chapter by Bernieri and Gillis (2001). These authors referred to this measure as the "18-Item Rapport Questionnaire", but it will be referred to here as the RQ-2001.

The original version of the RQ, here called the RQ-1988, was reported by Bernieri (1988a; Bernieri, 1988b) thirteen years before publication of the RQ-2001. In the years between 1988 and 2001, the RA underwent extensive development and major changes. Three distinct versions of the questionnaire were described by Bernieri and his colleagues, which are here called the RQ-1988, the RQ-1994, and the RQ-2001. These three versions of the RQ differ from each other in several key ways, most notably their number of items and the nature of the scales. The following section will provide a brief review of the three RQ versions.

### **RQ-1988**

The first version of the RQ was developed in 1988 at Harvard University as part of a doctoral dissertation by Bernieri (1988a, p. 22) that was supervised by Robert Rosenthal. This

version of the RQ – here called the RQ-1988 -- included 27 items asking individuals to rate “various dimensions of emotional affect and rapport” regarding a recent interaction. A sample of 38 participants rated their feelings during the interaction on 23 bi-polar adjective pairs (e.g., "not nervous" to "nervous," "not bored" to "bored") using an 8-point Likert-type scale ranging from 1 (indicating the absence of the quality) to 8 (indicating a high degree of the quality). Participants also responded to 4 broader questions regarding their perceptions of the interaction (e.g., “How much did you enjoy the interaction?”) using an 8-point Likert-type scale ranging from 1 (“not at all”) to 8 (“very much”). In development of the RQ-1988, Bernieri randomly assigned participants to the role of either teacher or student and had both participants complete the instrument after completing a 15-minute teaching exercise. In this initial description of the questionnaire, no underlying factor structure was hypothesized.

However, in order to reduce the number of RQ items to a "smaller, more manageable set of variables", Bernieri, (1988a, p. 31) conducted a principal components analysis of the responses produced by the 38 participants. The resulting 7-factor solution included the following "composite variables": Positivity, Negativity, Anxiety, Control, Talkative, Tiredness, and Sexuality (see Table 11 for items corresponding to these factors). Items that loaded most heavily onto each factor were averaged to calculate the score for each composite variable. The mean of the two interactants' scores was then taken for each composite variable, resulting in a single “dyad value” used for all subsequent analyses. In other words, the "dyad values" represented the combined perceptions of both interactants.

Bernieri’s further discussion of the psychometric properties of the RQ-1988 was limited but did provide some evidence of construct validity. Specifically, dyad scores of the various composite variables showed some conceptually meaningful correlations with each other.

Although Positivity and Negativity were not significantly and negatively correlated as would be expected, Tiredness was found to correlate positively with Negativity ( $r = .61, p < .01$ ) and negatively with Positivity ( $r = -.59, p < .01$ ). The variable Talkative was significantly and positively correlated with both Positivity ( $r = .55, p < .05$ ) and Control ( $r = .49, p < .05$ ). Comparison to an observational measure of synchronization in interactants' movements also provided some evidence of validity. Specifically, movement synchrony was highly correlated with the Positivity factor scale ( $r = 0.74, p < 0.001$ ) and negatively correlated, though not significantly, with the Negativity factor scale ( $r = -0.20$ ).

A modified version of the RQ-1988 was later developed by Bernieri et al. (1994) and is described in the next section. Additional publications on the RQ between 1988 and 1994 could not be located by the author of the present thesis proposal. Below, Table 11 lists the 27 original RQ-1988 items alongside their associated factors and compares these items to those of later versions.

Table 11. RQ-1988 items by factor compared with RQ-1994 and RQ-2001

<u>RQ-1988 Factor</u>	<u>RQ-1988 Item</u>	<u>RQ-1994 PoSelf Item</u>	<u>RQ-1994 PoInteraction Item</u>	<u>RQ-2001 Item</u>
Positive	Enjoyment			
	Excited			
	Interest			
	Enjoyment of Role			
	Enthusiastic			
	Motivated			
	Liked Partner			
	Humorous			
	Satisfied	Satisfied	Satisfaction	Satisfying
	Happy			
	Friendly	Friendly	Friendly	Friendly
	Easygoing			
	Cooperative	Cooperative	Cooperative	Cooperative
	Attentive			
Negative	Angry			
	Disgusted			
	Frustrated			
	Bored	Bored	Boredom	Boring
Control	Controlling			
	Dominant			
	Forceful			
Anxiety	Tense			
	Nervous			
	Self-conscious			
Tired	Tired			
Sexy	Sexy			
Talkative	Talkative			

## **RQ-1994**

In 1994, Bernieri and his colleagues presented a new version of the RQ, here called the RQ-1994, that was intended to reflect constructs similar to those measured by the RQ-1988. The RQ-1994 was developed by having randomly assigned high school and undergraduate students in mixed-gender pairs complete two tasks: one cooperative task (planning an imaginary trip around the world together) and one adversarial task (a short debate). The participants completed a series of rapport-related items after each task, and the results were used to develop the RQ-1994.

As can be seen in Tables 11 and 12, the items of the RQ-1994 were substantially different from the items of the RQ-1988. Specifically, (a) only 4 of the 27 items of the RQ-1988 were also included in the RQ-1994 and (b) only 8 of the 29 items of the RQ-1994 were also included in the RQ-1988. Furthermore, the factor structure of the RQ-1994 was much different from that of the RQ-1988. Specifically, Bernieri et al. (1996, p. 114) reported that the factor structure of the RQ-1994 was unifactorial, whereas Bernieri (1988) had reported that the RQ-1988 had seven factors. The article by Bernieri et al. (1996) did not provide any additional detail on the factor structure of the RQ-1994, beyond the bare statement that it was unifactorial, nor did the article include a table showing the factor loadings of the RQ-1994 items. So far as the author of the present study can determine, the details regarding the factor structure of the RQ-1994 have never been published.



Table 12. Items of the RQ-1994, by factor, compared with items of the RQ-1988 and RQ-2001

<u>RQ-1994 Factor</u>	<u>RQ-1994 Item</u>	<u>RQ-1988 Item</u>	<u>RQ-2001 Item</u>
	<u>Perceptions of Self</u>		
Involvement	Involvement Engrossed Boredom		Engrossing*
Emotional Positivity	Positive Friendly Cooperative		
Comfort	Comfortable Satisfied Awkward		
Activity Level	Activity Level		
Smoothness	Smoothness		
<u>RQ-1994 Factor</u>	<u>RQ-1994 Item</u>	<u>RQ-1988 Item</u>	<u>RQ-2001 Item</u>
	<u>Perceptions of Interaction</u>		<u>Perceptions of Interaction</u>
Involvement	Involvement Intense Active Dull		Involvement Intense Active Dull
Emotional Positivity	Positive Warm Friendly Worthwhile	Friendly	Positive Cold Friendly Worthwhile
Comfort	Comfortably Paced Awkward Slow		Comfortably Paced Awkward Slow
Harmony	Harmonious Cooperative Coordinated	Cooperative	Harmonious Cooperative Well-Coordinated

Table 12. *continued*

<u>RQ-1994 Factor</u>	<u>RQ-1994 Item</u>	<u>RQ-1988 Item</u>	<u>RQ-2001 Item</u>
	<u><i>Perceptions of Interaction</i></u>		<u><i>Perceptions of Interaction</i></u>
Boredom	Boredom	Bored	Boring
Satisfaction	Satisfaction	Satisfied	Satisfying
Focus	Focus		Focused

*Note. The RQ-2001 includes a Perception of Interaction subscale, but not a Perceptions of Self scale. Thus, the item "Engrossing" on the RQ-2001 Perception of Interaction subscale corresponded to the item "Engrossed" on the RQ-1994 Perception of Self scale.*

Bernieri et al. (1994) divided the 29 items of the RQ-1994 into 12 "composite variables," with each of these variables including 1, 2, 3 or 4 of the RQ-1994 items. Table 12 lists these composite variables and their associated items. It is unclear how Bernieri and his colleagues arrived at the decision to subdivide the RQ-1994 items into 12 composite variables since, according Bernieri et al. (1996), the RQ-1994 was unifactorial.

Five of the composite variables of the RQ-1994 reflected what Bernieri et al. (1994) called Perceptions of Self, that is, each interactant's emotional experience during the interaction. These five composite variables were: 1) involvement (items: *involved*, *engrossed*, *boredom* reversed), 2) emotional positivity (items: *positive*, *friendly*, *cooperative*), 3) comfort (items: *comfortable*, *satisfied*, *awkward* reversed), 4) activity level (item: *activity level*) and 5) smoothness (item: *smoothness*).

The seven remaining composite variables of the RQ-1994 reflected what Bernier et al. (1994) called Perceptions of the Interaction, that is, each interactant's perceptions of the emotional tone of the interaction itself. These seven composite variables were: 1) involvement

(items: *involving, intense, active, dull* reversed), 2) emotional positivity (items: *positive, warm, friendly, worthwhile*), 3) comfort (items: *comfortably paced, awkward* reversed, *slow* reversed), 4) harmony (*harmonious, cooperative, coordinated*), 5) boredom (item: *boredom*), 6) satisfaction (item: *satisfaction*), and 7) focus (item: *focus*).

Three points are worth noting regarding the composite variables reported by Bernieri et al. (1994). First, three of the composite variables -- involvement, emotional positivity, and comfort -- appeared twice in the RQ-1994, first as composite variables related to Perceptions of Self and second as composite variables related to Perceptions of the Interaction. Second, two composite variables (activity level; smoothness) were related to Perceptions of Self only and did not have any corresponding relationship to Perceptions of the Interaction. Similarly, four composite variables (harmony; boredom; satisfaction; focus) were related to Perceptions of the Interaction only and did not have any corresponding relationship to Perceptions of Self.

Third, three of the composite variables of the RQ-1994 -- Focus, Emotional Positivity, and Harmony -- closely corresponded to the three “essential components” of rapport as identified in the theory of rapport proposed by Tickle-Degnen and Rosenthal (1990) -- mutual attentiveness, positivity, and coordination, respectively.

Unlike the RQ-1988, the RQ-1994 included 2 distinct subscales (Bernieri et al. 1994, p. 306). One of these, *Perceptions of Self* (PoSelf subscale, 11 items), asked each interactant to report on their perception of their own emotional state during the interaction. The second subscale, *Perceptions of the Interaction* (PoInteraction subscale, 17 items), asked each interactant to report on their general feelings regarding the interaction as a whole. In the RQ-1994, both of these

subscales were treated with equal importance, with each supposedly representing a different aspect of rapport.

Bernieri et al. (1994) did not explain in detail how they calculated scale scores for the RQ-1994. However, in a later study that discussed the RQ-1994, Bernieri et al. (1996, p. 110) noted that “interactants evaluated their own rapport in a unidimensional fashion” suggesting that scores were calculated by averaging across all 29 items, thus combining the items from the Perceptions of Self subscale and items from the Perceptions of the Interaction subscales to yield one global score. Subsequent studies discussing the RQ-1994 also indicate that it yields a single score. For example, articles by Bernieri et al. (1996, p. 115) and Grahe and Bernieri (1999, p. 258) reported internal reliability of the RQ-1994 as a Cronbach’s  $\alpha$  of .94. The reporting of this single number apparently reflects the reliability of all 29 items when they are combined into a single scale. This scale, which includes all 29 items of the RQ-1994, will be referred to in the present thesis as the Global RQ-1994.

Bernieri et al. (1994) did not provide reliability estimates for the two subscales of the RQ-1994 or for its composite variables. However, they did report some evidence relevant to the construct validity of the composite variables and items of the RQ-1994. Specifically, the authors reported that most composite variables of the RQ-1994 were significantly correlated with the movement synchrony and posture similarity of female partners in cooperative interactions ( $r = .26 - .50$ ; Bernieri et al., 1994, p. 309). However, these findings did not generalize well across genders: Only one item (*activity*) correlated with movement synchrony ( $r = .31$ ) and two items (*comfort*, *smoothness*) correlated with posture similarity in male partners in cooperative contexts, ( $r = .28$  and  $.29$ , respectively). Furthermore, the validity of the RQ-1994 items did not generalize well across different contexts: Specifically, relatively few RQ-1994 items were significantly

correlated with behaviors in adversarial interactions, leading the authors to conclude that “it may well be that the conceptual definition of rapport is not well suited for some types of interaction activities” (Bernieri et al., 1994, p. 309).

Bernieri et al. (1996; p. 116) also reported some findings relevant to the construct validity of the Global RQ-1994 (based on all 29 items). In this study, Bernieri and his colleagues included an ad-hoc two-item observer rating scale intended to create a global measure of dyadic rapport. The researchers did not give this measure a formal name, but it will be referred to here as the Observer Rapport Questionnaire, Two-Item (ORQ-2I). The two items of the ORQ-2I asked judges to rate the interactants’ fondness of each other (i.e., Do they like each other?) and enjoyment of the interaction (i.e., Are they enjoying what they are doing?) on an 8-point Likert-type scale. Both the *liking* and *enjoyment* observer rating items were found to be weakly correlated with scores on the Global RQ-1994 ( $r = .14$  and  $r = .17$ , respectively), though Bernieri et al. (1996) did not indicate whether these correlations were statistically significant. Because the two rating items were significantly correlated with each other ( $r = .68$ ,  $p < .0001$ ), Bernieri et al. (1996) decided to average them to yield a single global rapport rating for the ORQ-2I. Bernieri et al. (1996) reported small to medium correlations of the ORQ-2I with the Global RQ-1994 ( $r = .24$  in study 1;  $r = .35$  in study 2).

No further psychometric properties of the RQ-1994, its factors, or subscales are presented by Bernieri et al. (1994), nor could the author of this thesis locate such information in later publications. To sum up, the publications of Bernieri and his colleagues (a) report high (.94) internal reliability of the global RQ-1994 scale (based on all 29 items), and report that scores on this scale correlated modestly ( $r = .24-.35$ ) with observer ratings on the ORQ-2I, (b) report limited validity findings for the RQ-1994 composite variables and items, but no internal

reliability, and (c) do not report any reliability or validity findings regarding the two RQ-1994 subscales.

### **RQ-2001**

In 2001 Bernieri and Gillis (2001; see also Bernieri, 2005) introduced a new, shortened version of the RQ. Although they referred to it as simply the "18-Item Rapport Questionnaire," for the sake of consistency the present thesis will refer to it as the RQ-2001. The RQ-2001 eliminated the entire *Perceptions of the Self* subscale that had been part of the RQ-1994. Like the RQ-1994 *Perceptions of the Interaction* subscale, the RQ-2001 asked each interactant to rate their perceptions of the interaction they had just participated in. In fact, the RQ-2001 and the RQ-1994 *Perceptions of the Interaction* subscale shared most of the same items (see Table 12 for comparison) and were generally similar, with the following differences: (1) five RQ-2001 items - *cold, well-coordinated, boring, satisfying, and focused* - were slightly modified from corresponding items of the RQ-1994 *PoInteraction* subscale, *warm, coordinated, boredom, satisfaction, and focus*, respectively; (2) one RQ-2001 item, *engrossing*, was not included in the RQ-1994 *PoInteraction* scale, but instead was modified from a corresponding item, *engrossed*, of the RQ-1994 *PoSelf* subscale; and (3) the remaining twelve RQ-2001 items were taken directly from the RQ-1994 *PoInteraction* subscale.

The "composite variables" that had been part of the RQ-1994 were completely eliminated from the RQ-2001. The differences between the RQ-1994 and RQ-2001 could be summarized as follows: (a) the RQ-1994 consisted of two subscales (*Perceptions of Self* and *Perceptions of the Interaction*), with the subscales divided into "composite variables," which in turn included 1 to 4 items; (b) in contrast, the RQ-2001 consisted of a single scale (*Perceptions of the Interaction*) which included 18 items and no "composite variables."

Only limited psychometric information on the RQ-2001 is available. In fact, Bernieri & Gillis (2001) did not explicitly list the scale's items, instead noting that "the 18 items can be found in Bernieri et al. (1994)", although as already noted, the RQ-1994 reported in that article contained 29 items (p. 72) and not simply 18. However, a later article by Bernieri (2005) provided a definitive list of the 18 items.

The factor structure of the RQ-2001 has apparently never been reported in a full form that shows the loadings of items on factors. Instead, Bernieri and Gillis (2001, p. 72) reported simply that a factor analysis had been performed and yielded a single factor solution, rather than the three factors that the authors had expected (i.e., *positive affect*, *mutual focus*, and *coordination and harmony*) based on the theories of Tickle-Degnen and Rosenthal (1990).

In support of the construct validity of the RQ-2001, Bernieri (2005, p.351) published a table showing 34 correlations of RQ self-report scores with non-verbal rapport-related behaviors during cooperative and adversarial interactions. The absolute values of these correlations ranged from .00 to .44, with a median of .20. However, there are two reasons why these correlations are not actually relevant to the validity of the RQ-2001. First, these same correlations were previously reported in an article by Bernieri et al. (1996, p. 117). That article indicated that the version of the RQ used was the 29-item Global RQ-1994, and not the RQ-2001 (or the RQ-1994 Perceptions of the Interaction subscale). Second, the Global RQ-1994 scores used in the article by Bernieri et al. (1996) were what those authors called "dyad" scores. Those scores were calculated by averaging the Global RQ-1994 scores from the two members of dyads who had just engaged in an interaction. Thus, the construct validity figures reported in the table by Bernieri (2005, p. 351) (a) are not based on the RQ-2001 (or the RQ-1994 POI scale which is

highly similar to it), and (b) are not based on the scores of individuals. For these two reasons, it would be an error to cite these validity figures in support of the RQ-2001.

Internal reliability of the RQ-2001 has been reported as high (Cronbach's  $\alpha = .88$ ; Bronstein, Nelson, Livnat, & Ben-Ari, 2012). Bernieri & Gillis (2001) reported that intradyad agreement (the correlation between the rapport ratings of the two interactants in a dyad) using the RQ-2001 was significant but low ( $r = .20$  when planning a trip;  $r = .37$  when debating) in a study of 60 dyads, leading Bernieri (2005, p. 348) to caution that the measure's reliability might be poor.

As indicated, Bernieri and his colleagues have not presented relevant evidence of the RQ-2001's construct validity. However, a later study by another group of researchers, Bronstein et al. (2012), examined the correlation of the RQ-2001 with other self-report measures of rapport. Relevant to convergent validity, Bronstein et al. found that rapport as measured by the 18-item RQ-2001 was highly correlated ( $r = .78$ ,  $p < .001$ ) with the Negotiation Emotional Satisfaction Scale (NESS), a three-item self-report scale created by Bronstein et al. (2012) to measure aspects of rapport during negotiations.

It should be noted that validity findings regarding the RQ-1988 and RQ-1994 have sometimes been cited in support of the RQ-2001, but in fact are irrelevant to its validity because (a) the RQ-2001 is much different from these earlier scales and (b) although the RQ-2001 does closely resemble the Perceptions of the Interaction subscale of the RQ-1994, no validity findings for that RQ-1994 subscale have ever been published.



## **Nonverbal correlates of RQ rapport**

Bernieri and his colleagues were primarily interested in identifying nonverbal indicators of rapport that could be rated by observers from video recordings. Bernieri developed the RQ, not because he had much interest in measuring rapport by self-report, but because he wanted a criterion measure with which he and his colleagues could correlate their observational ratings of non-verbal rapport-related behaviors. Throughout Bernieri's program of research (Bernieri, 1988b; Bernieri et al., 1994; Bernieri et al., 1996; Bernieri & Gillis, 2001), observers watched pairs of participants in interactions and coded specific nonverbal cues thought to be indicative of coordination, an essential component of rapport. Two general approaches – either *global ratings* or *count data* -- were taken to measuring interactional synchrony through observation. The Coordinated Movement Scales (CMS) used *global ratings* to measure interactional symmetry, whereas the Synchrony Cue Scales (SCS) used *count data* (e.g., frequency, duration) of specific behaviors to measure interactional symmetry. Each approach is described in the following sections.

### **Coordinated Movement Scales (CMS)**

Bernieri (1988a) had observers code interactions specifically for four aspects of interpersonal behavior thought to be associated with coordinated movement. The observer rating scales created for this study were never given a formal name, but will here be referred to as the CMS. Two versions of the CMS were published (Bernieri, 1988b; Bernieri et al., 1994).

### **Coordinated Movement Scales 1988 (CMS-1988)**

The CMS-1988 comprised four scales: 1) simultaneous movement, 2) tempo similarity, 3) coordination and smoothness, and 4) posture similarity. Each rating was made after viewing a thin slice (50 seconds) of a videotaped interaction. Observers were briefly introduced to the

concept of coordinated movement before making ratings and rated one scale at a time after viewing a 25-second clip of an interaction. Bernieri (1988a, p. 25) explains that the CMS-1988 scales required global ratings “made on a 9-point Likert scale” (p. 25). However, the rating instructions provided later by Bernieri (1988a, p. 60) indicate that each scale should be rated from 1 (indicating “little to none” of the quality) to 8 (indicating “very much” of the quality). Thus, the statements by Bernieri (1988a) on pages 25 and page 60 are inconsistent in this respect. Bernieri (1988a, p. 27) reported that “the intraclass  $r$ ’s of the four CMS-1988 scales ranged from .24 to .30 and yielded Spearman-Brown effective reliability coefficients... ranging from .75 to .86” but did not report the precise reliability figures for each individual scale. Intercorrelations of the four scales indicated that simultaneous movement, tempo similarity, and coordination/smoothness were so highly related that they should be collapsed into a single variable. Accordingly, a movement synchrony composite variable was calculated by averaging ratings of simultaneous movement, tempo similarity, and coordination and smoothness. Reliability of the composite scale was not reported; however, a principle components analysis (PCA) confirmed the decision to create the movement synchrony composite and analyze it separately from posture similarity (Bernieri, 1988a, p. 62).

Bernieri (1988a; 1988b) did not discuss any further psychometric properties of the CMS-1988. The scale intercorrelations and PCA mentioned above provided some evidence of the scales’ validity. The scales of dynamic / movement coordination were highly intercorrelated ( $r$ s = .78 - .84) and loaded almost perfectly on a single common component (Bernieri, 1988a). The single scale of static / configural coordination (posture similarity) was significantly associated with the dynamic scales ( $r$ s = .43 - .63), and the PCA indicated this scale loaded saliently onto a second component.

Bernieri (1988a) also provided limited evidence of the construct validity of the CMS-1998. Movement synchrony was rated significantly higher in true interactions than fabricated interactions ( $F(1,18) = 8.43, r = .56, p < .01$ ); however, this effect was not found for posture similarity ( $F(1,18) = 0.60, r = .18, p > .10$ ). The CMS-1988 movement synchrony composite was significantly correlated with the *positivity* ( $r = .74, p < .001$ ) and *talkative* ( $r = .48, p < .05$ ) composite variable dyad values (average scores of both interactants) of the RQ-1988. Movement synchrony was negatively, though not significantly, associated with *negativity*. Posture similarity showed a significant, positive relationship with dyad values on the *anxiety* scale ( $r = .50, p < .05$ ).

### **Coordinated Movement Scales (CMS)**

The scales of the CMS-1988 were slightly altered by Bernieri et al. (1994) creating a version that will be referred to simply as the CMS. The CMS originally contained scales identical to those of the CMS-1988 (Bernieri, 1988a). However, when Bernieri and his colleagues (1994) found high intercorrelation between the four scales (*median.  $r = .70$* ), they dropped simultaneous movement and tempo similarity because they “covaried nearly perfectly with coordination / smoothness” (p. 307). Rather than average these three scales as Bernieri (1988a) had, Bernieri et al. (1994) simply considered coordination / smoothness a strong enough indicator of movement synchrony. Ultimately, the CMS consists of two global ratings of movement synchrony and posture similarity on 8-point Likert-type scales (anchors not provided). To examine the relationship between CMS interactional synchrony and RQ-1994 rapport, Bernieri et al. (1994) videotaped 60 unacquainted mixed-gender dyads while they interacted in two different contexts. First, participants used a world map and \$20,000 in play money to plan an imaginary vacation anywhere in the world together. Participants then individually rated their

emotional state and perceptions of the interaction using the RQ-1994. After completing the self-report, dyads engaged in a debate over a topic on which they disagreed. Participants then completed the RQ-1994 once more in reference to the second interaction. Bernieri et al. (1994) edited the videos down to only include one thin slice (50s) of vacation planning and one thin slice of debating, for a total of 120 video clips. The authors then made copies of the video clips that were identical to the original clips in every way, but that also included a digital mosaic effect overlay that converted the video into monochrome pixels. The mosaic effect removed “all fine detail, including lines, curved surfaces, textures, and small objects” in order to remove affective displays (e.g., facial expression) that were not being rated by observers (Bernieri et al., 1994, p. 306). Twenty-seven undergraduates were recruited to rate the interactions using the two CMS scales after viewing each cooperative (planning a vacation) or adversarial (debating) interaction. Ratings for movement synchrony and posture similarity were made on only the video portion of the videos (no audio).

Interrater reliability was poor for movement synchrony and fair for posture similarity (intraclass  $r$ s = .19 and .40, respectively). Internal reliability was demonstrated to be high, with Spearman-Brown effective reliability coefficients for .81 for movement synchrony and .92 for posture similarity. While no evidence of factor structure is provided, Bernieri et al. (1994) do provide some evidence of the scales’ validity. CMS ratings of movement synchrony were significantly higher when trip-planning than when debating, in both normal ( $F(1, 59) = 14.29, r = .44, p < .0004$ ) and mosaic ( $F(1, 59) = 37.66, r = .62, p < .0001$ ) video conditions. Posture similarity also significantly differed between cooperative and adversarial interactions with nearly identical effects under both video conditions,  $F(1, 59) = 8.04, r = .35, p < .007$ . Despite the evidence of validity provided by the difference in CMS ratings between contexts, observer

ratings of movement synchrony and posture similarity were weakly related, if at all, to the 12 RQ-1994 composite variables or to the Perception of Self (PoSelf) and Perception of Interaction (PoInteraction) subscales on average. Table 13 presents the correlations between CMS ratings and RQ-1994 subscales by the context of the interaction (i.e., activity) and the type of video judges viewed to make CMS ratings (normal vs. digital mosaic overlay; summarized from Bernieri et al., 1994, pp. 309-310). Averages were computed for each across both video conditions for RQ subscale (Table 14). Movement synchrony ratings were significantly correlated with self-reported rapport on the PoSelf (avg.  $r = .28$ ,  $p < .05$ ) and PoInteraction (avg.  $r = .39$ ,  $p < .01$ ) in female interactants during the trip-planning activity across video conditions. Posture similarity was also significantly correlated with PoInteraction (avg.  $r = .32$ ,  $p > .05$ ) in female interactants during cooperative interactions. However, neither movement synchrony nor posture similarity were significantly associated with either RQ-1994 subscale for females during the debate activity. Further, CMS ratings were not correlated with RQ-1994 subscale scores for male participants. Although the correlations outlined above may appear promising, the fact that they do not generalize across contexts or genders raises concerns about the replicability and generalizability of the CMS.

Table 13. Correlations between CMS and RQ-1994 rapport subscales by type of video and interactant gender, summarized from Bernieri et al. (1994, pp. 309-310)

(n=60) Activity	Interactant Gender	RQ-1994 Subscale	Video Type	Posture Similarity	Movement Synchrony
Planning a Trip	Male	PoSelf	Normal	0.16	0.21
			Mosaic	0.18	0.12
		PoInteraction	Normal	0.10	0.09
			Mosaic	0.15	0.06
	Female	PoSelf	Normal	0.18	0.29*
			Mosaic	0.22	0.26*
		PoInteraction	Normal	0.28*	0.42***
			Mosaic	0.35**	0.35**
Debate	Male	PoSelf	Normal	0.09	0.20
			Mosaic	0.08	0.13
		PoInteraction	Normal	0.13	0.13
			Mosaic	0.15	0.11
	Female	PoSelf	Normal	-0.02	-0.07
			Mosaic	0.09	0.03
		PoInteraction	Normal	0.19	0.15
			Mosaic	0.29*	0.17

Note. \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$

Table 14. Average CMS & RQ-1994 Coordinated Movement and Rapport subscale correlations

(n=60) Activity	RQ-1994 Subscale	Posture Similarity	Movement Synchrony
Planning a Trip	PoSelf	0.19	0.22
	PoInteraction	0.22	0.23
Debate	PoSelf	0.06	0.07
	PoInteraction	0.19	0.14
Average Planning a Trip	Average PoSelf	0.12	0.15
	Average PoInteraction	0.21	0.19
		0.20	0.23
Average Debate		0.13	0.11

### **Synchrony Cue Scales (SCS)**

Bernieri et al. (1996) continued to explore the relationship between synchrony and rapport using the videotaped interactions and RQ-1994 data gathered by Bernieri et al. (1994). However, rather than measure synchrony through global impressions, such as with the CMS, Bernieri et al. (1996) instead developed a new tool that relied on the presence or frequency of specific behavioral cues – the Synchrony Cue Scales (SCS). Bernieri et al. (1996) coded over 70 “nonverbal, verbal, and paralinguistic” cues thought to be related to the dyadic expression of rapport (Bernieri, 2005, p. 349). Bernieri (2005) reports that this original set of variables was reduced to a smaller set of 17 cues (Table 15) “on the basis of zero order correlations, principal components analysis, and content validity” (p. 349). These 17 cues comprise what will here be referred to as the Synchrony Cue Scales (SCS). Cues are rated in ways relevant to the construct being measured and are meant to be interpreted as indicators of their presence and relative magnitude. That is to say, cues are coded for “extremity (e.g., proximity), frequency (e.g., gestures), or duration (e.g., mutual silence)” depending on what is most psychologically meaningful (Bernieri et al., 1996, p. 115). Bernieri et al. (1996, p. 128) provided a more detailed description of each cue and how it should be coded, though scaling for some cues (e.g., proximity) is not actually presented.

Bernieri et al. (1996) did not present any hypothesized factor structure or correlational pattern among these cues. Effective inter-coder reliabilities (computed as  $nr/[1 + (n - 1)r]$ , where  $r$  is the average inter-coder reliability over multiple trials and  $n$  is the total number of coders) for both individual and dyad-based behavioral cues have been reported as high, with the majority of cues having a reliability over .80 (Bernieri et al., 1996, p.127).

Bernieri et al. (1996, pp. 117-123) reported the correlations of the 17 behavioral cues with (a) self-reported rapport as measured by the Global RQ-1994, and (b) observer-rated rapport

as measured by the ORQ-2I. These correlations are summarized below in Table 15. As can be seen, two behavioral cues -- physical proximity and synchrony -- were consistently correlated with both self-reports of rapport (Global RQ-1994) and observer ratings of rapport (ORQ-21) across interactive contexts. Otherwise, no behavioral cues listed in Table 15 showed consistent significant correlations with rapport as measured by either self-reports (Global RQ-1994) or observer ratings (ORQ-2I). However, some of the correlations in Table 15 suggest that a few behavioral cues (e.g., expressivity; gesturing by female interactants) may be related to self-reported and observer-rated rapport under some circumstances.



Table 15. Correlations of Behavioral Cues of Rapport as measured by the Synchrony Cue Scales (SCS) with (a) Self-Reports of Rapport as Measured by the Global RQ-1994 (b) Observer Ratings of Rapport as Measured by the ORQ-2I ratings

	Debate	Planning a trip	Debate	Planning a trip
	Global RQ-1994 (self-report)	Global RQ-1994 (self-report)	ORQ-2I (observer)	ORQ-2I (observer)
<i>(n=50)</i>				
<b>Synchrony Cue Scale</b>				
Mean attractiveness	0.1	0.10	0.23	0.10
Attractiveness discrepancy	-0.15	-0.21	-0.02	-0.15
Orientation	-0.09	-0.11	-0.04	-0.01
Racial Similarity	-0.2	-0.25	-0.28*	0.01
Smiling	-0.03	0.13	0.51***	0.31*
Mutual Silence	-0.36*	-0.02	-0.14	-0.21
Back-channel Responses-	0.42**	0.06	0.48	0.01
Eye Contact	0.33*	0.06	0.22	0.27
Gestures (female)	0.44**	-0.16	0.16	0.71***
Gestures (male)	0.17	0.15	0.16	0.14
Adaptors-	-0.08	-0.01	0.05	-0.02
Nervous Behavior	-0.26	-0.19	-0.31*	-0.39*
Expressivity	0.17	0.26	0.61***	0.67***
Posture Shifts	-0.38**	-0.23	0.20	0.44**
Forward Lean	-0.28*	-0.06	0.10	-0.21
<b>Proximity</b>	<b>0.27*</b>	<b>0.32*</b>	<b>0.48***</b>	<b>0.41**</b>
<b>Synchrony</b>	<b>0.31*</b>	<b>0.40**</b>	<b>0.35*</b>	<b>0.50***</b>

*Note.* \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$ ; *Adaptors and backchannel responses displayed by sex in trip planning, so average was computed*

### Working Alliance Inventory (WAI)

The Working Alliance Inventory (WAI) was developed by Hovarth & Greenburg (1986) based Bordin's (1980) "pantheoretical, tripartite (bonds, goals, and tasks) conceptualization of the alliance" (p. 223) as a measure the quality of alliance between clinician and client in therapeutic settings. The 36-item WAI includes three subscales designed to reflect these aspects of the

clinical interaction: (1) the degree to which clinician and client agree regarding the goals of the interaction (goals), (2) the degree to which clinician and client agree regarding the activities that should be undertaken to achieve the goals of the interaction (task), and the quality of the bond between clinician and client during the interaction (bond). Each subscale includes 12 items rated on 7-point Likert-type scales. The WAI has been adapted for use by clients (WAI-C), therapists (WAI-T), and observers (WAI-O; Tichenor & Hill, 1989).

The WAI was later reduced to a 12-item short form (WAI-S; Tracey & Kokotovic, 1989) which was later further revised (WAI-SR; Hatcher & Gillaspay, 2006). Direct comparison of the WAI and WAI-S (Busseri & Tyler, 2003) indicated that “were highly correlated and had comparable descriptive statistics, internal consistencies, and subscale intercorrelations” (p. 193) suggesting that scores on the two scales are interchangeable. None of these scales have been sufficiently validated in a law enforcement or national security context.

Analysis has indicated that internal reliability of each of the scales is high, as indicated by coefficient alpha (WAI-C = .96, WAI-T = .95, and WAI-O = .98) as is interrater reliability of the WAI-O (ICC = .98; Tichenor & Hill, 1989, p. 198). Comparison of WAI versions to other measures of working alliance has supported the convergent validity of the observer version, but not the client nor therapist versions (Tichenor & Hill, 1989). Further, there is no significant intercorrelation between WAI versions (Tichenor & Hill, 1989). Analysis of the WAI factor structure has been inconsistent, with little support for the hypothesized three-factor structure (Hatcher & Gillespy, 2006; Tracy & Kokotovic, 1989). However, meta-analysis has indicated that favorable therapy outcomes are moderately related to assessments of using the WAI made by clients (average  $r = .27$ ) and observers (average  $r = .23$ ), though not to ratings made by therapists (average  $r = -.03$ ; Horvath & Symonds, 1991).

## Interaction Questionnaire (IQ)

Vallano and Schreiber Compo (2011; Kieckhafer, Vallano, & Schreiber Compo, 2014; Vallano & Schreiber Compo, 2015) have developed the Interaction Questionnaire (IQ) as a measure of rapport within the context of investigative interactions. Vallano and Schreiber Compo (2015, p. 87) state that the IQ was adapted from the measure that is here referred to as the RQ -1994, as it was described by Bernieri et al. (1996). The IQ has been presented in three versions, which will here be called the IQ, the IQ Short, and the IQ Long. The limited psychometric information that has been published for these scales is summarized below.

**IQ -** The 27-item IQ (Vallano & Schreiber Compo, 2011) includes two subscales (see Table 16). The first subscale, which will be referred to here as the "IQ Perceptions of Interaction subscale" (IQ-PoInteraction), closely resembles the Perception of Interaction (PoInteraction) subscale of the RQ-1994, as can be seen in Table 16. Both the IQ-PoInteraction subscale and the RQ-1994 PoInteraction subscale instruct the interviewee to rate their *interaction* with the interviewer, and both subscales have highly similar items.

The second subscale of the IQ, which will be referred to here as the "IQ Perception of Interviewer subscale" (IQ-PoInterviewer), resembles the RQ-1994 Perception of Self (PoSelf) subscale. These two subscales have nine identical items, as can be seen in Table 16, although the IQ-PoInterviewer omits two items ("Comfortable" and "Cooperative") that are part of the RQ-1994 PoSelf. However, the two subscales differ in an important respect: the IQ-PoInterviewer subscale instructs respondents to rate their *interviewer's behavior*, whereas the RQ-1994 PoSelf subscale instructs respondents to rate *their own behavior*.

The IQ (Vallano & Schreiber Compo, 2011) includes a total of 27 items to measure "rapport-related characteristics" (p. 4), including 18 items on the IQ-PoInteraction subscale and 9 items on

the IQ-PoInterviewer subscale. Interviewees are instructed to rate each item according to the presence or absence of the characteristic listed on a Likert-type scale ranging from 1 (lack of characteristic) to 7 (high amount of characteristic).

As already stated, the 18 items of the IQ-PoInteraction subscale closely resemble those of the RQ-1994 PoInteraction subscale. However, some differences can be noted (see Table 16). First, two items have been slightly re-worded. Specifically, the RQ-1994 items "Boredom" and "Coordinated" have been reworded on the IQ as "Boring" and "Well-Coordinated," respectively. Second, four RQ-1994 items have been reversed on the IQ. Specifically, the RQ items "Comfortably Paced," "Focus," "Satisfaction," and "Warm" have been reversed on the IQ as "Uncomfortably paced," "Unfocused", "Unsatisfying", and "Cold", respectively. Third, one item, "Engrossing", is not included in the RQ but has been added to the IQ PoInteraction subscale.

Though Vallano & Schreiber Compo (2015, p. 87) specifically state that the IQ was based on the RQ-1994 used by Bernieri et al. (1996), it appears likely that the authors also drew upon the RQ-2001 when developing the PoInteraction subscale. The IQ-PoInteraction and the RQ-2001 have 18 nearly identical items. The only difference between these scales is that the IQ-PoInteraction has reversed the wording of RQ-2001 items "Comfortably Paced", "Focused", and "Satisfying" to "Uncomfortably paced", "Unfocused", and "Unsatisfying", respectively. The IQ - PoInteraction subscale also includes an item, "Engrossing", that is present in the RQ-2001, but not the RQ-1994 subscales. The fact that the IQ-PoInteraction subscale items more closely resemble those of the RQ-2001 than the RQ-1994 seems to indicate that the IQ was at least partly based on the RQ-2001.

Limited psychometric information has been reported for any version of the IQ. No evidence of factor structure or reliability has been reported for the IQ or its subscales. Evidence of the scales' validity has not been systematically presented. However, given the IQ-PoInteraction subscale's similarity to the RQ-1994 PoInteraction subscale and to the RQ-2001, it is reasonable to conclude that all three of these scales share similar, although not identical, psychometric properties. A similar conclusion cannot be made regarding the IQ-PoInterviewer subscale, however. The IQ-PoInterviewer subscale and the RQ-1994 PoSelf subscale differ substantially from each other because they are intended to measure much different targets, i.e., the source's perceptions of the interviewer versus the source's perceptions of the self.

Regarding the validity of the IQ, Vallano and Schreiber Compo (2015, p. 87) have argued that research has consistently indicated that rapport as measured by the IQ differs significantly between rapport-based experimental conditions (i.e. rapport vs no rapport). However, to support this argument, Vallano and Schreiber point to alternate versions of the IQ that differ substantially from the 27-item version that has been discussed in this section. These alternate IQ versions are discussed in the following sections.

Table 16. Rapport Questionnaire 1994 (RQ-1994), RQ-2001, and Interaction Questionnaire (IQ): Comparison of Items and Subscales

RQ-1994	IQ	RQ-2001
<b><u>Perceptions of Self Subscale</u></b>	<b><u>Perceptions of Interviewer Subscale</u></b>	-
Activity Level	Active	-
Awkward	Awkward	-
Boredom	Bored	-
Comfortable	-	-
Cooperative	-	-
Engrossed	Engrossed	-
Friendly	Friendly	-
Involved	Involved	-
Positive	Positive	-
Satisfied	Satisfied	-
Smoothness	Smooth	-
		-
<b><u>Perceptions of Interaction Subscale</u></b>	<b><u>Perceptions of Interaction Subscale</u></b>	<b><u>Perceptions of Interaction</u></b>
Active	Active	Active
Awkward	Awkward	Awkward
Boredom	Boring	Boring
Comfortably Paced	Uncomfortably paced	Comfortably Paced
Cooperative	Cooperative	Cooperative
Coordinated	Well-coordinated	Well-Coordinated
Dull	Dull	Dull
Focus	Unfocused	Focused
Friendly	Friendly	Friendly
Harmonious	Harmonious	Harmonious
Intense	Intense	Intense
Involving	Involving	Involving
Positive	Positive	Positive
Satisfaction	Unsatisfying	Satisfying
Slow	Slow	Slow
Warm	Cold	Cold
Worthwhile	Worthwhile	Worthwhile
-	Engrossing	Engrossing

### **IQ-Short (IQS)**

An article by Vallano and Schreiber Compo (2011, p. 963) reported that analyses in that study were performed on a substantially shortened version of the IQ PoInteraction that included only 7 of its 18 items. The other 11 items of the IQPoInteraction were eliminated because they were not “relevant to the interview setting” or were “difficult for non-native English speakers to comprehend.” The article (Vallano & Schreiber Comp, 2011, p. 963) named one of the eliminated items (“Engrossing”) but not the other ten. The shortened (16-item) version of the IQ that includes the shortened (7-item) version of the IQ-PoInteraction and the full (9-item) version of the IQ-PoInterviewer will here be called the IQ Short (IQS). The shortened (7-item) version of the PoInteraction will here be called the IQS-PoInteraction.

Vallano and Schreiber Compo (2011) provide very little psychometric information regarding the IQS or the IQS-PoInteraction. For instance, they do not even indicate which of the 18 items of the IQ-PoInteraction were retained to form the IQS-PoInteraction. Relevant to construct validity Vallano and Schrieber (2011, p. 964) found some support for the validity of the IQS-PoInterviewer subscale as a rapport measure: Specifically, when the items of this subscale were used as the dependent variables in a MANOVA, they were found as a group to be significantly related to a rapport-building manipulation in simulated investigative interviews ( $F(18, 184) = 2.12, p = .007, \eta^2 = .17$ ). However, similar support was not found for the validity of the IQS-PoInteraction subscale: When the items of this subscale were used as the independent variables in a MANOVA, they were not found as a group to be significantly related to a rapport-building manipulation ( $F(14, 202) = 1.56, p = .09$ ).

## **IQ Long (IQL)**

Kieckhaefer (2014) created a new and longer version of the IQ with 32 items which will here be called the IQ Long (IQL). Table 17 compares the 27-item IQ with the 32-item IQL. As can be seen in Table 17, the IQ and IQL bear some similarity to each other but also differ in important respects. (1) The IQ and the IQL both include two subscales. The two IQ subscales seek to measure Perceptions of the Interviewer (PoInterviewer) and Perceptions of the Interaction (PoInteraction). The two corresponding IQL subscales seek to measure Perceptions of the Investigator (PoInvestigator) and Perceptions of the Interaction (PoInteraction). (2) The IQ-PoInteraction subscale is highly similar to the IQL-PoInteraction subscale. Each of these scales includes 18 items. The items are not identical, but the differences tend to be small. Specifically, the items “Friendly” and “Engrossing” in the IQ were changed to “Unfriendly” and “Engaging”, respectively, in the IQL. As with the IQ, the IQL-PoInteraction items are highly similar to the items of the RQ-2001. (3) The IQ-PoInterviewer subscale is substantially different from the IQL-PoInvestigator subscale. First, the two subscales have different names (“Interviewer” versus “Investigator”).

Second, whereas the IQ Perceptions of the Interviewer scale includes only 9 items, the IQL Perceptions of the Investigator subscale includes 14 items. Specifically, as can be seen in Table 17 of the present proposal, (a) two items of the IQ PoInterviewer scale (“Involved” and “Satisfied”) are completely eliminated from the IQL PoInvestigator subscale, (b) seven new items that are not included in the IQ PoInterviewer subscale (“Rude”, “Kind”, “Likeable”, “Trustworthy”, “Credible”, “Respectful”, and “Attentive”) have been added to the IQL PoInvestigator subscale, and (c) one item of the IQ PoInterviewer scale (“Engrossed”) has been changed to “Engaging” in the IQL PoInvestigator subscale.



Basic psychometric information on the IQL, such as reliability and factor structure, was not reported in Kieckhafer's (2014) dissertation. However, the dissertation (p. 47) used the IQL to check the rapport manipulation designed by the author and the results provided construct support for the validity of the IQL subscales. Specifically, when the items of the IQL PoInteraction and PoInvestigator subscales were entered into a MANOVA, nearly all items of each subscale were found to be significantly affected in the predicted direction by an experimental manipulation of rapport. Though no psychometric properties of the IQL have been published, it is likely that the IQL-PoInteraction subscale's properties closely resemble those of the RQ-2001, given their similar structure and items.

Table 17. Interaction Questionnaire (IQ) Interaction Questionnaire Long (IQL) Comparison of Items and Subscales

IQ	IQL
<u>Perceptions of Interviewer Subscale</u>	<u>Perceptions of Investigator Subscale</u>
Active	Active
Awkward	Awkward
Bored	Bored
Engrossed	Engaging
Friendly	Friendly
Involved	-
Positive	Positive
Satisfied	-
Smooth	Smooth
-	Rude
-	Kind
-	Likeable
-	Trustworthy
-	Credible
-	Respectful
-	Attentive
<u>Perceptions of Interaction Subscale</u>	<u>Perceptions of Interaction Subscale</u>
Active	Active
Awkward	Awkward
Boring	Boring
Uncomfortably paced	Uncomfortably paced
Cooperative	Cooperative
Well-coordinated	Well-coordinated
Dull	Dull
Unfocused	Unfocused
Friendly	Unfriendly
Harmonious	Harmonious
Intense	Intense
Involving	Involving
Positive	Positive
Unsatisfying	Unsatisfying
Slow	Slow
Cold	Cold

Worthwhile  
Engrossing

Worthwhile  
Engaging

### **Observing Rapport-Based Interpersonal Techniques (ORBIT)**

A set of rating scales to measure aspects of rapport building, the Observing Rapport-Based Interpersonal Techniques (ORBIT), has been developed by Alison and his colleagues at the University of Liverpool (Alison et al., 2013, Alison et al. 2014, Alison & Alison, 2017). Unlike self-report measures such as the RQ and the IQ that are completed by the participants in an interview, ORBIT is a rating tool that is filled out by observers who have watched the participants during the interview. Whereas the RQ and IQ attempt to measure the participants' experience of rapport during the interview, ORBIT focuses on evaluating the rapport-building skills of the interrogator/interviewer. Alison et al. (2013) argued that a new rapport instrument was necessary because existing instruments did not address differential expression of rapport due to contextual disparities. That is, much of the past literature concerning rapport focused on cooperative relationships (e.g., student and teacher, doctor and patient), and it may be that expression of rapport in these settings varies from adversarial interactions (e.g., investigative interviews, interrogations). Alison et al. (2013) further asserted that existing instruments focused on measuring rapport at the “micro-level,” without considering the wider context of interactions. Alison et al. (2013) sought to develop a more contextually-relevant instrument by focusing on the ““overall atmosphere of communication”” within an interaction (p. 412). The authors drew upon well-established theories and concepts developed by researchers in the fields of psychotherapy and counseling. Specifically, in developing the ORBIT, Alison et al. (2013) drew directly from the theories and measurement approaches 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).

MI is “a directive, client centered counseling style for eliciting behavior change by helping clients to explore and resolve ambivalence” (Rollnick & Miller, 1995, p.325). Within the MI framework, counselors work to develop an empathetic atmosphere that promotes behavior change through highlighting client autonomy, encouraging change talk, and developing discrepancy. The MISC was developed to gauge the degree to which counselors’ behaviors are representative of the core values and goals of MI (Miller, Moyers, Ernst, & Amrhein, 2008). The MISC (now in Version 2.1) requires two coding passes: (1) global ratings of counselor and client adherence to MI principles and (2) a count of specific counselor behaviors consistent or inconsistent with MI principles. A single global client rating of self-exploration is made on a Likert-type scale ranging from 1 (no personally relevant material discussed) to 7 (active intrapersonal discussion). Global counselor ratings are also made on 7-point Likert-type scales and measure the degree to which counselors adhere to the central MI principles of acceptance, empathy, and MI spirit. Acceptance refers to the counselor’s ability to communicate unconditional positive regard to the client. Empathy refers to the extent to which the counselor attempts to take the client’s perspective. The MI Spirit rating is intended to capture the degree to which the counselor is generally competent in applying MI principles. Spirit ratings should consider the counselor’s support of collaboration, evocation, and autonomy within the interaction. A second coding pass is made that consists of counting the presence of individual counselor behaviors such as emphasizing control, asking a question, or offering advice with the permission of the client.

The appropriateness of a counseling-centered instrument (the MISC) as a model for measuring rapport in interrogations (ORBIT) is not immediately obvious. Counseling sessions and law enforcement interrogations seemingly have little in common on the surface: Whereas counseling sessions are based on cooperation and helping, interrogations tend to be adversarial and focus on data gathering. However, the developers of ORBIT have maintained that clear parallels exist between the goals of MI and the goals of investigative interviewing. Alison et al. (2013) have emphasized the similarity between investigative interactions and MI, arguing that the goals of both are to develop “an empathetic, respectful, and nonjudgmental atmosphere” while maintaining focus on the search for truth and eliciting behavior change (p. 412). Alison et al. (2014) have highlighted the skill of establishing a “therapeutic alliance” that fosters collaboration between interactants and facilitates the active exploration of meaningful information – a skill that, the authors argue, is necessary for both counselors and investigators. Alison et al., (2013) have argued that, as the efficacy of MI in therapeutic interactions is well documented, police interviewer behaviors consistent with MI principles should contribute to increased rapport and greater source cooperation in investigative contexts.

In developing ORBIT, Alison et al. (2013) relied on the MISC v1.1 to identify the behaviors they have termed “rapport-based skills” (Glynn & Moyers, 2009). The ORBIT authors also drew upon the Interpersonal Behavioral Circle (IBC) model based on Leary and Coffey’s (1954) argument that personality should be considered within the context of interactions and Birtchnell’s (2002) addition of an adaptive/maladaptive component to behavioral styles. The IBC model locates dyadic interactants’ respective behavioral styles across 8 categories of behavior arranged octagonally (Figure 7). Each octant represents a behavioral style, and is located on a two-dimensional plane in which the horizontal axis represents a continuum from love to hate and

the vertical axis a continuum from dominance to submission. Two sets of categories are represented: one variant consisting of behavioral patterns considered adaptive to interpersonal interactions and another comprised of items behaviors considered maladaptive. Each variant includes the following eight dimensions: (1) authoritative, (2) authoritative/cooperative, (3) cooperative, (4) cooperative/passive, (5) passive, (6) passive confrontational, (7) confrontational, and (8) authoritative/confrontational.

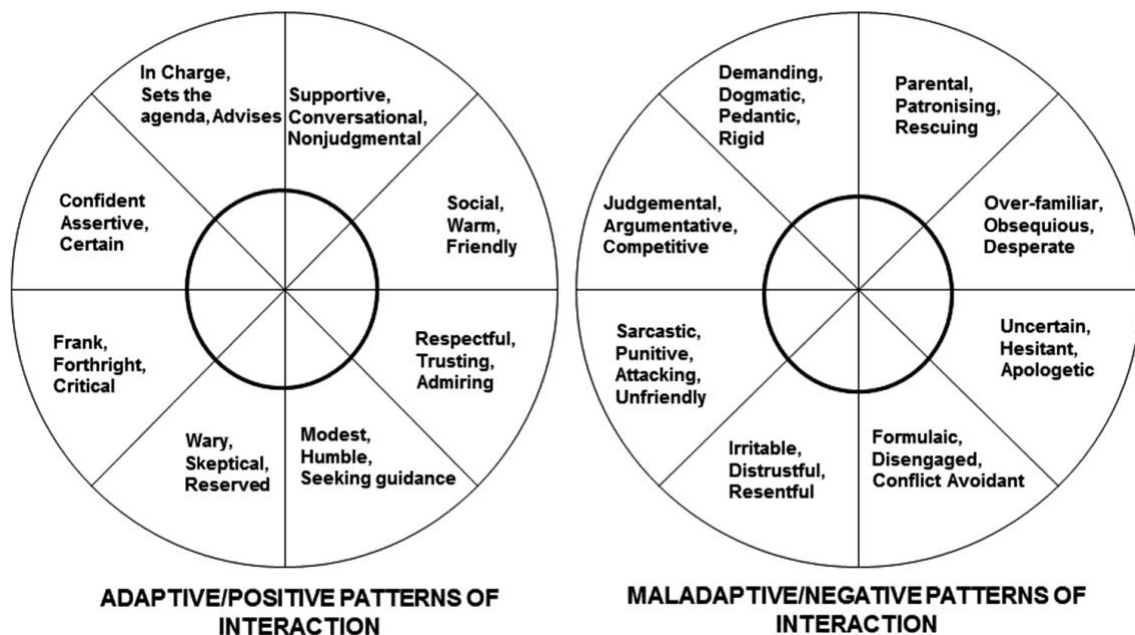


Figure 7. “Adaptive (left) and maladaptive (right) interpersonal behavior circles

(increasing intensity toward periphery, with center point behavior absent, first inner circle mild expression of behavior, mid circle moderate expression of behavior, outer circle persistent expression of behavior”; Alison et al., 2013, p. 418).

ORBIT is a multi-scale observational coding framework designed around two central measurements: (1) the degree to which interviewer behaviors are consistent with counseling techniques associated with rapport-building and (2) measurement of interpersonal communication styles adopted by both the interviewer and source. Within this framework, ORBIT includes three scales addressing interviewer behaviors and one addressing source

behavior. The ORBIT literature also includes three criterion scales by which the validity of the tool itself has been assessed. Thus, ORBIT consists of a total of seven sets of scales, including the criteria measures. The three sets of scales that measure interviewer behavior are the following: (1) G-MISC - Global Motivational Interviewing Skills: Investigative Interview Adaptation, (2) MIDAS – Motivational Interviewing of Detainees: Assessment of Skills and (3) IBC:I – Interpersonal Behavioral Circle style of the Interviewer. The ORBIT tool also includes a scale of source behavior, the IBC:D – Interpersonal Behavioral Circle style of the Detainee. These scales will be differentiated in the section to follow, describing the psychometric properties of each of the ORBIT scales.

The ORBIT coding manual (Alison & Alison, 2017) also includes three sets of scales that are not considered part of the ORBIT but can serve as validity criteria for the ORBIT scales. Addressing source behavior, these three criterion sets of scales are: (1) IYA – Interview Yield Assessment, (2) DES – Detainee Engagement Scale, and (3) CITS – Counter-Interrogation Tactics Scales. A description of each of these sets of scales and their coding strategies follows. It is important to note that much of the terminology used by ORBIT researchers has varied over the course of the tool’s publication history (Alison et al., 2013; Alison et al., 2014; Alison & Alison, 2017). The sections below discuss all scales and variables in the terms of the most recent ORBIT coding manual (Alison & Alison, 2017). Where discrepancies in descriptions occurred between the coding manual and earlier versions of the tool, the author will defer to the more recent publication.

### **Psychometric properties of ORBIT scales**

Alison et al. (2013) and Alison et al. (2014) report only limited information regarding the psychometric properties of some ORBIT scales. The following sections address the factor

structure, reliability, and validity of the four primary ORBIT measures: (1) G-MISC - IIA, (2) MIDAS, (3) IBC-I, and (4) IBC-D.

### **Global Motivational Interviewing Skills: Investigative Interview Adaptation (G-MISC)**

ORBIT includes a set of 5 observer rating scales that are intended to measure the degree to which an investigative interviewer's communication strategies are consistent with the counseling skills of Motivational Interviewing. These 5 rating scales, which are collectively called the Global Motivational Interviewing Skills: Investigative Interview Adaptation (G-MISC - IIA), are largely modeled on the MISC 1.1 (Glynn & Moyers, 2009), which, as already noted, is a set of rating scales designed to evaluate counselors' Motivational Interviewing skills. The five scales of the G-MISC - IIA are intended to measure the following interviewer skills: (1) Acceptance, (2) Empathy, (3) Adaptation, (4) Evocation, and (5) Autonomy (see Alison & Alison, 2017 for full description of each skill's operational definition). For the purpose of brevity, the G-MISC – IIA will be further abbreviated simply as G-MISC herein.

Observers using the G-MISC globally rate an interviewer on each of these five skills at the end of an interview or a lengthy interview sample. The ratings are intended to capture the observer's overall impression of the interviewer's adherence to MI principles over the course of the interview. Each skill is rated on a 7-point Likert-type scale ranging from 1 (poor skill) to 7 (positive skill), with 4 intended to represent a neutral point. Previous versions of ORBIT refer to the G-MISC set of scales as GMIS (Alison et al., 2013, p. 415; Alison et al., 2014, p. 424)

### **Motivational Interviewing of Detainees – Assessment of Skills (MIDAS)**

ORBIT also includes a set of eight observer rating scales called Motivational Interviewing of Detainees – Assessment of Skills (MIDAS). Each of these rating scales is intended to measure the degree to which an interviewer has shown certain interviewing strategies



consistent with MI. Specifically, the scales evaluate the degree to which the interviewer has used the following strategies: (1) rapport and resistance, (2) summaries/paraphrasing, (3) developing discrepancy, and (4) support of autonomy. Each of these four strategies or skills is rated on two 4-point Likert-type scales ranging from 0 (absence) to 3 (extreme). One of the two scales evaluates the degree to which the interviewer has shown behavior *consistent* with the MI strategy (MIC). The second of the two scales evaluates the degree to which the interviewer has shown behavior *inconsistent* with the MI Strategy (MIIN). Because there are four strategies, and each strategy is rated twice (once for consistency and once for inconsistency), the total number of MIDAS scales is eight.

Rating on the eight MIDAS scales are made by an observer after watching an entire interview or lengthy interview sample. Earlier versions of ORBIT (Alison et al., 2013, p. 426) included a fifth MIDAS scale: reflective listening. The most recent version of the ORBIT coding manual (Alison & Alison, 2017) does not include this scale. It is unclear whether the omission represents a deliberate elimination of the reflective listening scales or instead is due to a printing error.

Although MIDAS produces scores on eight separate scales, articles on ORBIT (Alison et al., 2013, p. 426; Alison et al., 2014, p. 425) have reported the results of structural equations models with only one MIC score and one MIIN score. It appears that the four MIC scale scores are combined into a single Global MIC score, and similarly the four MIIN scale scores are combined into a single Global MIIN score, possibly by averaging. However, the ORBIT manual and articles on the ORBIT do not explain how the various MIDAS scores are combined into two global scores.

### **Interpersonal Behavioral Circles – Interviewer & Detainee (IBC-I; IBC-D)**

Drawing upon the IBC model, ORBIT also attempts to measure the styles of communication adopted by the interviewer and the source (referred to in the ORBIT coding manual as *detainee*) throughout an interaction. As shown in Figure 7, the IBC model conceptualizes an individual's interpersonal style during an interaction as potentially expressing 16 different patterns. There are eight adaptive/positive patterns as represented by the eight sectors of the left circle in Figure 7, and eight corresponding maladaptive/negative patterns as represented by the eight sectors of the right circle. In ORBIT, each of the two participants in an interrogation, the "interviewer" and "detainee," is rated by an observer on each of the 16 IBC sectors using a 4-point Likert-type scale ranging from 0 (absent) to 3 (extreme). Thus, ORBIT yields 16 IBC ratings of the interviewer (IBC-I) and 16 IBC ratings of the detainee (IBC-D). These ratings are assigned by an observer at the end of an interview or a sample of interviews. The ORBIT coding manual (Alison & Alison, 2017) provides a list of suggested behaviors characteristic of each of the 16 behavioral patterns or styles. For example, interviewers who follow clear interview plans and clearly set out topics to be covered in the interview would be scored highly in the *positive authoritative* dimension, while maintaining rigid adherence to an interview plan and arguing over small points would earn the interviewer a higher rating on the *negative authoritative* dimension. Alison and his colleagues (2013, p. 425, Figure 25) have presented a structural equations model that shows IBC-I as a single manifest variable and likewise IBC-D as a single manifest variable. However, neither the article by Alison and his colleagues (2013) nor the ORBIT manual (Alison & Alison, 2017) explains how the 32 separate IBC ratings are combined to yield these two variables, IBC-I and IBC-D.

## Factor Structure

Alison et al. (2013, p. 423) report the results of a confirmatory factor analyses (CFA) as evidence of “construct validity of the latent variables created for Yield and MI”, that is the IYA and G-MISC, respectively. Overall model fit was determined to be adequate according to standards suggested by Schumacker and Lomax (2004) ( $\chi^2/df = 4.05$ , SRMR = .01, NFI = .99, TFI = .98, CFI = .99) although RMSEA was slightly over the limits considered acceptable (RMSEA = .089). In addition, Alison et al. (2013) presented indices of the fit of the hypothesized structural model as evidence of factorial validity. Indices indicated acceptable to good model fit, superior to all other alternate models tested ( $\chi^2/df = 2.76$ , SRMR = .04, NFI = .95, TFI = .95, CFI = .97, RMSEA = .067). Alison et al. (2014) also reported overall model fit as good (SRMR = .01, NFI = .99, TFI = .98, CFI = .99) despite some indices falling above the acceptable level (RMSEA = .09,  $\chi^2/df = 7.73$ ). As in earlier research, the authors found the hypothesized structural model including the G-MISC scales and latent MI variable as an acceptable to good fit for the data ( $\chi^2/df = 3.75$ , SRMR = .04, NFI = 0.93, TFI = 0.92, CFI = 0.95, RMSEA .06).

All factor loadings for G-MISC scales were statistically significant in models presented by Alison et al. (2013) and Alison et al. (2014). As shown in Table 18, all the G-MISC scales loaded onto a single factor, which Alison et al. (2013, p. 423) referred to as the "MI latent variable." This unifactorial factor solution suggests there is probably substantial redundancy among the G-MISC scales, and that most of the reliable variance of these scales reflects the same underlying construct, which Alison et al. (2013) call the MI latent variable.

So far as the author of this thesis proposal can determine, Allison and his colleagues have not reported factor analyses that show the factor structure or inter-relationship of ORBIT's MIDAS, IBC-I and IBC-D scales.

Table 18. Standardized loadings of ORBIT G-MISC scales on "MI latent variable" as reported by

Alison et al. (2013, p. 425) and Alison et al. (2014, p. 426)		
	2013	2014
Acceptance	.84	.81
Empathy	.91	.90
Adaptation	.86	.86
Evocation	.92	.93
Autonomy	.82	.73

### Reliability of ORBIT scales

An article by Alison et al. (2013) reported the inter-rater reliability of many, though not all, ORBIT scales. First, Table 10 of the article (p. 426) reported the reliability of the scales that the ORBIT manual and the present thesis proposal have referred to as the "MIDAS" scales (it should be noted, however, that the title of Table 10 in the article erroneously refers to these scales as belonging to the Global Motivational Interviewing Skills Code). The reliabilities of the MIDAS scales reported in Table 10 of that article are reproduced here in Table 19. As can be seen in Table 19, the article reported that kappas for the MI Consistent (MIC) scales of the MIDAS ranged between .07 and .27, with a median of .24, and that the reliability of the MI Inconsistent (MIIN) scales of the MIDAS ranged between .21 and .56, with a median of .24. The article by Alison et al. (2013, Table 9, p. 425) also reported interrater reliabilities for the 16 scales of the IBC-I and the 16 scales of the IBC-D. These reliabilities are reproduced here in Table 19. As can be seen in Table 19, for the 16 IBC-I scales, kappas ranged between .10 and .74, with a median of .47. For the 16 IBC-D scales, kappas ranged between .10 and .68, with a median of .365.

Alison et al. (2013, p.422) adopted the following standards for assessing the adequacy of kappa values: 0.00–0.20 = Poor; 0.21–0.40 = Poor; 0.41–0.60 = Moderate; 0.61–0.80 = Strong;

and  $> 0.80$  = Near Complete Agreement. Using these standards, the authors erroneously claimed (p. 422) that only 6 ORBIT scales fell below the threshold of moderate agreement ( $\kappa < .40$ ). In fact, however, as can be seen in Table 19, 23 (55%) of the kappa values for ORBIT scales as reported by Alison et al. (2013) fall below this threshold. In addition, no interrater reliability figures have been reported for the four scales of the G-MISC.

Table 19. ORBIT scale inter-rater reliability assessed from 26 random interviews (Alison et al., 2013)

Scale	Kappa	
IBC-I	Positive	Negative
Authoritative	0.10	0.71
Authoritative / Cooperative	0.54	0.51
Cooperative	0.37	0.43
Cooperative / Passive	0.26	0.42
Passive	0.43	0.66
Passive / Confrontational	0.33	0.68
Confrontational	0.21	0.74
Authoritative / Confrontational	0.51	0.71
IBC-D	Positive	Negative
Authoritative	0.47	0.24
Authoritative / Cooperative	0.66	0.24
Cooperative	0.39	0.36
Cooperative / Passive	0.57	0.37
Passive	0.10	0.44
Passive / Confrontational	0.23	0.20
Confrontational	0.57	0.14
Authoritative / Confrontational	0.68	0.15
MIDAS	MIC	MIIN
Reflective Listening*	0.12	0.21
Rapport and Resistance	0.27	0.53
Summaries	0.24	0.29
Developing Discrepancy	0.07	0.56
Autonomy	0.26	0.22

*Note. Reflective Listening does not appear in the ORBIT Coding Manual (Alison & Alison, 2017)*

## **Criteria for Assessing Validity of ORBIT Scales**

Evidence for the validity of the ORBIT scales consists primarily of findings that these scales are significantly related to three criteria that are designed to reflect desirable interview outcomes: the Detainee Engagement Scale (DES), the Counter-Interrogation Tactics Scale (CITS), and the Interview Yield Assessment (IYA). Each of these three validity criteria is described in the present section.

The Detainee Engagement Scale (DES) is an observer rating scale modeled on a similar client rating scale that is part of the MISC (Miller & Rollnick, 1992). At the end of an interview or interview sample, an observer makes a single, global rating on the DES to indicate the degree to which the source or interviewee has been actively engaged throughout the interview. The ORBIT manual (Alison & Alison, 2017) indicates that the DES rating is made on a 7-point Likert-type scale ranging from 1 (“the detainee says nothing at any point”) to 7 (“the detainee answers questions fully and thoroughly”).

The Counter-Interrogation Tactics Scale (CITS) is a set of observer rating scales that are intended to measure the frequency with which a source has employed counter-interrogation tactics (CITs; i.e. strategies meant to delay or disrupt the progression of an interview) while being interviewed. Ratings are made for three types of CITs: (1) distractions, (2) disengagement, and (3) provocation. Distractions include behaviors such as asking to be seen by a medical examiner and asking for food or drink. Disengagement is characterized by behaviors such as refusing to offer a comment or refusing to look at the interviewer. Provocations include being deliberately insulting or interrupting the interviewer.

A checklist of the three categories of CIT behaviors is provided in the ORBIT manual (Alison & Alison, 2017). Coders are instructed to first tally the presence of each of the listed behaviors while observing the interview. Then, at the end of the interview, each of the three

categories is scored on a 7-point Likert-type scale intended to convey intensity. The manual does not provide anchors for these three scales, nor does it explain the procedures that coders follow to calculate the Overall CIT Score.

Note that the CITS described in the ORBIT coding manual (Alison & Alison, 2017) is referred to in earlier publications as CIT (Alison et al., 2014), though the scales differ slightly. No publications in between these were located by the author of the current thesis. As CITS has not yet been employed in an empirical study, all further discussion of this set of scales will refer to the CIT scales.

The Interview Yield Assessment (IYA) is intended to capture the amount of relevant or valuable information that is shared by a source during an interview. It is comprised of a set of scales that measure the degree to which a source has shared information related to the source's (1) capability, (2) opportunity, (3) motive, and (4) descriptions (items, locations, individuals, and timing) related to the offense being discussed. Each of these information categories is rated by a coder on a global 4-point Likert-type scale ranging from 0 (absent) to 3 (high). Coders are instructed to score each information category and note specific details as they are discussed, stopping every 15 minutes to review the coding that has been completed. The global Likert-type score is assigned at the end of the interview.

### **Evidence of Validity for ORBIT Scales**

Alison and his colleagues have not formally reported validity coefficients showing the relationship of ORBIT scales to relevant criteria, such as DES, CITS, and IYA. However, validity coefficients for the ORBIT scales can sometimes be estimated by using information reported by Alison and his colleagues. First, validity coefficients can sometimes be estimated from the path coefficients of structural equations models (SEMs) presented in articles by Alison

and his colleagues (Alison et al. 2013; Alison et al, 2014). Second, validity coefficients can sometimes be estimated from the beta coefficients of regression or logistic regression equations presented in the same articles. The validity coefficients reported in the remainder of this section were calculated using these two information sources (i.e., SEMs or regression coefficients). First, regarding the validity of the G-MISC scales and the MI latent variable on which they load: All of these scales and the MI latent variable have moderate positive correlations with information yield as measured by IYA ( $r = .32$  to  $.39$ ) in findings reported by Alison et al. (2013, p. 425). Table 20 shows the estimated correlations of these scales with IYA, based on the structural equations model presented by Alison et al. (2013, p. 425). Overall, these correlations show that the G-MISC scales are related to total amount of relevant information shared by sources.

Table 20. Estimated correlation of G-MISC Scales with IYA

	IYA
Acceptance	.33
Empathy	.35
Adaptation	.34
Evocation	.36
Autonomy	.32
MI Latent Variable	.39

*Note. All estimates derived from Alison et al. (2013)*

In addition, regression coefficients provided by Alison et al. (2014, p. 426) allow estimation of validity coefficients for the relationship of G-MISC scales with CIT. Specifically, Alison et al. (2014, p. 426)) present unstandardized regression coefficients for G-MISC scales as predictors of counter-interrogation tactics. Simple calculations can convert these unstandardized coefficients into standardized coefficients, which serve as estimates of validity.



Table 21 presents the estimated validity coefficients for the relationship of G-MISC scales and the latent MI variable with the various CIT variables reported by Alison et al. (2014). As can be seen, the G-MISC scales have (c) a moderate relationship (values of  $r$  from  $-.28$  to  $-.38$ ) with one CIT variable (Passive), (b) a very weak relationship (all absolute values of  $r \leq .23$ ) with a second CIT variables (No Comment), and (c) virtually no relationship (all absolute values of  $r < .10$ ) with the remaining three CIT variables (Verbal, Passive Verbal, and Retraction). In other words, greater interviewer MI skill was (a) moderately associated with decreased use of Passive behaviors by the source (e.g., refusal to look, silence), (b) weakly associated with fewer "No Comment" responses from the source, and (c) virtually unrelated to the source's use of Verbal (e.g., relating unrelated, well-known, or scripted information), Passive Verbal (e.g., monosyllabic responses or claiming lack of memory), and Retraction counter-interrogation tactics.

Table 21. Approximated total effect of G-MISC Scales on detainee CIT use

	Verbal	Passive Verbal	Passive	Retraction	No Comment
Acceptance	-.073	.041	-.31	.057	-.19
Empathy	-.093	.045	-.34	.063	-.21
Adaptation	-.086	.043	-.33	.060	-.198
Evocation	-.09	.047	-.35	.065	-.21
Autonomy	-.081	.037	-.28	.051	-.17
Motivational	-.10	.05	-.38	.07	-.23

*Note. All estimates derived from Alison et al. (2014)*

The Passive variable used by Alison et al. (2014, p. 426) was comprised of two scales consisting of observed frequencies of source silence and refusal by the source to look at the interviewer. Of particular interest to the current study, these are behaviors that can potentially disrupt some aspects of interpersonal synchronization, even if intentional attempts to build

coordination are made by one interactant. These results suggest that some aspects of rapport-building strategies employed by interviewers could be practically useful in attempts to increase coordination with a source, an important component of rapport. Below, Table 22 lists each G-MISC scale's approximated correlation with the Passive scales (1) Silence and (2) Refuse to look.

Table 22. Approximated correlations of G-MISC scales on Passive variable scales in Alison et al. (2014)

	Passive	Silence	Refuse to look
Acceptance	-.31	-0.24	-0.20
Empathy	-.34	-0.27	-0.21
Adaptation	-.33	-0.26	-0.21
Evocation	-.35	-0.27	-0.22
Autonomy	-.28	-0.22	-0.18
Motivational	-.38	-0.30	-0.24

Only limited evidence of validity for the IBC-I and IBC-D scales has been presented in the published literature (Alison et al. 2013; Alison et al., 2014). Standardized parameter estimates of path coefficient from a structural equation model (SEM) presented by Alison et al. (2013, p. 425) show that Interviewer adaptive (-.08 ns) and Interviewer maladaptive (-.10) scales, as measured by IBC-I, were weakly or not at all associated with IYA. These scales were also weakly associated with CIT variables in a second SEM, with only one path coefficient achieving an absolute magnitude greater than .10 (Alison et al., 2014). The standardized parameter estimates for the SEM path coefficient of the total effect of Interviewer adaptive on the Passive Verbal variable was .28 and statistically significant.

To summarize these overall results: (1) Overall, the five G-MISC scales have been shown to be positively related to information yield (IYA) and negatively related to behaviors (i.e.,

refusal to look, silence) that indicate refusal to participate in the interaction (Passive CIT). It should be noted that the G-MISC scales are highly intercorrelated, and it is possible that some of these scales could be eliminated or collapsed with each other to create a more parsimonious tool without affecting its validity. (2) The validity of the IBC scales has been examined in the published literature and found to be low. (3) No evidence of validity for the MIDAS scales has been presented in the literature.

### **ORBIT Summary**

Several comments may be made about existing research on ORBIT. First details regarding the scoring of the interviewer-related scales is vague. Discussion of some scale calculations and other specific information is missing from the most recent version of the coding manual (Alison & Alison, 2017) and there are unexplained discrepancies between published accounts. Certain psychometric information is either omitted or presented unsystematically. When information is presented, it is often in a form that makes psychometric evaluation of ORBIT laborious and difficult. For instance, in studies of the ORBIT, unstandardized coefficients are sometimes inappropriately reported in place of standardized effect sizes for scales (e.g., Alison et al., 2014, pp.425 & 426). Because of this, readers must make laborious calculations to estimate relevant effect size statistics such as  $r$  or  $d$ .

Research on ORBIT seems to indicate the value of considering the global emotional atmosphere of an interrogation, along with more specific rapport-building behaviors of the interviewer. It is worth noting, though, that studies on ORBIT (Alison et al., 2013; Alison et al., 2014) have not attempted to measure the interviewer's and interviewee's perceptions of rapport using either self-report or observational coding. Instead, studies on the ORBIT have used the cooperativeness of the interviewee, as measured by information yield (IYA; Alison et al., 2013)

or frequency of CIT use (Alison et al., 2014) as a measure of rapport. Although it is probably true that yield and reduced CIT use are correlated with rapport, they can and probably should be distinguished from it. Rapport consists of the emotional and cognitive evaluations that the interviewer and interviewee have concerning the quality, depth, and ease of their communications. These evaluations may well influence the amount of information shared during an interview, but the relationship is not always straightforward.

For instance, an interviewer and interviewee may establish a very warm relationship during an interview, but the interviewee may still resist sharing crucial information that the interviewer seeks. Or conversely, an interviewee may decide to share crucial information (out of fear or self-interest), despite feeling little rapport with the interviewer. Thus, although studies on the ORBIT have reported a relationship between some interviewers' rapport-building behaviors and important interrogation outcomes (for instance, several MIDAS scores are related to IYA and some CITS), they have not established that the rapport-building behaviors influence participants' perceptions or experiences of rapport, or that this experience of rapport mediates the relationship between rapport-building behaviors and interrogation outcomes.

### **Rapport Scales for Investigative Interviews and Interrogations (RS3i)**

The Rapport Scales for Investigative Interviews and Interrogations (RS3i; Duke, Wood, Bollin, Scullin, & LaBianca, 2018) were developed with the intent of measuring Tickle-Degnen and Rosenthal's (1990) three essential components of rapport (e.g., mutual attentiveness, positivity, and coordination) while also accounting for other aspects of a source's perception of rapport (e.g., the source's perception of the interviewer's expertise) that may be experienced during an investigative interaction. The RS3i asks sources to rate their perceptions of the interviewer on 21 items using a 5-point Likert-type scale. Eighteen of the 21 items inquire about

rapport experienced as a result of interviewer behavior, and the remaining three items inquire about the source's engagement in the interaction. The items of the RS3i and the scales they belong to are presented below in Table 23.

Eleven of the RS3i items form three scales intended to measure the characteristics of rapport proposed by Tickle-Degnen and Rosenthal: Attentiveness (4 items, reflecting Tickle-Degnen and Rosenthal's "mutual attentiveness"), Trust/Respect (4 items, reflecting Tickle-Degnen and Rosenthal's "positivity"), and Connected Flow (3 items reflecting Tickle-Degnen and Rosenthal's "coordination"). Two additional RS3i scales measure aspects of rapport not included in Tickle-Degnen and Rosenthal's theory: Expertise (4 items, reflecting the interviewer's skill as an interviewer, as perceived by the interviewee) and Cultural Similarity (3 items, reflecting the interviewee's perception of shared cultural background with the interviewer). A sixth RS3i scale, Commitment to Communication (CtC; 3 items), is intended to measure, not rapport, but rather the outcome of the interview (i.e., increased sharing of information by the interviewee with the interviewer), and is closely related to the concept of *operational accord* proposed by Kleinman (2006). The CtC scale is thus intended to measure, via sources' self-report, a similar construct (interview outcome) as what ORBIT studies have attempted to measure using DES, CITS, and IYA.

Table 23. RS3i and Commitment to Communicate Items by Scale

<u>Scale</u>	<u>Item</u>
Attentiveness	The Interviewer really listened to what I had to say. The Interviewer paid careful attention to my opinion. The Interviewer was attentive to me. The Interviewer was interested in my point of view.
Trust/Respect	I think the Interviewer is generally honest with me. The Interviewer respects my knowledge. I think that the Interviewer can generally be trusted to keep his/her word. I feel I can trust the Interviewer to keep his/her word to me.
Connected Flow	The interviewer and I worked well together as a team. Communication went smoothly between the Interviewer and me. The Interviewer and I got along well during the interview.
Expertise	The Interviewer did his/her job with skill during interview. The Interviewer performed expertly during the interview. The Interviewer made an effort to do a good job. The Interviewer acted like a professional.
Cultural Similarity	The Interviewer probably shares my culture. The Interviewer and I probably share the same ethnicity. Interviewer shares my culture
Commitment to Communication	I was motivated to perform well during the interview. I wanted to do a good job during the interview. I felt committed to accomplishing the goals of the interview.

### RS3i Psychometric Properties

In two studies reported by Duke et al. (2018), participants took on the role of sources in simulated interrogations and afterwards rated their interviewers on the RS3i. In study one (n = 80), confirmatory factor analysis (CFA) and calculation of interscale correlations were used to test the hypothesized structure of the RS3i scales. Internal consistency for each scale was also calculated. Study two (n = 94) confirmed the identified factor structure with a second CFA.

Evidence of construct validity was demonstrated along with concurrent validity with multiple other relevant scales (see Duke et al., 2018, p. 7). Evidence of the RS3i scales' factor structure, reliability, and validity are summarized below.

### **Factor structure**

Duke et al. (2018) provided a detailed review of two CFAs each indicating a five-factor solution for the 18 items in the RS3i rapport subscales (Attentiveness, Trust/Respect, Expertise, Cultural Similarity, and Connected Flow). The factor loadings for each of the items on their RS3i subscale are reported by Duke et al. (2018, p. 4) and are recreated in Table 24 below. A CFA in study one tested the hypothesized five-factor structure and resulted in mixed indices of fit,  $\chi^2_{SB}(125) = 188.69$ ,  $p < .001$ ; CFI = .88; SRMR = .074. A second CFA in study two showed similar, though slightly improved fit,  $\chi^2_{SB}(125) = 169.85$ ,  $p = .005$ ; CFI = .94; SRMR = .067. The loadings for each item onto their hypothesized factor were all greater than .5 in both studies (Table 24). No evidence of factor structure for the CtC subscale has been published. However, it's relationship with other RS3i subscales has been clearly demonstrated.

Table 24. RS3i Item Loadings in Five-Factor Solution (Duke, et al. (2017, p. 4)

		Factor loadings	
		Study 1 (n=80)	Study 2 (n=94)
Item by Scale			
Attentiveness	Interviewer really listened to what I had to say	0.82	0.83
	Interviewer paid attention to my opinion	0.93	0.86
	Interviewer was attentive to me	0.57	0.75
	Interviewer interested in my point of view	0.79	0.78
Trust/respect	Interviewer honest with me	0.73	0.79
	Interviewer respects my knowledge	0.56	0.73
	Interviewer can generally be trusted to keep word	0.65	0.87
	I can trust interviewer to keep word to me	0.67	0.88
Expertise	Interviewer did job with skill	0.75	0.79
	Interviewer performed expertly	0.59	0.85
	Interviewer made effort to do good job	0.52	0.68
	Interviewer acted like a professional	0.61	0.50
Cultural similarity	We have our culture in common	0.79	0.82
	Interviewer and I share ethnicity	0.70	0.81
	Interviewer shares my culture	0.84	0.87
Connected flow	We work well as a team	0.58	0.87
	Communication went smoothly	0.65	0.83
	Interviewer and I got along well	0.77	0.69

Analysis of subscale intercorrelations produced further evidence of the RS3i's structure (Duke et al., 2018, p. 5). Table 25 reports the subscale intercorrelations for both study one and study two. The Interviewer Influence through Rapport model proposed by (Duke, 2013) hypothesized that Commitment to Communicate would be significantly positively correlated with each of the RS3i scales. This was supported by study two in which all RS3i scales, including CTC, were significantly correlated. Of particular interest to the current study, CtC was



consistently related to scales for Attentiveness Trust/Respect and Connected Flow. Every RS3i scale, with the exception of Expertise, was significantly correlated with similar measures to establish concurrent validity (Duke et al., 2018).

Table 25. Correlations among RS3i Scales. Study 1 (n = 80) source ratings are below the diagonal and Study 2 (n = 94) source ratings are above the diagonal.  
(Duke et al., 2018, p. 5)

Scales and Subscales	1	2	3	4	5	6
1. Attentiveness	---	.48*	.46*	.37*	.62*	.45*
2. Trust/Respect	.55*	---	.49*	.35*	.69*	.51*
3. Expertise	.50*	.59*	---	.36*	.42*	.60*
4. Cultural Similarity	.01	-.04	-.08*	---	.31*	.33*
5. Connected Flow	.48*	.52*	.58*	-.10	---	.62*
6. Commitment to Communication	.29*	.14	.13	.01	.40*	---

*Note.* \* $p < .05$

### Reliability

Duke et al. (2018, p. 5) also reported internal reliability figures indexed by Chronbach  $\alpha$  for each RS3i subscale, including CtC, for both studies. This information is summarized in Table 26 below. Chronbach  $\alpha$ 's were above .70 (.71-.88) for all scales except one, indicating adequate internal consistency.

Table 26. RS3i Scale Reliability Indices (Duke et al., 2018, p. 5)

Scale	(n = 80)	(n = 94)
	Study 1	Study 2
Attentiveness	.84	.88
Trust/Respect	.75	.84
Expertise	.71	.79
Cultural Similarity	.72	.87
Connected Flow	.69	.83
Commitment to communication	.74	.77

*Note. All figures provided are Cronbach's  $\alpha$*

### Validity

Validity of the RS3i scales was assessed through demonstrating convergent and discriminant validities using an additional series of related scales and concurrent validity through demonstrating relationships with theoretically related outcome variables. Convergent validity of each of the RS3i scales was demonstrated by significantly correlating each subscale with conceptually related scales that had been previously validated (see Duke et al., 2018, p. 6 for a detailed review of each scale used). For example, RS38 Attentiveness was correlated with other measures of attentiveness (average  $r = .51$ ) and RS3i Trust / Respect was correlated with other measures of trust (average  $r = .72$ ). Indices for RS3i Expertise (.44), RS3i Cultural Similarity (.75), and RS3i Connected Flow (.63) were similarly large. Neither convergent nor divergent validity for the CtC scale was demonstrated, as this scale was not significantly correlated with other measures of interviewee cooperation (Duke et al., 2018, p. 9). The average correlations of each scale with its respective external measures are presented below (Table 27).

Table 27. RS3i Scale Convergent Validity Indices (Duke et al., 2018)

Scale	Average <i>r</i>
Attentiveness	.51
Trust/Respect	.72
Expertise	.44
Cultural Similarity	.75
Connected Flow	.63
Commitment to Communication (CtC)	-

Concurrent validity for the RS3i scales was demonstrated through their relationship to interviewer styles and information gain. Specifically, every rapport scale other than Cultural Similarity was significantly higher in an interview condition manipulated such that the interviewer attempted to build rapport with the source. Duke et al. (2018) measured information gain using the Sharing Information Rating Scale (SIRS) to assess the total number of facts shared during the course of the interview. Only RS3i Attentiveness ( $r = .23$ ,  $p = .03$ , 95% CI: .03, .41) and RS3i Trust/Respect ( $r = .21$ ,  $p = .04$ , 95% CI: .01, .40) were significantly correlated with the total number of facts shared by sources.

### **Verbal Behavior and Rapport (the *Bronstein Suite*)**

Bronstein et al. (2012) argued that research involving rapport could greatly benefit from examining how rapport's many aspects are expressed through verbal communication. These researchers developed a tool intended to observationally measure the contributions of specific verbal behaviors to dyadic perceptions of rapport: the Verbal Rapport Assessment Scale (VeRAS). The authors highlighted the extant literature's focus on nonverbal behavior, noting that situations sometimes arise wherein rapport-building is highly important although the interactants do not have access to visual cues (e.g., law enforcement negotiations).

Bronstein et al. (2012) point out that behavioral indicators made purely through the verbal channel are indicative of the quality of communication in interactions (Archer & Akert, 1977), influence first impressions above and beyond nonverbal communication (Berry, Pennebaker, Mueller, & Hiller, 1977), and result in greater cooperation between negotiators compared to those with only visual channel access (Whichman, 1970). Linking research on the verbal expression of rapport in psychotherapeutic environments (e.g., Gfeller, Lynn, & Pribble, 1987; Sharpley, 1997; Sharpley, Fairnie, Tabaray-Collins, Bates, & Lee, 2000) to shared goals in the field of negotiations, Bronstein et al. (2012) examined "the role of the verbal channel in the production of rapport in negotiation interactions" (p. 1094).

Bronstein et al. (2012) had 292 Israeli undergraduates randomly assigned to the role of either interviewee or interviewer in a simulated pay negotiation wherein interactants only had access to the auditory channel (i.e., separated by an opaque screen). After a 15-minute interaction, participants completed the RQ-2001 (Bernieri & Gillis, 2001; Bernieri, 2005). Bronstein et al. (2012) reported that factor analysis indicated the RQ-2001 was unidimensional with high internal consistency (Cronbach's  $\alpha = .88$ ). In addition to the VeRAS tool itself, Bronstein et al. (2012) also constructed three measures to assess perceptions of rapport with global self-report and observational scales: (1) the Negotiation Emotional Satisfaction Scale (NESS), (2) the Interaction Rapport Scale (IRS), and (3) the Negotiators' Rapport Scale (NRS). That is, the Bronstein Suite includes a total of four independent measures (the NESS, IRS, NRS, and VeRAS). Detailed discussion of each measure along with available psychometric descriptions follow below.

### **Negotiation Emotional Satisfaction Scale (NESS),**

The Negotiation Emotional Satisfaction Scale (NESS) is a 3-item self-report questionnaire designed to assess interactants' feelings about their participation in an interaction. In fact, Bronstein and his colleagues (2012) did not give a name to this questionnaire, but for convenience it will herein be referred to as the NESS. This scale includes three items meant to assess unique dimensions of the interactants' feelings regarding the interaction: (1) their perceptions of a sense of chemistry with their partner, (2) satisfaction with the negotiation process, and (3) willingness to negotiate with their partner again. Respondents rate each item on an 8-point, Likert-type scale ranging from 0 (not at all) to 8 (very much). Bronstein et al. (2012) do not supply the precise wording of each item. They report that "rapport for each participant was calculated as the average of the three items" with higher NESS scores indicating higher levels of perceived rapport (p. 1096). Bronstein et al. (2012) do not explicitly state whether NESS scores should be averaged within dyads or should be considered as separate scores from each individual interactant. However, the article by Bronstein et al. (2012, p. 1102) seems to have used the NESS in both ways, both as an individual measure and (through averaging) as a dyadic measure. Bronstein et al. (2012) report that the three items of the NESS had high internal consistency (Cronbach's  $\alpha = .87$ ). However, when the NESS ratings by one member of an interaction were compared with the NESS ratings by the other member, intradyad agreement was found to be low ( $r = .20$ ,  $p < .05$ ). However, convergent validity was substantial: participants' self-report rapport scores on the NESS scores were highly correlated with their self-report rapport scores on the RQ-2001 ( $r = .78$ ,  $p < .001$ ).

### **Interaction Rapport Scale (IRS)**

The Interaction Rapport Scale (IRS) is an 11-item observer rating scale developed by Bronstein et al. (2012). In the study by these authors, judges read transcripts of recorded

negotiations and rated them on the IRS. The IRS items are intended to capture Tickle-Degnen & Rosenthal's (1990) three essential components of rapport: (1) mutual attention (items: *listening, tolerance, attentiveness*), (2) positivity (items: *pleasant atmosphere, absence of aggression, positivity, negativity*), and (3) coordination (items: *cooperation, like-mindedness among participants, synchrony, flexibility*). Each IRS item refers to the interaction as a whole and is rated on a 7-point Likert-type scale ranging from 1 (not at all) to 7 (very much). Item ratings are averaged to produce a single IRS score with high scores indicating higher judgements of expressed rapport. According to Bronstein et al. (2012), a factor analysis indicated that the IRS is unidimensional. However, these authors did not report the factor loadings of the items. In the same study, the IRS showed high internal consistency (Cronbach's  $\alpha = .96$ ). Relevant to construct validity, IRS scores, based on observer ratings by judges, were significantly correlated,  $r = .36$ ,  $p < .001$ , with dyad-level self-reported perceptions of rapport on the NESS (i.e. the average of both interactants' NESS scores).

### **Negotiators' Rapport Scale (NRS)**

The Negotiators' Rapport Scale (NRS) developed by Bronstein et al. (2012) is an observer rating scale highly similar to the IRS. However, whereas the IRS asks judges to rate *an interaction as a whole*, the NRS asks judges to make ratings *of the individual interactants (i.e., interviewer and interviewee)*. Judges using the NRS rate each interactant on the same 11-item, 7-point scales described above for the IRS. These observer ratings are then averaged to produce a single NRS score for each interactant, with higher scores indicating higher judgements of rapport. Factor analysis by Bronstein et al. indicated that the NRS, like the IRS, was unidimensional, although no factor loadings data were presented. Bronstein et al. reported that the NRS has high internal reliability (Cronbach's  $\alpha = .93$ ) similar to the IRS. Regarding validity,

NRS observer ratings of rapport characteristics for each interactant were found to significantly correlate with the self-reported experience of rapport by the interactant's partner, as reported on the NESS ( $r = .22, p < .05$ ).

### **Verbal Rapport Assessment Scale (VeRAS)**

Bronstein et al. (2012) developed the VeRAS tool “in order to assess the contributions of specific verbal behaviors to rapport” (p. 1097). The tool is meant to analyze indicators of rapport from observing purely verbal content (i.e., audio, text). Drawing on politeness theory (Brown & Levinson, 1987), the authors identified 13 verbal behaviors, organized into four categories, that they hypothesized would be relevant to the establishment of rapport in negotiations: (1) Behaviors related to the negotiation process (items: *negotiation facilitating acts*, *negotiation inhibiting acts*, *negotiation facilitating responses*, and *negotiation inhibiting responses*), (2) Behaviors related to the interaction process (items: *synchrony* and *asynchrony*), (3) Behaviors related to the interaction content (items: *direct positive content*, *direct negative content*, *indirect positive content*, and *indirect negative content*), and (4) Behaviors related to interpersonal relations (items: *politeness*, *positive interpersonal content*, and *negative interpersonal content*). Bronstein et al. (2012) provide very little information regarding the psychometric properties of these thirteen verbal behaviors measured by the VeRAS. The limited information provided in the article indicates considerable confusion regarding basic psychometric concepts. For instance, the article (p. 1099) includes the following statement about the scoring of verbal behaviors for the VeRAS:

In order to examine interjudge reliability, eighteen random simulations were coded by all judges, without their knowledge. Internal consistency analyses made on these simulations revealed high interjudge reliability (Cronbach's  $\alpha > .97$ ).

As may be seen, in this statement, the article (a) confused the concepts of internal consistency reliability and interrater reliability; (b) made the incoherent assertion that "internal consistency analyses.... revealed high interjudge reliability," (c) inappropriately presented Cronbach's alpha as a measure of interrater reliability, and (d) presented only a single reliability coefficient to represent the reliability of 13 different VeRAS scores.

In respect to validity, the article by Bronstein et al. (2013, p. 1102) presents the correlations of the 13 VeRAS verbal variables with interactants "level of felt rapport." Interactants' "level of felt rapport" was apparently measured by the NESS and perhaps by the NRS, although the article is somewhat ambiguous on this point. The correlation of the 13 VeRAS verbal variables with the three measures of "felt rapport" are shown in Table 28. As may be seen the 39 validity coefficients tended to be quite small, with a minimum absolute value of .01, a maximum absolute value of .21, and a median absolute value of 0.05. Furthermore, even though the verbal behaviors occasionally showed significant correlations with one of the "felt rapport" measures, these correlations were generally isolated and did not replicate from one measure of "felt rapport" to the next.



Table 28. HLM Coefficients for VeRAS behaviors on Experience (NESS) & Judgement (NRS) of Rapport (n = 280)

VeRAS behaviors	NESS		NRS	
	Individual level (df = 272)		Dyad level (df = 138)	Dyad level
	Actor behavior	Partner behavior	Actor x Partner Behavior	Actor x Partner Behavior
Behaviors related to the negotiation process				
Negotiation facilitating acts	.06	.16*	.01	.30 - .31***
Negotiation inhibiting acts	-.03	-.12*	-.02	.15 - .17***
Negotiation facilitating responses	.11	.13*	-.16*	-.19 - -.26***
Negotiation inhibiting responses	-0.05	-.15*	.04	-.18 - -.25***
Behaviors related to the interaction process				
Synchrony	.17**	.06	-.04	.13 - .14*
Asynchrony	.08	-.08	-.1	-.26***
Behaviors related to the interaction content				
Direct positive content	.13	.03	-.03	-.1-
Direct negative content	.06	-.20*	.15*	.05
Indirect positive content	.04	.02	-.02	-.02
Indirect negative content	-.05	.05	.14*	-.17**
Behaviors related to interpersonal relations				
Politeness	.03	.18*	.03	.00
Positive interpersonal content	.10	.05	.12*	.03
Negative interpersonal content	.06	-.09	.21**	.00

Note. \* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$

### Contrasting Existing Instruments

The preceding sections have described many scales that seek to measure different aspects of rapport. The following section compares these measures in respect to several characteristics and dimensions.

## **Conceptual**

### **1. Constructs measured**

First, when comparing each of these tools reviewed above, it is imperative to address the constructs that each scale is intended to measure. This is best done by first organizing each instrument's scales according to broad categories. Some instruments (e.g., RQ-2001) are comprised of single scales, while others (e.g., ORBIT) are conglomerations of several scales intended to capture various aspects of an interaction.

Table 28 below identifies each rapport-based scale (and those intended to measure synchrony or other aspects of an interaction) among 6 distinct categories: 1) interactants' experience of rapport, 2) interactants' expression of rapport, 3) interactant's rapport-building behaviors, 4) interactional synchrony, 5) interactant's cooperation, and 6) other interpersonal attributes. In the case of instruments with multiple iterations (e.g. RQ-1994, RQ-2001), the most recent published version of that instrument will be listed, unless the versions measure different constructs.

Table 29. Scales by Construct Category

Experience of Rapport	Expression of Rapport	Rapport-building behaviors	Interactional Synchrony	Cooperation	Other Interpersonal
RQ-2001	ORQ-2I	ORBIT G-MISC	CMS	RS3i CtC	RS3i Expertise
IQ	IRS	ORBIT MIDAS	SCS	ORBIT IYA	ORBIT IBC-I
NESS	NRS			ORBIT DES	ORBIT IBC-D
RS3i	VeRAS			ORBIT CITS	RS3i Cultural Similarity
Attentiveness					
RS3i Trust /					
Respect					
RS3i					
Connected					
Flow					

*Note.* RQ-2001 - Rapport Questionnaire 2001 Version (Bernieri & Gillis, 2001); IQ - Interaction Questionnaire (Vallano et al., 2015); NESS - Negotiation Emotional Satisfaction Scale (Bronstein et al., 2012); RS3i - Rapport Scales for Interrogations and Investigative Interviews (Duke et al., 2018); ORQ-2I - Two-Item Observer Rapport Questionnaire (Bernieri et al., 1996); IRS - Interaction Rapport Scale (Bronstein et al., 2012); NRS - Negotiators' Rapport Scale (Bronstein et al., 2012); VeRAS - Verbal Rapport Assessment Scale (Bronstein et al., 2012); ORBIT G-MISC - Observing Rapport-Based Interpersonal Techniques Global Motivational Interviewing Skills: Investigative Interview Adaptation (Alison & Alison, 2017); ORBIT MIDAS - ORBIT Motivational Interviewing of Detainees - Assessment of Skills Alison & Alison, 2017); CMS - Coordinated Movement Scales (Bernieri et al., 1994); SCS - Synchrony Cue Scales (Bernieri et al., 1996); RS3i CtC - RS3i Commitment to Communication scale (Duke et al., 2018); ORBIT IYA - ORBIT Interview Yield Assessment (Alison & Alison, 2017); ORBIT DES - ORBIT Detainee Engagement Scale (Alison & Alison, 2017); ORBIT CITS - ORBIT Counter Interrogation Tactics Scale (Alison & Alison, 2017); ORBIT IBC-I - ORBIT Interpersonal Behavioral Circles – Interviewer (Alison & Alison, 2017); ORBIT IBC-D - Interpersonal Behavioral Circles – Detainee (Alison & Alison, 2017).

It is then useful to consider the theoretical perspectives guiding the development of each scale designed to capture some aspect related to dyadic rapport. This information can be found summarized below in Table 30. By and large, scales intended to measure the experience or expression of rapport have drawn on the tripartite model of rapport proposed by Tickle-Degnen & Rosenthal (1990). Most of these tools employ single scales that include items intended to capture aspects of each “essential component”. Other tools draw on other sources of interpersonal theory (e.g., NESS draws on Politeness Theory), and others do not explicitly state the theoretical basis of the scales (e.g., ORQ-2I).

Table 30. Rapport Scales by Theory

TD&R "Essential 3"	Motivational Interviewing	Other / Unstated
RQ-2001	ORBIT G-MISC	ORQ-2I
IQ	ORBIT MIDAS	NESS
RS3i Attentiveness		VeRAS
RS3i Trust / Respect		
RS3i Connected Flow		
IRS		
NRS		

*Note. "TD&R 'Essential 3'" refers to the 3 "essential components" of rapport noted by Tickle-Degnen & Rosenthal (1990): Mutual Attentiveness, Positivity, and Coordination.; RQ-2001 - Rapport Questionnaire 2001 Version (Bernieri & Gillis, 2001); IQ - Interaction Questionnaire (Vallano & Schreiber Compo, 2011); RS3i - Rapport Scales for Interrogations and Investigative Interviews (Duke et al., 2018); IRS - Interaction Rapport Scale (Bronstein et al., 2012); NRS - Negotiators' Rapport Scale (Bronstein et al., 2012); ORBIT G-MISC - Observing Rapport-Based Interpersonal Techniques Global Motivational Interviewing Skills: Investigative Interview Adaptation (Alison & Alison, 2017); ORBIT MIDAS - ORBIT Motivational Interviewing of Detainees - Assessment of Skills Alison & Alison, 2017); ORQ-2I - Two-Item Observer Rapport Questionnaire (Bernieri et al., 1996); NESS - Negotiation Emotional Satisfaction Scale (Bronstein et al., 2012); VeRAS - Verbal Rapport Assessment Scale (Bronstein et al., 2012);*

The RS3i is unique in its inclusion of three separate scales to measure each of the three components of rapport proposed by Tickle-Degnen & Rosenthal (1990). The ORBIT tool also stands out among the other measures, as the theory of rapport driving development of its rapport-based scales was based in the efficacy of Motivational Interviewing (MI), a counseling style that emphasizes the role of rapport in dyadic interactions. It is then of utmost importance here to draw a distinction between the theoretical approaches taken by rapport-based scales included in the RS3i (Attentiveness, Trust/Respect, & Connected Flow) and ORBIT (G-MISC & MIDAS). Rather than directly measuring the construct of perceived rapport itself, ORBIT researchers

(Alison et al., 2013; Alison et al., 2014) compared *rapport-based skills* measured by adherence to MI principles (G-MISC) and techniques (MIDAS) in interviewer behavior to the cooperativeness of a source. The experience of rapport between the interviewer and the interviewee (as measured by the RS3i) is conceptually distinct from the behaviors that may enhance rapport (as measured by ORBIT). Further, in these studies on the ORBIT, cooperation was assessed by measuring information gain and the source's use of CITs. It may be that rapport and cooperation are also two distinct constructs, and that one can exist without the other. Coercion often results in cooperation. Indeed, Duke (2013) draws this distinction, placing rapport building and increased cooperation in two separate categories. It must be said though, that in some studies the degree of cooperation in dyadic negotiations has been found to correlate with perceived rapport (Drolet & Morris, 1995; Drolet & Morris, 2000).

## **2. Intended Context for Measurement**

Next, it is important when examining these instruments to consider the setting or context in which they are intended to be used. While the RQ-2001 and ORQ-2I are intended to measure general rapport across a variety of settings, the IQ focuses the respondent on interviewer qualities specifically in the context of investigative interactions (e.g., witness interviews, interrogations). Similarly, ORBIT and the RS3i were both developed with the intent of being employed specifically within investigative interactions and this is reflected in the behavioral and emotive channels the respondent is directed to attend to. The Bronstein et al. (2012) suite of tools offer a means of investigating rapport in adversarial interactions; however, they have not been specifically applied to investigative scenarios. Instead, the measures have been developed within the context of negotiations in the conflict resolution literature

### 3. Validation Criteria

Lastly, the means of validation differ significantly across measures. Bernieri (1988a) correlated perceptions of rapport on the RQ-1988 with observed interactional synchrony in genuine and pseudosynchronous interactions using coordinated movement measured by the CMS-1988. Bernieri et al. (1994) also used coordinated movement (using the CMS) to validate the RQ-1994 as a measure of rapport. Later research also attempted to validate the RQ-1994 using correlations with the SCS, a measure of synchrony (Bernieri et al., 1996). Overall, the RQ-2001 has shown the strongest evidence of validity of the observational measures reviewed here. It is very similar in structure to the RQ-1994 PoInteraction, and though no evidence of that subscale's validity has been demonstrated, evidence has been presented for the Global RQ-1994 as a whole. It is likely, therefore, that the RQ-2001 measures at least some aspect of rapport shared by the RQ-1994 PoInteraction. Convergent validity has also been presented for the RQ-2001, as Bronstein et al. (2012) reported that it significantly correlated with another self-report measure of rapport, the NESS. The ORQ-2I presented by Bernieri et al. (1996) has not been thoroughly validated in the published literature.

The Interaction Questionnaire (IQ) was developed by emulating versions of the RQ, though it was adapted for investigative interactions by having interactants rate the interviewer rather than themselves. Limited evidence of validity has been reported for the various versions of the IQ, though it has been somewhat validated through significant differences between experimental rapport manipulations. However, only the IQS-PoInterviewer and the IQL have been shown to differ by rapport conditions.

The RS3i has been examined in relation to rapport-based experimental manipulation as well as to source cooperation (CtC, information gain). Each RS3i scale was also compared with scales designed to measure similar and distinct constructs in order to demonstrate their

convergent and discriminant validity. Convergent validity of all five RS3i scales has been clearly demonstrated, as Duke et al. (2018) found that each scale correlated significantly with other measures intended to measure similar constructs. Convergent validity was not demonstrated for the CtC scale. Concurrent validity has been demonstrated for all of the RS3i scales except for Cultural Similarity. The Attentiveness, Trust/Respect, Connected Flow, and Expertise scales were significantly higher in rapport interviews than in no rapport interviews (Duke et al., 2018). Only the Attentiveness and Trust / Respect scales were significantly correlated with the total number of facts shared by sources as measured by the Sharing Information Rating Scale (SIRS). With regard to concern for cooperation, ORBIT and the RS3i are the more conceptually similar pair due to their shared investigative nature.

The validity evidence of the ORBIT tool varies by scale. The G-MISC has shown the strongest evidence of validity. ORBIT researchers used information gain (IYA) and increased cooperation (CITs) as criteria for their rapport-based scales. There is validity evidence showing that the ORBIT G-MISC scale was significantly related to increased information gain. The correlations for G-MISC items with IYA ranged from absolute values of  $r = .32$  to  $r = .36$  with an absolute value median of  $r = .35$ . The MI Latent variable calculated using all G-MISC scales was correlated with IYA ( $r = .39$ ). The G-MISC was also correlated with some CIT use. No significant relationship was found between G-MISC items (or the Latent MI Variable) and verbal, passive verbal, retraction, or no comment CITs. However, the correlations between G-MISC items and passive CITs were between absolute values of  $r = .28$  and  $r = .35$  with an absolute value median of  $r = .33$ . The Latent MI variable was also correlated with use of passive CITs (absolute value of  $r = .38$ ).

No evidence of validity has been presented for the MIDAS scales, and only very weak validity evidence has been presented for the IBC-I and IBC-D. The IBC-I has shown little to no relationship with IYA and CITS. The IBC-D was not compared to CIT use, but showed a negligible relationship with IYA. The G-MISC scales are certainly the most strongly validated ORBIT measures.

Lastly, Bronstein et al. (2012) specifically set out to measure the dyadic expression of rapport via the verbal channel with the VeRAS tool, but created several scales measuring experience (NESS) and expression (IRS, NRS) of rapport in the process. The authors have presented poor evidence of the VeRAS scale's validity using the relationship between verbal behaviors and these self-report and observational rapport measures (Table 30). The evidence of validity for the NESS, IRS, and NRS, is much stronger. Self-reports of rapport on the NESS were highly correlated with another self-report measure of rapport, the RQ-2001. Observational ratings of the dyad-level rapport in an interaction (IRS) were significantly correlated with observational ratings of rapport at the individual level (NRS) averaged to create a dyad-level measurement. Evidence of construct validity for the NRS was also strong, as observed ratings of rapport for each individual were found to be correlated with their partners' self-reported experience of rapport using the NESS.

The main validation criteria in studies on the RQ is interactional synchrony, which is a product of the researchers' desire to examine rapport across a variety of interactions. These findings have indicated weak validity of the RQ-1988 and RQ-1994, but the RQ-2001 was never examined in this way. Differences in rapport between experimental rapport conditions have supported the validity of the IQS-PoInterviewer and the IQL; however, extending these conclusions to the IQ is not appropriate as there are significant differences between versions. The



IQ was developed to measure rapport in investigative interviews, though its validity as a measure of rapport has not been effectively demonstrated. In contrast, the ORBIT G-MISC scale has been shown to be related to some outcomes in actual terrorist interrogations. Specifically, it has been shown to be related to information gain and the use of passive counter interrogation tactics. The RS3i scales lie between – they were validated in *simulated* interviews about terrorism but have not yet been validated in actual investigative interviews. Similarly, the Bronstein Suite of scales was designed for use in adversarial, though not investigative, interactions.

## **Mechanical**

### **1. Self-Report vs. Observer Ratings**

Measurement methods also differ substantially across instruments. The various versions of the RQ and IQ, along with the RS3i, employ self-report scales, while the ORQ-2I and ORBIT are observational rating scales. The Bronstein Suite includes mostly observational scales (IRS, NRS, VeRAS), but also includes one self-report scale as a validity criterion (NESS). The source for each scale (self-report vs. observer) can impact not only the time it takes to complete each tool, but also when researchers are able to employ the tool (during vs. after an interaction). Self-report measures such as the RS3i and RQ can be completed rather quickly at the conclusion of an interaction (perhaps even during an interaction), whereas observer ratings, such as those used in ORBIT require careful observation of entire interactions to complete and are generally more time-consuming. Observational VeRAS scales may be completed more quickly than ORBIT as they were designed to be used with written material rather than audiovisual, but the nature of the measure requires that there is considerable preparation of interview materials. However, both ORBIT and the Bronstein Suite do include scales that are less cumbersome. The G-MISC scale can be completed quickly and has demonstrated the strongest evidence of validity

among ORBIT scales. The Bronstein Suite includes three scales (NESS, IRS, NRS) that can all be completed very quickly. In fact, the three-item NESS has shown high correlation with much more involved scales (RQ-2001;  $r = .78$ ,  $p < .001$ ) and may be an appropriate substitute for longer self-report measures of rapport. However, it is important to note that the exact wording of the three NESS items was not published. Instead, Bronstein et al. (2012) provide a brief summary description of each item's purpose and content (p. 1097).

Observer rating scales can also sometimes be made more efficient by the use of "thin slice" sampling, in which observers rate only brief "slices" of time in an interaction, rather than the entire interaction. However, this approach has not yet been tried with ORBIT nor Bronstein Suite instruments. The importance of the concept of behavioral adaptability is key to ORBIT, thus observations of thin slices of interactions may not provide enough time to observe a sufficient range of behavior. Similarly, scales that rely on behavior counts (e.g., VeRAS) may not be appropriate for use in "thin slice" observation.

## **2. Global vs. Count Ratings**

Many of the rapport instruments that have been reviewed here use some combination of either counts (the frequency, magnitude, or duration of displaying a cue) or global ratings (gestalt judgements related to the expression of a construct). Scales are categorized below into global and frequency count rating schemes (Table 31). The RQ and IQ require individuals to make global ratings of specific aspects of their partners' behaviors (e.g., the degree to which a partner is *engrossed* or an interaction is *comfortable*), but do not require individual cue counts. The ORQ-2I requires observers make global ratings as to how much the interactants *like* and enjoy the *interaction*. The ORBIT tool uses both global and count methods: G-MISC is a global scale to calculate rapport-based skills, the MIDAS relies on cue frequencies, and the IBC-I&D use cue

ratings. The RS3i uses a combination of behavioral cue and global ratings. The VeRAS tool is unique among these measures, in that it requires frequency counts of individual verbal behaviors. The CMS and SCS were both developed to measure interactional synchrony though they each take a different approach. The CMS directs observers to make global ratings of the interactants' movement synchrony and posture similarity while the SCS approaches the same construct with a cue frequency strategy.

Table 31. Scales by Rating Method

Global	Frequency Count
RQ-2001	SCS
ORQ-2I	ORBIT MIDAS
CMS	ORBIT IBC-I
IQ	ORBIT IBC-D
ORBIT G-MISC	ORBIT IYA
ORBIT DES	ORBIT CITS
RS3i Attentiveness	VeRAS
RS3i Trust / Respect	
RS3i Connected Flow	
RS3i Expertise	
RS3i Cultural Similarity	
RS3i CtC	
IRS	
NRS	
NESS	

*Note. RQ-2001 - Rapport Questionnaire 2001 Version (Bernieri & Gillis, 2001); ORQ-2I - Two-Item Observer Rapport Questionnaire (Bernieri et al., 1996); CMS - Coordinated Movement Scales (Bernieri et al., 1994); IQ - Interaction Questionnaire (Vallano & Schreiber Compo, 2011); ORBIT G-MISC IIA - Observing Rapport-Based Interpersonal Techniques Global Motivational Interviewing Skills: Investigative Interview Adaptation (Alison & Alison, 2017); ORBIT DES - ORBIT Detainee Engagement Scale (Alison & Alison, 2017); RS3i - Rapport Scales for Interrogations and Investigative Interviews (Duke et al., 2018); RS3i CtC - RS3i Commitment to Communication scale (Duke et al., 2018); IRS - Interaction Rapport Scale (Bronstein et al., 2012); NRS - Negotiators' Rapport Scale (Bronstein et al., 2012); NESS - Negotiation Emotional Satisfaction Scale (Bronstein et al., 2012); SCS - Synchrony Cue Scales (Bernieri et al., 1996); ORBIT MIDAS - ORBIT Motivational Interviewing of Detainees - Assessment of Skills Alison & Alison, 2017); ORBIT IBC-I - ORBIT Interpersonal Behavioral Circles – Interviewer (Alison & Alison, 2017); ORBIT IBC-D - Interpersonal Behavioral Circles – Detainee (Alison & Alison, 2017); ORBIT IYA - ORBIT Interview Yield Assessment (Alison & Alison, 2017); ORBIT DES - ORBIT Detainee Engagement Scale (Alison & Alison, 2017).*

## Psychometric

### 1. Reliability

The internal reliability of the RQ-2001 has not been reported by the scale's author, but research has indicated it is high (Bronstein et al., 2012). Bernieri (2005) notes that intradyad correlations using the RQ-2001 have been poor (.37 in adversarial and .20 cooperative interactions). This may suggest that perception of rapport is a uniquely individual experience

even within the same interaction and even though rapport exists at the dyadic level. No reliability figures have been presented for two of the scales that accompany the RQ research (ORQ and CMS), though reliability of SCS cues has been reported as high (Chronbach  $\alpha$  > .80).

Evidence of ORBIT interrater reliability has varied by scale. Interrater reliability was generally above the accepted threshold ( $\kappa = .40$ ) for the IBC-I maladaptive subscale, but only for 3 (38%) of the adaptive subscale items. Whereas 5 (62%) of the IBC-D adaptive subscale item  $\kappa$  values were above the acceptable level, only 1 (13%) of the IBC-D maladaptive subscale items demonstrated acceptable interrater reliability. Interrater reliability for the MIDAS scale was also very weak. Two (40%) MIDAS MIIN subscale items demonstrated acceptable  $\kappa$  values, but no MIC subscale items did. No interrater reliability has been presented for the G-MISC. Neither has any evidence of internal reliability been presented for the G-MISC scales, though their factor structure indicates that these scales intercorrelate highly, and therefore a measure that combines these scales into one single score would probably have very high internal reliability. While no internal reliability has been reported for the remaining ORBIT scales, the inter-rater reliability indexed by  $\kappa$  for each scale has been variable. Internal reliability of all RS3i and Bronstein Suite scales, with the exception of the VeRAS scale itself, has been demonstrated. There is some confusion regarding the reliability of the VeRAS scale, as Bronstein et al. (2012) report Chronbach  $\alpha$  as a measure of inter-rater-reliability.

## **2. Validity**

As already noted, several different versions of the RQ, along with several RQ "composites" and subscales, have been presented over the years. Because these different versions, "composites" and subscales often differ from each other, sometimes in major ways, the

validity evidence for one RQ version or subscale does not necessarily apply to other versions or subscales. The validity evidence for the various versions, "composites," and subscales can be summarized as follows: (a) No RQ "composite" has well-demonstrated, replicated validity; (b) The validity evidence for the RQ-1994 Perception of the Interviewer subscale has not been adequately demonstrated, correlating almost exclusively with female interactant behavior and most strongly when videos of the interaction had been edited to remove most of the visual content; (c) Evidence of the validity evidence for the RQ-1994 Perception of the Interaction subscale, and the RQ-2001 which closely resembles it, has also been weak and has failed to replicate across varying contexts. However, Bronstein et al. (2012) reported that the RQ-2001 was significantly correlated with another self-report measure of rapport (the NESS), providing some evidence of the scale's convergent validity. Though the authors of the RQ-2001, as well as Bronstein et al. (2012), report that this scale is unidimensional, it is important to note that the RQ-2001's factor structure has never been reported nor has construct validity been systematically demonstrated (Bernieri, 2005). No evidence that the ORQ-2I is associated with the RQ-2001 has been demonstrated.

With regard to the observational rating scales of synchrony that Bernieri et al. (1994) and Bernieri et al. (1996) employed alongside the RQ-1994, evidence of validity is limited. While the factor structure of the CMS has been clearly indicated (Bernieri, 1988a), no evidence of convergent nor discriminant validity has been presented. The Movement Synchrony subscale has significantly differed between observations of real and fabricated interactions, though Posture Similarity has not. The validity of most SCS cues has not been adequately established, as they have not consistently been related to self-reported rapport. However, two cues (Proximity &

Synchrony) have consistently been associated with significantly higher levels of self-report rapport across cooperative and adversarial interactions.

Similar to the RQ, the IQ has taken three distinct forms: the IQ, the IQS and the IQL. Very limited information regarding these scales' validity has been published. Only the IQL has been shown to be significantly related to experimental rapport manipulation.

The demonstration of the validity of ORBIT scales has been similarly variable. Research has demonstrated the factorial validity of the G-MISC scales and supported these scales' convergent validity through significant relationships with information gain (IYA). Findings also indicate that the G-MISC is related to decreased use of passive CITs, though not verbal, passive verbal, retraction, or no comment CITs. Validity has not been examined for the MIDAS MIC and MIN subscales. Further only very weak evidence of validity has been presented for the Positive and Negative IBC-I & D subscales as they have been weakly associated with desired outcomes (Alison et al., 2013; Alison et al., 2014). No factor structure of these scales has been published, as they have been treated as observed scores rather than indicators of a latent factor in all analyses.

The factorial validity of RS3i scales have been clearly indicated. Furthermore, convergent and discriminant validity has been found to be adequate to excellent for nearly all RS3i scales (though evidence for Expertise is weaker), demonstrated by correlations with scales designed to measure similar constructs. Construct validity of RS3i scales has also generally been supported by fit with hypothesized inter-factor correlations and the significant relationship between RS3i self-reported rapport and desirable investigative outcomes (i.e., CtC, information gain).

Lastly, evidence of the Bronstein Suite scales' validity has been mixed. The global scales (NESS, IRS, and NRS) have all shown evidence of convergent validity. The self-report NESS was significantly correlated with RQ-2001 self-reported rapport. The IRS and NRS, based on observational ratings, were both significantly correlated with the NESS. IRS scores were also correlated with dyad-averaged NRS scores. Weak evidence of the VeRAS scale's validity has been presented. Few VeRAS cues were significantly related to rapport at the individual or dyad level using the NESS (Bronstein et al., 2012). Ratings of individuals on the VeRAS Behaviors Related to the Negotiation Process subscale were significantly related with the NESS rapport reported by the individuals' partners. Further evidence of this subscale's validity was demonstrated as all of its items were correlated with observed rapport on the dyad level (NRS). The VeRAS Behaviors Related to the Interaction Process subscale (items: *synchrony* and *asynchrony*) was also significantly related to NRS dyad-level rapport, but not with individual or dyad-level self-report rapport. The VeRAS Behaviors Related to the Interaction Content subscale and the VeRAS Behaviors Related to Interpersonal Relations subscale were generally not related to self-report or observed rapport at either the individual or dyad levels.



## **Appendix B. Rapport Scales for Interrogations and Investigative Interviews**

### **(RS3i)**

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 C. Negotiator Rapport Scale (NRS) Attentiveness

Consider the interview you just watched. Read each of the characteristics below and rate how much the **INTERVIEWER** displayed these characteristics on a scale from Not at all (1) to Very much (7).

**1. Listening** – How much did the interviewer demonstrate they were listening to the source? Did the interviewer pay attention to what the source had to say overall? Were the interviewer's responses appropriate to what the source previously said? Did the interviewer correctly summarize or refer to what their partner had said previously?

Listening

1    2    3    4    5    6    7

Not at all

Very Much

**2. Tolerance** – How much did the interviewer demonstrate they accepted the source and what they had to say? Did they accept the source's points of view as valid? (Remember: interactants can disagree while remaining tolerant.)

Tolerance

1    2    3    4    5    6    7

Not at all

Very Much

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

Attentiveness

1    2    3    4    5    6    7

Not at all

Very Much

## NOTES

Please use the following document to take notes while observing the interaction you are rating. You are not required to take notes but may find doing so helpful in making ratings. You may pause the video while taking notes if necessary. Notes should include relevant behaviors or observations to aid you in making ratings after you have finished viewing the interaction.

### 1. Listening

---

---

---

---

---

---

---

### 2. Tolerance

---

---

---

---

---

---

---

### 3. Attentiveness

---

---

---

---

---

---

---

## Appendix D. Negotiator Rapport Scale (NRS) Positivity

Consider the interview you just watched. Read each of the characteristics below and rate how much the **INTERVIEWER** displayed these characteristics on a scale from Not at all (1) to Very much (7).

**1. Pleasant atmosphere** – How much did the interviewer contribute to creating a pleasant atmosphere? Did the interviewer convey trust and/or respect for the source?

Pleasant atmosphere

1    2    3    4    5    6    7

Not at all

Very Much

**2. Absence of aggression** – How much did the interviewer contribute to creating a peaceful and calm (rather than hostile or argumentative) interview? Did the interviewer express their point of view without insulting, intimidating, or threatening the source?

Absence of aggression

1    2    3    4    5    6    7

Not at all

Very Much

**3. Positivity** – How much did the interviewer convey positivity toward the source during the interaction? Did the interviewer demonstrate that they trusted and/or respected the source? Did they encourage the source to participate and/or try to make the source feel accepted?

Positivity

1    2    3    4    5    6    7

Not at all

Very Much

**4. Negativity** – How much did the interviewer convey negativity toward the source during the interaction? Did the interviewer convey distrust and/or disrespect toward the source? Did they behave in a hostile manner toward the source?

Negativity

1    2    3    4    5    6    7

Not at all

Very Much

## NOTES

Please use the following document to take notes while observing the interaction you are rating. You are not required to take notes but may find doing so helpful in making ratings. You may pause the video while taking notes if necessary. Notes should include relevant behaviors or observations to aid you in making ratings after you have finished viewing the interaction.

### 1. Pleasant Atmosphere

---

---

---

---

### 2. Absence of Aggression

---

---

---

---

### 3. Positivity

---

---

---

---

### 4. Negativity

---

---

---

---

## **Appendix E. Modified Global Motivational Interviewing Skills Code –**

### **Investigative Interview Adaptation (G-MISC – IIA) Instructions**

The scoring instructions used for the G-MISC Acceptance / Empathy scale in the present thesis are copyrighted by the instrument's authors and will not be included herein. To locate these instructions please refer to the following publication:

Alison, L., & Alison, E. (2017). ORBIT: *Observation of rapport based interview techniques: Coding manual*. Liverpool: University of Liverpool.



## Appendix F. Modified G-MISC Rating Form

1. **Acceptance** - unconditional positive regard

1	2	3	4	5	6	7
Poor skill			Neutral		Positive skill	

2. **Empathy** - extent to which the interviewer understands the detainee's perspective

1	2	3	4	5	6	7
Poor skill			Neutral		Positive skill	

## NOTES

Please use the following document to take notes while observing the interaction you are rating. You are not required to take notes but may find doing so helpful in making ratings. You may pause the video while taking notes if necessary. Notes should include relevant behaviors or observations to aid you in making ratings after you have finished viewing the interaction.

### 1. Acceptance

---

---

---

---

---

---

---

---

---

---

---

---

### 2. Empathy

---

---

---

---

---

---

---

---

---

---

---

---

## Appendix G. Interaction Rapport Scale (IRS) Coordination

Consider the interview you just watched. Read each of the characteristics below and rate how much the **INTERACTION AS A WHOLE** displayed these characteristics from Not at all (1) to Very much (7).

**1. Cooperation** – How much did the interactants cooperate with each other throughout the interaction? Did they work together toward the same goal? If not, did each interactant help their partner to achieve their respective goal?

Cooperation

1    2    3    4    5    6    7

Not at all

Very Much

**2. Like-mindedness among participants** – How much were the interactants “on the same page”? Did the interactants know what to expect from each other? Did they understand each other?

Like-mindedness among participants

1    2    3    4    5    6    7

Not at all

Very Much

**3. Synchrony** – How much were the interactants “in sync”? Did the conversation exchanges switch from one speaker to the next without interruptions? Did the interactants avoid awkward silent periods? Did the interactants’ postures and physical expressions seem appropriate compared to their partner’s?

Synchrony

1    2    3    4    5    6    7

Not at all

Very Much

**4. Flexibility** – How much flexibility did the interactants convey during the interaction? Were they willing to set aside what they wanted to talk about in order to address their partner’s concerns?

Flexibility

1    2    3    4    5    6    7

Not at all

Very Much

## NOTES

Please use the following document to take notes while observing the interaction you are rating. You are not required to take notes but may find doing so helpful in making ratings. You may pause the video while taking notes if necessary. Notes should include relevant behaviors or observations to aid you in making ratings after you have finished viewing the interaction.

### 1. Cooperation

---

---

---

---

### 2. Like-mindedness Among Participants

---

---

---

---

### 3. Synchrony

---

---

---

---

### 4. Flexibility

---

---

---

---

## **Appendix H. 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](mailto:bcbollin@miners.utep.edu) and to the UTEP Institutional Review Board (IRB) at (915-747-8841) or [irb.orsp@utep.edu](mailto: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 I. Negotiator Rapport Scale (NRS) Attentiveness Comprehension

### Check

Ratings should be made based on whose behavior?

---

---

Provide a definition of **Listening** in your own words:

---

---

Rate the degree to which the following characteristics are related to **Listening** on a scale of 1 (Completely unrelated) to 5 (Highly related):

3.1 Attentive

1	2	3	4	5
Completely unrelated				Highly related

3.2 Bored

1	2	3	4	5
Completely unrelated				Highly related

3.3 Fixated

1	2	3	4	5
Completely unrelated				Highly related

3.4 Interested

1	2	3	4	5
Completely unrelated				Highly related

3.5 Aware

1	2	3	4	5
Completely unrelated				Highly related

Describe, in your own words, behaviors that indicate to you that an individual is displaying **Listening**:

---

---

---

Provide a definition of **Tolerance** in your own words:

---

---

Rate the degree to which the following characteristics are related to **Tolerance** on a scale of 1 (Completely unrelated) to 5 (Highly related):

6.1 Accepting

1	2	3	4	5
Completely				Highly
unrelated				related

6.2 Thoughtful

1	2	3	4	5
Completely				Highly
unrelated				related

6.3 Kind

1	2	3	4	5
Completely				Highly
unrelated				related

6.4 Accusatory

1	2	3	4	5
Completely				Highly
unrelated				related

6.5 Understanding

1	2	3	4	5
Completely				Highly
unrelated				related

Describe, in your own words, behaviors that indicate to you that an individual is displaying **Tolerance**:

---

---

---

---

Provide a definition of **Attentiveness** in your own words:

---

---

Rate the degree to which the following characteristics are related to **Attentiveness** on a scale of 1 (Completely unrelated) to 5 (Highly related):

9.1 Focused



	1	2	3	4	5
	Completely				Highly
	unrelated				related
9.2 Aggressive					
	1	2	3	4	5
	Completely				Highly
	unrelated				related
9.3 Pleasant					
	1	2	3	4	5
	Completely				Highly
	unrelated				related
9.4 Considerate					
	1	2	3	4	5
	Completely				Highly
	unrelated				related
9.5 Uninterested					
	1	2	3	4	5
	Completely				Highly
	unrelated				related

Describe, in your own words, behaviors that indicate to you that an individual is displaying **Attentiveness**:

---



---



---



---

## Appendix J. Negotiator Rapport Scale (NRS) Positivity Comprehension

### Check

Ratings should be made based on whose behavior?

---

Provide a definition of a **Pleasant Atmosphere** in your own words:

---

---

Rate the degree to which the following characteristics are related to a **Pleasant Atmosphere** on a scale of 1 (Completely unrelated) to 5 (Highly related):

3.1 Friendly

1	2	3	4	5
Completely				Highly
unrelated				related

3.2 Tense

1	2	3	4	5
Completely				Highly
unrelated				related

3.3 Peaceful

1	2	3	4	5
Completely				Highly
unrelated				related

3.4 Enjoyable

1	2	3	4	5
Completely				Highly
unrelated				related

3.5 Hostile

1	2	3	4	5
Completely				Highly
unrelated				related

Describe, in your own words, behaviors that indicate to you that an individual is contributing to a **Pleasant Atmosphere**:

---

---

---

Provide a definition of **Absence of Aggression** in your own words:

---

---

Rate the degree to which the following characteristics are related to **Absence of Aggression** on a scale of 1 (Completely unrelated) to 5 (Highly related):

6.1 Intimidating

1	2	3	4	5
Completely				Highly
unrelated				related

6.2 Relaxing

1	2	3	4	5
Completely				Highly
unrelated				related

6.3 Rude

1	2	3	4	5
Completely				Highly
unrelated				related

6.4 Forceful

1	2	3	4	5
Completely				Highly
unrelated				related

6.5 Sociable

1	2	3	4	5
Completely				Highly
unrelated				related

Describe, in your own words, behaviors that indicate to you that an individual is displaying **Absence of Aggression**:

---

---

---

Provide a definition of **Positivity** in your own words:

---

---

Rate the degree to which the following characteristics are related to **Positivity** on a scale of 1 (Completely unrelated) to 5 (Highly related):

9.1 Certain

1	2	3	4	5
---	---	---	---	---

Completely unrelated					Highly related
9.2 Agreeable					
1	2	3	4	5	
Completely unrelated					Highly related
9.3 Unhappy					
1	2	3	4	5	
Completely unrelated					Highly related
9.4 Genuine					
1	2	3	4	5	
Completely unrelated					Highly related
9.5 Apathetic					
1	2	3	4	5	
Completely unrelated					Highly related

Describe, in your own words, behaviors that indicate to you that an individual is displaying **Positivity**:

---



---



---

Provide a definition of **Negativity** in your own words:

---



---

Rate the degree to which the following characteristics are related to **Negativity** on a scale of 1 (Completely unrelated) to 5 (Highly related):

12.1 Argumentative					
1	2	3	4	5	
Completely unrelated					Highly related
12.2 Accepting					
1	2	3	4	5	
Completely unrelated					Highly related

12.3 Unpleasant

1	2	3	4	5
Completely				Highly
unrelated				related

12.4 Happy

1	2	3	4	5
Completely				Highly
unrelated				related

12.5 Antagonistic

1	2	3	4	5
Completely				Highly
unrelated				related

Describe, in your own words, behaviors that indicate to you that an individual is displaying **Negativity:**

---

---

---

## Appendix K. Modified G-MISC Comprehension Check

Ratings should be made based on whose behavior?

---

Provide a definition of **Acceptance** in your own words:

---

---

---

Rate the degree to which the following interviewer characteristics are related to the skill of **Acceptance** on a scale of 1 (Completely unrelated) to 5 (Highly related):

3.1 Respectful

1	2	3	4	5
Completely unrelated				Highly related

3.2 Open

1	2	3	4	5
Completely unrelated				Highly related

3.3 Judgmental

1	2	3	4	5
Completely unrelated				Highly related

3.4 Warm

1	2	3	4	5
Completely unrelated				Highly related

3.5 Neutral

1	2	3	4	5
Completely unrelated				Highly related

3.6 Disapproving

1	2	3	4	5
Completely unrelated				Highly related

3.7 Positive

1	2	3	4	5
Completely unrelated				Highly related

3.8 Constructive

1	2	3	4	5
Completely unrelated				Highly related

3.9 Confrontational

1	2	3	4	5
Completely unrelated				Highly related

3.10 Supportive

1	2	3	4	5
Completely unrelated				Highly related

Describe, in your own words, interviewer behaviors that indicate to you that they are displaying high skill in **Acceptance**:

---

---

---

Provide a definition of **Empathy** in your own words:

---

---

Rate the degree to which the following interviewer characteristics are related to the skill of **Empathy** on a scale of 1 (Completely unrelated) to 5 (Highly related):

6.1 Critical

1	2	3	4	5
Completely unrelated				Highly related

6.2 Understanding

1	2	3	4	5
Completely unrelated				Highly related

6.3 Reflective

1	2	3	4	5
Completely unrelated				Highly related

6.4 Condescending

1	2	3	4	5
Completely unrelated				Highly related

6.5 Interested

1	2	3	4	5
Completely				Highly
unrelated				related

6.6 Insightful

1	2	3	4	5
Completely				Highly
unrelated				related

6.7 Communicative

1	2	3	4	5
Completely				Highly
unrelated				related

6.8 Negative

1	2	3	4	5
Completely				Highly
unrelated				related

6.9 Rigid

1	2	3	4	5
Completely				Highly
unrelated				related

6.10 Thoughtful

1	2	3	4	5
Completely				Highly
unrelated				related

Describe, in your own words, interviewer behaviors that indicate to you that they are displaying high skill in **Empathy**:

---

---

---

---



## Appendix L. Interaction Rapport Scales (IRS) Coordination Comprehension

### Check

Ratings should be made based on whose behavior?

---

---

---

Provide a definition of **Cooperation** in your own words:

Rate the degree to which the following characteristics are related to **Cooperation** on a scale of 1 (Completely unrelated) to 5 (Highly related):

#### 3.1 Respectful

1	2	3	4	5
Completely unrelated				Highly related

#### 3.2 Hostile

1	2	3	4	5
Completely unrelated				Highly related

#### 3.3 Shared goals

1	2	3	4	5
Completely unrelated				Highly related

#### 3.4 Trusting

1	2	3	4	5
Completely unrelated				Highly related

#### 3.5 Intimidating

1	2	3	4	5
Completely unrelated				Highly related

Describe, in your own words, behaviors that indicate to you that an interaction is high in **Cooperation**:

---

---

---

---

Provide a definition of **Like-mindedness Among Participants** in your own words:

---

---

Rate the degree to which the following characteristics are related to **Like-mindedness Among Participants** on a scale of 1 (Completely unrelated) to 5 (Highly related):

6.1 Communicative

1	2	3	4	5
Completely unrelated				Highly related

6.2 Agreement

1	2	3	4	5
Completely unrelated				Highly related

6.3 Cooperative

1	2	3	4	5
Completely unrelated				Highly related

6.4 Incompatible

1	2	3	4	5
Completely unrelated				Highly related

6.5 Amicable

1	2	3	4	5
Completely unrelated				Highly related

Describe, in your own words, behaviors that indicate to you that an interaction is high in **Like-mindedness Among Participants**:

---

---

---

---

Provide a definition of **Synchrony** in your own words:

---

---

---

---

Rate the degree to which the following characteristics are related to **Synchrony** on a scale of 1 (Completely unrelated) to 5 (Highly related):

9.1 Awkward  
 1      2      3      4      5  
 Completely      Highly  
 unrelated      related

9.2 Uncooperative  
 1      2      3      4      5  
 Completely      Highly  
 unrelated      related

9.3 Coordination  
 1      2      3      4      5  
 Completely      Highly  
 unrelated      related

9.4 Peaceful  
 1      2      3      4      5  
 Completely      Highly  
 unrelated      related

9.5 Smooth  
 1      2      3      4      5  
 Completely      Highly  
 unrelated      related

Describe, in your own words, behaviors that indicate to you that an interaction is high in **Synchrony**:

---

---

---

---

Provide a definition of **Flexibility** in your own words:

---

---

---

Rate the degree to which the following characteristics are related to **Flexibility** on a scale of 1 (Completely unrelated) to 5 (Highly related):

12.1 Rigid  
 1      2      3      4      5  
 Completely      Highly  
 unrelated      related

12.2 Flexible  
1 2 3 4 5  
Completely Highly  
unrelated related

12.3 Authoritative  
1 2 3 4 5  
Completely Highly  
unrelated related

12.4 Adaptive  
1 2 3 4 5  
Completely Highly  
unrelated related

12.5 Hostile  
1 2 3 4 5  
Completely Highly  
unrelated related

Describe, in your own words, behaviors that indicate to you that an interaction is high in  
**Flexibility:**

---

---

---

---

---

## 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 and began undergraduate studies at Arkansas Tech University in Fall 2006. He earned a Bachelor of Arts degree in Sociology in 2011. Justin then started a Master of Science program in Intelligence and National Security at the University of Texas at El Paso in Spring 2014, where his research focused on law enforcement and national security investigative interviews and interrogations. After completing his M.S. INSS in Fall 2015, Justin continued pursuing this research and began a Master of Arts program in Psychology at UTEP, which he completed in Fall 2018. He is currently pursuing his Ph.D. in Psychology at UTEP. Justin has co-authored an article on the effectiveness of Army Field Manual interrogation approaches published in *Law and Human Behavior*.

Contact Information: [jlmagee@utep.edu](mailto:jlmagee@utep.edu)