1-1-2014

The Reliability And Validity Of The Thin Slice Technique: Observational Research On Video Recorded Medical Interactions

Tanina Suzanne Foster
Wayne State University,

THE RELIABILITY AND VALIDITY OF THE THIN SLICE TECHNIQUE: OBSERVATIONAL RESEARCH ON VIDEO RECORDED MEDICAL INTERACTIONS

by

TANINA S. FOSTER

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2014

MAJOR: EVALUATION AND RESEARCH

Approved by:

____
Advisor Date

____

____

____

____

DEDICATION

I dedicate this dissertation to:

My grandmother, Mary Eileen Hall. Your love and support throughout this process has been immeasurable. You are a symbol of strength and endurance and I cherish every moment together.

In memory of my father, David Charles Foster. I know you saw this day coming LONG before I did, thank you for believing in me.

In memory of Dr. Gail Fahoome. Your knowledge and teachings from the Masters program through the Doctoral program have provided me with the skills necessary to continue this journey and see it through to completion.

In memory of Dr. Nalini Ambady. A recognized leader in research using the thin slice technique who provided a foundation for this dissertation. Thank you!

ACKNOWLEDGEMENTS

I wish to formally acknowledge the following individuals who have been a part of this process and have provided support and encouragement through my entire academic and research career. First, I would like to thank my dissertation committee. My advisor and committee chair, Dr. Shlomo Sawilowsky, for providing guidance, expertise and patience during this journey. Dr. Irwin Jopps, Dr. Cathryn Bock and the late Dr. Gail Fahoome for agreeing to serve on my dissertation committee. Dr. Barry Markman for stepping in for Dr. Fahoome during the final stages of this project. And a special thank you to Dr. Louis Penner, who not only agreed to serve on my committee, but also provided the observational data used for this study. I thank you all for your time, expertise and helpful suggestions that have continued to enhance and strengthen the quality of my dissertation.

I would like to thank my colleagues, Dr. Terrance Albrecht and Lisa Berry-Bobovski, for their continued support and encouragement; and allowing me to utilize space in the Behavioral and Field Research Core to conduct training and coding sessions. I would also like to thank the following students at Wayne State University for helping to code the observational data used in the analysis: Samantha Cloutier, Jamie Hoffman, Jonathan Kama, Erica Montgomery, Jasminder Phalore, Mary Shultz and Christopher Yoo.

Finally, I would like to thank my family and friends who provided encouragement and moral support, always believing in me even when I may not have believed in myself. I am truly thankful for each and every one of you. I hope you continue to reach for your dreams, no matter how difficult the path!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

The Reliability and Validity of the Thin Slice Technique: Observational Research on Video Recorded Medical Interactions

Chapter 1

Introduction

Psychometric properties of data collection instruments, including reliability and validity, are critical to the research process. Reliability is a psychometric measurement property assessing the general ability to reproduce results consistently over multiple iterations or by independent observational ratings (Cook & Beckman, 2006; Kundle & Polansky, 2003). Validity refers to a psychometric property assessing the precision of accuracy in the interpretation of results, or more pointedly, if the results have meaning (Cook & Beckman, 2006). Different types of validity include construct validity (are measures or variables operationally defined), concurrent validity (are measures correlated with previous validated measures), convergent validity (are two or more measures that should be related, actually related) and predictive validity (do measures predict some future measure or outcome).

Only when reliability and validity of methods are established in the research process can the research community be certain research findings are appropriate, relevant and meaningful. Therefore, reliability and validity are vital aspects of data collection instruments and consequently research outcomes are dependent on them. As research is published, it helps to inform practice, attitudes and beliefs, and can lead to intervention development and further research based on reported findings. Research results can also assist in defining future research priorities, (Hootman, Driban, Sitler, Harris, & Cattano, 2011) making it imperative to ensure reliability and validity in selected data collection instruments.

Observational Research

Observational research is a type of non-experimental research that is used to study behavior. Observational research, in this context, involves observing human interaction and rating the behaviors observed based on a predefined coding scheme. Observational research methods are frequently used in the behavioral and social sciences to collect and analyze data on human behavior. Mays and Pope (1995) described observational research methods as the systematic, detailed observation of communication behavior, observing and recording how people communicate (verbally and non-verbally). Interpersonal interaction can be observed and measured using the scientific techniques of observational research (Mays & Pope, 1995).

One technique of observational research is the coding or rating of behavior in pre-recorded video interactions. When utilizing video recordings of real interactions, researchers are allowed to return to the video to code and analyze behavior following the interaction. This method of observational analysis permits the opportunity to refine coding scheme(s) and assess reliability through independent coding and multiple reviews of the video recording (Murphy, 2005).

Another method of observational analysis using video recorded interactions is the thin slice technique. Thin slices are small pieces of an interaction, extracted and edited from a longer behavioral interaction. Depending on the research question and analysis plan, the editing process may be limited to specific extraction (i.e. cutting video to a specific time identified) or may involve advanced methods, such as the blurring of faces to protect participant identities. The resulting video clips are subsequently coded for targeted behaviors and the results are generalized to the interaction as a whole. Research has been conducted using the thin slice technique in various areas of interest, including criminal justice, education, advertising/marketing, and medical/health sciences (Ambady, Koo, Rosenthal, & Winograd, 2002a; Ambady, Krabbenhoft, & Hogan, 2006; Ambady et al., 2002b; Ambady & Rosenthal, 1993; Grahe & Bernieri, 1999; Lippa & Dietz, 2000; Murphy, 2005; Murphy, Hall, & Colvin, 2003; Peracchio & Luna, 2006).

Murphy (2005) defined the thin slice as an excerpt of behavior sampled from a longer stream of consistent behavior. Before this methodology was developed and tested, observational research was conducted utilizing full interactions, sometimes ranging over an hour long. Research conducted using the full interaction has been demonstrated to be costly, time consuming and resource intensive (Ambady, LePlante, & Johnson, 2001; Ambady & Rosenthal, 1997; Murphy, 2005). Murphy estimated each behavior coded could take up to two passes per interaction. Therefore, if an interaction were 45 minutes in length, it would take an hour and a half to code one interaction for one behavior. If the dataset included 100 video recorded interactions, averaging 45 minutes in length, this would equate to approximately 150 hours to code one behavior.

As research usually involves the coding of more than one behavior, time and resources can quickly become exhausted. In addition, the calculation does not include the time invested in training or the time invested in establishing reliability between coders (Murphy, 2005). This identified a need to develop an alternative, resource effective, way to capture data from behavioral interactions and subsequently led to the development of the thin slice method, or coding brief segments of an interaction as a representative of the whole interaction. This method was developed to address the challenge of using observational research to answer important research questions that could be gleaned from pre-existing video recorded interactions.

Thin slices, derived from the longer video recorded interactions, can range from ten second slices to slices as long as five minutes. Research has previously demonstrated that slices can be coded reliably and can accurately predict subsequent behavior (Ambady, Hallahan, & Conner, 1999; Ambady et al., 2002a; Ambady et al., 2006; Ambady et al., 2002b; Ambady et al., 2001; Ambady & Rosenthal, 1993; Grahe & Bernieri, 1999; Kraus & Keltner, 2009; Lippa & Dietz, 2000). Recently, the thin slice technique has been used to rate rapport in video recorded interactions with medical students and standardized patients. Standardized patients are actors who are given a script to perform as a patient presenting pre-defined symptoms; they are used primarily in training medical students. Results of this study indicated that thin slice measurements, specifically in the case of standardized patients, (1) correlated highly across slices, demonstrating convergent validity, and (2) were able to predict subsequent behavior, demonstrating predictive validity (Roter, Hall, Blanch-Hartigan, Larson, & Frankel, 2011). Research has not been published however, demonstrating the reliability of coding thin slice judgments and validating this methodology within a sample of actual patient/physician encounters.

Racial Disparities in Medical Outcomes

Racial/ethnic health disparities, specifically in medical outcomes, continue to be a problem across multiple diseases. Those who self-identify as Black or African American routinely have comparatively poorer health outcomes across diseases including cardiovascular disease, diabetes, cancer and stroke (Centers for Disease Control, 2005; Smedley, Stith, & Nelson, 2003). Multiple factors contribute to poorer health outcomes including socioeconomics, cultural practices and access to healthcare services (Braveman, Cubbin, Egerter, Williams, & Pamuk, 2010; Copeland, 2005; Kirby, Taliaferro, & Zuvekas, 2006; LaVeist, Thorpe, Galarraga, Bower, & Gary-Webb, 2009; Quinn et al., 2011; Smedley et al., 2003; Tian, Goovaerts, Zhan, Chow, & Wilson, 2012).

Research has indicated that racial/ethnic health differences in outcomes can be partially explained by differences in communication patterns between the physician and patient during the medical interaction and, from the patient perspective, a general distrust of healthcare including distrust of the healthcare team (physicians, nurses, etc.) and the healthcare system (Casagrande, Gary, LaVeist, Gaskin, & Cooper, 2007; Dovidio et al., 2008; Eggly et al., 2011; Penner et al., 2009; Sheppard, Zambrana, & O'Malley, 2004; Thrasher, Earp, Golin, & Zimmer, 2008). Studies have also suggested that unintentional bias, or bias that an individual is unaware or unconscious of, can lead to poorer communication patterns, less patient satisfaction, lower adherence to recommendations and ultimately affect short and long term health outcomes (Dovidio et al., 2008; Richeson & Shelton, 2005; Stepanikova, 2006; Stepanikova, Mollborn, Cook, Thom, & Kramer, 2006).

Medical Interactions: Background on Bias, Satisfaction and Outcomes

Building on the need to understand health care interactions and associated outcomes, a study was recently conducted to assess variables associated with patient care, satisfaction, health outcomes and both implicit and explicit bias in health care

(Penner et al., 2009; Penner et al., 2010). Participants in the study included African-American patients seen by participating healthcare providers practicing at a Midwestern urban primary care clinic. During data collection, both the patient and physician completed a pre-interaction questionnaire, and the medical interaction was video-recorded. The video recordings were captured using a portable system with digital processing technology that allows simultaneous recording of both the patient and physician during the interaction. Following the interaction the patient and physician completed a post interaction survey.

Clinic visits were video-recorded using a unique method developed for real time medical interaction research (Albrecht et al., 2005), which was demonstrated to be non-intrusive in the medical setting (Penner et al., 2007). The video capture system included high-resolution, digital video cameras with wide-angle lenses housed in custom made cylinders with external microphones, one camera capturing the physician and one camera capturing the patient simultaneously within the interaction. The system was controlled remotely with tilt/pan/zoom capabilities and monitored (real time) from another private, secure location in the clinic. Once captured, the recordings were processed and edited with AVID Media Composer software, resulting in a single image split screen encompassing the patient/physician medical interaction. This resulting file was converted to MPEG format for subsequent coding and analysis. This technology is currently being used to collect data in multiple nationally funded research projects by the National Cancer Institute, involving real time medical interactions at Karmanos Cancer Institute, Children's Hospital of Michigan, St Jude Children's Research Hospital and the Josephine Ford Cancer Institute at Henry Ford Health System.

The video recorded interactions collected as part of Penner and colleagues' study of primary care interactions (N=133) were further analyzed to assess patient-physician communication during the medical interaction, specifically when pain was discussed. The results of the sample selected (n=113) indicated the majority (69%) of interactions included some discussion of pain, and 63% of participants indicated their pain was moderate to severe (Henry & Eggly, 2012). These results led to further evaluation and assessment of how pain related discussions affect the quality of patient/physician communication and the patient/physician relationship in the primary care setting using the thin slice technique.

Henry and Eggly (2013) extracted three thirty second slices from the beginning, middle and end of the full video recorded interactions in which pain was discussed (n=85) and two thirty second slices from the beginning and end of the full video recorded interactions in which there was no discussion of pain (n=48). Slices were randomized and coded by independent research assistants using elements of the Roter Interaction Analysis System (RIAS) developed for capturing information and analyzing didactic relationships (Roter & Larson, 2002). Specifically, thin slices were rated on a nine point Likert scale ranging from "no" to "high" on variables including rapport, trust, liking, attention and coordination. Based on the results of this study, Henry and Eggly found no evidence to support the relationship between discussions about pain and patient-physician rapport during medical interactions. However, the study results did suggest an association between discussions about pain and an increase in patient unease and patient positive engagement, also assessed using the thin slice technique, compared to other topics of discussion (Henry & Eggly, 2013).

Purpose of the study

It is well known in psychometric research that reliability can be demonstrated without establishing validity, however research cannot be valid without establishing reliability. As previous research has demonstrated the reliability of the thin slice technique, this research will attempt to replicate these results. Additionally, although the thin slice technique is routinely incorporated in observational research methods, there is little empirical evidence supporting or validating this technique in rating behavior in lieu of rating behavior on the full interaction. Furthermore, no literature has been found testing the validity of this method in an actual physician/patient interaction in an urban medical setting serving low income African Americans.

The purposes of the study are (1) to determine if thin slices sampled from the beginning, middle and end of an actual medical interaction in an urban medical setting serving low income African Americans can be reliably coded by independent raters using a validated coding system (Price, Windish, Magaziner, & Cooper, 2008; Roter & Larson, 2002), (2) to determine if the rating of the three slices obtained are associated with each other demonstrating convergent validity over time, and (3) to determine if the ratings of the three slices obtained are associated with ratings from the whole interaction demonstrating construct validity.

Research Aims

- Aim 1: To determine if independent raters can reliably code relational variables (liking, attention, coordination, trust and rapport) between patients and physicians using thin slices sampled from an actual patient/physician interaction.

- Aim 2: To determine if there is a significant difference in ratings between the first, second and third slice of the interaction.
- Aim 3: To determine if there is a significant difference in thin slice ratings compared to ratings from the interaction as a whole.

Assumptions and Limitations

This study was based on the assumption that data collected in previous research (real-time video capture) was an unbiased, representative random sample of the clinic population during the time of data collection (June 2006-February 2008). In addition, it is assumed that the homogeneity of the patient population and the physician population did not affect the results obtained and that the process of video capture did not affect physician/patient communication patterns.

Three limitations have been identified and are acknowledged for this study: 1) data was collected in one urban clinic population, 2) data was collected during a specific timeframe and 3) this is a secondary data analysis using previously collected video-recorded data, these limitations limit the generalizability of results outside these parameters.

Definition of Terms

Convergent Validity	A validity measurement that refers to the degree that multiple variables or ratings correlate or converge on the same construct. For example, the degree to which thin slices sampled at the beginning of an interaction correlate with thin slices sampled in the middle and the end of the interaction.

Full Interaction	A complete video recorded interaction capturing the verbal and nonverbal behavioral patterns of participants. In this context, the entire patient/physician medical encounter.
Predictive Validity	A validity measurement that refers to the ability of the measure to predict a future behavior or outcome. For example, a test of the thin slice ratings to predict ratings obtained from a full interaction.
Thin Slice	A sample or brief segment of video recorded behavior obtained from the full interaction or larger behavioral stream (Murphy, 2005).

Chapter 2

Literature Review

Observational Research

Psychology, sociology, medicine, and other areas of research have benefited from observational research methods. Directly observing, monitoring and analyzing human interaction helps to understand reality in context, and allows researchers and specialists to develop and test appropriate training and/or interventions to improve such outcomes as quality of care and patient satisfaction.

In the medical field specifically, researchers using observational research methods have increased knowledge on how parental styles and behaviors can impact child coping during painful pediatric procedures (Cline et al., 2006; Penner et al., 2008; Peterson et al., 2007) and understanding the communication process and how it may contribute to differences in health care behavior that may lead to poorer patient outcomes (Beck, Daughtridge, & Sloane, 2002; Eggly et al., 2011; Penner et al., 2009). In addition, this research method has been utilized by researchers to gain insight into how trust in a medical interaction can influence patient satisfaction, adherence to medical recommendations and ultimately impact overall medical outcomes (Albrecht et al., 2008; Eggly et al., 2008; Fiscella et al., 2004; Penner et al., 2009).

Video recording methods have been utilized to capture data and allow subsequent observation and coding to be done at a later time. This technique allows many research questions to be addressed as well as the opportunity to review the interaction to look for multiple behaviors and develop appropriate coding schemes (Murphy, 2005). However, as mentioned, observational research can be time and resource intensive. Coding behavior during lengthy video recorded interactions can take many hours to reach acceptable reliability depending on the nature of the question being asked and then may take multiple passes through the video to capture information needed to address research questions (Murphy, 2005). This problem has directly led to the rationale for the thin slice technique: that brief samples taken from the interaction may be representative of the interaction as a whole, thus saving significant time and resources in the observational coding process.

Thin Slice Research

Research using thin slices has been documented throughout the literature over the past 20 years. Relevant studies and findings are presented to demonstrate the current knowledge of utilization, reliability and validity of this observational technique. Ambady and Rosenthal (1992) conducted a meta-analysis of studies that utilized thin slices. Thirty-eight independent studies, ranging from 1970-1990, were reviewed that indicated a significant result in predicting behavior. Findings indicate that all studies included in the analysis showed a positive effect size for accuracy in predictions with a significant overall mean effect size (r=. 39) and significant mean effect sizes in the following categories: clinical psychology (r=. 41), social psychology (r=. 47) and deception (r=. 31).

Following the findings of their meta-analysis, Ambady and Rosenthal (1993) explored the feasibility of making accurate judgments based on minimal information obtained from observations of teachers teaching in the classroom. They conducted a series of studies to determine if behavior can be accurately coded using a very brief exposure (10-30 sec) to an interaction. The criterion used to establish the validity of this method included using the ratings obtained to predict future behavior. In the first study, the video recorded participant sample included 13 teachers who were graduate teaching fellows that taught in the Teaching Laboratory at Harvard University, seven of the 13 teachers (53%) were male. The sample was selected out of a larger database of teaching fellows video recorded for instructional purposes and feedback as part of their training program. Participants agreed to be video recorded teaching an undergraduate course that included a diverse curriculum and lasted approximately an hour. The sample, used to test thin slices, was selected by the members of the Teaching Center. Members of the Teaching Center who were asked to select a wide range of teacher effectiveness, basing their decision on the average measure overall. (Ambady and Rosenthal, 1993).

Ambady and Rosenthal (1993) selected thin slices as follows: 10 sec from the first 10 minutes of the interaction, 10 sec from the middle of the interaction and 10 sec from the last 10 minutes of the interaction. All slices were chosen with the teacher as the focus (no students in the clip). Nine female students were then asked to rate these brief samples (three 10 sec clips from each of the 13 participating teachers) on nonverbal behavior focusing on fifteen dimensions (e.g. accepting, competent, likable, professional, supportive, etc.). Ratings ranged from 1 (not at all) to 9 (very) on all dimensions.

Reliability between judges was computed using intra-class correlations. The effective reliabilities reported ranged from .60 - .89. Because the means were inter-correlated, a principal component analysis was completed which yielded a single composite variable. The composite variable included all variables with the exception of anxiety, which was removed due to low reliability on this dimension. Results were then compared to the end of semester student evaluations of the participating teachers after a completed semester of coursework. Findings showed that ratings of brief segments completed by independent raters who were unfamiliar with the course or the teacher correlated highly with end of semester student evaluations. The global composite score yielded a significant correlation of .76 (Ambady & Rosenthal, 1993).

A follow-up study conducted by Ambady and Rosenthal (1993) was done to attempt to replicate findings from their previous study. The sample for the second study included a new sample of 13 high school teachers, five of whom were male (38%). Video recorded participants were consented and agreed to participate in the study. The criterion variable used for this study was supervisory, specifically principal performance ratings of the teachers. Thin slices were sampled using the same procedure as the previous study (10 sec from beginning with no student in clip, 10 sec from middle with no student in clip, 10 sec from end with no student in clip). Eight independent raters (all female) rated the identical 15 dimensions described in the first study. Similar to the first study, a composite variable was identified that included 14 of the dimensions coded (again anxiety was dropped from the final measure).

Results of this comparison showed that brief excerpts of a longer interaction coded by independent blind raters correlated with supervisory ratings of performance ($r=.68$). Conclusions drawn from these two studies indicate that thin slices can be rated accurately and can be validated using future evaluation outcomes as predictors of construct validity (Ambady & Rosenthal, 1993).

In a third study, Ambady and Rosenthal (1993) examined the length of the thin slice to determine if different lengths influenced reliability and/or validity. To test this assumption, 5 sec slices and 2 sec slices were created by randomly selecting excerpts from the previous 10 sec slices used for the first two investigations. Videos were created based on each sample (5 sec clips and 2 sec clips). Thirty-two female undergraduates were recruited to rate the video clips (8 raters per video). Results of this third study were similar to their first two studies. Mean overall effect size reported was significant ($r=.59$) and no significant difference was found based on length of clips (5 seconds vs 2 seconds).

Ambady and Rosenthal (1993) acknowledged the implications of their findings and the impact specifically on the field of education, such as identifying the importance of teacher affect and nonverbal behavior on teaching satisfaction and performance ratings and suggest that appropriate training on affect and nonverbal behavior may be beneficial to educators. Generalizing their results to a wider population, it is easy to see how this methodology can be useful in predicting behaviors and outcomes in other areas of interest, specifically medical interactions where interpersonal communication patterns can affect medical decisions and ultimately health related outcomes.

Ambady et al. (2002b), continued research on thin slices of behavior using audio recorded interactions between patients and community surgeons, attempting to replicate results previously reported on thin slice technique outside of the field of education. Using the thin slice technique, Ambady and colleagues conducted a secondary analysis of previously collected data and attempted to determine if thin slices of behavior could be coded accurately and ultimately predict a future behavioral outcome, predicting instances of malpractice litigation (Ambady et al., 2002b). Sixty-five surgeons participated in the original study and an average of 10 patients per physician were recruited and consented for their medical interaction to be audio recorded (Levinson, Roter, Mullooly, Dull, & Frankel, 1997). For this secondary analysis only two interactions were used per physician (the patient with the highest self-reported satisfaction and the patient with the lowest self-reported satisfaction). Eight physicians were excluded from final analysis due to quality of audio recording or quality of the patient interaction (i.e. patient didn't participate in discussion) leaving a sample size of 57 surgeons in 114 interactions. Thin slices were extracted from the interactions (10 sec from the first minute of the interaction and 10 sec from the last minute of the interaction) resulting in 228 audio clips. These clips were further edited to remove recognized speech, leaving only intonation, speed, pitch and rhythm creating an additional 228 audio clips for review which included audio clips with recognized content and audio clips with unrecognized content (Ambady et al., 2002b).

Independent raters were assigned to each condition, thin slices with content and thin slices without content (tone only). Thin slices were rated on a seven point Likert scale ranging from "not at all" to "extremely" on ten variables including warmth, interest, professionalism, competence, dominance, satisfaction and genuine behavior. A principle component analysis using varimax rotation resulted in four distinct categories: 1) warm/professional, 2) concern/anxious, 3) hostile and 4) dominant. The hostile category, however was dropped from final analysis as it demonstrated the lowest reliability and correlated highly with anxiety in both conditions (content r=.72; tone r=.83). Logistic regression was then used to determine if results from thin slice coding predicted future malpractice claims. Results indicated that surgeons who were rated high on dominance and low on concern/anxiety were more likely to have malpractice claims after controlling for speech content (Ambady et al., 2002b). These results demonstrated the significance of information obtained from thin slices of a full interaction; however does not take into account information that could be gleaned from the full interaction. Specifically, if ratings obtained from thin slices sampled from a larger behavioral stream correlate positively with ratings obtained from the full interaction.

Murphy, Hall and Colvin (2003) used thin slices to determine if raters can accurately assess intelligence from a brief exposure to a stranger and if gender influences accuracy of judgments. The study sample consisted of both video/audio recorded subjects and independent raters. Subjects were assessed on intelligence measures (IQ, grade point average and SAT score) and discussions between subject pairs were video recorded for five minutes and sessions were transcribed for analysis. The second full minute of interaction was extracted for the thin slice ratings. Five minute discussions were then randomized to one of three categories for rating purposes – audio/video, video only and transcript (Murphy et al., 2003).

Forty-four interactions were video recorded yielding 88 target/pairs (subjects/raters). A composite score on intelligence was obtained by averaging the z-score for each measure (IQ, GPA and SAT score). Subjects missing two of the three measures were dropped from further analysis yielding 79 target/pairs in the sample to be coded. Subjects were rated for intelligence measures (IQ, GPA, SAT score) by 415 undergraduate psychology students (124 male, 291 female) who were randomly assigned to each measure and condition. Raters were trained by describing the assigned rating measure, providing the average score for college students on that measure and reviewing the range of all possible scores on the measure. A composite score was calculated for each subject in each condition similar to the actual composite intelligence score by averaging the z-score for each measure. Results between subjects' measured intelligence and perceived intelligence correlated significantly in audio/video ($r=.37$) and video only ($r=.23$) conditions, however was not significant in the transcript only condition ($r=.04$). Results indicate that ratings of perceived intelligence, using one minute thin slices from video recorded interactions, significantly predicted measured intelligence (Murphy et al., 2003). This study provides further evidence that brief expose or thin slices of video recorded interactions can be used to accurately predict specific outcomes.

Thin Slices vs. Full Interaction

Murphy (2005) recognized the need to establish the validity of the thin slice technique in relation to the full behavioral interaction. Specifically, the examination of ratings obtained from thin slices of an interaction and determining if thin slices can be substituted for full length interactions should explore measuring the same variables on each length and conducting a statistical comparison of the results obtained. In the meta-analysis conducted by Ambady and Rosenthal (1992), the average reported effect size was $r=.39$. Murphy argued that while this is acceptable for predictive validity, comparing thin slices to the larger behavioral stream should yield a stronger reliability coefficient for justifying this methodology (Murphy, 2005).

Murphy conducted a study comparing shorter excerpts to longer interactions. Fifty undergraduates were consented to be video recorded while participating in a group exercise (two participants at a time) lasting approximately 15 minutes. Thin slices, three 60 sec slices, were randomly selected using a random number generator from the larger video recorded interaction and edited to be additive in nature (i.e. the first slice included only the first randomized 60 second slice, the second slice combined the first and second randomized slice yielding a 2 minute length clip, and the third slice combined all three randomized slices yielding a 3 minute length clip). Thin slices obtained and the full interactions were subsequently coded on five behaviors: number of gestures, nods, self-touches, smiles and time spent gazing at partner. Independent raters were used for each condition (length of slice). Reliability was calculated using Pearson's r. Reliability for thin slices ranged from r=.83 to r=.99; reliability for the full interaction ranged from r=.61 to r=.95 (Murphy, 2005).

Results were stratified into two categories, results from comparison when the thin slice was included in the larger interaction and results from comparison when the thin slice was removed from the interaction. The first scenario, when the thin slice was included, resulted in high positive correlations across four of the five behaviors (gesture, gaze, nod, smile). The second scenario, when the slice was removed, resulted in variability across the three slices using the part-whole correlation formula (Cohen and Cohen, 1983, cited by Murphy, 2005). Ratings obtained from the first slice (1 minute of interaction) resulted in high positive correlations with four of the five behaviors (gesture, gaze, nod, and smile). Ratings obtained from the second slice (2 minutes of interaction) resulted in high positive correlations with all five identified behaviors. Ratings from the third slice (3 minutes of interaction) resulted in high positive correlations with four of the five behaviors (gesture, gaze, self-touch, smile). Findings from this study demonstrated that ratings of thin slices can correlate with the larger interaction, providing support for using this observational technique. However, no research has been found in a published literature search to demonstrate these findings can be replicated using actual patient/physician video recorded medical interactions.

Roter et al. (2011) examined the relationship between thin slice segments and the full interaction as well as predictive validity of slices to ratings of rapport in a simulated medical setting. Participants included the entire third year class of medical students at a Midwestern school of medicine (N=253). As part of the training program, students participate in a simulated patient examination known as an objective structured clinical examination or OSCE. These sessions were video recorded and analyzed using the RIAS coding scheme. Thin slices, 60 seconds, were extracted from the full interaction at three time points: 1 minute, 5 minutes and 9 minutes. Relational rapport, between students and simulated patients, was rated by trained research assistants on a subset of the interactions (n=141) using a nine point Likert scale ranging from no rapport to high rapport. Multivariate analysis and correlations were used to analyze differences in ratings. Findings indicate ratings of one-minute slices correlated with the ratings of the full interaction and global affect showed a consistent pattern of relationship (Roter et al., 2011). Results of this investigation indicated both concurrent (thin slices correlation with full interaction) and predictive validity (thin slices prediction of global affect). The limitation of this study, however is that the research was conducted in a simulated setting and it is unknown if results can be generalized to another population including actual patient/physician medical encounters.

Summary of Findings/Next Steps

Research described consistently provides support and demonstrates that thin slices can be reliably and accurately coded and can be used to predict behavior. Murphy further demonstrated convergent validity by comparing thin slices to the larger interaction of behavior and Roter et al. (2011) investigated both convergent and predictive validity using standardized patients in an objective structured clinical examination. As the thin slice technique continues to be utilized by researchers, it is important to establish reliability and validity of this research technique in a sample of real life interactions. In the medical context, this is critical as continued observational research is conducted and published relating to patient experiences and associated outcomes.

Chapter 3

Methodology

Design

The research design is an archival non-experimental, observational analysis of previously recorded video interactions of primary care patient visits in a low-income urban clinic (see Appendix A for Institutional Review Board approvals). The video recorded interactions were collected as part of a R21 funded interventional study by the National Institutes of Health (L. Penner 1 R21 HD050450-01). One hundred thirty three participants were video recorded during the study. Video recordings were captured using a portable system that allows simultaneous recording of both the patient and physician during the interaction. The individual recordings were processed and edited, resulting in a single image split screen encompassing the patient/physician medical interaction used in subsequent coding and analysis (Albrecht et al., 2005).

Participants and Data

Study participants (video recorded in previous research) included both patients and medical residents recruited using IRB approved recruitment and consent procedures. Seventeen family medicine residents and 126 patients comprise the study sample (video recorded interactions) used in this research. Patients included 96 females (76.2%) and 30 males (23.8%). Although race was not an inclusion criteria for the study, all self-identified as African Americans. Patients' average age was 44.14 years (range 18-64; standard deviation 14.45). Family medicine residents self-identified primarily as Indian, Pakistani or Asian and were evenly split in regards to gender (8 male, 8 female, 1 declined to provide).

Data used for this study include 126 of the 133 available patient/physician interactions. Seven of the original interactions were excluded due to audio/technical issues with recording or length of interaction (not able to extract three thin slices). Sixty-one interactions comprised the control condition in the original study (phase 1) and 65 interactions comprised the intervention condition in the original study (phase 2). For the purposes of collecting data, each phase will be randomized independently.

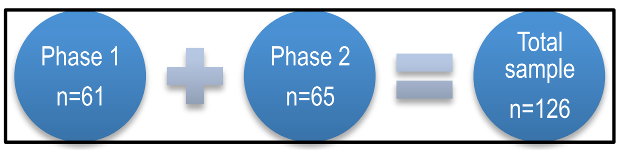


Figure 1. Study Sample

Selection of Thin Slice Observations

Three slices were selected and extracted from the 126 available interactions. Slices (30 seconds in length) were sampled from the beginning, the middle and the end of the interaction. Selections were sampled using the procedure developed for a current nationally funded grant (N. Hagiwara 1 R03 NR013249-01). Specifically, the total length of the interaction was measured (when the patient and physician are in the room together), the first minute and the last minute of the interaction were subtracted from this measurement and the interaction was subsequently divided into three equal segments. Three thin slice samples were then created from each interaction comprised of the first 30 seconds of interaction in each of the three segments (see Figure 2). Full interactions were also edited to remove the first and last minute of the interaction to be consistent with the method used to select thin slices.

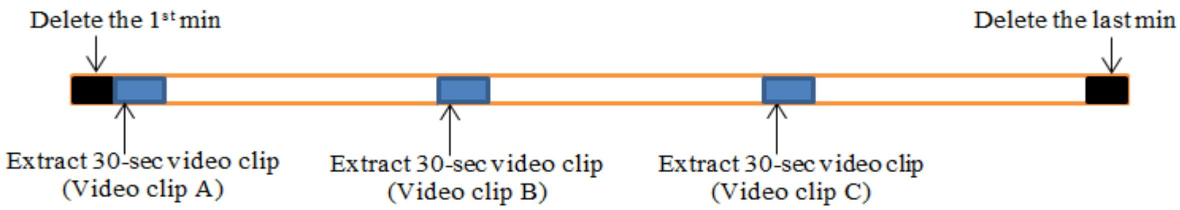


Figure 2. Creation of Thin Slices

Data Collection

Interactions will be split into two groups representing the phase of the original study. Each group will be randomized and coded independently by four trained undergraduate coders from Wayne State University. Volunteers will be recruited through advertisement and word of mouth in the following departments: Psychology, Sociology, Communication, Education or a related field to serve as coders (see Appendix B). The coding procedure will use elements of global attributes of the RIAS coding system consistent with previous research using thin slice observation and coders were trained using the same procedure (Henry & Eggly, 2012). Coders will be asked to rate the specific elements of the patient/physician interaction. Specific elements of the data collection instrument will include rapport, liking, attention, coordination and trust. Ratings of each element will be based on a nine point Likert scale ranging from ‘none’ to ‘high’. Full interactions were independently coded by the same eight trained coders using the same elements of global attributes of the RIAS coding system as stated above. Specific elements of the data collection instruments again included rapport, liking, attention, coordination and trust. Ratings were based on a nine point Likert scale ranging from ‘none’ to ‘high’ to ensure consistency between the thin slice coding and the full interaction (see Appendix C for data collection instruments).

Data Analysis

Descriptive statistics will be calculated for Likert scale item (liking, attention, coordination, trust, rapport) including frequencies, means and standard deviations where appropriate. Descriptive statistics will include measures of central tendencies of for each item (liking, attention, coordination, trust and rapport) stratified by phase (one, two). The data will be assessed following the reliability analysis for normality. Data transformation will be attempted and non-parametric statistics will be applied as appropriate. In addition, an exploratory factor analysis will be conducted for thin slice ratings and for full interaction ratings using principle components extraction and varimax rotation to determine if the measured variables converge on a single construct. Factors retained will have eigenvalues greater or equal to 1.0 and a favorable visual examination of the scree plot. Weights will be sorted and presented with magnitude greater than or equal to |.4|.

Aim 1: *To determine if independent coders can reliably code relational variables (liking, attention, coordination, trust and rapport) between patients and physicians using thin slices sampled from an actual patient/physician interaction.* Inter-rater reliability will be assessed using the interclass correlation coefficient (ICC). This method incorporates the sum of unique pairwise comparisons (Landis, King, Choi, Chinchilli, & Koch, 2011). Reliability will be calculated following coder training and after coding completion. Once acceptable reliability is established, a random selection between coders will yield a single rating for each thin slice/full interaction. This rating will be used for subsequent analysis.

Aim 2: *To determine if there is a significant difference in ratings between the first, second and third slice of the interaction*. Repeated measures analysis of variance (ANOVA) will be used to assess if there is any difference between the first, second and third slices of the interaction. This repeated measure ANOVA model has general linear model components and takes into account the lack of independence (physicians seen by multiple patients) controlling for individual level differences (phase and condition) that may affect the within group variance (Hair, Black, Babin, Anderson, & Tatham, 2006; Hinkle, Wiersma, & Jurs, 1998).

Aim 3: *To determine if there is a significant difference in thin slice ratings compared to ratings from the interaction as a whole*. After determining if there is a significant difference between thin slice ratings (Aim 2), a composite variable for each item will be created that represents the average thin slice rating for the specific item. This composite variable will be used to assess if thin slice ratings significantly differ from ratings obtained from the full interaction. A series of paired t tests will be conducted to assess if the thin slices ratings on specific items are associated with the ratings of the same items obtained from coding the full interaction.

Validity of thin slice ratings will also be assessed in a correlation matrix and visual model describing how each slice and a composite score across thin slices compare to the full interaction. Data was analyzed using IBM SPSS 22.0 statistical package (SPSS Inc., Chicago, IL). All analyses will be conducted at the $p \leq .05$ significance level.

Chapter 4

Results

Patient/Physician and Coder Demographics

Of the 126 interactions in our sample, one interaction was not coded by all coders, thus the results will be presented with a sample size (n=125). Patients included 95 females (76%) and 30 males (24%) who self-identified as African Americans. Patients' average age was 44.29 years (range 18-82; standard deviation 14.42). Seventeen family medicine residents remained in the sample. These physicians self-identified primarily as Indian, Pakistani or Asian and were evenly split in regards to gender (8 male, 8 female, 1 declined to provide). Six physicians declined to provide age; average reported age of physician (n=11) was 29.91 years (range 26-35; standard deviation 2.66).

Eight undergraduate students were recruited through the Psychology and Sociology Departments at Wayne State University. Coders were primarily female (62.5%) upper class students intending to pursue graduate education; average age was 21.75 (range 19-28; standard deviation 2.77) and self-identified as Caucasian (50%), African American (12.5%) and Asian/Pacific Islander (37.5%).

Randomization and Coder Assignment

Randomization occurred on multiple levels. Video interactions in each phase, defined by the original study (control or intervention), were randomized twice (see Appendix D). With eight coders participating, coders were randomized to a specific phase and condition (see Figure 3). Condition was dichotomized, two of the four coders began with the full interaction and progressed to the thin slices, while two of the four coders began with the thin slices and progressed to the full interactions. As a result, 61 phase one interactions were coded by four independent coders and 65 phase two interactions were coded by four independent coders, yielding 500 ratings for each variable measured (liking, attention, coordination, trust and rapport).

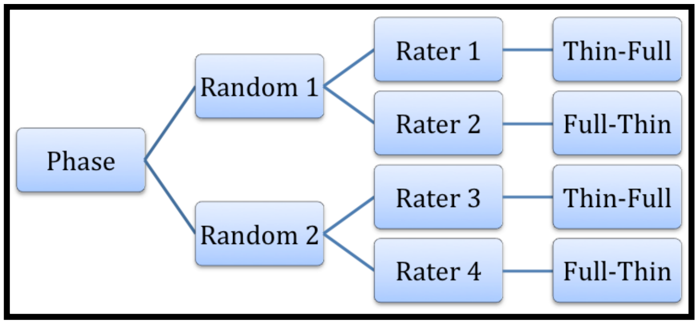


Figure 3. Randomization and Coder Assignment Within Each Phase

Training and Initial Reliability

One 3-hour training session was conducted; all coders participated in the training session. A brief overview of observation research using previously recorded data was provided and specific research using medical interactions and thin slice methodology was reviewed. General coder instructions and the instrument that would be used during coding were also reviewed. Concepts (i.e. specific scale items) were defined and group coding was incorporated with open discussion to ensure concepts were uniformly being applied. Session agenda, PowerPoint presentation slides and coder instruction sheet are presented in Appendix E.

Initial reliability was assessed following training. All coders coded an independent set of thin slice interactions (n=20) during training to assess initial reliability and understanding of scale items. These interactions were specifically developed for training purposes. Means and standard deviations are presented in Table 1.

Table 1

Mean and Standard Deviations for Initial Reliability (n=20)

Rater	Liking	Attention	Coordination	Trust	Rapport
1	6.20 (2.118)	6.30 (2.494)	5.35 (2.346)	5.70 (2.386)	5.00 (2.938)
2	6.35 (1.040)	7.10 (0.968)	5.90 (1.252)	7.25 (0.967)	6.60 (1.142)
3	6.00 (1.835)	7.30 (1.455)	6.90 (1.447)	6.40 (1.536)	6.75 (1.585)
4	6.70 (1.174)	7.60 (0.940)	6.65 (1.137)	6.95 (1.099)	6.65 (1.309)
5	6.50 (1.357)	7.10 (1.373)	6.60 (1.188)	6.50 (1.539)	6.25 (1.618)
6	7.10 (1.119)	7.65 (0.671)	6.70 (1.031)	6.95 (1.276)	6.45 (0.999)
7	6.15 (1.387)	7.00 (1.717)	6.30 (1.625)	6.35 (1.137)	5.90 (1.619)
8	7.60 (1.142)	7.25 (1.410)	7.30 (1.559)	7.95 (1.317)	7.60 (1.046)

Likert scale variables were rated 1-9 (higher ratings=more liking, attention, coordination, trust and rapport)

Descriptive statistics for individual items are presented in Appendix F. Individual item and average inter-rater reliability (IRR) across all coders calculated from the average measure intra-class correlation (ICC 2,8) ranged from .952 to .972 (see Table 2). Using the Spearman Brown Correction, taking into account eight independent coders, the average Interrater Reliability (IRR) was .960.

Table 2

Individual Item Average Measures ICC and IRR - Initial Reliability (n=20)

Measures	Average Measures ICC	Interrater Reliability IRR
Liking	0.813	0.972
Attention	0.728	0.955
Coordination	0.738	0.957
Trust	0.762	0.962
Rapport	0.713	0.952

IRR calculated from ICC measure using Spearman Brown correction.
Likert scale variables were rated 1-9 (higher ratings=more liking, attention, coordination, trust and rapport)

Aim 1: Assessment of Coder Reliability

Full interactions: One hundred twenty five full interactions were coded by four independent coders. Means and standard deviations are presented in Table 3.

Table 3

Mean and Standard Deviations for Full Interactions (n=125)

Rater	Liking	Attention	Coordination	Trust	Rapport
Phase 1					
1	7.02 (1.372)	8.10 (1.017)	7.02 (1.372)	7.46 (1.385)	7.30 (1.145)
2	7.07 (1.250)	7.64 (1.065)	7.00 (1.080)	7.03 (1.612)	6.70 (1.716)
3	6.07 (1.850)	6.10 (2.407)	5.34 (2.218)	5.75 (1.972)	4.79 (2.310)
4	7.31 (1.246)	8.02 (1.072)	7.18 (0.992)	7.49 (1.164)	6.93 (1.289)
Phase 2					
5	7.11 (1.861)	7.08 (2.095)	6.36 (2.522)	7.13 (2.149)	6.09 (2.700)
6	6.23 (1.520)	7.78 (0.967)	7.25 (1.309)	7.30 (1.136)	7.09 (1.281)
7	6.19 (1.355)	7.48 (1.469)	6.94 (1.754)	6.84 (1.556)	6.34 (1.766)
8	6.86 (1.125)	7.27 (1.116)	6.80 (1.211)	6.98 (1.420)	6.91 (1.165)

Phase 1 included 61 interactions; Phase 2 included 64 interactions
Likert scale variables were rated 1-9 (higher ratings=more liking, attention, coordination, trust and rapport)

Descriptive statistics for individual items are presented in Appendix G. Individual item inter-rater reliability for full interaction coding was calculated from the average measure intra-class correlation (ICC 2,8) and ranged from .831 to .909 (see Table 4) with an average IRR across phases of .873.

An exploratory principal component analysis (PCA) on full interaction coding yielded a single construct (Appendix H). The Kaiser-Meyer-Olkin Measure of Sampling Adequacy (testing whether partial correlations among variables are small), KMO=.883, and the Bartlett's Test of Sphericity (testing if there are correlations among the variables present) was significant (χ^2 (10) = 1793.08, p < .001) indicating the sample was appropriate for factor analysis. The PCA, using varimax rotation, was conducted with one construct identified with an Eigenvalue greater than 1.0 explaining 75.70% of the variance. All five scale items (liking, attention, coordination, trust and rapport) were

Table 4

Individual Item Average Measures ICC and IRR - Full Interactions (n=125)

Measures	Average Measures ICC	Interrater Reliability IRR
Phase 1		
Liking	0.712	0.908
Attention	0.618	0.866
Coordination	0.580	0.847
Trust	0.669	0.890
Rapport	0.551	0.831
Phase 2		
Liking	0.623	0.869
Attention	0.607	0.861
Coordination	0.625	0.870
Trust	0.648	0.880
Rapport	0.715	0.909

IRR calculated from ICC measure using Spearman Brown correction.
Phase 1 included 61 interactions; Phase 2 included 64 interactions.
Likert scale variables were rated 1-9 (higher ratings=more liking, attention, coordination, trust and rapport)

represented in the emerged component for full interaction coding (see Table 5). Factor loadings and communalities are presented in Table 6.

Table 5

PCA: Eigenvalues and Percent Variance - Full Interactions

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% Variance	Cumulative %	Total	% Variance	Cumulative %
1	3.785	75.700	75.700	3.785	75.700	75.700
2	0.459	9.170	84.870			
3	0.297	5.934	90.804			
4	0.250	4.992	95.796			
5	0.210	4.204	100.000			

Extraction Method: Principal Component Analysis - Rotation was not completed as only one component was extracted.

Table 6

Factor loading and communalities - Full interactions

	Loading	Communality
Full Liking	0.837	0.701
Full Attention	0.882	0.778
Full Coordination	0.891	0.794
Full Trust	0.864	0.747
Full Rapport	0.900	0.811

Thin Slices: Three hundred and seventy five thin slices were coded by four independent coders (three slices for each of the 125 full interactions). Means and standard deviations are presented in Table 7. Descriptives for individual items are presented in Appendix I. Individual item inter-rater reliability for thin slice coding

Table 7

Mean and Standard Deviations for Thin Slices (n=375)

Rater	Liking	Attention	Coordination	Trust	Rapport
Phase 1					
1	6.47 (1.083)	7.57 (1.136)	6.84 (1.438)	6.98 (1.307)	6.93 (1.359)
2	5.71 (1.693)	7.27 (1.468)	6.24 (1.582)	6.13 (1.790)	5.66 (1.903)
3	6.17 (1.210)	7.19 (1.647)	5.96 (1.612)	5.97 (1.522)	5.62 (1.917)
4	6.44 (1.420)	7.54 (1.194)	6.98 (1.170)	7.12 (1.270)	6.58 (1.264)
Phase 2					
5	6.95 (1.324)	6.93 (1.388)	6.60 (1.628)	6.83 (1.682)	6.56 (1.763)
6	5.92 (1.505)	8.04 (0.894)	7.09 (1.258)	7.21 (1.500)	6.51 (1.476)
7	6.06 (1.110)	7.29 (1.571)	6.49 (1.535)	6.79 (1.265)	6.40 (1.425)
8	6.02 (0.805)	6.41 (0.781)	6.13 (0.897)	6.49 (1.028)	6.14 (0.922)

Phase 1 included 183 thin slices; Phase 2 included 192 thin slices
Likert scale variables were rated 1-9 (higher ratings=more liking, attention, coordination, trust and rapport)

was calculated from the average measure intra-class correlation (ICC 2,8) and ranged from .762 to .910 (see Table 8) with an average IRR of .850. Individual item inter-rater reliability for thin slice coding was calculated from the average measure intra-class correlation (ICC 2,8) and ranged from .762 to .910 (see Table 8) with an average IRR of .850.

Table 8

Individual Item Average Measures ICC and IRR - Thin Slices (n=375)

Measures	Average Measures ICC	Interrater Reliability IRR
Phase 1		
Liking	0.717	0.910
Attention	0.637	0.875
Coordination	0.618	0.866
Trust	0.695	0.901
Rapport	0.682	0.897
Phase 2		
Liking	0.626	0.870
Attention	0.444	0.762
Coordination	0.461	0.774
Trust	0.498	0.799
Rapport	0.576	0.845

IRR calculated from ICC measure using Spearman Brown correction.
Phase 1 included 183 thin slices; Phase 2 included 192 thin slices.
Likert scale variables were rated 1-9 (higher ratings=more liking, attention, coordination, trust and rapport)

An exploratory principal component analysis (PCA) on thin slice coding yielded a single construct (Appendix J). The Kaiser-Meyer-Olkin Measure of Sampling Adequacy indicated this sample was also appropriate to conduct a factor analysis (KMO=.865 and the Bartlett's Test of Sphericity was significant (χ^2 (10) = 4673.893, p <.001). The PCA, using varimax rotation was conducted, with one construct identified with an Eigenvalue greater than 1.0 explaining 71.72% of the variance. All five items (liking, attention, coordination, trust and rapport) were again represented in the PCA component for thin slice coding (see Table 9). Factor loadings and communalities are presented in Table 10.

Table 9

PCA: Eigenvalues and Percent Variance - Thin Slices

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% Variance	Cumulative %	Total	% Variance	Cumulative %
1	3.586	71.719	71.719	3.586	71.719	71.719
2	0.559	11.188	82.907			
3	0.355	7.105	90.013			
4	0.274	5.475	95.488			
5	0.226	4.512	100.000			

Extraction Method: Principal Component Analysis - Rotation was not completed as only one component was extracted.

Table 10

Factor loading and communalities - Thin interactions

	Loading	Communality
Thin Liking	0.806	0.650
Thin Attention	0.794	0.631
Thin Coordination	0.858	0.737
Thin Trust	0.882	0.777
Thin Rapport	0.889	0.791

Aim 2: Assessment of Difference in Thin Slice Ratings

Variables for each slice (liking, attention, coordination, trust, rapport) across the three time points in the interaction (1, 2, 3) were tested for normality and found to be non-representative of a normal population. Shapiro-Wilk statistics for all variables yielded a p-value of .000, results are presented in Table 11. After several attempts to transform the data, including calculating the Z score, square root, square, Log10 and reciprocal measures of each variable, normality tests continued to show a deviation from normality. See example for the first slice coding of the variable Liking in Table 12.

Table 11

Thin Slice Test for Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Liking.Slice1	0.193	125	0.000	0.923	125	0.000
Liking.Slice2	0.161	125	0.000	0.941	125	0.000
Liking.Slice3	0.205	125	0.000	0.911	125	0.000
Attention.Slice1	0.203	125	0.000	0.893	125	0.000
Attention.Slice2	0.206	125	0.000	0.889	125	0.000
Attention.Slice3	0.165	125	0.000	0.924	125	0.000
Coordination.Slice1	0.196	125	0.000	0.930	125	0.000
Coordination.Slice2	0.168	125	0.000	0.932	125	0.000
Coordination.Slice3	0.162	125	0.000	0.946	125	0.000
Trust.Slice1	0.187	125	0.000	0.924	125	0.000
Trust.Slice2	0.152	125	0.000	0.934	125	0.000
Trust.Slice3	0.201	125	0.000	0.923	125	0.000
Rapport.Slice1	0.151	125	0.000	0.944	125	0.000
Rapport.Slice2	0.208	125	0.000	0.935	125	0.000
Rapport.Slice3	0.176	125	0.000	0.927	125	0.000

a. Lilliefors Significance Correction

Due to this deviation from normality, the non-parametric version of the repeated measures ANOVA, or the Friedman test, was used to test the significance between the ratings of thin slices over the course of the interaction. Freidman tests conducted on

Table 12

Example of Transformed Data - Test for Normality

	Shapiro-Wilk		
	Statistic	df	Sig.
Liking.Slice1	0.923	125	0.000
Zscore (Liking.Slice1)	0.923	125	0.000
SquareRoot (Liking.Slice1)	0.928	125	0.000
Square (Liking.Slice1)	0.897	125	0.000
Log10 (Liking.Slice1)	0.928	125	0.000
Reciprocal (Liking.Slice1)	0.909	125	0.000
Log10 K-X (Liking.Slice1)	0.835	125	0.000

scores obtained from the three slices of the interaction on each of the variables are as follows: Liking (χ^2 (2) = .095, *p* =.953), Attention (χ^2 (2) = 5.211, *p* =.074), Coordination (χ^2 (2) = 3.803, *p* =.149), Trust (χ^2 (2) = 3.775, *p* =.151) and Rapport (χ^2 (2) = 3.236, *p* =.198). Results failed to reject the null, indicating no significant difference between the first, second and third slice of the interaction.

Aim 3: Assessment of Difference between Thin Slice & Full Interaction Ratings

Assessment of differences between a thin slice composite variable (i.e. an average rating across three thin slices) and the full interaction was computed using the non-parametric version of the paired t test or the Wilcoxon Signed Ranks Test (Table 13).

Table 13

Wilcoxon Signed Ranks - Thin Average Ratings Vs Full Interaction Ratings

	Z Statistic	Asymp. Sig. (2 tailed)
Full Liking - Thin Liking	-5.138	0.000
Full Attention - Thin Attention	-2.606	0.009
Full Coordination - Thin Coordination	-3.456	0.001
Full Trust - Thin Trust	-2.563	0.010
Full Rapport - Thin Rapport	-2.222	0.026
Full Composite - Thin Composite	-3.846	0.000

The Wilcoxon Signed-ranks test indicated that full interaction ratings were consistently higher across all variables: full interaction liking (Mdn=7.00) versus thin slice liking (Mdn=6.33), Z=5.14, p<.001; full interaction attention (Mdn=8.00) versus thin slice attention (Mdn=7.33), Z=2.61, p<.01; full interaction coordination (Mdn=7.00) versus thin slice coordination (Mdn=6.67), Z=3.46, p<.01; full interaction trust Mdn=7.00) versus thin slice trust (Mdn=6.67), Z=2.56, p<.05; full interaction rapport

(Mdn=7.00) versus thin slice rapport (Mdn=6.33), Z=2.22, p<.05; full interaction composite (Mdn=7.20) versus thin slice composite (Mdn=6.67), Z=3.85, p<.001.

A correlation matrix was also developed to see how individual ratings and composite scores for thin slices compared to the full interaction ratings (see Appendix K). The results indicate a significant correlation among all rating pairs: thin slice liking and full interaction liking, r=.468, p<.01; thin slice attention and full interaction attention, r=.335, p<.01; thin slice coordination and full interaction coordination, r=.390, p<.01; thin slice trust and full interaction trust, r=.325, p<.01; thin slice rapport and full interaction rapport, r=.363, p<.01. In addition, the correlation of the composite scores yielded a significant result, r=.493, p<.01 indicating a shared variance (R^2) of 24.3%.

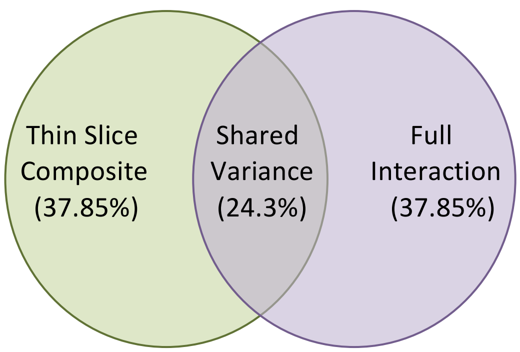


Figure 4. Thin Slice Composite and Full Interaction – Shared Variance

Chapter 5

Discussion

The purposes of the study were to determine if thin slices sampled from the beginning, middle and end of an actual medical interaction in an urban medical setting serving low income African Americans can be reliably coded by independent raters, if the rating of the three slices obtained are associated with each other, and if the ratings of the three slices obtained are representative of ratings from the whole interaction.

Aim 1: Assessment of Coder Reliability

Inter-rater reliability (IRR) was assessed using a two-way mixed, consistency, average-measures ICC to assess the degree to which coders were similar in rating relational items (liking, attention, coordination, trust and rapport) in human interaction. Results indicate that independent raters can reliably code relational variables (liking, attention, coordination, trust and rapport) between patients and physicians using thin slices sampled from an actual patient/physician interaction. This was demonstrated in the initial reliability calculation with an average IRR=.960 across all eight coders (n=20), and in the coding of thin slices (average IRR=.850) and coding of full interactions (average IRR=.873) calculated across four coders (n=375 and n=125 respectively), indicating a high degree of agreement across coders. In addition, findings indicate that regardless of the length of video (30 second slices vs. a full interaction) the instrument was found to be reliable in coding human interactions.

Once reliability was established, a random selection of thin slices was compiled to determine if the assessments between the first, second and third slice of the interaction ratings were similar. Specifically, for each interaction where three thin slices were coded by four independent coders, one coder was randomly selected for each interaction to represent the ratings in the analysis of the second aim.

Aim 2: Assessment of Difference in Thin Slice Ratings

Before assessing the difference between the thin slice ratings, the data was tested for normality. The Shapiro-Wilk statistic indicated the data was not normally distributed. Therefore, instead of using repeated measures ANOVA, the non-parametric equivalent or the Friedman test was used to determine if there was a significant difference between the means. This tested the null hypothesis that the means of each slice were equal ($\mu_{slice1}=\mu_{slice2}=\mu_{slice3}$). The Friedman test results on all Likert scale variables (liking, attention, coordination, trust and rapport) failed to reject the null hypothesis, providing support that thin slices sampled at the beginning, middle and end of an interaction are not statistically significant from each other.

Although the Friedman test results for all variables failed to reject the null hypothesis, there was a variation in these results. Some variables were more consistent across the three slices while others were less consistent. This is seen in the not significant liking rating resulting in an extremely low probability of a type II error (failing to reject the null when in fact it should be rejected) at p=.953 and the not significant attention rating resulting in a statistic that is approaching significance at p=.074.

Aim 3: Assessment of Difference between Thin Slice & Full Interaction Ratings

Once the results from the second aim indicated that ratings between the first, second and third slice were comparable, a composite variable for the thin slice ratings was computed (an average rating for each variable by case) and a random selection of full interactions were compiled. Specifically, for each full interaction coded by four independent coders, one coder was randomly selected for each interaction to represent the ratings in the analysis of the third aim. These two variables (average thin slice rating and randomly selected full interaction rating) were used to determine if there is a significant difference in thin slice ratings compared to ratings from the interaction as a whole.

The Wilcoxon Signed-ranks test, or the non-parametric version of the paired t test, was used to test the hypothesis that the means between the thin slice composite variable and the full interaction composite variable were equal ($\mu_{\text{thin slice}} = \mu_{\text{full interaction}}$). Results of this test supported the alternative hypothesis that the means were not equal, or that thin slices ratings in this sample were not comparable to full interaction ratings.

Summary and Conclusions

Results from the first two aims provided support to the concept that thin slices could indeed be reliably coded and furthermore that each slice within a specific interaction was representative of other slices within the same interaction. The result of the third aim did not provide support to the hypothesis that thin slice ratings were comparable to full interaction ratings. Specifically, the ratings from the thin slice composite were not representative of a random sample rating of the same full interaction. Murphy (2005) indicated that thin slice methodology should have a stronger reliability coefficient when compared to full interactions in order to justify this methodology. In this study, the conclusion would be that thin slices are not a suitable substitute for full slice ratings. This result may be a factor of study limitations (sample clinic, sample timeframe of data collection) or it may be a factor of more information being providing in a full interaction that cannot be captured in a sample of thin slices.

A quick analysis using parametric tests found, similar to previous research, that the ratings from thin slice interactions for each Likert scale variable were significantly correlated with ratings from the corresponding full interactions and the thin slice composite variable correlated with the full interaction (r=.493) indicating a shared variance of 24.3%, this was above the average reported effect size (r=.39) and slightly above the reported effect size in social psychology research studies (r=. 47) in the meta-analysis conducted by Ambady and Rosenthal (1992). Additionally, it was found though a simple linear regression that thin slice ratings did in fact predict the ratings obtained from the full interaction (see Appendix K) with the regression equation equal to Y=.794(X)+1.714; where Y= the full interaction composite variable and X= the thin slice composite variable.

Parametric tests were used in the above testing, as the software selected for analysis (SPSS) does not have a mechanism to analyze the non-parametric version of a simple regression, so care should be taken when interpreting the results. In this study, the third aim was to specifically test the hypothesis that thin slice ratings were representative of full interaction ratings. The research question, along with the results of the normality tests led to the decision to analyze the data using the Wilcoxon Sign-ranked test comparing the thin slice composite variable to the full interaction variable. If the research question was to determine if there was an association between the two measures, another statistic such as the Pearson correlation or the Spearman correlation would be utilized. Parametric testing, using the Pearson correlation and simple linear regression, did in fact find an association, both correlative and predictive, between the thin slice composite variable and the full interaction variable, replicating previous research results reporting correlations and predictions.

As noted, many areas of research (psychology, sociology, medicine, etc.) have benefited from observational research methods such as the thin slice technique. By using this technique, researchers can study human behavior, while saving valuable resources (time, money). The study of human behavior leads to increased knowledge, ultimately leading to the testing and development of future interventions to improve outcomes. The research presented here utilized a specific sample (data was collected in one urban clinic population during a specific timeframe) and rated patient/physician relational components (liking, attention, coordination, trust and rapport), which has been previously reported to have an effect on quality of care and patient satisfaction (Dovidio et al., 2008; Stepanikova, 2006; Stepanikova et al., 2006). Again, only when reliability and validity of methods are established in the research process can the research community be certain research findings are appropriate, relevant and meaningful. However it is also imperative that researchers keep the end result in mind when designing and evaluating research, selecting the best statistical test to answer their specific research question.

Future research on thin slices may address the affect of condition on ratings. Does it make a difference if a coder begins with full interactions and proceeds to thin slices compared to if a coder begins with thin slices and proceeds to full interactions? What is the impact of time between coding assignments? Could more or less time between coding assignments yield different results? Finally, what is the impact and/or what are the differences between naïve coders and trained experience coders on thin slice ratings using real time video recorded interactions.

APPENDIX A – INSTITUTIONAL REVIEW BOARD - INITIAL APPROVAL AND CONTINUATION

WAYNE STATE UNIVERSITY

IRB Administration Office
87 East Canfield, Second Floor
Detroit, Michigan 48201
Phone: (313) 577-1628
FAX: (313) 993-7122
http://irb.wayne.edu

NOTICE OF EXPEDITED APPROVAL

To: Tanina Foster
Oncology
Karmanos Cancer Institute

From: Dr. Scott Millis ____________________________________
Chairperson, Behavioral Institutional Review Board (B3)

Date: January 25, 2013

RE: IRB #: 129312B3E

Protocol Title: Reliability and Validity of the Thin Slice Technique - Observational Research on the Patient/Physician Medical Information

Funding Source: Sponsor: NATIONAL SCIENCE FOUNDATION

Protocol #: 1301011593

Expiration Date: January 24, 2014

Risk Level / Category: Research not involving greater than minimal risk

---

The above-referenced protocol and items listed below (if applicable) were **APPROVED** following *Expedited Review* Category (#7)* by the Chairperson/designee *for* the Wayne State University Institutional Review Board (B3) for the period of 01/25/2013 through 01/24/2014. This approval does not replace any departmental or other approvals that may be required.

- Revised Protocol Summary Form (received in the IRB Office 1/22/13)
- Protocol (received in the IRB Office 12/18/12)
- A waiver of consent has been granted according to 45CFR 46 116(d) and justification provided by the Principal Investigator in the Protocol Summary Form (this study is an analysis of an existing data archive and no new patient or physician data is being collected). This waiver satisfies: 1) risk is no more than minimal, 2) the waiver does not adversely affect the rights and welfare of research participants, 3) the research could not be practicably carried out without the waiver, and (4) providing participants additional pertinent information after participation is not appropriate.
- Data collection tools: Rating Sheet

- Federal regulations require that all research be reviewed at least annually. You *may* receive a *"Continuation Renewal Reminder"* approximately two months prior to the expiration date; however, it is the Principal Investigator's responsibility to obtain review and continued approval ***before*** the expiration date. Data collected during a period of lapsed approval is unapproved research and can never be reported or published as research data.
- All changes or amendments to the above-referenced protocol require review and approval by the IRB **BEFORE** implementation.
- Adverse Reactions/Unexpected Events (AR/UE) must be submitted on the appropriate form within the timeframe specified in the IRB Administration Office Policy (http://www.irb.wayne.edu//policies-human-research.php).

NOTE:

1. Upon notification of an impending regulatory site visit, hold notification, and/or external audit the IRB Administration Office must be contacted immediately.
2. Forms should be downloaded from the IRB website at **each** use.

*Based on the Expedited Review List, revised November 1998

WAYNE STATE UNIVERSITY

IRB Administration Office
87 East Canfield, Second Floor
Detroit, Michigan 48201
Phone: (313) 577-1628
FAX: (313) 993-7122
http://irb.wayne.edu

NOTICE OF EXPEDITED CONTINUATION APPROVAL

To: Tanina Foster
Oncology
Karmanos Cancer Institute

From: Dr. Deborah Ellis ____
Chairperson, Behavioral Institutional Review Board (B3)

Date: December 20, 2013

RE: IRB #: 129312B3E

Protocol Title: Reliability and Validity of the Thin Slice Technique - Observational Research on the Patient/Physician Medical Information

Funding Source: Sponsor: NATIONAL SCIENCE FOUNDATION

Protocol #: 1301011593

Expiration Date: December 19, 2014

Risk Level / Category: Research not involving greater than minimal risk

---

Continuation for the above-referenced protocol and items listed below (if applicable) were APPROVED following Expedited Review by the Chairperson/designee of the Wayne State University Institutional Review Board (B3) for the period of **12/20/2013 through 12/19/2014**. This approval does not replace any departmental or other approvals that may be required.

- Closed to accrual and active intervention completed.

- Federal regulations require that all research be reviewed at least annually. You *may* receive a *"Continuation Renewal Reminder"* approximately two months prior to the expiration date; however, it is the Principal Investigator's responsibility to obtain review and continued approval ***before*** the expiration date. Data collected during a period of lapsed approval is unapproved research and can never be reported or published as research data.
- All changes or amendments to the above-referenced protocol require review and approval by the IRB **BEFORE** implementation.
- Adverse Reactions/Unexpected Events (AR/UE) must be submitted on the appropriate form within the timeframe specified in the IRB Administration Office Policy (http://www.irb.wayne.edu//policies-human-research.php).

NOTE:

1. Upon notification of an impending regulatory site visit, hold notification, and/or external audit the IRB Administration Office must be contacted immediately.
2. Forms should be downloaded from the IRB website at **each** use.

*Based on the Expedited Review List, revised November 1998

APPENDIX B – CODER RECRUITMENT

Volunteer Undergraduate Coders

Needed for Dissertation Research

Opportunity to Gain Valuable Experience in Behavioral Research Methods and Techniques

- Volunteers should be undergraduate students in anthropology, psychology, sociology, communication, education or related field.
- Individuals should be highly-motivated undergraduate students interested in applying to graduate school in the future.
- Anticipated 25-30 hour commitment with flexible scheduling.
- Training will be provided and volunteers will gain knowledge and insight into behavioral research and research methodology using real time medical interactions.

APPENDIX C – RATING SHEETS

Full Interaction

Rater:

PT ID: Length of interaction:

Instructions:

Watch the entire video segment without stopping. Focus on the general overall relationship between the patient and the doctor. When the segment ends, pause/stop the video to complete the ratings. Watch each segment only once.

Ratings are an ‘average’ over the entire segment. There is no “correct” or “incorrect” answer. You should not take more than 30 seconds to complete these ratings.

1. Rate how much the patient and doctor like each other:

No 1 2 3 4 5 6 7 8 9 High

2. Rate how much the patient and doctor are paying attention to each other.

No 1 2 3 4 5 6 7 8 9 High

3. Rate how much the patient and doctor were coordinated with each other in their movements, speech, and posture.

No 1 2 3 4 5 6 7 8 9 High

4. Rate how much the patient and doctor trust each other:

No 1 2 3 4 5 6 7 8 9 High

5. Rate the level of overall rapport between the patient and doctor:

No 1 2 3 4 5 6 7 8 9 High

Thin Slice

Rater:

PT ID: Slice number:

Instructions:

Watch the entire video segment without stopping. Focus on the general overall relationship between the patient and the doctor. When the segment ends, pause/stop the video to complete the ratings. Watch each segment only once.

Ratings are an 'average' over the entire segment. There is no "correct" or "incorrect" answer. You should not take more than 30 seconds to complete these ratings.

1. Rate how much the patient and doctor like each other:

No 1 2 3 4 5 6 7 8 9 High

2. Rate how much the patient and doctor are paying attention to each other.

No 1 2 3 4 5 6 7 8 9 High

3. Rate how much the patient and doctor were coordinated with each other in their movements, speech, and posture.

No 1 2 3 4 5 6 7 8 9 High

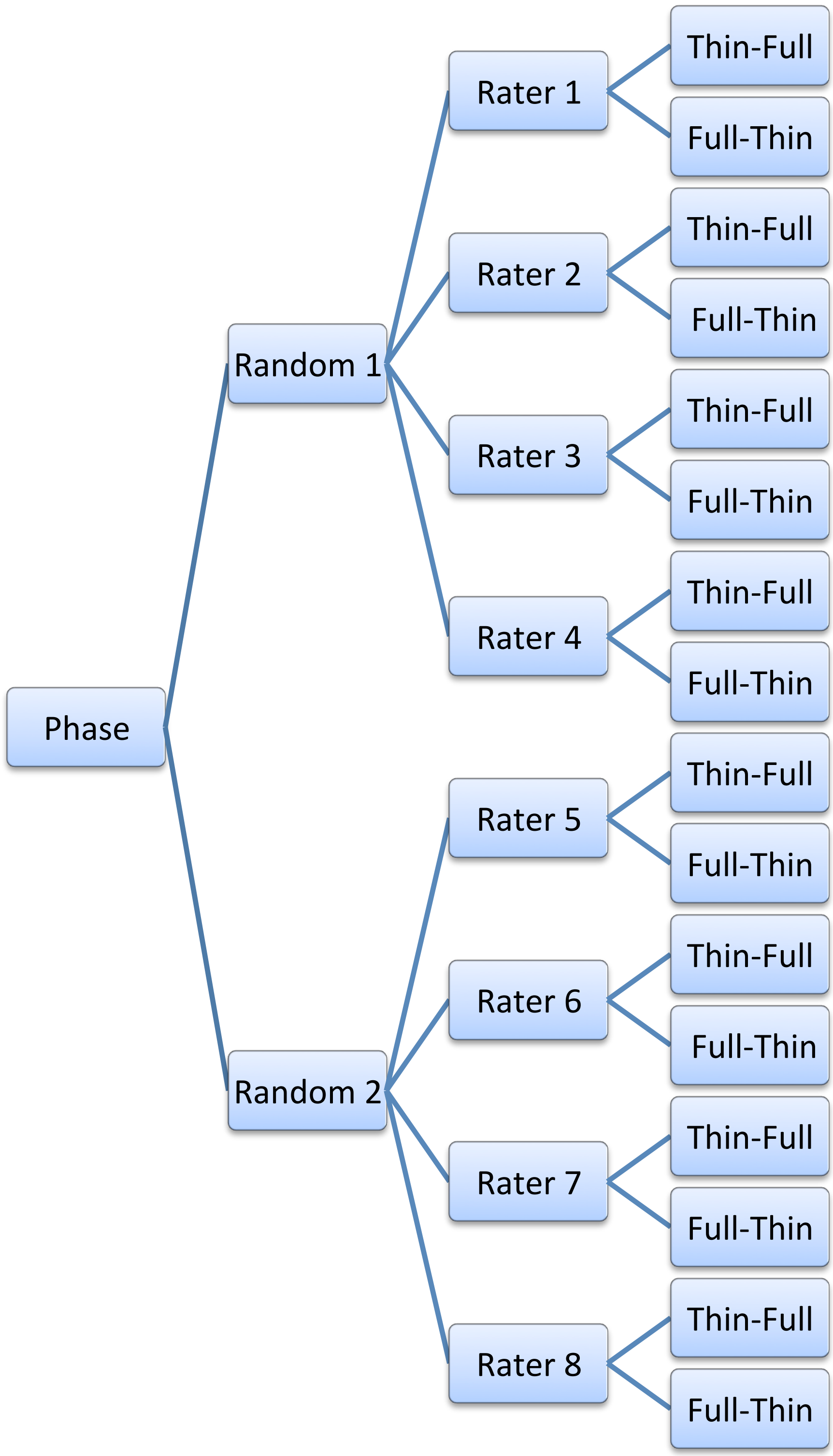
4. Rate how much the patient and doctor trust each other:

No 1 2 3 4 5 6 7 8 9 High

5. Rate the level of overall rapport between the patient and doctor:

No 1 2 3 4 5 6 7 8 9 High

APPENDIX D – RANDOMIZATION BY PHASE



Appendix E

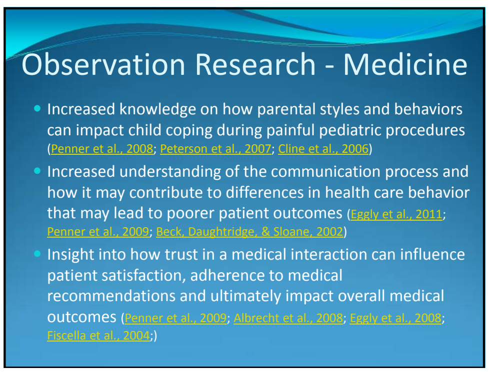
APPENDIX E – TRAINING MATERIALS

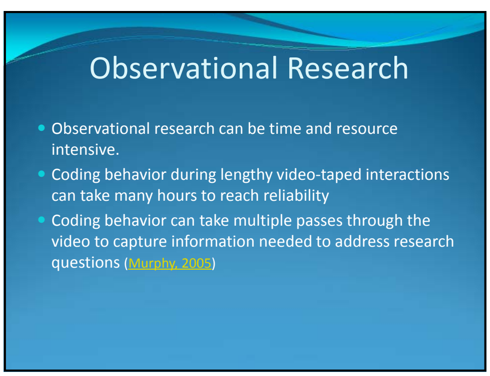
Training Agenda

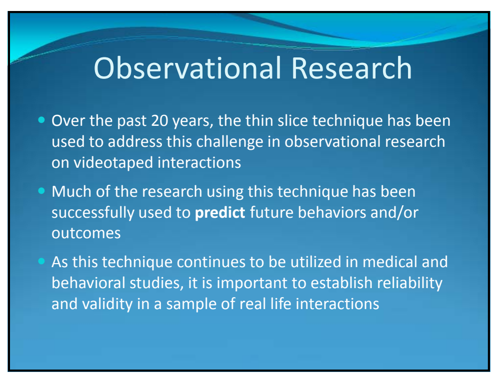
June 5, 2013

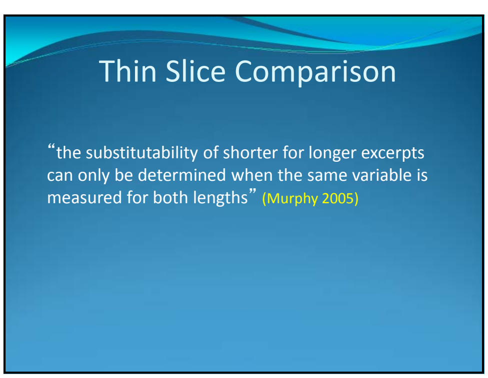
1. Welcome, Introductions, and Overview
2. Behavioral Research
3. Medical Interaction and Thin Slice Research
4. Instrument Review
5. Open Discussion and Practice
6. Questions and Answers
7. Application / Initial Group Coding
8. Session Review and Next Steps

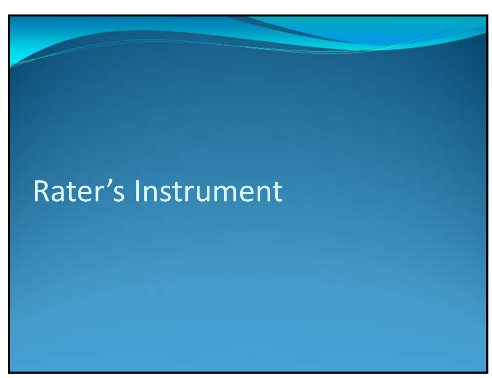
Training - Power Point











Rater: **<Coder ID>**

PT ID: **<Patient ID>** Slice number: **<Slice>**

Instructions:
Watch the entire video segment without stopping. Focus on the general overall relationship between the patient and the doctor. When the segment ends, pause/stop the video to complete the ratings.

There is no "correct" or "incorrect" answer. You should not take more than 30 seconds to complete these ratings. Watch each segment only once.

Ratings are an 'average' over the entire segment. Start at a rating of 5 for each rating. 'Slide' your rating up and/or down as you watch the interaction. Don't use the rating you land on at the end of the interaction, but rather decide which number you slide over the most and use that average as your final rating.

1. Rate how much the patient and doctor **like** each other:

   no 1 2 3 4 5 6 7 8 9 high liking

2. Rate how much the patient and doctor are **paying attention** to each other.

   no 1 2 3 4 5 6 7 8 9 high attention

3. Rate how much the patient and doctor were **coordinated** with each other in their movements, speech, and posture.

   no 1 2 3 4 5 6 7 8 9 high coordination

4. Rate how much the patient and doctor **trust** each other:

   no 1 2 3 4 5 6 7 8 9 high trust

5. Rate the level of overall **rapport** between the patient and doctor:

   no 1 2 3 4 5 6 7 8 9 high rapport

General Instructions

- Raters will be asked to rate both 30-sec video segments and full interaction segments between patients and their primary care physicians
- Raters and videos will be randomized within and between methods
- The task will be to watch the video segments in the order randomized and rate the relationship between patient and physician on five characteristics based on your average impression over time using a scale ranging from 1 (not observed) -9 (high observed)

5 Characteristics

- Liking
- Attention
- Coordination
- Trust
- Rapport

Review of Characteristics

And Practice Coding

Review

- The higher the rating given, the more the characteristic is seen/felt.
- Indicators that may assist in determining the degree of each rating can include:
  - Verbal expressions: words, tone of voice, etc.
  - Nonverbal behaviors: eye contact, facial expressions, orientation, etc.

Review

- Ratings should indicate an 'average' over the entire video segment.
- Imagine starting each observation at the midpoint (5 for each characteristic) and throughout the interaction, mentally 'sliding' your rating up and/or down as you watch the interaction.
- Don't automatically use the rating you land on at the end of the interaction, but rather decide which number you slide over the most and use that average as your final rating.

Questions?

Scheduling Sessions

Training - Rater Instructions

General Instructions for Raters

All volunteers will be asked to rate both 30-sec video segments and full interactions between patients and their primary care physician. Each volunteer will be randomly assigned to one of the following scenarios:

- Full Interactions - Period of Time - Thin Slices
  - or -
- Thin Slices – Period of Time - Full Interactions

Your task is watch each video segment and following the segment rate the relationship between the patient and physician on several characteristics based on your average impression over time, using a scale ranging from 1 to 9.

There are five characteristics we will be looking at including liking, attention, coordination, trust and rapport. The higher the number, the more the characteristic is seen/felt.

- Liking – Degree of warmth, friendliness, sincerity, sociability, understanding.
- Attention – Degree of interest or consideration.
- Coordination – Degree of synchronicity or orientation in movement, speech, posture.
- Trust – Degree of relational confidence.
- Rapport – Degree of relational connection.

Some indicators that will assist in determining the degree rating in each of the five characteristics can include: verbal expressions, nonverbal behavior, eye contact, facial expressions, etc.

In this study, we are interested in how your global perceptions of the patient/physician relationship. There is no right or wrong answer. Some coders are reluctant to use extreme numbers (e.g., 1 and 9), but please try to use the entire range (from 1 to 9).

Ratings should indicate an ‘average’ over the entire video segment. Imagine starting each observation at the midpoint (5 for each characteristic). Mentally ‘sliding’ your rating up and/or down as you watch the interaction. Don’t automatically use the rating you land on at the end of the interaction, but rather decide which number you slide over the most and use that average as your final rating.

APPENDIX F – INITIAL RELIABILITY INDIVIDUAL ITEM DESCRIPTIVE STATISTICS

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
R1Liking	20	6	3	9	6.20	2.118	4.484	-.069	.512	-1.116	.992
R2Liking	20	5	4	9	6.35	1.040	1.082	.445	.512	2.118	.992
R3Liking	20	6	3	9	6.00	1.835	3.368	-.057	.512	-.916	.992
R4Liking	20	4	5	9	6.70	1.174	1.379	.221	.512	-.917	.992
R5Liking	20	5	4	9	6.50	1.357	1.842	.070	.512	-.383	.992
R6Liking	20	5	4	9	7.10	1.119	1.253	-.966	.512	1.794	.992
R7Liking	20	5	4	9	6.15	1.387	1.924	.230	.512	-.225	.992
R8Liking	20	4	5	9	7.60	1.142	1.305	-.508	.512	-.192	.992
R1Attention	20	8	1	9	6.30	2.494	6.221	-.448	.512	-.861	.992
R2Attention	20	4	4	8	7.10	.968	.937	-1.764	.512	4.681	.992
R3Attention	20	5	4	9	7.30	1.455	2.116	-1.268	.512	1.176	.992
R4Attention	20	3	6	9	7.60	.940	.884	.101	.512	-.798	.992
R5Attention	20	5	4	9	7.10	1.373	1.884	-.605	.512	.053	.992
R6Attention	20	3	6	9	7.65	.671	.450	-.613	.512	.736	.992
R7Attention	20	6	3	9	7.00	1.717	2.947	-1.040	.512	.419	.992
R8Attention	20	4	5	9	7.25	1.410	1.987	-.243	.512	-1.116	.992
R1Coordination	20	8	1	9	5.35	2.346	5.503	-.007	.512	-.571	.992
R2Coordination	20	5	4	9	5.90	1.252	1.568	.743	.512	.754	.992
R3Coordination	20	6	3	9	6.90	1.447	2.095	-1.312	.512	1.966	.992
R4Coordination	20	5	4	9	6.65	1.137	1.292	-.174	.512	.725	.992
R5Coordination	20	5	4	9	6.60	1.188	1.411	-.369	.512	.341	.992
R6Coordination	20	4	4	8	6.70	1.031	1.063	-.922	.512	1.203	.992
R7Coordination	20	6	3	9	6.30	1.625	2.642	-.377	.512	-.455	.992
R8Coordination	20	5	4	9	7.30	1.559	2.432	-.742	.512	-.501	.992
R1Trust	20	8	1	9	5.70	2.386	5.695	-.043	.512	-.752	.992
R2Trust	20	3	5	8	7.25	.967	.934	-.947	.512	-.320	.992
R3Trust	20	6	3	9	6.40	1.536	2.358	-.469	.512	-.227	.992
R4Trust	20	4	4	8	6.95	1.099	1.208	-.950	.512	.934	.992
R5Trust	20	5	4	9	6.50	1.539	2.368	-.193	.512	-.690	.992
R6Trust	20	4	5	9	6.95	1.276	1.629	.103	.512	-.728	.992
R7Trust	20	5	4	9	6.35	1.137	1.292	.174	.512	.725	.992
R8Trust	20	4	5	9	7.95	1.317	1.734	-1.282	.512	.748	.992
R1Rapport	20	8	1	9	5.00	2.938	8.632	.263	.512	-1.437	.992
R2Rapport	20	5	3	8	6.60	1.142	1.305	-1.449	.512	4.286	.992
R3Rapport	20	7	2	9	6.75	1.585	2.513	-1.393	.512	3.005	.992
R4Rapport	20	5	4	9	6.65	1.309	1.713	-.521	.512	.173	.992
R5Rapport	20	5	4	9	6.25	1.618	2.618	.129	.512	-.896	.992
R6Rapport	20	3	5	8	6.45	.999	.997	-.024	.512	-.933	.992
R7Rapport	20	6	3	9	5.90	1.619	2.621	.179	.512	-.719	.992
R8Rapport	20	4	5	9	7.60	1.046	1.095	-.600	.512	.602	.992
Valid N (listwise)	20										

APPENDIX G – FULL INTERACTION INDIVIDUAL ITEM DESCRIPTIVE STATISTICS

Phase 1

Descriptives

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
R1Liking	61	8	1	9	7.02	1.372	1.883	-1.831	.306	6.188	.604
R2Liking	61	6	3	9	7.07	1.250	1.562	-.710	.306	1.226	.604
R3Liking	61	8	1	9	6.07	1.825	3.329	-.696	.306	.007	.604
R4Liking	61	5	4	9	7.31	1.246	1.551	-.516	.306	-.246	.604
R1Attention	61	5	4	9	8.00	1.017	1.033	-1.575	.306	3.710	.604
R2Attention	61	4	5	9	7.64	1.065	1.134	-.249	.306	-.456	.604
R3Attention	61	8	1	9	6.10	2.047	4.190	-.619	.306	-.237	.604
R4Attention	61	4	5	9	8.02	1.072	1.150	-1.040	.306	.547	.604
R1Coordination	61	7	2	9	7.02	1.372	1.883	-1.471	.306	3.489	.604
R2Coordination	61	4	5	9	7.00	1.080	1.167	.410	.306	-.422	.604
R3Coordination	61	8	1	9	5.34	2.128	4.530	-.458	.306	-.533	.604
R4Coordination	61	5	4	9	7.18	.992	.984	-.694	.306	.838	.604
R1Trust	61	7	2	9	7.46	1.385	1.919	-1.351	.306	3.098	.604
R2Trust	61	8	1	9	7.03	1.612	2.599	-.993	.306	2.196	.604
R3Trust	61	8	1	9	5.75	1.972	3.889	-.548	.306	-.120	.604
R4Trust	61	6	3	9	7.49	1.164	1.354	-1.325	.306	2.839	.604
R1Rapport	61	6	3	9	7.30	1.145	1.311	-1.437	.306	3.765	.604
R2Rapport	61	8	1	9	6.70	1.716	2.945	-.812	.306	1.458	.604
R3Rapport	61	8	1	9	4.79	2.310	5.337	-.184	.306	-.981	.604
R4Rapport	61	5	4	9	6.93	1.289	1.662	-.646	.306	-.407	.604
Valid N (listwise)	61										

Phase 2

Descriptives

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
R1Liking	64	8	1	9	7.11	1.861	3.464	-1.247	.299	2.087	.590
R2Liking	64	7	2	9	6.23	1.520	2.309	-.439	.299	-.133	.590
R3Liking	64	6	3	9	6.19	1.355	1.837	-.036	.299	-.657	.590
R4Liking	64	4	5	9	6.86	1.125	1.266	-.336	.299	-1.109	.590
R1Attention	64	8	1	9	7.08	2.095	4.391	-1.357	.299	1.645	.590
R2Attention	64	5	4	9	7.78	.967	.936	-1.388	.299	3.352	.590
R3Attention	64	7	2	9	7.48	1.469	2.158	-1.611	.299	3.313	.590
R4Attention	64	4	5	9	7.27	1.116	1.246	-.339	.299	-.763	.590
R1Coordination	64	8	1	9	6.36	2.522	6.361	-.819	.299	-.284	.590
R2Coordination	64	5	4	9	7.25	1.309	1.714	-.701	.299	.055	.590
R3Coordination	64	7	2	9	6.94	1.754	3.075	-.722	.299	-.095	.590
R4Coordination	64	5	4	9	6.80	1.211	1.466	-.648	.299	-.394	.590
R1Trust	64	8	1	9	7.13	2.149	4.619	-1.326	.299	1.237	.590
R2Trust	64	5	4	9	7.30	1.136	1.291	-.683	.299	.202	.590
R3Trust	64	7	2	9	6.84	1.556	2.420	-.646	.299	.654	.590
R4Trust	64	5	4	9	6.98	1.420	2.016	-.624	.299	-.415	.590
R1Rapport	64	8	1	9	6.09	2.700	7.293	-.582	.299	-.815	.590
R2Rapport	64	5	4	9	7.09	1.281	1.642	-.648	.299	-.443	.590
R3Rapport	64	7	2	9	6.34	1.766	3.118	-.277	.299	-.843	.590
R4Rapport	64	4	5	9	6.91	1.165	1.356	-.124	.299	-.936	.590
Valid N (listwise)	64										

APPENDIX H – FULL INTERACTION EXPLORATORY PRINCIPLE ANALYSIS

Factor Analysis

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.886
Bartlett's Test of Sphericity	Approx. Chi-Square	1863.927
	df	10
	Sig.	.000

Communalities

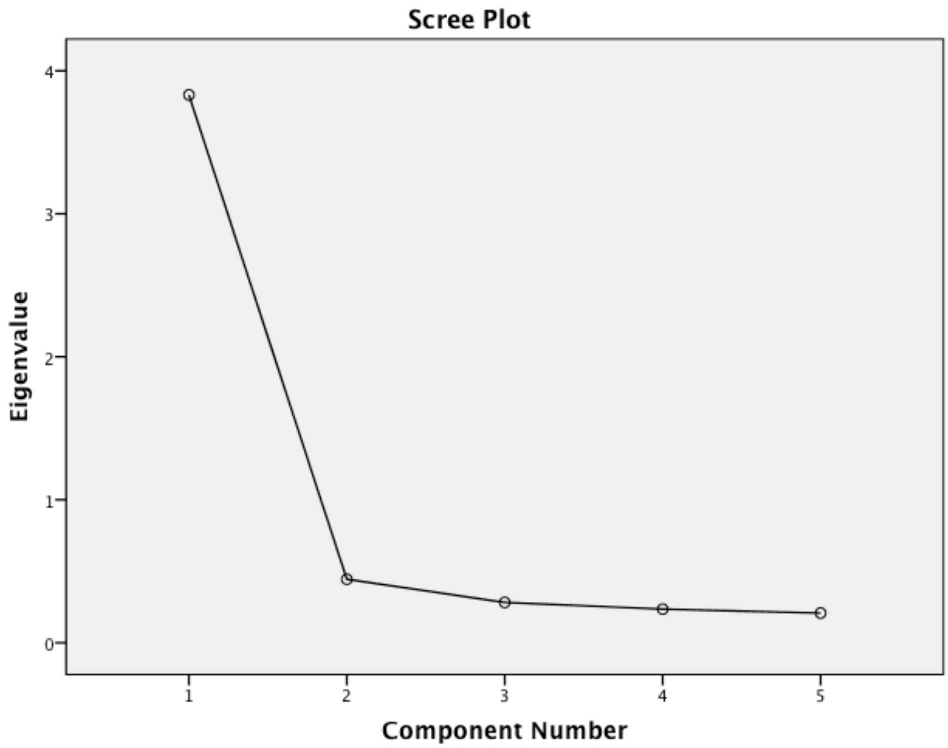
	Initial	Extraction
Liking_4	1.000	.701
Attention_4	1.000	.778
Coordination_4	1.000	.794
Trust_4	1.000	.747
Rapport_4	1.000	.811

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.831	76.622	76.622	3.831	76.622	76.622
2	.444	8.883	85.505			
3	.282	5.642	91.147			
4	.235	4.704	95.851			
5	.207	4.149	100.000			

Extraction Method: Principal Component Analysis.



Component Matrix[a]

	Component
	1
Liking_4	.837
Attention_4	.882
Coordination_4	.891
Trust_4	.864
Rapport_4	.900

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

APPENDIX I – THIN SLICE INDIVIDUAL ITEM DESCRIPTIVE STATISTICS

Phase 1

Descriptives

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
R1Liking	183	5	4	9	6.47	1.083	1.174	.104	.180	-.373	.357
R2Liking	183	8	1	9	5.71	1.693	2.866	-.512	.180	.198	.357
R3Liking	183	6	3	9	6.17	1.210	1.464	.280	.180	.270	.357
R4Liking	183	6	3	9	6.44	1.420	2.017	-.435	.180	-.184	.357
R1Attention	183	5	4	9	7.57	1.136	1.291	-.682	.180	-.019	.357
R2Attention	183	8	1	9	7.27	1.468	2.156	-1.420	.180	3.399	.357
R3Attention	183	8	1	9	7.19	1.647	2.713	-1.294	.180	2.123	.357
R4Attention	183	6	3	9	7.54	1.194	1.426	-.852	.180	.735	.357
R1Coordination	183	6	3	9	6.84	1.438	2.068	-.469	.180	-.553	.357
R2Coordination	183	8	1	9	6.24	1.582	2.502	-.841	.180	1.364	.357
R3Coordination	183	8	1	9	5.96	1.612	2.597	-.415	.180	.031	.357
R4Coordination	183	7	2	9	6.98	1.170	1.368	-.801	.180	1.592	.357
R1Trust	183	6	3	9	6.98	1.307	1.709	-.372	.180	-.434	.357
R2Trust	183	8	1	9	6.13	1.790	3.202	-.473	.180	.007	.357
R3Trust	183	8	1	9	5.97	1.522	2.318	-.615	.180	.929	.357
R4Trust	183	7	2	9	7.12	1.270	1.612	-.766	.180	.749	.357
R1Rapport	183	6	3	9	6.93	1.359	1.847	-.561	.180	.252	.357
R2Rapport	183	8	1	9	5.66	1.903	3.623	-.502	.180	.294	.357
R3Rapport	183	8	1	9	5.62	1.917	3.676	-.287	.180	-.391	.357
R4Rapport	183	6	3	9	6.58	1.264	1.597	-.437	.180	-.124	.357
Valid N (listwise)	183										

Phase 2

Descriptives

Descriptive Statistics

	N	Range	Minimum	Maximum	Mean	Std. Deviation	Variance	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
R1Liking	192	7	2	9	6.95	1.324	1.752	-.981	.175	1.534	.349
R2Liking	192	8	1	9	5.92	1.505	2.265	-.369	.175	.023	.349
R3Liking	192	6	3	9	6.06	1.110	1.232	.270	.175	-.280	.349
R4Liking	192	4	4	8	6.02	.805	.649	.084	.175	-.545	.349
R1Attention	192	7	2	9	6.93	1.388	1.927	-.720	.175	.420	.349
R2Attention	192	4	5	9	8.04	.894	.800	-.693	.175	-.028	.349
R3Attention	192	7	2	9	7.29	1.571	2.467	-1.246	.175	1.354	.349
R4Attention	192	4	4	8	6.41	.781	.610	.065	.175	-.033	.349
R1Coordination	192	8	1	9	6.60	1.628	2.649	-.667	.175	.214	.349
R2Coordination	192	6	3	9	7.09	1.258	1.583	-.816	.175	.772	.349
R3Coordination	192	7	2	9	6.49	1.535	2.356	-.573	.175	-.158	.349
R4Coordination	192	5	4	9	6.13	.897	.805	.268	.175	.173	.349
R1Trust	192	8	1	9	6.83	1.682	2.831	-1.087	.175	1.035	.349
R2Trust	192	6	3	9	7.21	1.500	2.250	-.888	.175	.084	.349
R3Trust	192	5	4	9	6.79	1.265	1.600	-.212	.175	-.494	.349
R4Trust	192	5	4	9	6.49	1.028	1.058	.204	.175	-.662	.349
R1Rapport	192	8	1	9	6.56	1.736	3.012	-.844	.175	.510	.349
R2Rapport	192	7	2	9	6.51	1.476	2.178	-.682	.175	.138	.349
R3Rapport	192	7	2	9	6.40	1.425	2.032	-.450	.175	.033	.349
R4Rapport	192	4	4	8	6.14	.922	.851	.334	.175	-.440	.349
Valid N (listwise)	192										

APPENDIX J – THIN SLICE EXPLORATORY PRINCIPLE COMPONENT ANALYSIS

Factor Analysis

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.865
Bartlett's Test of Sphericity	Approx. Chi-Square	4673.893
	df	10
	Sig.	.000

Communalities

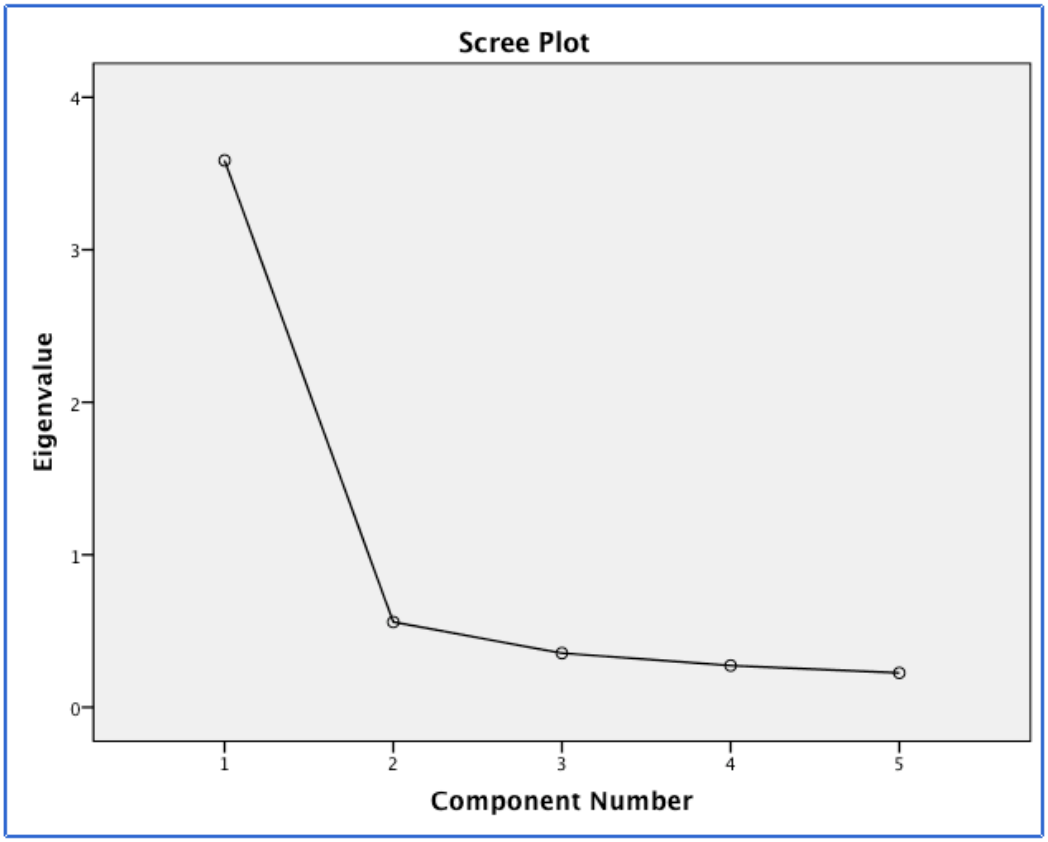
	Initial	Extraction
Liking_TH	1.000	.650
Attention_TH	1.000	.631
Coordination_TH	1.000	.737
Trust_TH	1.000	.777
Rapport_TH	1.000	.791

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3.586	71.719	71.719	3.586	71.719	71.719
2	.559	11.188	82.907			
3	.355	7.105	90.013			
4	.274	5.475	95.488			
5	.226	4.512	100.000			

Extraction Method: Principal Component Analysis.



Component Matrix[a]

	Component
	1
Liking_TH	.806
Attention_TH	.794
Coordination_TH	.858
Trust_TH	.882
Rapport_TH	.889

Extraction Method: Principal Component Analysis.

a. 1 components extracted.

APPENDIX K – CORRELATIONS BETWEEN THIN SLICE AND FULL INTERACTION RATINGS

Correlations

		FullLiking	FullAttention	FullCoordination	FullTrust	FullRapport	FullComposite
ThinLiking	Pearson Correlation	**.468****	.354**	.379**	.385**	.373**	.457**
	Sig. (2-tailed)	**.000**	.000	.000	.000	.000	.000
	N	**125**	125	125	125	125	125
ThinAttention	Pearson Correlation	.424**	**.335****	.341**	.205*	.256**	.359**
	Sig. (2-tailed)	.000	**.000**	.000	.022	.004	.000
	N	125	**125**	125	125	125	125
ThinCoordination	Pearson Correlation	.485**	.371**	**.390****	.343**	.383**	.459**
	Sig. (2-tailed)	.000	.000	**.000**	.000	.000	.000
	N	125	125	**125**	125	125	125
ThinTrust	Pearson Correlation	.423**	.289**	.371**	**.325****	.324**	.405**
	Sig. (2-tailed)	.000	.001	.000	**.000**	.000	.000
	N	125	125	125	**125**	125	125
ThinRapport	Pearson Correlation	.494**	.384**	.399**	.332**	**.363****	.458**
	Sig. (2-tailed)	.000	.000	.000	.000	**.000**	.000
	N	125	125	125	125	**125**	125
ThinComposite	Pearson Correlation	.527**	.398**	.433**	.368**	.392**	**.493****
	Sig. (2-tailed)	.000	.000	.000	.000	.000	**.000**
	N	125	125	125	125	125	**125**

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

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

APPENDIX L – SIMPLE LINEAR REGRESSION: PREDICTING FULL INTERACTION RATINGS FROM THIN SLICE RATINGS

Regression

Variables Entered/Removed[a]

Model	Variables Entered	Variables Removed	Method
1	ThinComposite[b]	.	Enter

a. Dependent Variable: FullComposite

b. All requested variables entered.

Model Summary[b]

					Change Statistics					
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change	df1	df2	Sig. F Change	Durbin-Watson
1	.493[a]	.243	.237	1.14310	.243	39.428	1	123	.000	1.700

a. Predictors: (Constant), ThinComposite

b. Dependent Variable: FullComposite

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	51.520	1	51.520	39.428	.000[b]
	Residual	160.721	123	1.307		
	Total	212.241	124			

a. Dependent Variable: FullComposite

b. Predictors: (Constant), ThinComposite

Coefficients[a]

		Unstandardized Coefficients		Standardized Coefficients			95.0% Confidence Interval for B	
Model		B	Std. Error	Beta	t	Sig.	Lower Bound	Upper Bound
1	(Constant)	1.714	.845		2.028	.045	.041	3.387
	ThinComposite	.794	.127	.493	6.279	.000	.544	1.045

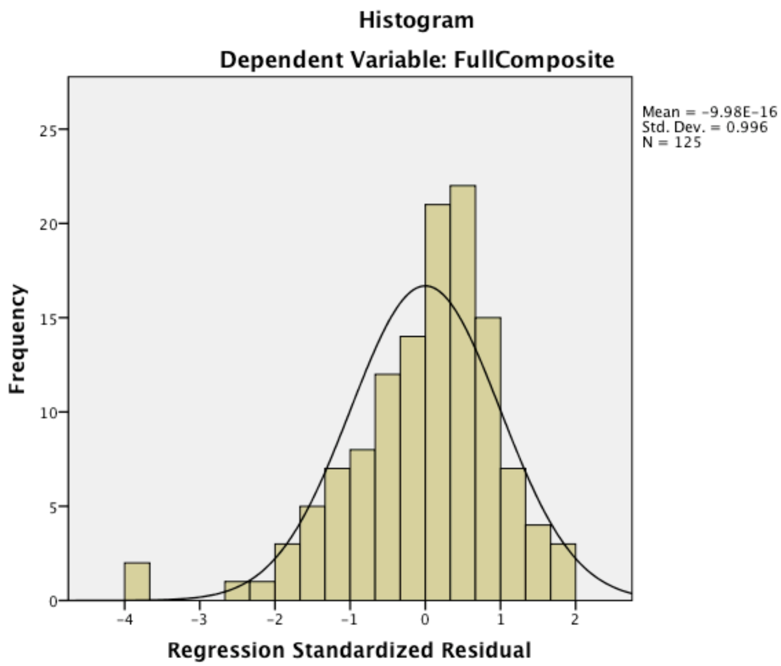
a. Dependent Variable: FullComposite

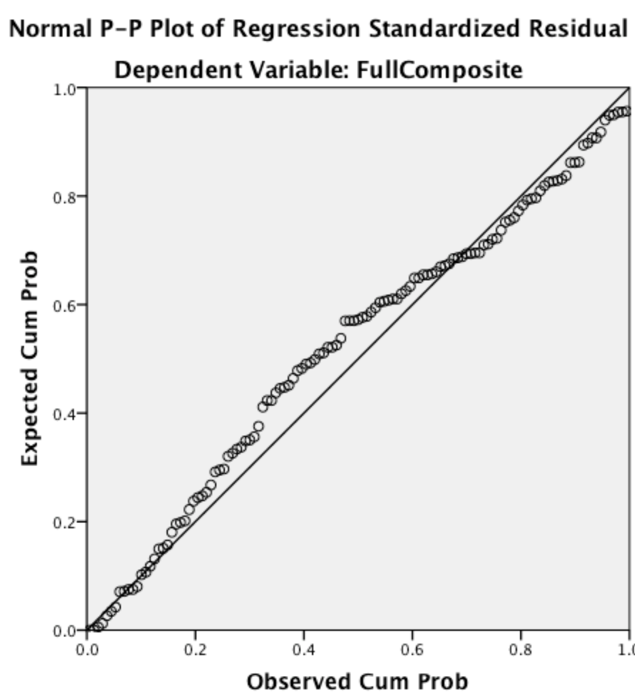
Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	5.3149	8.5982	6.9824	.64458	125
Residual	-4.31491	1.95553	.00000	1.13848	125
Std. Predicted Value	-2.587	2.507	.000	1.000	125
Std. Residual	-3.775	1.711	.000	.996	125

a. Dependent Variable: FullComposite

Charts


Histogram
Dependent Variable: FullComposite
Mean = -9.98E-16
Std. Dev. = 0.996
N = 125
Frequency
Regression Standardized Residual
0
5
10
15
20
25
-4
-3
-2
-1
0
1
2


Normal P-P Plot of Regression Standardized Residual
Dependent Variable: FullComposite
Expected Cum Prob
Observed Cum Prob
0.0
0.2
0.4
0.6
0.8
1.0

REFERENCES

Albrecht, T. L., Eggly, S. S., Gleason, M. E., Harper, F. W., Foster, T. S., Peterson, A. M., . . . Ruckdeschel, J. C. (2008). Influence of clinical communication on patients' decision making on participation in clinical trials. *Journal of Clinical Oncology, 26*(16), 2666-2673.

Albrecht, T. L., Ruckdeschel, J. C., Ray, F. L., 3rd, Pethe, B. J., Riddle, D. L., Strohm, J., . . . Blanchard, C. G. (2005). A portable, unobtrusive device for videorecording clinical interactions. *Behavioral Research Methods, 37*(1), 165-169.

Ambady, N., Hallahan, M., & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *J Pers Soc Psychol, 77*(3), 538-547.

Ambady, N., Koo, J., Rosenthal, R., & Winograd, C. H. (2002a). Physical therapists' nonverbal communication predicts geriatric patients' health outcomes. *Psychol Aging, 17*(3), 443-452.

Ambady, N., Krabbenhoft, M. A., & Hogan, D. (2006). The 30-sec sale: Using thin-slice judgments to evaluate sales effectiveness. *Journal of Consumer Psychology, 16*(1), 4-13.

Ambady, N., Laplante, D., Nguyen, T., Rosenthal, R., Chaumeton, N., & Levinson, W. (2002b). Surgeons' tone of voice: a clue to malpractice history. *Surgery, 132*(1), 5-9.

Ambady, N., LePlante, D., & Johnson, E. (2001). Thin slice judgements as a measure of interpersonal sensitivity. In J. A. Hall & F. J. Bernieri (Eds.), Interpersonal sensitivity theory and measurement (pp. 89-103). Mahwah, N.J.: L. Erlbaum Associates. Retrieved from http://proxy.lib.wayne.edu/login?url=http://www.netLibrary.com/urlapi.asp?action=summary&v=1&bookid=63044.

Ambady, N., & Rosenthal, R. (1992). Thin slices of expressive behavior as predictors of interpersonal consequences: A meta-analysis. *Psychol Bull, 111*(2), 256-274. doi: 10.1037/0033-2909.111.2.256.

Ambady, N., & Rosenthal, R. (1993). Half a minute: Predicting teacher evaluations from thin slices of nonverbal behavior and physical attractiveness. *J Pers Soc Psychol, 64*(3), 431-441. doi: 10.1037/0022-3514.64.3.431

Ambady, N., & Rosenthal, R. (1997). Judging social behavior using "thin slices". *Chance, 10*(4), 12-18.

Beck, R. S., Daughtridge, R., & Sloane, P. D. (2002). Physician-patient communication in the primary care office: a systematic review. *Journal of the American Board of Family Practice, 15*(1), 25-38.

Braveman, P. A., Cubbin, C., Egerter, S., Williams, D. R., & Pamuk, E. (2010). Socioeconomic disparities in health in the United States: What the patterns tell us. *Am J Public Health, 100*(S1), S186-S196. doi: 10.2105/ajph.2009.166082

Casagrande, S. S., Gary, T. L., LaVeist, T. A., Gaskin, D. J., & Cooper, L. A. (2007). Perceived discrimination and adherence to medical care in a racially integrated community. *J Gen Intern Med, 22*(3), 389-395. doi: 10.1007/s11606-006-0057-4

Centers for Disease Control. (2005). Health Disparities Experienced by Black or African Americans - United States. *MMWR 2005, 54*(1), 1-3. http://www.cdc.gov/mmwr/preview/mmwrhtml/mm5401a1.htm.

Cline, R. J., Harper, F. W., Penner, L. A., Peterson, A. M., Taub, J. W., & Albrecht, T. L. (2006). Parent communication and child pain and distress during painful pediatric cancer treatments. *Social Science and Medicine, 63*(4), 883-898.

Cook, D. A., & Beckman, T. J. (2006). Current concepts in validity and reliability for psychometric instruments: theory and application. *American Journal of Medicine, 119*(2), 166.e167-166.e116. doi: 10.1016/j.amjmed.2005.10.036

Copeland, V. C. (2005). African Americans: Disparities in health care access and utilization. *Health and Social Work, 30*(3), 265-270.

Dovidio, J. F., Penner, L. A., Albrecht, T. L., Norton, W. E., Gaertner, S. L., & Shelton, J. N. (2008). Disparities and distrust: the implications of psychological processes for understanding racial disparities in health and health care. *Social Science and Medicine, 67*(3), 478-486. doi: 10.1016/j.socscimed.2008.03.019.

Eggly, S., Albrecht, T. L., Harper, F. W., Foster, T., Franks, M. M., & Ruckdeschel, J. C. (2008). Oncologists' recommendations of clinical trial participation to patients. *Patient Educ Couns, 70*(1), 143-148.

Eggly, S., Harper, F. W., Penner, L. A., Gleason, M. J., Foster, T., & Albrecht, T. L. (2011). Variation in question asking during cancer clinical interactions: a potential source of disparities in access to information. *Patient Educ Couns, 82*(1), 63-68.

Fiscella, K., Meldrum, S., Franks, P., Shields, C. G., Duberstein, P., McDaniel, S. H., & Epstein, R. M. (2004). Patient trust: is it related to patient-centered behavior of primary care physicians? *Med Care, 42*(11), 1049-1055.

Grahe, J. E., & Bernieri, F. J. (1999). The importance of nonverbal cues in judging rapport. *Journal of Nonverbal Behavior, 23*(4), 253-269. doi: 10.1023/a:1021698725361.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis (6th ed.)*. Upper Saddle River, NJ: Pearson Prentis Hall.

Henry, S. G., & Eggly, S. (2012). How much time do low-income patients and primary care physicians actually spend discussing pain? A direct observation study. *J Gen Intern Med, 27*(7), 787-793. doi: 10.1007/s11606-011-1960-x

Henry, S. G., & Eggly, S. (2013). The effect of discussing pain on patient-physician communication in a low-income, black, primary care patient population. *J Pain, 14*(7), 759-766. doi: 10.1016/j.jpain.2013.02.004

Hinkle, D. E., Wiersma, W., & Jurs, S. G. (1998). *Applied statistics for the behavioral sciences (4th ed.)*. Boston, MA: Houghton Mifflin Company.

Hootman, J. M., Driban, J. B., Sitler, M. R., Harris, K. P., & Cattano, N. M. (2011). Reliability and validity of three quality rating instruments for systematic reviews of observational studies. *Research Synthesis Methods, 2*(2), 110-118. doi: 10.1002/jrsm.41

Kirby, J. B., Taliaferro, G., & Zuvekas, S. H. (2006). Explaining racial and ethnic disparities in health care. *Med Care, 44*(5), I-64-72.

Kraus, M. W., & Keltner, D. (2009). Signs of socioeconomic status: a thin-slicing approach. *Psychol Sci, 20*(1), 99-106. doi: 10.1111/j.1467-9280.2008.02251.x.

Kundle, H. L., & Polansky, M. (2003). Measurement of Observer Agreement. *Radiology, 228*, 303-308. doi: 10.1148/radiol.2282011860

Landis, J. R., King, T. S., Choi, J. W., Chinchilli, V. M., & Koch, G. G. (2011). Measures of agreement and concordance with clinical research applications. *Statistics in Biopharmaceutical Research, 3*(2), 185-209. doi: 10.1198/sbr.2011.10019.

LaVeist, T., Thorpe, R., Galarraga, J., Bower, K., & Gary-Webb, T. (2009). Environmental and socio-economic factors as contributors to racial disparities in diabetes prevalence. *J Gen Intern Med, 24*(10), 1144-1148. doi: 10.1007/s11606-009-1085-7

Levinson, W., Roter, D. L., Mullooly, J. P., Dull, V. T., & Frankel, R. M. (1997). Physician-patient communication - The relationship with malpractice claims among primary care physicians and surgeons. *Jama-Journal of the American Medical Association, 277*(7), 553-559. doi: DOI 10.1001/jama.277.7.553

Lippa, R. A., & Dietz, J. K. (2000). The relation of gender, personality, and intelligence to judges' accuracy in judging strangers' personality from brief video segments. *Journal of Nonverbal Behavior, 24*(1), 25-43. doi: 10.1023/a:1006610805385.

Mays, N., & Pope, C. (1995). Qualitative research: Observational methods in health care settings. *Br Med J, 311*(6998), 182-184. doi: 10.1136/bmj.311.6998.182

Murphy, N. A. (2005). Using thin slices for behavioral coding. *Journal of Nonverbal Behavior, 29*(4), 235-246. doi: 10.1007/s10919-005-7722-x.

Murphy, N. A., Hall, J. A., & Colvin, C. R. (2003). Accurate intelligence assessments in social interactions: mediators and gender effects. *J Pers, 71*(3), 465-493.

Penner, L. A., Cline, R. J., Albrecht, T. L., Harper, F. W., Peterson, A. M., Taub, J. M., & Ruckdeschel, J. C. (2008). Parents' empathic responses and pain and distress in pediatric patients. *Basic Appl Soc Psych, 30*(2), 102-113.

Penner, L. A., Dovidio, J. F., Edmondson, D., Dailey, R. K., Markova, T., Albrecht, T. L., & Gaertner, S. L. (2009). The Experience of Discrimination and Black-White Health Disparities in Medical Care. *Journal of Black Psychology, 35*(2), 180-203. doi: Doi 10.1177/0095798409333585

Penner, L. A., Dovidio, J. F., West, T. V., Gaertner, S. L., Albrecht, T. L., Dailey, R. K., & Markova, T. (2010). Aversive racism and medical interactions with black patients: A field study. *J Exp Soc Psychol, 46*(2), 436-440. doi: 10.1016/j.jesp.2009.11.004.

Penner, L. A., Orom, H., Albrecht, T. L., Franks, M. M., Foster, T. S., & Ruckdeschel, J. C. (2007). Camera-related behaviors during video recorded medical interactions. *Journal of Nonverbal Behavior, 31*(2), 99-117. doi: DOI 10.1007/s10919-007-0024-8

Peracchio, L. A., & Luna, D. (2006). The role of thin-slice judgments in consumer psychology. *Journal of Consumer Psychology, 16*(1), 25-32.

Peterson, A., Cline, R., Foster, T., Penner, L., Parrott, R., Keller, C., . . . Albrecht, T. (2007). Parents’ interpersonal distance and touch behavior and child pain and distress during painful pediatric oncology procedures. *Journal of Nonverbal Behavior, 31*(2), 79-97.

Price, E. G., Windish, D. M., Magaziner, J., & Cooper, L. A. (2008). Assessing validity of standardized patient ratings of medical students' communication behavior using the Roter interaction analysis system. *Patient Educ Couns, 70*(1), 3-9. doi: 10.1016/j.pec.2007.10.002.

Quinn, S. C., Kumar, S., Freimuth, V. S., Musa, D., Casteneda-Angarita, N., & Kidwell, K. (2011). Racial disparities in exposure, susceptibility, and access to health care in the US H1N1 influenza pandemic. *Am J Public Health, 101*(2), 285-293. doi: 10.2105/ajph.2009.188029

Richeson, J. A., & Shelton, J. N. (2005). Brief report: Thin slices of racial bias. *Journal of Nonverbal Behavior, 29*(1), 75-86. doi: 10.1007/s10919-004-0890-2.

Roter, D., Hall, J. A., Blanch-Hartigan, D., Larson, S., & Frankel, R. M. (2011). Slicing it thin: new methods for brief sampling analysis using RIAS-coded medical dialogue. *Patient Educ Couns, 82*(3), 410-419.

Roter, D., & Larson, S. (2002). The Roter interaction analysis system (RIAS): Utility and flexibility for analysis of medical interactions. *Patient Educ Couns, 46*(4), 243-251.

Sheppard, V. B., Zambrana, R. E., & O'Malley, A. S. (2004). Providing health care to low-income women: a matter of trust. *Fam Pract, 21*(5), 484-491. doi: 10.1093/fampra/cmh503

Smedley, B. D., Stith, A. Y., & Nelson, A. R. (Eds.). (2003). *Unequal treatment: Confronting racial and ethnic disparities in health care*. Washington, D.C.: National Academy Press.

Stepanikova, I. (2006). Patient-physician racial and ethnic concordance and perceived medical errors. *Social Science and Medicine, 63*(12), 3060-3066. doi: 10.1016/j.socscimed.2006.08.015.

Stepanikova, I., Mollborn, S., Cook, K. S., Thom, D. H., & Kramer, R. M. (2006). Patients' race, ethnicity, language, and trust in a physician. *J Health Soc Behav, 47*(4), 390-405.

Thrasher, A. D., Earp, J. A., Golin, C. E., & Zimmer, C. R. (2008). Discrimination, distrust, and racial/ethnic disparities in antiretroviral therapy adherence among a national sample of HIV-infected patients. *Journal of Acquired Immune Deficiency Syndrome, 49*(1), 84-93. doi: 10.1097/QAI.0b013e3181845589

Tian, N., Goovaerts, P., Zhan, F. B., Chow, T. E., & Wilson, J. G. (2012). Identifying risk factors for disparities in breast cancer mortality among African-American and Hispanic women. *Women's Health Issues, 22*(3), e267-e276. doi: 10.1016/j.whi.2011.11.007

ABSTRACT

THE RELIABILITY AND VALIDITY OF THE THIN SLICE TECHNIQUE: OBSERVATIONAL RESEARCH ON VIDEO RECORDED MEDICAL INTERACTIONS

by

TANINA S. FOSTER

May 2014

Advisor: Dr. Shlomo Sawilowsky

Major: Evaluation and Research

Degree: Doctor of Philosophy

Introduction: Observational research using the thin slice technique has been routinely incorporated in observational research methods, however there is limited evidence supporting use of this technique compared to full interaction coding. The purpose of this study was to determine if this technique could be reliability coded, if ratings are consistent between the first, second and third slice, and if they are representative of full interactions.

Methods: Three 30-second thin slices were sampled from the beginning, middle and end of a full-length video-recorded patient/physician interaction collected a part of a larger research study in a low income urban primary care clinic. Thin slice excerpts and full interactions were rated on five dimensions (liking, attention, coordination, trust and rapport) using a nine point Likert scale ranging from 'none' to 'high' by eight independent coders. Reliability was assessed using the intraclass correlation measure, validity of thin slices was assessed using the Friedman test (non-parametric equivalent of the Repeated measures ANOVA), and the comparison of thin slice coding to full

interaction coding was assessed using the Wilcoxon Sign Ranks test (nonparametric version of the Paired t-test).

Results: Thin slice reliability on Likert scale items ranged from .762-.910 with an average IRR of .850. Friedman tests conducted on all five variables (liking, attention, coordination, trust and rapport) comparing the rating of the three slices of the interaction were non-significant. Results of the Wilcoxon Signed Ranks test indicated there was a significant difference between the composite thin slice rating (average across three slices) and the full interaction ratings with full interaction variables rated consistently higher than their respective thin slice composite.

Conclusion: Results indicate that thin slices can be reliability coded by independent coders with a high degree of agreement across coders. Observational ratings across thin slices sampled at the beginning middle and end of an interaction were not significantly different demonstrating convergent validity. However, there was a significant difference between ratings obtained from thin slices and ratings obtained from the full interaction, indicating care should be taken when thin slices are used to represent the interaction as a whole.

AUTOBIOGRAPHICAL STATEMENT

Tanina S. Foster

Professional Preparation

B.S.	Public Health Education and Family Studies	Central Michigan University	1997
M.Ed.	Educational Evaluation & Research	Wayne State University	2005
GCPHP	Graduate Certificate – Public Health Practice	Wayne State University	2012

Professional Appointments

1997-2004	Project/Program Coordinator, Karmanos Cancer Institute
2004-2009	Laboratory Manager, Karmanos Cancer Institute
2009-Present	Laboratory Manager, Behavioral and Field Research Core, Population Studies and Disparities Program, Department of Oncology, Wayne State University

Professional Societies

American Statistical Association

Publications

1. Eggly, S., Penner, L., Albrecht, T, Cline, R., **Foster, T**., Naughton, M., Peterson, A., & Ruckdeschel, J. (2006) Discussing 'bad news' in the outpatient oncology clinic: Rethinking current communication guidelines. *Journal of Clinical Oncology, 24*, 716-719.
2. Cline, R. J. W., Penner, L. A., Harper, F. W. K., **Foster, T. S**., Ruckdeschel, J. C., & Albrecht, T. L. (2007). Patients' Internet use for cancer information and socioeconomic status as factors in oncologist-patient communication. *Journal of Oncology Practice, 3*, 167-171.
3. Penner, L. A., Orom, H., Albrecht, T., Franks, M., **Foster, T**., & Ruckdeschel, J. C. (2007). Camera-related behaviors during video recorded medical interactions. *Journal of Nonverbal Behavior, 31,* 99-117.
4. Peterson, A. M., Cline, R. J. W., **Foster, T. S**., Penner, L. A., Parrott, R. L., Keller, C. M., Naughton. M. C., Taub, J. W., Ruckdeschel, J. C., & Albrecht, T. L. (2007). Parents' interpersonal distance and touch behavior and child pain and distress during painful pediatric oncology procedures. *Journal of Nonverbal Behavior, 31,* 79-97.
5. Albrecht, T., Eggly, S., Gleason, M., Harper, F., **Foster, T**., Peterson, A., Orom, H., Penner, L., & Ruckdeschel, J. (2008). The influence of clinical communication on patients' decision making about clinical trials. *Journal of Clinical Oncology, 26(16)*, 2666-2673.
6. Eggly, S., Albrecht, T. L., Harper, F. W. K , **Foster, T**., Franks, M. M., Ruckdeschel, J. C. (2008). Oncologists' recommendation of clinical trial participation to patients. *Patient Education and Counseling, 70,* 143-148.
7. Eggly, S., Harper, F. W., Penner, L.A., Gleason, M. J., **Foster, T.**, Albrecht, T.L. (2011). Variation in question asking during cancer clinical interactions: A potential source of disparities in access to information. *Patient Education and Counseling 82*, 63-68.
8. Carcone, A. I., Naar-King, S., Brogan, K. E., Albrecht, T., Barton, E., **Foster, T.**, Martin, T., & Marshall, S. (2013). Provider Communication Behaviors that Predict Motivation to Change in Black Adolescents with Obesity. *Journal of Developmental and Behavioral Pediatrics 34,* 599-608.