A STUDY OF PHYSICIANS' SERENDIPITOUS KNOWLEDGE DISCOVERY: AN EVALUATION OF SPARK AND THE IF-SKD MODEL IN A CLINICAL SETTING

Mark E. Hopkins

Dissertation Prepared for the Degree of

DOCTOR OF PHILOSOPHY

UNIVERSITY OF NORTH TEXAS

May 2018

APPROVED:

Oksana L. Zavalina, Major Professor
T. Elizabeth Workman, Committee Member
LeRoy Southmayd III, Committee Member
Richard S. Herrington, Committee Member
Yunfei Du, Interim Chair of the Department of Information Science
Kinshuk, Dean of the College of Information
Victor Prybutok, Dean of the Toulouse Graduate School

Hopkins, Mark E. *A Study of Physicians' Serendipitous Knowledge Discovery: An Evaluation of Spark and the IF-SKD Model in a Clinical Setting*. Doctor of Philosophy (Information Science), May 2018, 190 pp., 29 tables, 28 figures, references, 88 titles.

This research study is conducted to test Workman, Fiszman, Rindflesch and Nahl's information flow-serendipitous knowledge discovery (IF-SKD) model of information behavior, in a clinical care context. To date, there have been few attempts to model the serendipitous knowledge discovery of physicians. Due to the growth and complexity of the biomedical literature, as well as the increasingly specialized nature of medicine, there is a need for advanced systems that can quickly present information and assist physicians to discover new knowledge. The National Library of Medicine's (NLM) Lister Hill Center for Biocommunication's Semantic MEDLINE project is focused on identifying and visualizing semantic relationships in the biomedical literature to support knowledge discovery. This project led to the development of a new information discovery system, Spark. The aim of Spark is to promote serendipitous knowledge discovery by assisting users in maximizing the use of their conceptual short-term memory to iteratively search for, engage, clarify and evaluate information presented from the biomedical literature. Using Spark, this study analyzes the IF- SKD model by capturing and analyzing physician feedback. The McCay-Peet, Toms and Kelloway's Perception of Serendipity and Serendipitous Digital Environment (SDE) questionnaires are used. Results are evaluated to determine whether Spark contributes to physicians' serendipitous knowledge discovery and the ability of the IF-SKD ability to capture physicians' information behavior in a clinical setting.

ACKNOWLEDGEMENTS

For my wife Shokooh and my daughter Veyana. The pursuit of knowledge is an endeavor with few worthy equals. Yet, it is not an adequate substitute for a life well lived with those we love. I cannot find the words to express how deeply I love you both. I feel so thankful to share this life with you. It is a grace I did not deserve, but for which I will always be grateful. You are both amazing, remarkable and strong and I look forward to a long life with you both so that I can hopefully give you back a fraction of what you've given me.

This journey would not have been possible without the support of others. I would like to thank my Committee Chair, Dr. Zavalina, for your support and confidence. Thank you for helping me move forward while on this arduous path! Drs. Workman and Southmayd, I also want to thank you both for sticking with me the whole way through this process. Your kind words, help and support have meant more than I can say. Dr. Herrington, I thank you for a whirlwind of support with the data analysis. It wouldn't have been possible without you and UNT's Research & Statistical Support Services. It was an exhausting journey, but a rewarding one!

I have always been a fan of Victor Hugo. And, while I don't have a favorite quote, I'm very fond of these words: "Idleness is the heaviest of all oppressions".

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION

Introduction and Background

This research study is conducted to test Workman, Fiszman, Rindflesch and Nahl's (2014) information flow – serendipitous knowledge discovery (IF-SKD) model of information behavior, in a clinical care context. To date, there have been few attempts to model the serendipitous knowledge discovery of physicians. Due to the growth and complexity of the biomedical literature, as well as the increasingly specialized nature of medicine, there is a need for advanced systems that can quickly present information and assist physicians to discover new knowledge. The National Library of Medicine's (NLM) Lister Hill Center for Biocommunication's Semantic MEDLINE project is focused on identifying and visualizing semantic relationships in the biomedical literature to support knowledge discovery. This project led to the development of a new information discovery system, Spark. The aim of Spark is to promote serendipitous knowledge discovery by assisting users in maximizing the use of their conceptual short-term memory to iteratively search for, engage, clarify and evaluate information presented from the biomedical literature. Using Spark, this study analyzes the IF-SKD model by capturing and analyzing physician feedback. The McCay-Peet, Toms and Kelloway's (2015) Perception of Serendipity and Serendipitous Digital Environment questionnaires are used. Results are evaluated to determine whether Spark contributes to physicians' serendipitous knowledge discovery and the ability of the IF-SKD ability to capture physicians' information behavior in a clinical setting

The concept, study and application of serendipitous knowledge discovery (SKD) are not new to the field of information science. Over the years, the idea of discovery through serendipity, or "information encountering" has been studied and understood using a variety of methods with varied outcomes, though some common themes (Erdelez, 1997). However, few formal models exist that are derived from information science and its literature. Foster and Ford (2003) noted that due to serendipity's "elusive, unpredictable" nature, SKD is challenging to understand within existing information behavior models (p. 321). Often, technical and psychosocial factors are at the heart of understanding this behavior. Intervening variables such as age, education, task, personality, information need, prior knowledge, etc. play a role as well (Burkell, Quan-Hase & Rubin, 2012; Heinström, 2006; Spink, 2004). Yet, despite these fundamental complexities, it is paramount that in today's richly complex information world, the study of serendipitous knowledge discovery remains a priority. The IF-SKD model is a step in this direction, and its further analysis only aids in its development and refinement.

In contrast to the study of SKD, physicians' information behavior is quite rich. Because some aspects of clinical workflow, and the often-required information technologies that drive it are process driven, there are numerous studies that review the utility of an array of information resource solutions within those workflows. For example, Del Fiol et al. (2012) and others have shown how context driven infobuttons information have helped meet clinical information needs. For years, electronic medical records (EMRs) have contained clinical alerting mechanisms designed to provide safety precautions within the use of certain activities, such as drug administration, when known issues exist that could cause harm, such as drug interactions and black box warnings. A strong motivation in the literature has been to maximize the utility, automation and breadth of content available to physicians, and then to study how that content (or system) was used, and whether access to the information impacted their clinical decision making. One study explored physicians' questions within the EMR workflow to ascertain the situational factors likely to resolve unmet information needs through the implementation of solutions that can address them specific points in the workflow (Cimino, Li, Bakken & Patel, 2002). Another collected and categorized the types of unmet information needs (Currie et al., 2003).

These are important topics yet are all predicated on a known (or anticipated) user information need. Accounting for the unformed and unknown needs of physicians, including how best to model those needs as well as apply them to the design of new tools and system, is an area needing further exploration. General models of SKD that have been developed are relatively new, especially in their application to system design and even more particularly in their lack of application to physicians within the clinical setting.

Essential to both physicians' information behavior and the idea of SKD is an understanding of existing information resources and content (including the rich taxonomies and controlled vocabularies contributing to it) that comprise the biomedical literature. Within the biomedical information space, there are numerous information resources. The National Library of Medicine (NLM) has been central to the creation and growth of these online databases and resources, which offer unique and powerful access to information. Years of careful and meaningful curation of underlying data has, in large part, made this possible. However, for many resources, there is the inherent assumption of a goal, or known information need by the user. Only recently have tools been designed to support serendipitous knowledge discovery for situations where a goal (or information need) is not present, or potentially unknown by the user.

These rich information resources, and their underlying metadata provide the ideal springboard from which to build new systems that can promote serendipitous discovery. Through improved system design, the meaningful identification of semantic relationships, and the use of information visualization, these new tools can assist in modeling an iterative information seeking process that improves not only outcomes, but also “reduc[es] the cognitive demands of information organization” by ultimately increasing the chance for serendipitous knowledge discovery (Workman, Fiszman, Rindflesch & Nahl, 2014). New systems should be built to support SKD within the clinical setting. The task of future researchers is to better understand how these systems should be examined in order to explain how system design equates with the discipline’s understanding of serendipitous knowledge discovery as a type of information behavior. In turn, this helps address another major challenge, which is the growth and specialization of biomedical information.

Statement of the Problem

The increased specialization of the medical field, along with the enormous growth of biomedical information, pose unique challenges for how best to identify, present and use information in effective ways within the clinical care setting. Physicians, who can benefit greatly through the consumption and application of relevant information, are often challenged with effective ways to discover it. Systems are just beginning to incorporate SKD design principles. Of those, few are effectively integrating rich data structures, such as semantic predications with effective visualization and refinement techniques. The Spark application, which is designed based on the recently developed IF-SKD model, provides opportunities to increase moments of serendipitous knowledge discovery.

At the present time, there is no understanding of Spark's efficacy to address these issues. A thorough analysis of Spark's ability, within an actual clinical setting, to promote SKD could be beneficial. By studying the IF-SKD model and analyzing Spark, this study extends recently published findings and poses new research questions that provide a better understanding of the use of Spark and the degree to which it can promote serendipitous discoveries within the clinical context.

Purpose of the Study

The purpose of this research study is to evaluate the online system Spark using the Information Flow - Serendipitous Knowledge Discovery (IF-SKD) model of information behavior in a clinical setting, developed by Workman, Fiszman, Rindflesch and Nahl (2014). The IF-SKD model is used to evaluate physicians' use of Spark, a tool designed to promote serendipitous knowledge discovery (SKD) using both the organization and visualization of semantic relationships derived from the biomedical literature. The design of Spark was done in consideration of "four core principles of SKD" derived from the information science literature (p. 24).

The core principles include: 1) SKD is an iterative process; 2) SKD often involves change or clarification of initial information interests, which may involve integrating new topics; 3) SKD is grounded in the user's prior knowledge; 4) Information organization and presentation have fundamental roles in SKD. These principles were also central to the IF-SKD's development. A major aim of this study is to explore the utility of the model in representing physicians' serendipitous knowledge discovery in a clinical setting.

Definitions

The following terms represent key concepts of interest to this research and assist in the understanding of how they are operationalized within the context of the study, as well as providing a general conceptual introduction.

- *Clinical care setting* – In this study, the clinical care setting is purposefully broad and could include a physician's office, the patient's room, the physician's home, the physicians' lounge(s), or other settings. Because workflow surrounding the acquisition of information can differ between providers, the goal is not to assume where a serendipitous event should occur, but rather understand how physicians' information behavior in using Spark correlated to the clinical care setting.

- *Information behavior* – In the context of this study, information behavior refers to two different aspects. First, it refers to the historical and studied information seeking behavior, information needs, and gaps encountered by physicians in their information acquisition activities. Second, it is a reference for the feedback from the questionnaire provided by physician participants regarding their experience with serendipitous knowledge discovery, utilizing Spark and generally. Together these present an unique viewpoint for how physicians engage in information behavior, and of a specific type of information acquisition relevant to the goals of the study.

- *Information flow-serendipitous knowledge discovery (IF-SKD) model* – An information behavior model, developed by Workman, Fiszman, Rindflesch and Nahl (2014), which outlines the stages of initial information engagement, through the visual representation of retrieved information that supports conceptual short-term memory evaluation, including the iterative clarifications or refinements of that searching, ultimately resulting in knowledge discovery. Four components underpin this model: 1) SKD is an iterative process; 2) SKD often involves change or clarification of initial information interests, which may involve integrating new topics; 3) SKD is grounded in the user's prior knowledge; 4) Information organization and presentation have fundamental roles in SKD (Workman, Fiszman, Rindflesch & Nahl 2014).
- *Physicians (MD/DO)* – The population identified for study includes physicians, with Doctor of Medicine (MD) or Doctor of Osteopathic Medicine (DO) credentials, working for the INTEGRIS system.
- *Semantic MEDLINE* – A project (and resource) created by the National Library of Medicine (NLM) that encompasses the application of natural language processing (NLP) to the identification of semantic predications derived from the MEDLINE database, as well as the use of those semantic predications in other applications; in particular the visual representation of predications to engage users in more effective information seeking behavior and knowledge discovery. Spark is an application created alongside, and makes use of the underlying aspects of Semantic MEDLINE.
- *Serendipitous knowledge discovery (SKD)* – This refers to the chance, or accidental discovery of new knowledge, where its encountering is done so without this being the express or known information of interest at the time of initial searching/browsing.

- *Spark* – An online system designed to support serendipitous knowledge discovery. Spark is constructed to support an iterative step process that maximizes users' conceptual short-term memory (CTSM). Through an initial search, or topic of interest, the user can refine and visually explore semantic relationships found within the biomedical literature. Users can adjust the presentation of these relationships using a set of retrieval affordance options by selecting for frequency of occurrence in the literature (*rare*, *common* or *all*), and by relation or concept (e.g. *therapy* and *drugs or chemical*).

- *Spark system factors* – These refer to the core components, or features, that make up the Spark application and which is studied as part of this research. They include: work space, graph presentation and retrieval affordance mechanisms.

    - *Work space* – This is the layout of Spark, in particular, the major left and right pane sections that permit information organization geared to support the CTSM process. This includes the radial connected graph in the left pane and the saved connections of interest in the right pane.

    - Graph presentation – This refers to the structure and visual layout of the results from an information search. The uses of colors and lines, as well as graph type are considered.

    - Retrieval Affordance Mechanisms – These represent options related to:

        - Frequency occurrence in the literature: All, common, rare

        - Concept type**:** Disorder, drugs genes, etc.

        - Relation type: Therapy, diagnosis, comorbidity

Research Questions

This study addresses two key questions.

R1: Does Spark successfully contribute to physicians' serendipitous knowledge discovery?

H1$_O$: Spark does not contribute to physicians' serendipitous knowledge discovery.

R2: Does the IF-SKD model reflect physician serendipitous knowledge discovery information behavior in the clinical care setting?

H2$_O$: The IF-SKD model does not reflect physician serendipitous knowledge discovery information behavior in the clinical care setting.

Assumptions

An important assumption about this study is that there is significant value in the serendipitous discovery of knowledge in the clinical setting. Another important assumption is that existing tools, and workflow, are unable to induce SKD events meaningfully. A third assumption is that users within the clinical setting have an interest in facilitating more SKD opportunities and would therefore be strongly vested in providing feedback that would be valuable to the overall interpretation of results.

Limitations

There are several limitations to this study. For one, it introduces potentially unknown environmental factors that could influence results, such as interruptions due to patient care, participants' technology familiarity, unknown biases to this type of information behavior among participants, and the study duration. In selecting to analyze Spark and the IF-SKD model within a context that considers system design aspects and underlying assumptions governing the model, other salient influencing variables could be missed.

While an enhanced understanding of how to operationalize the concept of serendipity, and better measure it, are anticipated products of this study, the concept of serendipity itself remains challenging to convey and measure in practice, and therefore serves as a limitation to the study. Aspects of this research are grounded in an understanding of the study of SKD to date. Nonetheless, it remains a challenging aspect of information behavior to measure and therefore could act as a limitation to the effectiveness of the study. Through analysis of the research methods and instruments used, including their ability to successfully measure SKD, improvements to future research could be possible.

Significance of the Study

This study has the potential to contribute to the understanding of the information seeking behavior of physicians. It provides an opportunity to test a new model of information behavior dealing with the complex topic of serendipitous knowledge discovery. Likewise, it allows for the assessment of Spark, *a* new information resource designed in consideration of this model. Results derived from this study assist in understanding the IF-SKD's general applicability to a clinical care setting and could spur further targeted research. Finally, results could serve to improve the field's overall understanding of serendipitous knowledge discovery and to suggest improvements to Spark that could help physicians and other medical providers in the future, which could in turn lead to improved patient outcomes. Relevant findings from the study could later be incorporated into the development of new research tools and avenues for future research.

Summary of the Chapter

The current state of biomedical information is vast and complex. The study of information behavior models, especially in the clinical care setting, to support serendipitous knowledge discovery is an area where research is needed. This understanding is paramount to improving the development of future information resources.

This research investigates the IF-SKD model's viability within a clinical environment and what factors contribute to that understanding. Additionally, it assesses Spark's system functionality and how it contributes to serendipitous knowledge discovery through the analysis of physician feedback.

The need to continually challenge existing methods of information behavior is important. This study aims to build upon existing work and to support the development and understanding of information resources that promote serendipitous knowledge discovery.

CHAPTER 2

RELATED LITERATURE

Introduction

The expansive scope of information seeking behavior is impressive. Many theories and models exist, which seem to grow in relative parallel with changes in technology and information. Yet, within the field of information science, there are few models specifically focused on serendipitous knowledge discovery (SKD), and within the context of the clinical setting, they are almost non-existent. As the information landscape, its systems and resources continue to grow, there is an increased need to study this type of information behavior.

One major reason for this is the difficulty in measuring the central concept, serendipity. Additional challenges exist in the administration, acquisition and collection of information from physicians engaged in patient care. This is due in part to physicians' routines which are complex and busy. There is a need to explore the theories and models that can explain, and moreover reinforce, the conceptual framework of serendipitous knowledge discovery. Research in the environments that users, in this case physicians, engage in as part of their normal information behavior is critical to capturing real world variables that can influence models within the field.

Traditional information resources, such as point of care tools, support relatively fixed types of information behavior in online environments and are often driven by specific known workflow. These work well for context driven types of questions, but are predicated on the user already having an idea of their information need, or a system being able to anticipate one based on workflow. Many information resources can also confine information presentation and limit how the user can explore relationships within the literature. Even Boolean logic searches can become unwieldy and unnecessarily limiting. This is not conducive to the discovery of information for which the user is unable to articulate a need, or perhaps is unaware of altogether, and limits the discovery of relationships that could have led to more serendipitous knowledge discoveries.

A careful review and analysis of the historical and theoretical origins of Spark and serendipitous knowledge discovery is significant to showing how this research can help address the aforementioned challenges. Before starting, however, some context is warranted. First, the idea or role of serendipity in discovery generally is explored. This is followed by a brief explanation and background on the state of the biomedical literature and the development of the Spark system, which encompasses how Spark searches and derives relationships from the biomedical literature. This awareness is important for understanding how Spark facilitates the meaningful retrieval and presentation of information for users. Next, an overview of serendipitous knowledge discovery, alongside a review of related information behavior theories and models, shows the origins and uses of this concept within the field of information science. It also highlights small, yet significant distinctions in how SKD is interpreted and operationalized. Then, a review of user characteristics associated with serendipitous knowledge discovery in the literature are noted with consideration to their relevance and influence on the research. Finally, the IF-SKD model is explored in detail. Analysis of the four core components underpinning its design are reviewed with attention to how the Spark system's design is influenced by the application of the IF-SKD model.

The Role of Serendipity in Discovery

While the study of serendipity can and has been challenging, the significance of serendipity as a component of information behavior is singularly valued as an integral component towards fostering future fortuitous, lucky, or accidental knowledge discovery. In a recent review that looked at the opportunities to utilize existing scientific knowledge to assist with the identification of new drugs to treat disease, and the costs often associated with these endeavors, the authors noted that serendipity was, and remains, an integral factor in many major drug discoveries (Prasad, Gupta & Agarwal 2016). While this may be due in part to how knowledge is absorbed and integrated, and the extent to how and when information is encountered and processed to add value, it demonstrates that serendipity still plays a major role in discovery.

Considering the role of serendipity in the larger context of information science, Agarwal (2015) discussed, using Wilson's (1999) nested model of information behavior, how serendipitous knowledge discovery is a logical extension of this model, placing the concept of serendipity within, but extended outside as well, the spectrum of traditional nested assumptions. In effect, Agarwal's (2015) framework visually captures the overlapping nature serendipitous events have on a traditional view of information behavior.

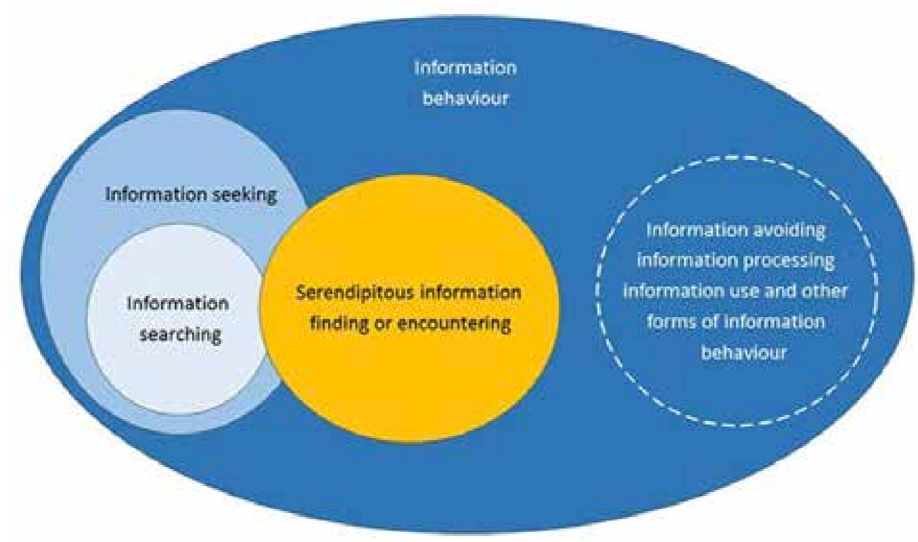

Figure 2-1. Agarwal's framework.

Working towards a definition of serendipity, Agarwal (2015) looked across the bow at the landscape of information science addressing the topic of serendipity to provide a view into the varying system, user, environmental and other factors that influence the occurrence of serendipity, and its place within the larger information science sphere. While covered in more detail in the review of information science models relevant to the concept of serendipity below, Agarwal (2015) presented a well cultivated set of contributions from the literature, demonstrating that serendipity is driven by numerous, often competing, facets of information seeking behavior. It is this confluence of variables that make honing in on a research approach that can be broadly applicable, even in the same domain, challenging.

The information science literature has shown that serendipity influences different aspects of the information environment. For example, the user, their disposition, or how prone they are to rely on serendipity, generally, have all been shown to correlate to how likely one might report experiencing a serendipitous information encounter (Heinström, 2006). McCay-

Peet & Toms (2010) and Cunha (2005) all noted that there is an apparent social aspect and activity orientation that lends itself to a higher amount of serendipity in information behavior.

Recently, Erdelez et al. (2016) participated in a panel that evaluated the concept of Serendipity in Information Science. They showed, notably, that while the concept of serendipity has been present in the literature for multiple decades, its targeted study has shown enormous growth in the literature in the past two decades. Figure 2-2 provides a summary of their findings.

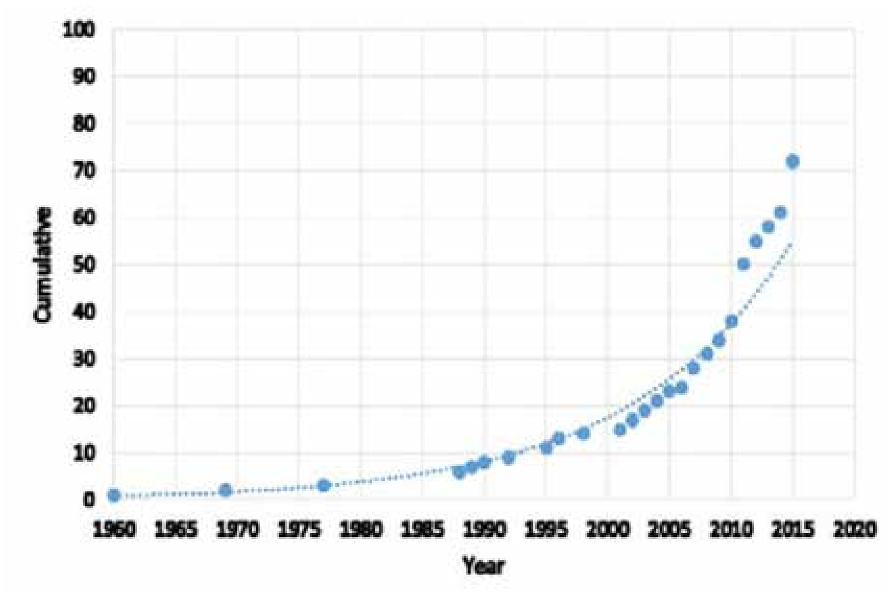

Figure 2-2. “Serendipity” information science research.

Physician Information Seeking Behavior

The application of serendipitous knowledge discovery to physician information seeking behavior is not expressly clear. While the preceding section expresses the important role and impact that serendipity can have on information discovery, the information needs and seeking behavior of physicians are complex and not always well-oriented towards being captured or expressed through information systems. Understanding the relationship and opportunities to bridge the application of serendipity and physicians' information seeking behavior is, in part, a goal of this study.

The following sections are designed to provide context surrounding physicians' information behavior. In presenting the literature on this topic, Case (2016) noted that the "overwhelming emphasis has been on how providers learn about things like treatment modalities, procedures, equipment, and medication" (p. 296). Capturing how physicians find this type information, and use it, is also challenging despite efforts made toward its automated study (Chen, Bakken, Currie, Patel & Cimino, 2006).

Gorman (1995) looked at both "information used" and "information need", which helped present a general understanding and growth on the subject. For "information used", Gorman (1995) pointed to five pieces of data that play a role in physician information behavior: patient data, population statistics, medical knowledge, logistical information, social influences. The information need that the physician might experience were captured by the following four types: recognized, pursued, satisfied, unrecognized (1995). This last type, unrecognized, is a type of need that relates to general concepts of information needs as described in the literature related to serendipitous knowledge discovery.

Other researchers have evaluated physicians' awareness of information resources, in addition to their use to better understand how to improve usage (Lialiou & Mantas, 2016). Le et al. (2016) conducted research that evaluated general practitioners' information behavior, awareness of resources, demographics and other user characteristics and found that there were not differences in how often physicians sought information based on gender. One interesting study looked at information use related to “patient care, knowledge development and research activities” and found that patient care information was positively associated with their perceived medical competence (Mikalef, Kourouthanassis & Pateli, 2017 p.58).

Case (2016) points out that physicians’ questions are challenging not only due to the nature of their work, but also due to how the field thinks about and attempts to measure them, which is complicated by the use of different study designs, sampling sizes, and types of providers. Adding to this is the fact that many physician questions are not followed up on, as Gorman and Helfand (1995) noted. However, Gorman (1999) later goes on to clarify that the way information need is defined impacts how we assess questions as not pursued. There is an inherently tangible component across studies that demonstrates that physician information needs are complex and rely on varied and equally complex types of data.

In addition to the categorizations and nuanced complexities of physicians’ information needs, the use of electronic resources and access to information is yet another piece of the puzzle that complicates physicians’ information seeking behavior. The format, presentation, access and modes of using information have changed greatly over the past 30 years, with a strong move towards utilizing electronic resources to answer clinical questions. Different medical information resources exist, such as clinical decision support systems that operate within a workflow for which specific likely questions are relatively known, to systems designed to support general information searching, to newer systems that seek to tap into those non-pursued, less easily articulated information needs.

Huang (1997) noted that colleagues and textbooks were preferred sources for many physicians, and while this study predates some of the improvements to online information resources, Younger (2010) also noted that the challenges related to usability and ease of access persist. Gorman (1995) noted that since physician questions are complex, "and often narrative in nature", that this may be part of the reason they "rely on human sources of information" over other electronic information resources (p. 734).

Another challenge for physicians is the idea of information overload. Bawden, Holtham and Courtney (1999) pointed out that "information overload occurs when information received becomes more of a hindrance rather than a help when the information is potentially useful" (p. 4). Information overload can be compounded by the challenge physicians sometimes face in being able "to convert their information need into a query that can be understood by the retrieval system (Clarke et al., 2013 p. 179). Davies and Harrison (2007), in reviewing a 10-year span of the literature, looked at "barriers to information searching", in addition to other challenges faced (p. 78). Bennet, Casebeer, Kristofco and Collins (2005) also looked at barriers among family physicians.

The different approaches to studying physicians' information seeking behavior and how to categorize their information needs has provided opportunities to assess the application of systems designed to promote serendipitous knowledge discovery to meet these needs. A study by Arborlelius and Timpka (1990) looked at the dilemmas or perplexing questions that physicians face. These types of studies that focus on the less known, more challenging types of information behavior, opened the door towards understanding how information systems could meet these types of needs.

Physician information seeking behavior, while well studied in the literature, has demonstrated that there is significant complexity to their information needs, and that physicians rely on a variety of resources in attempting to address their questions. There is an apparent need to improve information resources to better address the less defined, and often not pursued, clinical questions that arise. Focusing on systems that promote serendipity is one way to do this. An understanding of the challenges that exist in the biomedical literature and how these have translated into the development of new systems, in this case Spark, is important towards adding perspective that shows how the information needs of physician and the advent of new ways to represent and visualize information may help improve and address these needs. Moreover, this background helps provide useful information to consider alongside the development of new models of information behavior focused on capturing how serendipitous knowledge discovery occurs in electronic information systems.

Spark and the Biomedical Literature

From a high-level view, the field of information science is broadly concerned with the meaning of information, its definition, how it's organized, etc., as well as the information behavior of users. This research is primarily focused on understanding a subset of this behavior through the application of the IF-SKD model. The goal is to measure and report on the ability of the Spark system to promote serendipitous knowledge discovery among physicians, and the IF-SKD model's applicability to the clinical care setting.

Spark employs a unique approach of presenting summarized relationships within the biomedical literature to facilitate serendipitous knowledge discovery. Reflection on the contributing factors specific to the biomedical literature provides a broader understanding of this research and the implications these factors could have on the interpretation of results.

Consider that the most comprehensive quality database of medical information in the world, MEDLINE, which is used to support a variety of information resources, is made up of over 23 million articles. The magnitude of medical knowledge is so vast and diverse that its discovery and utility can easily be overshadowed by an inability to effectively evaluate it. In fact, in 2012, Goodwin, Cohen and Rindflesch succinctly noted that a "known contributor to knowledge deficiency in science is the body of scientific knowledge itself" (p. 232). Wilson (1995) also noted how "specialization, deferral, oversupply" were contributing negatively to providers being able to locate relevant information in a system (p. 47). It is this growth of information that complicates the application and development of tools for users and which has ultimately led to rich collaborations and research aimed at working to address this problem.

The origin of Spark is encapsulated in the storied history of the National Library of Medicine (NLM). The NLM's strategic initiatives focused on maintaining rich metadata on the biomedical literature is a major reason why Spark is possible. For example, the application of controlled terminologies, applied by experts, like Medical Subject Headings (MeSH), allow for semantically meaningful hierarchical relationships to be derived through natural language processing (NLP) techniques. These relationships are central to Spark's ability to present refined and meaningful information to users. Without these contributions by the NLM, it would be exceptionally more challenging to address the complex nature of medical information, in particular, the ability to effectively design for SKD information behavior.

Spark is then, in essence, an online information resource that is designed to support the serendipitous discovery of knowledge by allowing users to iteratively browse, refine and review the extracted meaningful relationships, semantic predications, from the biomedical literature.

Before Spark, there was *Semantic MEDLINE*. As its name implies, Semantic Medline resource was an initial application of these NLP derived semantic relationships within the literature. An enormous amount of research on the development of the associated resources and tools that were created and studied to ascertain the ability of these relationships to be accurately identified has been conducted. Rindflesch, Kilicoglu, Fiszman, Rosemblat and Shin (2011), aptly summarized how "automatic semantic interpretation is intended to augment document retrieval systems by manipulating information, not just documents, and thereby bridge the gap between text and meaning" (p. 15). Numerous Unified Medical Language System (UMLS) tools and resources have been tapped to support the automation and extraction of meaningful information from the biomedical literature.

Related research has considered the role of visualization and graphical representations of these derived relationships. For instance, the role in the application of graph theory and degree centrality and its effectiveness at identifying and presenting relationships to users has been studied (Zhang, Fiszman, Shin, Miller, Rosemblat & Rindflesch, 2011). Other studies have considered potential end-user applications, including the benefit to literature based discovery research (Workman, Fiszman, Rindflesch & Nahl, 2014). These brief examples demonstrate the principled and structural soundness that was crucial to the development Spark and the role and development of the information curation process that powers the system.

The Concept of Serendipitous Knowledge Discovery (SKD)

It is difficult to apply a model to any system development when the concept at the center of that model is conceptually vague. This vagueness also complicates its measurement.

Foster and Ford (2003) noted this inherent issue, stating that serendipity is “elusive, unpredictable” explaining why it does not fit prominently into existing models (p. 321). To understand serendipity’s use within this study, a review of the relevant related information behavior literature is presented. The general meaning of serendipity is identified across several studies, with some notably impactful theories and models discussed. In addition, principle contributors who have demonstrated the operationalization of serendipity in other studies, and for system development, are considered.

At a general level, the idea of serendipity in the information discovery process is often presented in the context of actions in which a user is engaged. In many cases, it is evaluated as a relevant finding, separate from the core objective of the study. This is not to say that these evaluations have been unhelpful; in fact, quite the contrary is true. Marchionini (1995) looked at serendipity, in the context of browsing, through a task-oriented design. He noted that design strategies are “well-advised to build on human capabilities and propensities first” (p. 161). Erdelez (1999) noted that fully understanding the accompanying aspects of human behavior that impact a system are challenging, and that many users move both laterally, between topics, and vertically, within topics and across time, as they use systems. In this way, SKD must optimize users’ ability to capture and quickly jump amongst these approaches while maintaining some control over the relationships encountered along the way. Workman, Fiszman, Rindflesch and Nahl’s (2014) framework, particularly around conceptual short-term memory, is a strong supporting design mechanism to support SKD regardless of the task or any other prospective facets that might influence the users’ use of the system to discover information serendipitously.

Literature Path Research Summary

Before looking further at the relevant theories and models, including the IF-SKD Model, the Figure 2-3 provides a visual summary of the literature path and background that led to the decision to study the topic of serendipitous knowledge discovery in a clinical care context with physician participants and the Spark system.

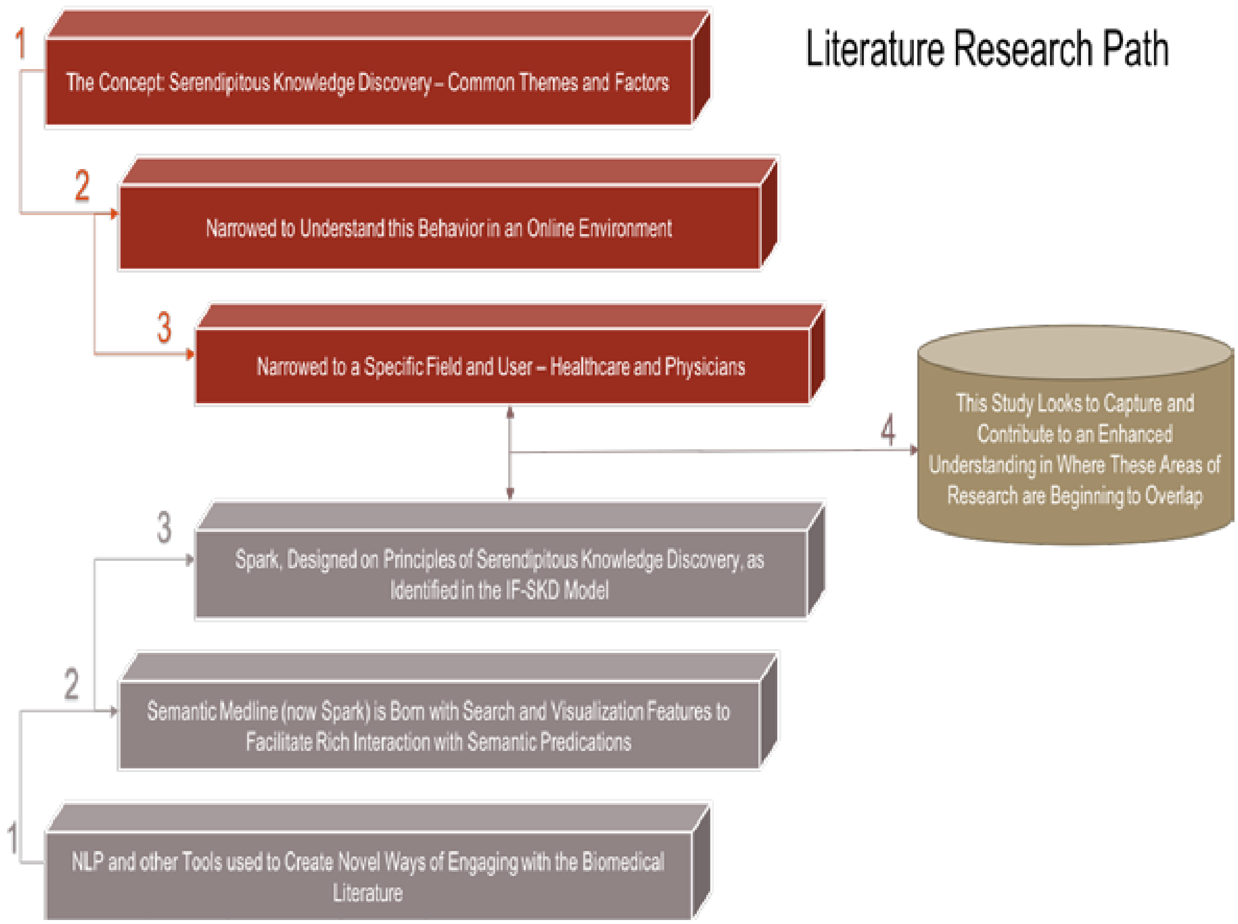

Figure 2-3. Literature research path.

Related Information Science Theories and Models

The theories and models mentioned below are summarized to demonstrate the way in which serendipitous knowledge discovery, or serendipity more specifically, has have used in previous studies. If one considers each study's individual application of serendipity as a series of overlapping circles highlighting its conceptual operationalization to date, it is possible to see how aspects of each study share some interpretive similarities, while deviating fundamentally on other aspects of the concept's use.

Information Encountering

Information encountering is a topic that strongly relates to the idea of serendipitous knowledge discovery. Erdelez (1997, 1999, 2004) addressed this related term and presented a variety of rationale collected using qualitative research methods in an academic setting using employees and students as respondents. Data gathering focused on capturing rich descriptive data based on participant's memory of information encounters experienced (1997). The analysis of these data illuminated many aspects of this complex type of information behavior and paved the way for future research focused on specific user traits or characteristics. These studies also made broad recommendations for system development, such as facilitating better browsing features for users (1997).

One aspect of information encountering that is distinguishable from SKD is that the measurement of serendipity is attributable to its recognition by users. SKD considers the possibility that discovery can occur, even if not recognized by the user.

Literature Based Discovery

The concept of literature-based discovery is one where users are able to explore a vast amount of information with the expectation that certain expected relationships will be proven out or that new relationships will surface as a result of engaging with, processing and summarizing findings. This process has been, traditionally, inherently complicated. It involves extensive time and expertise of both content and search techniques to find, review and summarize vast amounts of information. Notwithstanding these concerns, its efficacy as a type of information seeking behavior is a proven concept.

Swanson (1986a) was the pioneer of this technique, referring to the link associating Raynaud's Syndrome as being unassociated with existing studies discussing the value of fish-oil in reducing blood viscosity and reactivity, which he referred to as "undiscovered public knowledge" (Swanson, 1968b). Literature-based discovery findings do not guarantee that research or knowledge exists on a topic, but rather that it may be an ongoing private or public pursuit. Nonetheless, the early and easy identification of unexplored relationships are paramount to the concept of this type of information seeking behavior.

Different literature-based discovery techniques exist. The two most common are open discovery involving users exploring presumed relationships among concepts in the literature, and closed discovery involving looking for relationships between concepts with no presumed relationships. Miller et al. (2012) considered these as two altogether unique paradigms.

In an open discovery situation, there are relationships that are known, or accepted (A-B and B-C). The goal of such an approach would be to explore whether an A-C relationship also exists (Miller et al., 2012). Confirmation of such findings could provide sufficient knowledge to justify further research on the A-C relationship.

In the closed discovery model, the A-C relationship above is assumed. This assumption may or may not be based on existing knowledge. The user in such an instance would be looking for the respective A-B and B-C relationships that might also exist.

Information Foraging Theory (IFT)

The term information foraging theory (IFT), and its application to system design, borrows much of its core meaning from a related theory in another discipline, optimal foraging theory, which is derived from "ethological studies of food seeking and prey selection among animals" (Pirolli & Card, 1999, p. 644). The application of IFT to information behavior is centered on promoting effective information discovery encounters by effectively capturing the user's information "scent", also referred to as "expected utility", presented during the search process (p. 1-2). In its application to a biomedical information search system, Goodwin, Cohen and Rindflesch (2012) discussed how such a system could improve recommendations and information presentation through the effective measurement of these scent items along the way.

At a glance, this shares many goals and hallmarks of serendipitous knowledge discovery in that the goal is to provoke SKD through effective system design and fluid user interaction. However, there is a core difference. In IFT theory, there is an assumption that the user has a question in mind for which they seek information, and the goal of the system is to effectively guide the user to that answer and promote as much serendipitous discovery along the way. Within the IF-SKD model, Workman, Fiszman, Cairelli, Nahl and Rindflesch (2016) note that information foraging theory "is orthogonal" to the IF-SKD model for this specific reason (p. 25).

By requiring the user to have an expected information need already articulated, for which the system evaluates the utility of the information scent along the way, in an effort to better promote serendipitous interactions, this approach restricts system design and effective measurement of user actions because it does not permit random exploration or browsing activities that could be considered curiosity driven.

User Characteristics

So far serendipity has been presented from the vantage of the researcher through a consolidated analysis of existing knowledge shown in the literature, with emphasis placed on individuals who have been highly influential in shaping the understanding of serendipitous knowledge discovery. It is important to also consider specific aspects that highlight trends of the user characteristics commonly associated with SKD.

Effective analysis of user behavior and characteristics associated with serendipitous knowledge discovery is important for a couple of reasons. First, while some of the bedrock information science literature is fundamentally good at providing a structure and framework for the application and operationalize of serendipity, there are additional studies that provide solid analysis and noteworthy examples of SKD user characteristics (Burkell, Quan-Hase & Ruin, 2012; Erdelez, 1997; Spink, 2004; Workman, Fiszman, Rindflesch & Nahl, 2014). Second, some studies also show how they have attempted to quantify and administer research instruments for the purpose of measuring SKD events (Fine & Deegan, 1996). And even though not all of these studies portend to the same exact operationalization of terms, they demonstrate a foundation from which the research instruments proposed as part of this study were conceived.

While the goals of this research are specific and narrower than a full examination of all possible physician user characteristics associated with their SKD activities in the clinical care setting, the outcome of this research looks to any associated statistical significance found from analysis of the data collected. This may help provide context in understanding findings related to the core research questions posed. It could also yield ideas for future research or more immediate refinements to Spark benefiting its users.

While many enjoy and may even revel in the discovery of information by chance, there is a characteristic of user behavior that could predispose them with a higher likelihood to discover knowledge by serendipity. Erdelez (1997) noted how some users are predisposed to rely on serendipity "as an integral part of their information behavior" (p. 417). In a later study of users Heinström (2006) found many links between this preference to rely on serendipity with specific user personality traits within the literature. Nahl (2004) pointed to motivation as a driving force in SKD, while others addressed the traits of curiosity, enthusiasm, spontaneous and adventure driven (Costa & McCrae, 1992; Erdelez, 1997; Heeter & Greenberg 1985; Heinström, 2003; Roberts, 1989). Even though some of these characteristics may seem obvious, they underscore the integral nature users as individuals bring with them, and they are not an aspect of information behavior to which one can reasonably rely on to be present in every situation.

It would seem plausible that SKD would be more likely among users having a clear understanding of their information need, yet, Nutefall and Ryder (2010) found the opposite to be true as well with some users benefiting from having no question in mind. Perhaps this is attributable to context, or the topic with which the user is engaged. Lawley and Tompkins

(2008) presented a model with an implied observation for a user's insight being present based on capturing the outcome of long term value being recognized. Heinström (2006) noted the general agreement in the literature, indicating the largest number of serendipitous discoveries being associated with normal daily activities, such as "reading a newspaper" (p. 581). In other contexts, such as a clinical setting, the formulation of an information need, even if not required, may carry more weight in the number of serendipitous knowledge discovery events for a user. Solomon (1997) pointed out that context, and also timing, is a crucial aspect of this type of information behavior. In a review of existing literature on serendipity, as well as through qualitative research undertaken, Makri and Blandford (2012) found that the event, or trigger, for serendipity and the outcome often overlapping, which can make it challenging to measure.

Whether a user has a question in mind or not, the ability for them to articulate their information need within an online system is not necessarily simple. Belkin, Oddy and Brooks (1982) noted a unique aspect of information behavior, which is that often users can explain better what they don't know as compared to what they do know. This has some bearing on another user characteristic noted by Erdelez (1997), which is that some users "'shifted' to other dimensions of information needs" while in the process of "information encountering" (p. 416). This latter aspect of user behavior is a strong consideration for system designers, who should be careful to allow for a variety of ways for users to engage with and discover information, including the refinement of their information needs along the way. It is worth noting that Spark accounts for this situation in part by allowing users to filter results for common vs. rare occurrences of information, which might be useful in promoting triggers throughout the iterative discovery process.

The user characteristics outlined above show another layer that makes up this mosaic-like type of information behavior. Successfully identifying every type of user characteristic that could impact a study or influence a system design for facilitating serendipitous knowledge discovery is at best unlikely. Workman, Fiszman, Rindflesch and Nahl (2014) addressed the significance of this type of ambiguity on system design, stating that “in the context of SKD, the user’s intentions act as a fluid agent that the online system must accommodate” (p. 502). Bates (1989) in striking similarity, stated that “system should be sufficiently flexible to allow the user to adapt the information-seeking process to his own current needs” (p. 421). In each instance, it is important to note that these statements are not meant to derail the meaningful discernment and study of user characteristics critical to SKD, but that instead both information models of serendipitous information behavior and related system design err on the side of being able to accommodate multiple types of users to optimize the opportunity for serendipity to be achieved.

The IF-SKD Model

The information flow-serendipitous knowledge discovery model (IF-SKD) developed by Workman, Fiszman, Rindflesch and Nahl (2014), represented in Figure 2-4, captures the information behavior process within an online system.

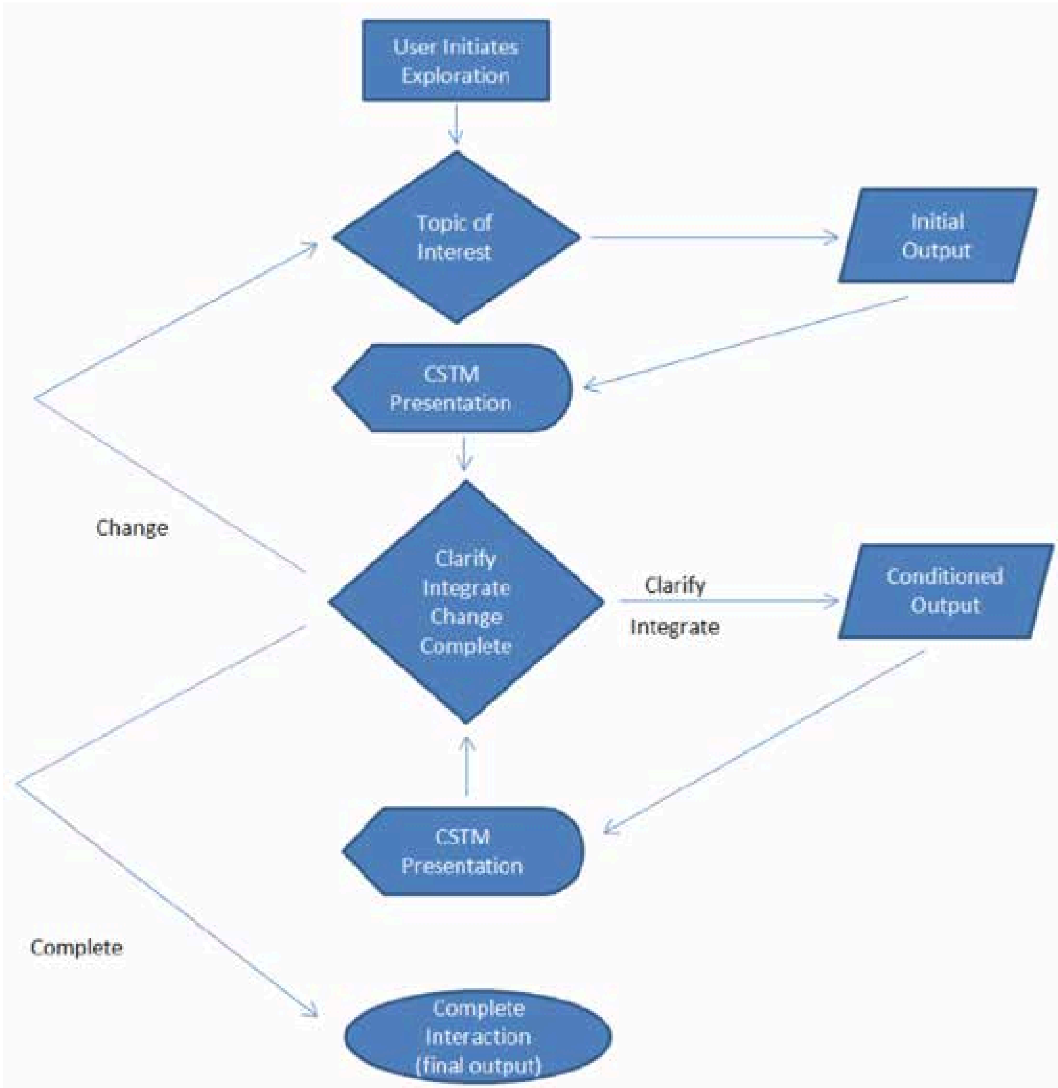

Figure 2-4. IF-SKD model.

This model captures the authors' representation of how information behavior, as understood from a consolidated reflection from the literature, represents a user's actions whilst engaged in a process that could lend itself to serendipitous knowledge discovery. Reflecting on the four key aspects of this model helps to put the goals of its authors' intentions in this being applicable to application for system design and reflects the inherent nature of serendipity being a process, rather than an outcome. Furthermore, the key components of the

IF-SKD model bear strong relationships to recent summarized findings of how serendipity is understood and operationalized within the information science literature.

Four Core SKD Components

As a process, the IF-SKD model presents four concepts or themes, derived from the literature, that help "maps the SKD process within electronic environments" (Workman, Fiszman, Cairelli, Nahl and Rindflesch, 2016 p. 4):

1. SKD is an iterative process
2. SKD often involves change or clarification of initial information interests, which may involve integrating new topics
3. SKD is grounded in user's prior knowledge
4. Information organization and presentation have fundamental roles in SKD

Hider (2006) noted that the information science literature has challenged "the system based model of classical IR", with Belkin (1982) being one of the first to do so, noting that information needs and how they are expressed are inherently complex (p. 354). Hider (2006) further pointed out, in reference to Bates' berrypicking model, how the need for a system to be iterative in nature is necessary to support the varied tactics users employ in their regular information seeking behavior. In essence, if a system is not iterative and engaging for the user, it could make it more challenging to be successful in serendipitous knowledge discovery.

The second concept related to SKD involving change or clarification is more straightforward. It assumes that users will refine, change or state their information needs differently, and iteratively, to an information system as they are exposed to new information.

Third is the concept of how a user's prior knowledge relates to SKD. Workman, Fiszman, Rindflesch & Nahl (2014) note that a previous problem may represent prior knowledge, or information need, that is known to the user. Additionally, the user's "expertise or prior experience" has significance on how the user interacts, absorbs and engages with the system, which ultimately could impact their SKD (p. 24).

Information organization and presentation is the fourth core concept of the IF-SKD model. With respect to system design and the study of serendipity, information organization and presentation is perhaps the most fluid, and to some degree measureable, aspect in consideration for understanding the SKD process and what components of system design influence, or take into account, the preceding three concepts.

The incorporation of these themes as part of the IF-SKD model's process flow presents an ongoing process of discovery and re-discovery – an open loop. While the model does allow for a stopping point where output can be captured for reference, or future use, it does not ever fully require, or expect, an information need, or serendipitous knowledge discovery to be closed or wholly met. This has relevance to the way in which potential system designers interpret the model, with the implication being that the process is more important than the outcome.

The challenge is that in order to assess vitality of system design, as well as the significance of the underlying model that drives it, measurement is required. This is precisely why studies such as this are necessary, both in the immediate tactical sense to support future system development and end-user feedback, but also to the general ongoing understanding of how to interpret concepts, such as serendipity, and further the understanding and development and refinement of knowledge.

Designed for SKD vs. Designing SKD

Erdelez (1999) noted that the “development of information systems that would support information encountering” is an important area of application, in particular, in helping the “non-encounters make better use of encountering” (p. 6). This idea is underscored in the IF-SKD’s four core concepts and is central to Spark’s main focus, which is to support SKD by promoting interaction with the biomedical literature through the manipulation of the user’s concepts or ideas, which the user may either know, or come to know, through a variety of tools and system features. The more flexibility and refinement that can be done quickly and iteratively, the more likely it is presumed the user will be able to engage in discoveries of a serendipitous nature.

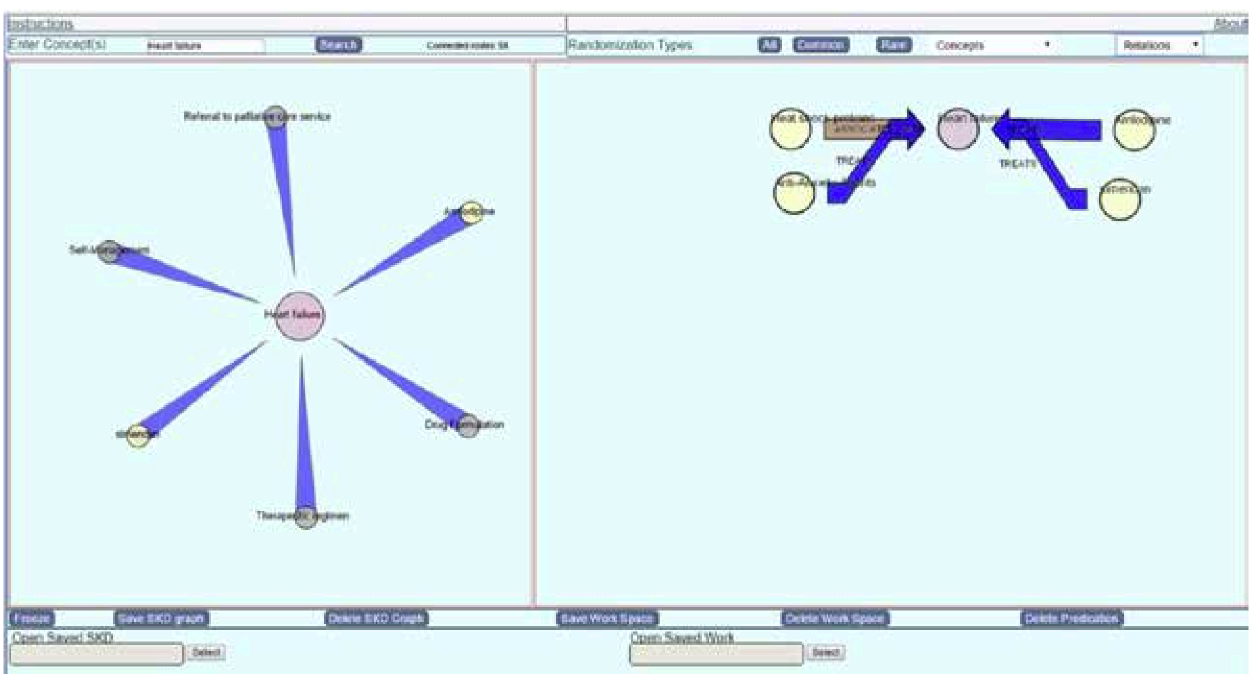

Figure 2-5. Spark system.

Summary

As demonstrated in this chapter, there is nothing simple about applying serendipitous knowledge discovery for modeling users' information behavior or packaging it into a one size fits all system design. What this chapter does show is that despite some nuanced, yet important distinctions, serendipity is inherently more a broad process of information behavior, rather than a statement outlining a specific information seeking behavior. As a process, it is constrained by the context and unique aspects of its users. The IF-SKD model was demonstrated and shown to capture the richness of this broad process. As such, it provides an equally broad likelihood of finding utility in disparate contexts and in its application to system design. It also provides the structure to incorporate and review the impact of individual user characteristics relative to serendipitous knowledge discovery.

To enhance and deepen the understanding around the IF-SKD model, and the study of serendipity, and its application, it is necessary to measure, relate and assess its application in the real world. The research goals and objectives of this study are to do these exact things, providing context and data to evaluate the IF-SKD model's relevance to system design and the influence of the clinical context. It also helps establish whether Spark contributes to serendipitous knowledge discovery. The furthered pursuit of measuring serendipity using quantitative tools also assists in the ongoing understanding and refinement of research instruments capable of reflecting the nuanced aspects of serendipity, which may in turn lead to a deeper understanding of how to reflect and model this type of information behavior in ways that increasingly positively influence future system development and study.

CHAPTER 3

METHODOLOGY

Introduction

The purpose of this study is to explore whether Spark contributes to physicians' serendipitous knowledge discovery (SKD) and to understand to what degree the IF-SKD model reflects physicians' SKD in a clinical context. By using the McCay-Peet (2013) *Serendipitous Digital Environment (SDE) Questionnaire* and *Perception of Serendipity Scale*, the researcher hopes to demonstrate Spark's capacity to positively contribute to physician SKD. An analysis mapping the IF-SKD model's core components to that of the SDE questionnaire is made, which allows the IF-SKD model to be studied using confirmatory factor analysis.

This research employs a pre-experimental design. Feedback on the research instrument was collected using expert review. A single treatment sample group was provided a video introduction on the use of the Spark, then asked to complete the research instrument. As there is no known established quantitative approach for measuring serendipitous knowledge discovery in a clinical setting, this method is favored. This method is also preferred due to the sensitive nature of a clinical setting as well as the challenge associated with the participant time constraints and accessibility.

Because the purposeful, direct and intentional study of serendipity within the information science literature is relatively early in its development, the furthering of new models to explain this behavior, coupled alongside with research tools, is imperative. Studies by Erdelez (2004), Bjorneborn (2008) and McCay-Peet and Toms (2010) all have contributed to the development of the research tool employed in this study. These same studies help illustrate the need to develop and understand the quantitative driven tools that] assist in measuring serendipity and how to relate those to system design. While some recent research, such as Sun, Sharples and Makri's (2011) quick diary technique and Jiang, Zhang, Li, Fan and Yang's (2018) diary process using critical incident technique may be a path towards a middle ground between quantitative and qualitative methods, it is not effective for consistent, ongoing, organizational independent data collection, particularly in a clinical setting.

Dantonio, Makri and Blandford (2012) note that serendipity is non-reproducible in a controlled setting. This sentiment reinforces the need to evaluate tools such as the questionnaire employed here, despite any limitations it may pose. This evaluation helps to better understand what aspects of serendipity measurement can withstand cross-organization use and assist in paving the generalized role serendipity plays in today's information-rich world.

The primary goal of this methodology is to facilitate the application of a new research instrument, McCay-Peet (2013) *Serendipitous Digital Environment questionnaire* and *Perception of Serendipity Scale* to evaluate Spark and the IF-SKD Model. Both the IF-SKD model and the development of the research instrument used here seek to address the complexity associated with defining the way in which serendipity is understood and applied in an effort to improve on the reliability in measuring serendipity and the generalizability of findings.

Research Design

Participants and Hospitals

This research employs participant self-selection as a means of identifying participants for inclusion. Physicians (MD and DO) currently working for INTEGRIS Health of any specialty, across the state of Oklahoma, are candidates for inclusion. INTEGRIS Health operates numerous hospitals, standalone primary and specialty clinics throughout Oklahoma, as well as specialty facilities, such as Jim Thorpe Rehabilitation, Lakeside Women's Hospital and the INTEGRIS Cancer Institute. Figure 3-1 shows the location of INTEGRIS' seven multispecialty hospitals in the State of Oklahoma.

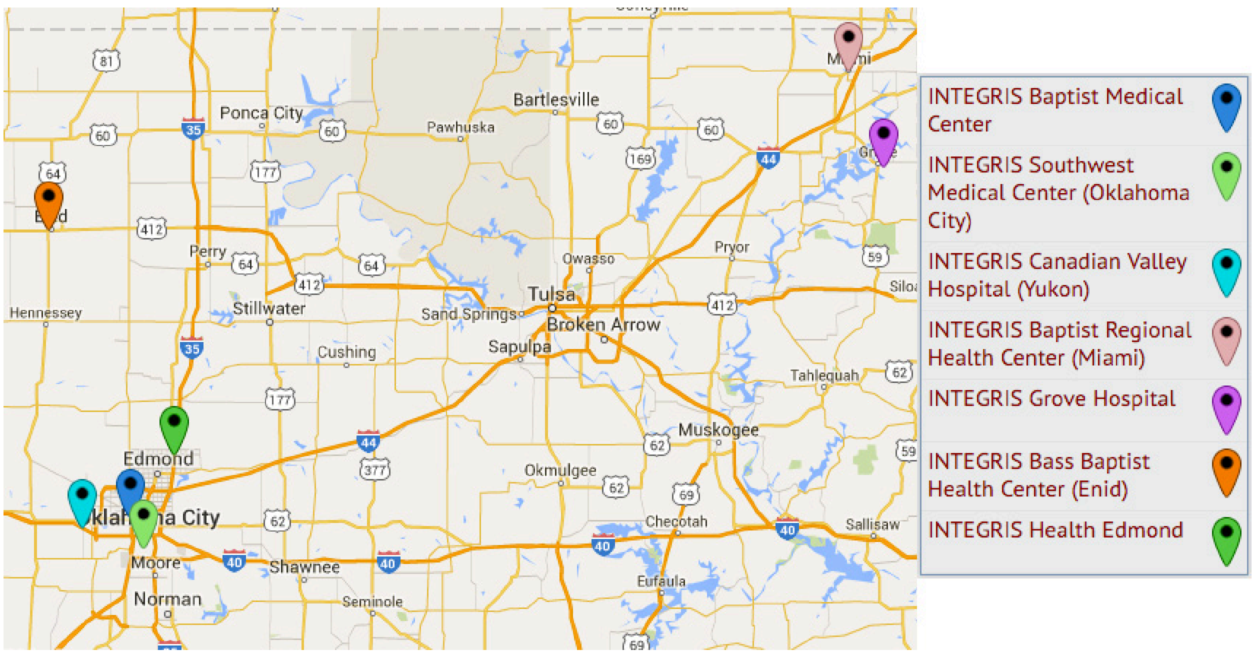

Figure 3-1. Map of INTEGRIS hospitals.

INTEGRIS' major hospitals share similarities in the core patient services they provide. Figure 3-2 illustrates the physician population at these hospitals and the associated specialties represented.

	Organization						
Specialty	Baptist	Edmond	Enid	Grove	Miami	Southwe..	Yukon
Allergy	5					3	
Allergy/Immunology	1		1			1	
Anesthesia	69	7	3		1	28	35
Burn and Wound Care	2						
Cardiac Transplant Surger..	1					1	
Cardiology	6			4		5	2
Cardiovascular Disease	46	15	7	2	1	30	15
Cardiovascular Surgery	7	4				6	
Colon and Rectal Surgery							1
Critical Care Medicine	3	4				3	
Dermatology	5					2	
Electrophysiology	1						
Emergency Medicine	23	27	5	14	11	44	23
Endocrinology, Diabetes, ..	5	2				3	
Endovascular Surgical Ne..	1	1					1
Family Medicine	68	10	17	20	10	38	21
Gastroenterology	21	5	2			9	11
Gynecologic Oncology	2						
Gynecology	4	4				1	
Hematology	7					4	1
Hematology/Oncology	1	2	6			1	
Hepatology	3					1	
Hospitalist	28	1	2			31	3
Hospitalist-Pediatric	2						
Infectious Disease	11	2	1			5	
Infertility	1						
Internal Medicine	52	10	4	1	4	20	11
Interventional Cardiology	4	1	3	1	1	3	1
Interventional Neuroradi..	1						
Interventional Pain Mana..						1	
Maternal/Fetal Medicine	2						
Medical Oncology	7	4				1	1
Neonatology	6					6	6
Nephrology	25	13	1	1		16	11
Neurological Surgery	10		1			3	
Neurology	11		1	5	5	9	4
Neurosurgery	1					1	
Obstetrics and Gynecology	65	13	5	3	3	11	8
Oncology				1			
Ophthalmology	23	6		1	1	7	5
Orthopedic Surgery	1			1			1

Number of Records

1 69

Sum of Number of Records broken down by Organization vs. Specialty. Color shows sum of Number of Records. The marks are labeled by sum of Number of Records. The view is filtered on Organization, which excludes Lakeside Women's Hospital.

Number of Records: 1 – 69

Specialty	Baptist	Edmond	Enid	Grove	Miami	Southwe..	Yukon
	Organization						
Otolaryngology	18	6	1	1	1	3	1
Pain Management	6	1	3			4	2
Pathology	11	8		5	5	12	9
Pathology, Anatomic	1	1				1	1
Pathology, Anatomic and ..	1	2	1	6	5	2	2
Pathology, Clinical				1	1		
Pediatric Cardiology	6					5	1
Pediatric Cardiovascular S..	1						
Pediatric Critical Care Me..	1						
Pediatric Emergency Medi..						1	
Pediatric Gastroenterology	1						
Pediatric Neurology	3	1				2	
Pediatric Ophthalmology	1						
Pediatric Otolaryngology	3						
Pediatric Rehabilitation	1						
Pediatric Surgery	4						
Pediatric Urology	3						
Pediatrics	48	15	4		1	13	5
Pediatrics/Internal Medici..	1						
Perinatology/Neonatology		2					
Physical Medicine & Reha..	6	1				5	
Psychiatry	13	2	1		2	5	1
Psychiatry, Child & Adoles..	5						
Psychiatry, Geriatric	1						
Pulmonary Disease	12	5		1		8	2
Pulmonary/Critical Care	1		3	1			1
Radiation Oncology	12	9	1			10	
Radiology	10	17			8	7	18
Radiology, Diagnostic	4		11	3		1	1
Radiology, Interventional	4	2					
Retina and Vitreous Disea..	1						
Rheumatology	4	2				1	
Shoulder & Elbow Surgery	1						
Sleep Medicine					1	1	
Spine Surgery	3						
Sports Medicine	1						
Surgery, Bariatric	4						
Surgery, Cardiothoracic	1						
Surgery, Colon/Rectal	1	1					
Surgery, General	20	10	4	5	2	11	6
Surgery, Hand	8	2				2	

Sum of Number of Records broken down by Organization vs. Specialty. Color shows sum of Number of Records. The marks are labeled by sum of Number of Records. The view is filtered on Organization, which excludes Lakeside Women's Hospital.

Specialty	Baptist	Edmond	Enid	Grove	Miami	Southwe..	Yukon
				Organization			
Surgery, Oral & Maxillofac..	1						
Surgery, Orthopedic	39	26	3	1	1	18	17
Surgery, Plastic	13	3	1			2	2
Surgery, Thoracic			1				
Surgery, Vascular	5	1				3	
Telemedicine-Neurology		8		3	3		3
Teleradiology		32	32	51	30	31	
Transplant Assisting/Org..	2						
Transplant Surgery	3					2	
Transplantation Medicine	3					2	
Urogynecology/Female Pe..	1						
Urology	27	12	4	1	1	4	5
Wound Care						1	

Number of Records 1 – 69

Sum of Number of Records broken down by Organization vs. Specialty. Color shows sum of Number of Records. The marks are labeled by sum of Number of Records. The view is filtered on Organization, which excludes Lakeside Women's Hospital.

Figure 3-2. INTEGRIS hospital specialty breakdown.

Setting and Data Collection Process

Setting

The setting for the study, described as the *clinical setting* is inclusive of the locations and of the workflows used by the providers participating in the study. This research does not mean to denote what constitutes a specific clinical setting, but rather is constructed to inform participants of the goal to understand how Spark contributes to SKD, wherever and however they would normally choose to incorporate the resource.

The questionnaire is administered online using Qualtrics provided by the University of North Texas. A link was created and emailed to participants. Qualtrics is able to provide a breakdown of participant demographics and other data collected on the research instrument. Figure 3-3 provides a view of how this questionnaire appears to participants if accessed over a web browser or mobile phone for completion.

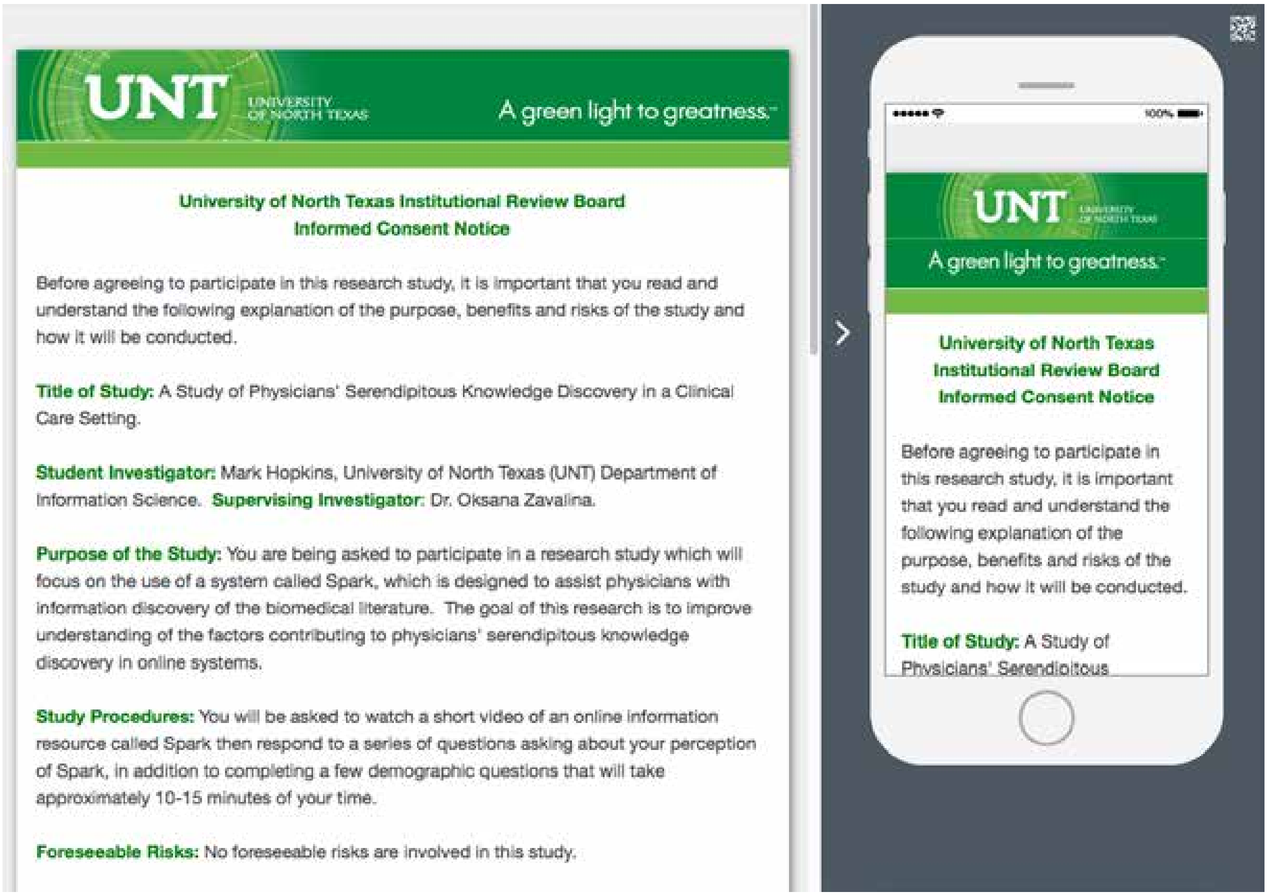

Figure 3-3. Research instrument visual using UNT Qualtrics.

Data Collection Process

At a high level, the methodology for data collection includes a summary of the research goals, implications of the research and an assurance of anonymity for any contributed feedback.

As previously noted, an introduction to Spark is provided to participants using a brief, yet meaningful, summary video of Spark being used to explore a medical question. The anticipated time to complete the questionnaire is 10-15 minutes. Should any variations to the length or presentation change as part of the research instrument expert review feedback, the expected completion time could change.

The research was distributed to physicians through email and word of mouth. Additionally, the Medical Director, Inpatient Informatics for INTEGRIS Health, Dr. LeRoy Southmayd III, helped communicate with providers regarding the opportunity to participate in this research.

Serendipitous Knowledge Discovery

Since the concept of serendipity is a central aspect of the study, it is carefully communicated to participants. This is vital not only to explaining the purpose and goals of the study, but also to assist in providing participants with a shared conceptual understanding that is important to informing their ability to complete the research instruments. At the top of the questionnaire, the operationalized definition of serendipity provided in Chapter 1 is provided to assist with the instrument's completion.

Research Instrument

The research instrument used in this study is a variation of the McCay-Peet (2013) 37-item Serendipitous Digital Environment questionnaire (Figure 3-4) and 4-item Perception of Serendipity Scale (Figure 3-5). The questionnaire represents a consolidated pluralistic approach to measuring serendipity. It also accounts for varying definitions of serendipity, determined by an analysis of the literature, in an effort to capture the presence of serendipity based upon the differing ways it has been presented and discussed (2013).

In her research, McCay-Peet (2013) conducted content validity testing on the SDE questionnaire to evaluate the questions, their meaning and wording, and the appropriateness of their facet assignments. This testing took two forms. First, there was a review of the questions and the underlying meaning behind them performed by experts in the field, including: "Paul André, Lennart Bjorneborn, Jose Campos, Nigel Ford, Jannica Heinström, Stephann Makri, Anabel Quan-Haase, and Borchuluun Yadamsuren" (McCay-Peet, 2013 p. 90).

In addition to this, the author utilized an analysis of variance (ANOVA) approach to evaluate responses collected through an online survey. This survey asked participants to rate how well an item matched the definition provided of its facet, where the relationship of item-to-facet differed between surveys. The analysis considered "items that have the highest mean rating on their posited facet; and items that have a significantly higher mean rating ($p < .05$) on their posited facet" (McCay-Peet, 2013 p.98). This provided a mechanism to evaluate how well the proposed item-to-facet relationships could potentially work as a model of information systems' serendipitous characteristics.

This research study aims to employ the research instrument, presented in the same outlined manner as McCay-Peet (2013); however, this research also intends to take a confirmatory factor analysis (CFA) approach to analyze proposed item-to-facet relationships in consideration of the IF-SKD model.

The outcome of this analysis helps support the evaluation of the proposed item-to-facet mappings and the utility of the IF-SKD model. It also provides an opportunity to consider, separate from this primary CFA model fit analysis, the conceptual space of the item questions and how they relate to how systems' serendipitous characteristics match to broader facets, or components, identified in the research literature.

Enables Exploration: A user's assessment of the degree to which a digital environment supports exploration and examination of its information, ideas, or resources

1. E1 It is easy to explore [the digital environment]'s content
2. E2 [The digital environment] supports exploration
3. E3 It is easy to wander around in [the digital environment]
4. E6 There are many ways to explore information in [the digital environment]
5. E7 [The digital environment] invites examination of its content
6. E8 [The digital environment] is an instrument for discovery
7. E9 [The digital environment] is a tool for exploration

Trigger-Rich: A user's assessment of the degree to which a digital environment contains a variety of information, ideas, or resources that is interesting and useful to the user

8. T1 The content contained in [the digital environment] is diverse
9. T2 [The digital environment] is rich with interesting ideas
10. T3 [The digital environment] offers a wide variety of information
11. T4 There is a depth of information in [the digital environment]
12. ***T5 [The digital environment] is full of information useful to me***
13. ***T6 I find information of value to me in [the digital environment]***
14. ***T7 [The digital environment] is a treasure trove of information***

Enables Connections: A user's assessment of the degree to which a digital environment makes relationships or connections between information, ideas, or resources apparent

15. ***C1 [The digital environment] enables me to make connections between ideas***
16. C2 Associations between ideas become obvious in [the digital environment]
17. ***C3 I can see connections between topics in [the digital environment]***
18. C4 It is easy to see links between information in [the digital environment]
19. C6 I make useful connections in [the digital environment]
20. C8 The features of [the digital environment] help me see connections between its content
21. ***C9 I come to understand relationships between ideas in [the digital environment]***

Highlights Triggers: A user's assessment of the degree to which a digital environment brings interesting and useful information, ideas, or resources to the user's attention

22. H1 I am directed toward valuable information in [the digital environment]
23. ***H2 [The digital environment] has features that ensure that my attention is drawn to useful information***
24. H3 Information that interests me is highlighted in [the digital environment]
25. H4 The way that [the digital environment] presents content captures my attention
26. H5 I am alerted to information in [the digital environment] that helps me
27. H7 I notice content I wouldn't normally pay attention to in [the digital environment]
28. ***H8 [The digital environment] has features that draw my attention to information***
29. ***H9 I am pointed toward content in [the digital environment]***
30. ***H10 [The digital environment] has features that alert me to information***

Leads to the Unexpected: A user's assessment of the degree to which a digital environment provides opportunities for unexpected interactions with information, ideas, or resources

31. ***U1 I bump into unexpected content in [the digital environment]***
32. ***U2 I encounter the unexpected in [the digital environment]***
33. U3 I am surprised by what I find in [the digital environment]
34. ***U4 I come across topics by chance in [the digital environment]***
35. ***U5 [The digital environment] exposes me to unfamiliar information***
36. U6 My interactions in [the digital environment] are unexpectedly valuable
37. ***U7 I stumble upon information in [the digital environment]***

Note. E = **SDE-Enables Exploration** items; T = **SDE-Trigger-Rich** items; C = **SDE-Enables Connections** items; H = **SDE-Highlights Triggers** items; U = **SDE-Leads to the Unexpected** items.

Figure 3-4. Serendipitous Digital Environment questionnaire (McCay-Peet, Toms & Kelloway, 2015).

Perception of serendipity scale

Specific Digital Environment (S-SpecificDE)	Digital Environments in General (S-DEs)	General (S-Gen)
[S-SpecificDE-1] **In the digital environment I selected,** I experience serendipity that has an impact on my everyday life	[S-DEs-1] **In digital environments** I experience serendipity that has an impact on my everyday life	[S-Gen-1] I experience serendipity that has an impact on my everyday life
[S-SpecificDE-2] **In the digital environment I selected,** I experience serendipity that has an impact on my work	[S-DEs-2] **In digital environments** I experience serendipity that has an impact on my work	[S-Gen-2] I experience serendipity that has an impact on my work
[S-SpecificDE-3] I encounter useful information, ideas, or resources that I am not looking for **when I use the digital environment I selected**[a]	[S-DEs-3] I encounter useful information, ideas, or resources that I am not looking for **when I use digital environments**[a]	[S-Gen-3] I encounter useful information, ideas, or resources that I am not looking for[a]
[S-SpecificDE-4] **In the digital environment I selected,** I experience mixes of unexpectedness and insight that lead to valuable, unanticipated outcomes[b]	[S-DEs-4] **In digital environments** I experience mixes of unexpectedness and insight that lead to valuable, unanticipated outcomes[b]	[S-Gen-4] I experience mixes of unexpectedness and insight that lead to valuable, unanticipated outcomes[b]

[a] Items adapted from Erdelez's (2005) definition of information encountering.
[b] Items adapted from Makri and Blandford's (2012) elements of serendipity.

Figure 3-5. Perception of Serendipity Scale (McCay-Peet, Toms & Kelloway, 2015).

Incorporating the IF-SKD Model

As the IF-SKD model is also a construct of the literature on serendipity and information behavior, it is important to note that aspects of the questionnaire do not reflect the exact question grouping mix as laid out by the IF-SKD model. However, it is worth noting that there are several seemingly logical mappings between this questionnaire and the four core components of the IF-SKD model.

Another reason to compare the IF-SKD model to the questionnaire is to help reflect the aspect of serendipity as a process, which is central to the IF-SKD model. This helps, during data analysis, broaden the consideration for any variables that might correlate with the refinement and understanding of the core meaning of serendipity as used throughout the questionnaire. This also assists in better understanding what characteristics influence the concept of serendipity in the clinical setting.

In addition, the study of the IF-SKD model as it relates to system design may help demonstrate significance in how the IF-SKD model is interpreted in consideration of the design of Spark from a process orientation perspective, more than a specific aspect of system design.

Table 3-1 reflects high-level conceptual mappings of the IF-SKD model to the Serendipitous Digital Environment questionnaire groupings while Table 3-2 shows the specific question mappings within each grouping.

Table 3-1

IF-SKD Concept Mappings to Questionnaire

Mc-Cay-Peet, Tom & Kelloway (2015) Concepts	Workman et al. (2014) IF-SKD Model Proposed Mappings
Enables Exploration	· Iterative Process · Change/Clarification/Integration · Information Organization and Presentation have Fundamental Role
Trigger-Rich	· Grounded in Prior Knowledge
Enables Connections	· Iterative Process · Change/Clarification/Integration · Grounded in Prior Knowledge
Highlights Triggers	· Grounded in Prior Knowledge · Information Organization and Presentation have Fundamental Role
Leads to the Unexpected	· Iterative Process · Change/Clarification/Integration · Grounded in Prior Knowledge · Information Organization and Presentation have Fundamental Role

Table 3-2 presents the specific SDE questionnaire items mapped to the IF-SKD model. The following key is used for the IF-SKD model specified in the right column of the table:

1. Iterative process
2. Change/clarification/integration
3. Grounded in prior knowledge
4. Information organization and presentation have fundamental role

Table 3-2

IF-SKD Individual Question Concept Mappings

Enables Exploration: A user's assessment of the degree to which a digital environment supports exploration and examination of its information, ideas, or resources (A)	**IF-SKD Question Mapping**
1. E1 It is easy to explore [the digital environment]'s content **2.** E2 [The digital environment] supports exploration **3.** E3 It is easy to wander around in [the digital environment] **4.** E6 There are many ways to explore information in [the digital environment] **5.** E7 [The digital environment] invites examination of its content **6.** E8 [The digital environment] is an instrument for discovery **7.** E9 [The digital environment] is a tool for exploration	§ **[E6]** Question 4: **2** § **[E7]** Question 5: **2** § **[E8]** Question 6: **2** § **[E9]** Question 7: **2**
Trigger-Rich: A user's assessment of the degree to which a digital environment contains a variety of information, ideas, or resources that is interesting and useful to the user (B)	**IF-SKD Question Mapping**
8. T1 The content contained in [the digital environment] is diverse **9.** T2 [The digital environment] is rich with interesting ideas **10.** T3 The digital environment] offers a wide variety of information **11.** T4 There is a depth of information in [the digital environment] **12.** T5 [The digital environment] is full of information useful to me **13.** T6 I find information of value to me in [the digital environment] **14.** T7 [The digital environment] is a treasure trove of information	§ **[T5]** Question 12: **3** § **[T6]** Question 13: **3**
Enables Connections: A user's assessment of the degree to which a digital environment makes relationships or connections between information, ideas, or resources apparent (C)	**IF-SKD Question Mapping**
15. C1 [The digital environment] enables me to make connections between ideas **16.** C2 Associations between ideas become obvious in [the digital environment] **17.** C3 I can see connections between topics in [the digital environment] **18.** C4 It is easy to see links between information in [the digital environment] **19.** C6 I make useful connections in [the digital environment] **20.** C8 The features of [the digital environment] help me see connections between its content **21.** C9 I come to understand relationships between ideas in [the digital environment]	§ **[C1]** Question 15: **1** § **[C2]** Question 16: **1** § **[C3]** Question 17: **4** § **[C4]** Question 18: **4** § **[C6]** Question 19: **3** § **[C8]** Question 20: **4** § **[C9]** Question 21: **3**
Highlights Triggers: A user's assessment of the degree to which a digital environment brings interesting and useful information, ideas, or resources to the user's attention (D)	**IF-SKD Question Mapping**
22. H1 I am directed toward valuable information in [the digital environment] **23.** H2 [The digital environment] has features that ensure that my attention is drawn to useful information **24.** H3 Information that interests me is highlighted in [the digital environment] **25.** H4 The way that [the digital environment] presents content captures my attention **26.** H5 I am alerted to information in [the digital environment] that helps me **27.** H7 I notice content I wouldn't normally pay attention to in [the digital environment]	§ **[H3]** Question 24: **4** § **[H4]** Question 25: **4** § **[H5]** Question 26: **4** § **[H7]** Question 27: **4** § **[H2]** Question 28: **4** § **[H9]** Question 29: **4** § **[H10]** Question 30: **4**

28. H8 [The digital environment] has features that draw my attention to information **29.** H9 I am pointed toward content in [the digital environment] **30.** H10 [The digital environment] has features that alert me to information	
Leads to the Unexpected: A user's assessment of the degree to which a digital environment provides opportunities for unexpected interactions with information, ideas, or resources (E)	**IF-SKD Question Mapping**
31. U1 I bump into unexpected content in [the digital environment] **32.** U2 I encounter the unexpected in [the digital environment] **33.** U3 I am surprised by what I find in [the digital environment] **34.** U4 I come across topics by chance in [the digital environment] **35.** U5 [The digital environment] exposes me to unfamiliar information **36.** U6 My interactions in [the digital environment] are unexpectedly valuable **37.** U7 I stumble upon information in [the digital environment]	§ **[U1]** Question 31: **1** § **[U2]** Question 32: **2** § **[U3]** Question 33: **3** § **[U6]** Question 36: **1** § **[U7]** Question 37: **1**

Table 3-3 shows original questions on the SDE Questionnaire according to the proposed IF-SKD groupings.

Table 3-3

Questions Grouped by Proposed IF-SKD Mappings

SKD Is an Iterative Process
C1 [The digital environment] enables me to make connections between ideas C2 Associations between ideas become obvious in [the digital environment] U1 I bump into unexpected content in [the digital environment] U6 My interactions in [the digital environment] are unexpectedly valuable U7 I stumble upon information in [the digital environment]
SKD Often Involves Change or Clarification of Initial Information Interests, Which May Involve Integrating New Topics
E6 There are many ways to explore information in [the digital environment] E7 [The digital environment] invites examination of its content E8 [The digital environment] is an instrument for discovery E9 [The digital environment] is a tool for exploration U2 I encounter the unexpected in [the digital environment]
SKD Is Grounded in the User's Prior Knowledge
T5 [The digital environment] is full of information useful to me T6 I find information of value to me in [the digital environment] C6 I make useful connections in [the digital environment] C9 I come to understand relationships between ideas in [the digital environment] U3 I am surprised by what I find in [the digital environment]
Information Organization and Presentation Have Fundamental Roles
C3 I can see connections between topics in [the digital environment] C4 It is easy to see links between information in [the digital environment] C8 The features of [the digital environment] help me see connections between its content H3 Information that interests me is highlighted in [the digital environment] H4 The way that [the digital environment] presents content captures my attention

H7 I notice content I wouldn't normally pay attention to in [the digital environment] H2 Spark has features that ensure that my attention is drawn to useful information H9 I am pointed toward content in [the digital environment] H5 I am alerted to information in [the digital environment] that helps me H10 [The digital environment] has features that alert me to information

Table 3-4 highlights the final 21-item SDE questionnaire used in the study. As part of the committee review, four questions were removed, leaving 21 final question that were used in the study.

Table 3-4

Revised 21-Item Questionnaire

SKD Is an Iterative Process (1)
C1 Spark enables me to make connections between ideas U1 I bump into unexpected content i Spark U6 My interactions in Spark are unexpectedly valuable U7 I stumble upon information in Spark
SKD Often Involves Change or Clarification of Initial Information Interests, Which May Involve Integrating New Topics (2)
E6 There are many ways to explore information in Spark E7 Spark invites examination of its content E8 Spark is an instrument for discovery E9 Spark is a tool for exploration U2 I encounter the unexpected in Spark
SKD Is Grounded in the User's Prior Knowledge (3)
T5 Spark is full of information useful to me T6 I find information of value to me in Spark C9 I come to understand relationships between ideas in Spark U3 I am surprised by what I find in Spark
Information Organization and Presentation Have Fundamental Roles (4)
C3 I can see connections between topics in Spark C4 It is easy to see links between information in Spark H3 Information that interests me is highlighted in Spark H4 The way that Spark presents content captures my attention H7 I notice content I wouldn't normally pay attention to in Spark H2 Spark has features that ensure that my attention is drawn to useful information H9 I am pointed toward content in Spark H5 I am alerted to information in Spark that helps me

Data Analysis

Data analysis focuses on the use of descriptive statistics, frequency analysis and confirmatory factor analysis (CFA) for both the questionnaire's original groupings as well as the proposed mappings to the IF-SKD model. Data is analyzed using RStudio and various R packages, which is discussed further in Chapter 4, along with SPSS Statistics.

This technique is beneficial because it allows this research to explore earlier studied and explored theories, as well as identified relationships within the literature, from an a priori perspective. This allows for the examination of latent constructs to determine appropriateness of fit with respect to the IF-SKD model.

Results from the questionnaire support the evaluation of these mappings, but also determine how well the questionnaire captures the meaning and significance of serendipity and the aspects of it that contribute to system design. This aids future research and could potentially highlight improvements that could be made to the research tool.

A goal of this analysis, beyond answering the core research questions, is to determine in what ways the questionnaire could be improved in the future. The comparison of findings in this research, along with the analysis of the IF-SKD mappings, may help present valuable insights into refinements to better capture the meaning of serendipity, as well as improve its utility within the clinical setting.

Summary

This chapter highlights the purpose of this study, which is to understand if Spark contributes to physicians' serendipitous knowledge discovery, and to assess the relevance of the IF-SKD model towards capturing physicians' SKD in a clinical setting. The McCay-Peet (2013) Serendipitous Digital Environment questionnaire and Perception of Serendipity Scale are used to capture participant feedback on their use of the Spark system and then paired with the IF-SKD model to evaluate the efficacy of this model in representing physician SKD.

This research offers insight into a relatively new research tool aimed at studying serendipity in digital environments. Its application to physicians, and to Spark, should provide a rich analysis and opportunity to both evaluate the potential applicability of the IF-SKD model to represent physicians' serendipitous knowledge discovery, as well as highlight future improvements to the Serendipitous Digital Environment questionnaire. In addition, it contributes to the understanding of how serendipity is perceived and measured.

CHAPTER 4

DATA ANALYSIS

This chapter summarizes the data and statistics derived from the research instruments. Information about the survey, its completion, and the participants' demographics are provided. Additionally, descriptive statistics are presented and evaluated to provide context prior to the confirmatory factor analyses.

Following an analysis of these data, there is an overview of the confirmatory factor analysis (CFA) approach as well as discussion on sample size and methods utilized to enhance the existing data to support the data analysis. An overview of the software packages and processes used to conduct the CFA are presented, along with a visual presentation of the models. For each model, the same CFA process, fit statistics and output are analyzed to support individual and between model comparisons.

Survey Distribution and Data Collection

The initial survey was distributed via email in late December 2017, with the primary wave of emails being sent in January and February 2018 to INTEGRIS physicians. In total, there were 30 responses to the survey during the collection period. However, five of the total responses were entirely blank, leaving only 25 responses containing data. Of the 25 responses, 14 participants completed the entire survey, whereas nine participants completed part of the survey, leaving some questions blank. For the SDE questionnaire portion, only 23 responses were fully or partially complete.

Of the respondents, six (24%) were female and 19 (76%) were male. Responses to this question can be seen in Figure 4-1.

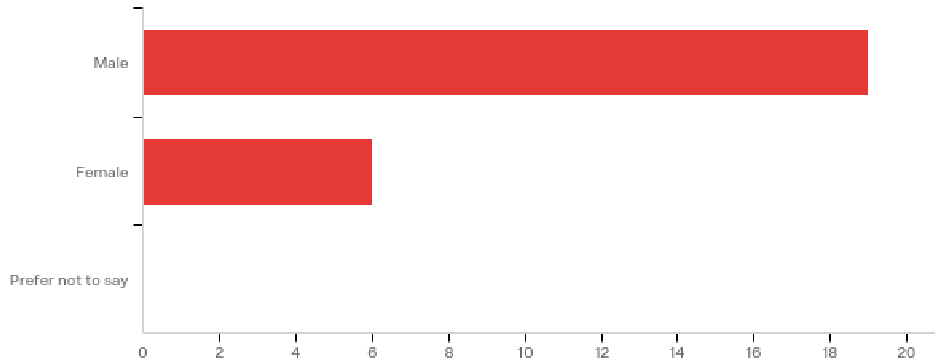

Figure 4-1. Gender summary.

An overview of responses by age bracket is shown in Figure 4-2. The highest number of responses came from the age bracket 41-45 years old, and the lowest number of responses came from the age bracket 65+ years old. No responses were collected from anyone under the age of 30.

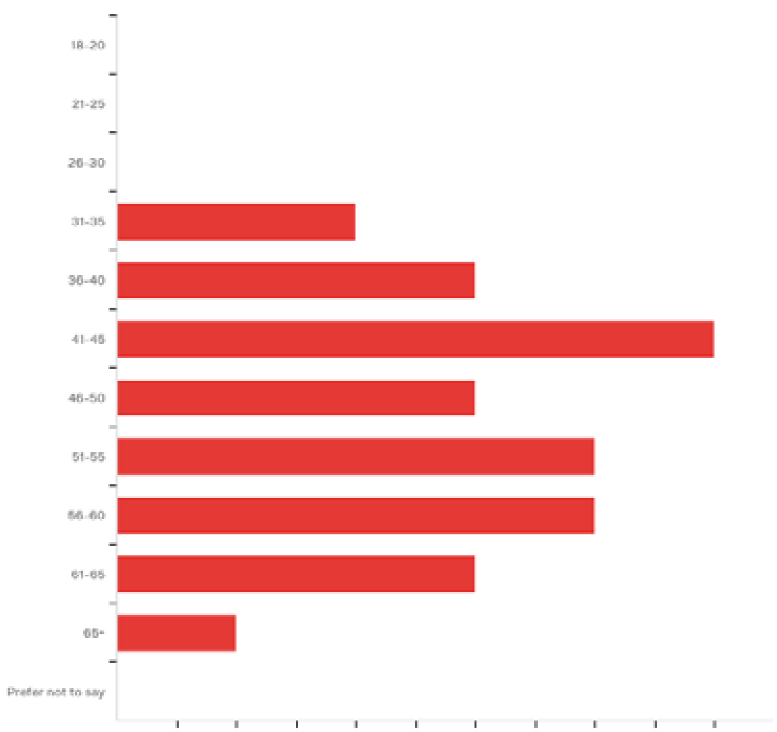

Figure 4-2. Age bracket summary.

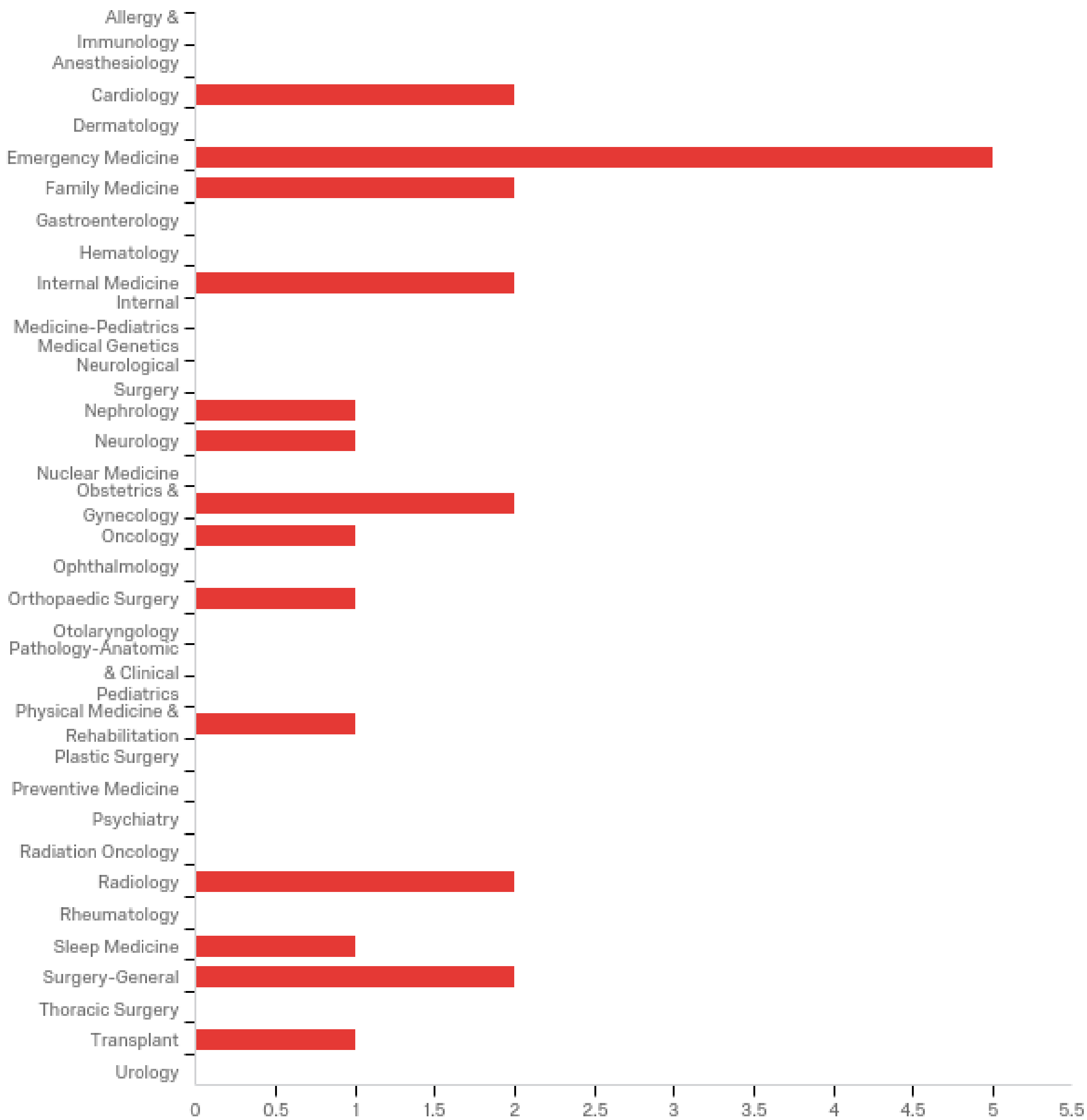

Figure 4-3. Physician specialty summary.

The number of unique specialties who participated in the survey were 14. Figure 4-3 provides a visual summary of the specialties represented in the responses and the total participants within each specialty. Interestingly, emergency medicine was the specialty that had the highest (21%) participation, which given the overall demanding nature of this role was surprising. It is likely that part of the high participation rate was due to some of the respondents acting as gatekeepers for other respondents by passing on the survey and encouraging them to participate. While several specialties participated, there were many that were not represented. For example, gastroenterology, hematology, dermatology, surgery, and psychiatry were some of the specialties for which there were no participants.

In addition, the following figures provide information about survey use by participants. Figure 4-4 provides information on the time participants took in completing the survey. For example, the average amount of time spent participants spent taking the survey was less than 15 minutes, which aligns with expectations set for the anticipated amount of time it would take to complete the research instrument. Even though a few participants took longer than 30 minutes to complete the survey, a majority completed the survey between 8-12 minutes, which includes the time taken to watch the video overview.

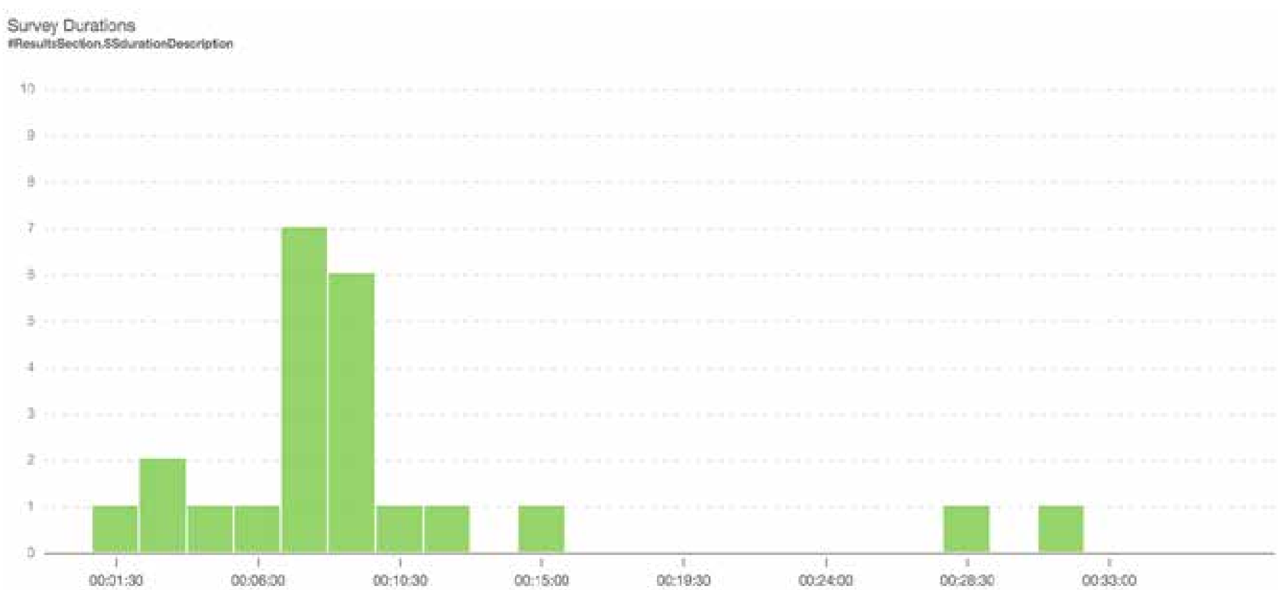

Figure 4-4. Survey durations.

Figure 4-5 illustrates the start times, over an entire day, when participants clicked on the link to initiate the survey. As can be seen, start times ranged across the entire day. Several responses occurred in the late evening.

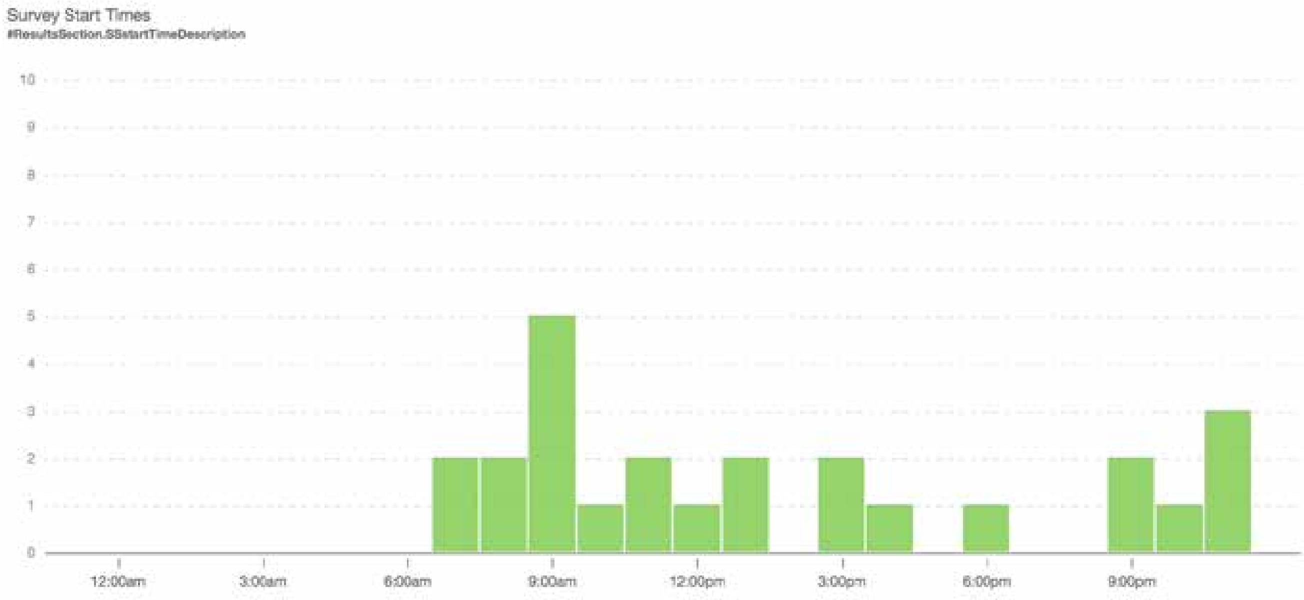

Figure 4-5. Survey start times.

Descriptive Statistics

The following tables present summary descriptive statistics for both the Perception of Serendipity questionnaire as well as the SDE questionnaire. Values for each question, where responses were not provided, are indicated by NA's.

Table 4-1

Survey Data Summary Statistics – Perception of Serendipity Scale

Q1_1	Q1_2	Q1_3	Q1_4
Min. :1.000	Min. :2.000	Min. :2.000	Min. :2.000
1st Qu.:2.000	1st Qu.:3.000	1st Qu.:3.000	1st Qu.:3.000
Median :3.000	Median :3.000	Median :4.000	Median :4.000
Mean :2.857	Mean :3.286	Mean :3.944	Mean :3.706
3rd Qu.:3.000	3rd Qu.:4.000	3rd Qu.:5.000	3rd Qu.:4.000
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
NA's :11	NA's :11	NA's :7	NA's :8

The initial data set for the SDE questionnaire contained some missing values. These along with the means, standard deviations and other values are summarized in Table 4-2.

Table 4-2

Survey Data Summary Statistics – SDE Questionnaire

X1	X2	X3	X4	X5	X6	X7
Min. :3	Min. :2.000	Min. :3.000	Min. :3.000	Min. :3.000	Min. :2	Min. :3.0
1st Qu.:4	1st Qu.:3.00	1st Qu.:4.00	1st Qu.:3.00	1st Qu.:3.00	1st Qu.:3	1st Qu.:4.0
Median :4	Median :4.0	Median :4.0	Median :4.0	Median :4.0	Median :4	Median :4.0
Mean :4	Mean:3.889	Mean:4.053	Mean:3.882	Mean:3.947	Mean :4	Mean :4.2
3rd Qu.:4	3rd Qu:4.75	3rd Qu:4.50	3rd Qu:4.00	3rd Qu:4.50	3rd Qu.:5	3rd Qu.:5.0
Max. :5	Max. :5.00	Max. :5.00	Max. :5.00	Max. :5.00	Max. :5	Max. :5.0
NA's :5	NA's :5	NA's :4	NA's :6	NA's :4	NA's :2	NA's :3
X8	X9	X10	X11	X12	X13	X14
Min. :3.000	Min. :2.00	Min. :3.00	Min. :3.0	Min. :3.00	Min. :3.0	Min. :3.000
1st Qu.:3.00	1st Qu.:4.0	1st Qu.:4.	00 1st Qu.:4.0	1st Qu.:4.00	1st Qu.:4.0	1st Qu.:3.000
Median :4.00	Median :4.0	Median :4.	Median :4.0	Median :4.0	Median :4.0	Median :4.0
Mean :3.737	Mean:4.304	Mean :4.25	Mean :4.2	Mean :3.95	Mean :4.1	Mean:3.882
3rd Qu.:4.00	3rd Qu.:5.0	3rd Qu.:5.0	3rd Qu.:5.0	3rd Qu:4.00	3rd Qu.:5.0	3rd Qu:4.00
Max. :5.00	Max. :5.00	Max. :5.	Max. :5.0	Max. :5.00	Max. :5.0	Max. :5.00
NA's :4		NA's :3	NA's :3	NA's :3	NA's :3	NA's :6
X15	X16	X17	X18	X19	X20	X21
Min. :2.00	Min. :2.00	Min. :2	Min. :3.000	Min. :3.0	Min. :3.000	Min. :3.00
1st Qu.:3.00	1st Qu.:3.	1st Qu.:4	1st Qu:4.00	1st Qu.:4.0	1st Qu:4.00	1st Qu.:3.75
Median :4.0	Median :4.0	Median :4	Median :4.0	Median :4.0	Median :4.0	Median :4.0
Mean :3.833	Mean:3.647	Mean :4	Mean:4.286	Mean :4.2	Mean:4.111	Mean :3.85
3rd Qu.:4.750	3rd Qu.:4.0	3rd Qu.:4	3rd Qu:5.00	3rd Qu.:5.0	3rd Qu:5.00	3rd Qu:4.00
Max. :5.00	Max. :5.00	Max. :5	Max. :5.00	Max. :5.0	Max. :5.00	Max. :5.00
NA's :5	NA's :6	NA's :4	NA's :2	NA's :3	NA's :5	NA's :3

Software

SPSS was used to conduct the frequency analysis that addressed the first hypothesis of whether or not Spark contributed to physicians' serendipitous knowledge discovery.

For the analysis of the second hypothesis, the R environment (R Version 3.4.3, 2017-11-30) was used to compute all the statistics used in the confirmatory factor analyses and descriptive statistics. The RStudio editing environment was used to manage the R scripts and required libraries. Additionally, the following R packages were used:

- Lavaan – used to specify and fit CFA models.
- missForest – used to impute missing data found in the participant responses.
- corpcor – used to estimate shrinkage based positive-definite correlation and covariance matrices.
- MASS – used to perform draws from multivariate normal distribution with a specified mean and covariance matrix (uses function *mvrnorm* to do this).
- Psych – used to calculate estimates of Omega coefficient and to estimate winzorized Monte Carlo estimates.
- Foreign – this package was used to read in the raw survey data so that it could be imputed using missForest.

Perception of Serendipity Questionnaire

The Perception of Serendipity questionnaire, comprised of four questions, was used to evaluate the first research hypothesis that asks whether Spark contributes to physicians' serendipitous knowledge discovery. Similar to the SDE questionnaire, the Perception of Serendipity questionnaire also presented with some missing data.

CFA Overview and the SDE Questionnaire

Confirmatory factor analysis was used to analyze the Serendipitous Digital Environment (SDE) questionnaire. This technique looks at three different models to evaluate how well they represent the data collected from physicians as a part of this study. As previously indicated in Chapter 3, this approach was specifically chosen because the CFA approach is especially useful when the overall study of a topic has a strong conceptual underpinning and initial efforts to measure it are in the early development stages. As Brown (2015) has stated "CFA is almost always used during the process of scale development to examine the latent structure of a test instrument (e.g., a questionnaire)" (p. 1). Work by McCay-Peet (2013) in evaluating this topic using exploratory factor analysis (EFA) was a precursor to the use of a CFA here. McCay-Peet's (2013) work pointed toward a likely 4-factor model, though a 5-factor model was proposed.

In effect, this approach allows for the evaluation of the second research hypothesis, of whether the IF-SKD model reflect physicians' serendipitous knowledge discovery in a clinical setting.

This chapter focuses on the presentation of the models to be analyzed using confirmatory factor analysis and delves into each model's fit statistics to help evaluate them. Moreover, these findings are evaluated in consideration of the SDE questionnaire to assess how well the questions capture aspects of serendipity among respondents and how effectively the instrument performed.

The overall process performed to support the data analysis is captured in Figure 4-6.

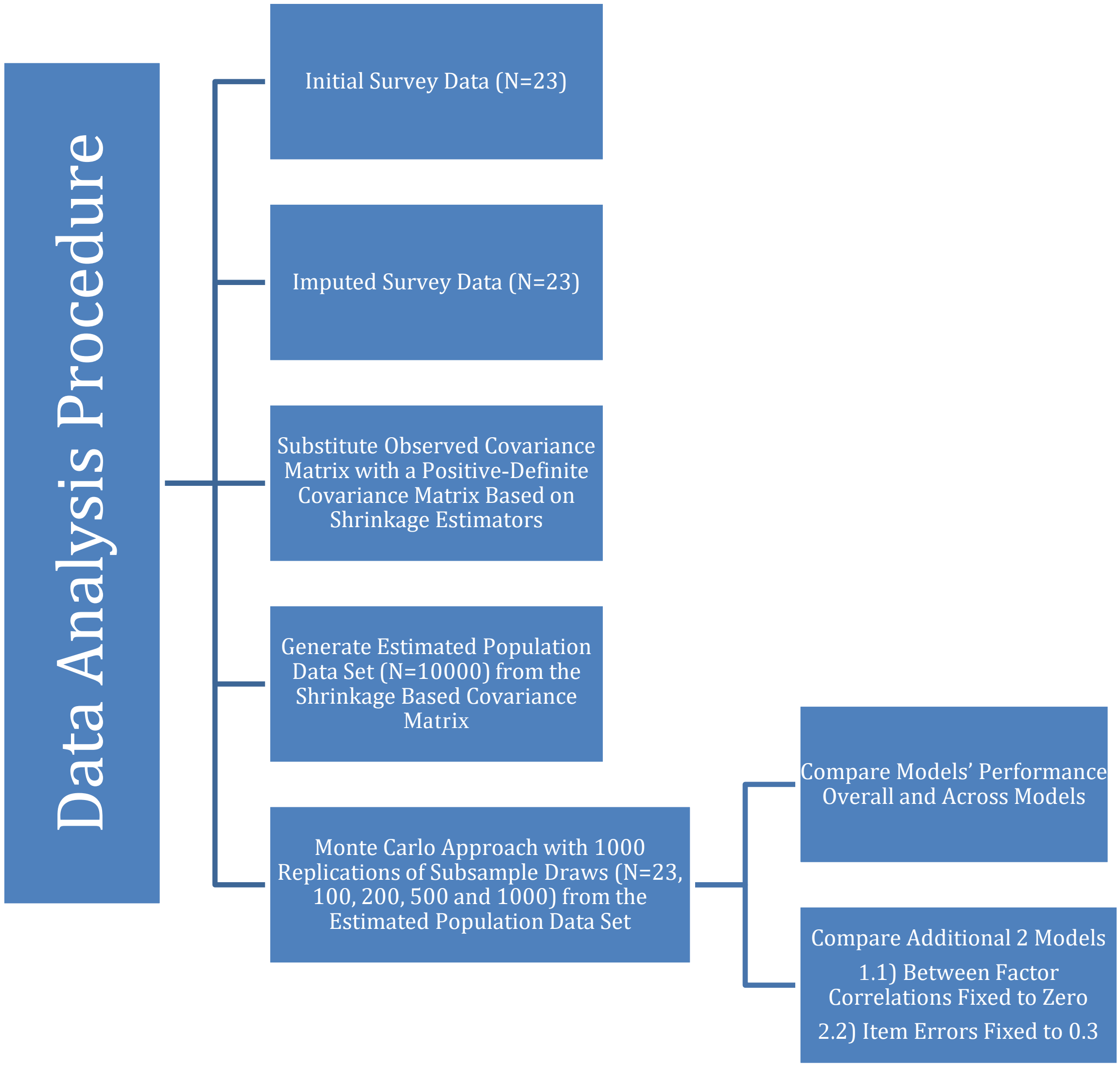

Figure 4-6. Data analysis procedure overview.

Data Imputation , Covariance Shrinkage, Population Estimation and Monte Carlo

Due to the fact that some of the participant survey responses contained missing data, additional steps were necessary to allow for an effective set of confirmatory factor analyses to be possible. This required that an estimated population be generated following data imputation. The following steps outline the tools, strategy and mathematical approaches used to arrive at a final data set that could be studied with the proposed models. However, it is important to first point out that all these steps were not undertaken simply to get the data in a functionally usable state, but rather the literature has supported the use of this type of approach in producing viable data for this type of analysis. Specifically, Krinsky and Robb (1986) demonstrated that the use of Monte Carlo simulations to describe the mean and variances of a random sample was as effective as other methods at representing reliable standard errors.

Data imputation involves an estimation of the raw data set to approximate what values should be selected to replace missing data. The missForest R package was chosen for this task because it handles "categorical data including complex interactions and nonlinear relations" (Stekhoven & Buhlmann, 2011). The bootstrap approach was not selected for use because it is not recommended with sample sizes that are less than 200 (Nevitt & Hancock, 2001).

The data imputation resulted in a new data set with the following statistics for the SDE questionnaire (Table 4-3).

Table 4-3

Imputed Survey Data Summary Statistics – SDE Questionnaire

X1	**X2**	**X3**	**X4**	**X5**	**X6**	**X7**
Min. :3.000	Min. :2.00	Min. :3.000	Min. :3.000	Min. :3.000	Min. :2.000	Min. :3.000
1st Qu.:3.837	1st Qu.:3.50	1st Qu.:3.960	1st Qu.:3.313	1st Qu.:3.440	1st Qu.:3.327	1st Qu.:4.000
Median :4.00	Median :4.00	Median :4.00	Median :4.00	Median :4.00	Median :4.00	Median :4.00
Mean :3.970	Mean :3.92	Mean :4.033	Mean :3.881	Mean :3.948	Mean :3.992	Mean :4.173
3rd Qu.:4.00	3rd Qu.:4.06	3rd Qu.:4.03	3rd Qu:4.023	3rd Qu.:4.00	3rd Qu.:5.00	3rd Qu.:5.00
Max. :5.000	Max. :5.00	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
X8	**X9**	**X10**	**X11**	**X12**	**X13**	**X14**
Min. :3.000	Min. :2.000	Min. :3.000	Min. :3.000	Min. :3.000	Min. :3.000	Min. :3.000

1st Qu.:3.000	1st Qu.:4.000	1st Qu.:4.000	1st Qu.:4.000	1st Qu.:3.825	1st Qu.:4.000	1st Qu.:3.320
Median :3.82	Median :4.00	Median :4.00	Median :4.00	Median :4.00	Median :4.00	Median :4.00
Mean :3.738	Mean :4.304	Mean :4.243	Mean :4.172	Mean :3.924	Mean :4.086	Mean :3.815
3rd Qu.:4.00	3rd Qu.:5.00	3rd Qu:4.755	3rd Qu.:4.55	3rd Qu.:4.00	3rd Qu:4.505	3rd Qu.:4.00
Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000	Max. :5.000
X15	**X16**	**X17**	**X18**	**X19**	**X20**	**X21**
Min. :2.000	Min. :2.000	Min. :2.000	Min. :3.000	Min. :3.000	Min. :3.000	Min. :3.000
1st Qu.:3.000	1st Qu.:3.000	1st Qu.:3.953	1st Qu.:4.000	1st Qu.:3.992	1st Qu.:3.950	1st Qu.:3.910
Median: 4.00	Median3.793	Median :4.00	Median :4.00	Median :4.00	Median :4.00	Median :4.00
Mean :3.832	Mean :3.629	Mean :3.977	Mean :4.263	Mean :4.138	Mean :4.046	Mean :3.862
3rd Qu.:4.04	3rd Qu.:4.00	3rd Qu.:4.00	3rd Qu.:5.00	3rd Qu.:5.00	3rd Qu.:4.51	3rd Qu.:4.00
Max. :5.000	Max. :5.00	Max. :5.00	Max. :5.00	Max. :5.00	Max. :5.00	Max. :5.00

The following steps were used specifically for the data set related to the SDE questionnaire. In addition to data imputation using missForest, the R package corpcor was used to assist with the MASS package in creating the estimation population data by ensuring that the covariance matrix used as input with the MASS function was positive definite. This process would ensure that subsequent samples drawn from that population would generate a positive definite covariance matrix by lavaan when performing the CFA analysis. More specifically, corpcor performs the following steps:

1. Each random variable's empirical variance is calculated and shrunken toward the mean.
2. The shrinkage intensity is then computed using the following formula by Opgen-Rhein and Strimmer (2005). In the formula the median refers to the median of the empirical variances.

$$\lambda^*_{var} = (\sum_{k=1}^{p} Var(s_{kk})) / \sum_{k=1}^{p} (s_{kk} - median(s))^2$$

3. The covariance matrix shrinkage is calculated towards the identity matrix using the following formula by Shafer and Strimmer (2007).

$$\lambda^* = \sum_{k \neq l} Var(r_{kl}) / \sum_{k \neq l} r_{kl}^2$$

Regularization is intended to minimize the variance in the small imputed data set so that the implied covariance matrix produced is still representative of the underlying data, and capable of being analyzed in a CFA framework. The concept of regularization within the literature has taken different forms and matured over time to account for different types of data, such as normal vs. non-normal. Ridge regression is one of the ways regularization has been employed. For example, "in the case of severe multicollinearity in a regression model, without imposing a bit of bias on the regression coefficient estimates via ridge regression, it would be impossible to obtain estimates of these coefficients" (Mooney & Duval, 1993 p. 44). Another way to envision regularization is as a process whereby additional new information is introduced in an effort to address an ill-posed question (Neumaier, 1998).

It is not possible to always have an ideal sample from which to run a set of statistics. Evaluating the least impactful approach to regularizing data to support the goals of the research, within realistic bounds of interpretation, is the goal of this proposed approach. For this research, due to the low sample size, data imputation, along with covariance shrinkage, was used to obtain an estimated population (N=10,000) from which Monte Carlo simulation of subsamples were drawn and then fitted to each model to support fit statistic comparisons and to understand the changes of sample size occurring on each of the models. Tofighi and MacKinnon (2016) have noted that while there are different approaches to performing summary analysis in structure equation modeling (SEM), "the Monte Carlo method produces more accurate results especially for smaller sample sizes" (p. 194).

The rationale underlying the use of the Monte Carlo method in this study, is to generate many Monte Carlo replications (e.g. 1000 replications) of subsample size draws of N=23, 100, 200, 500 and 1000 from the estimated populations. This allows evaluation of the confirmatory factor analysis (CFA) models, and the CFA fit statistics, as the sample size increases. Moreover, this method allows valid estimates of standard errors for factor loadings and factor correlations for the original small sample size of N=23. The diagonally-weighted least squares (DWLS) parameter estimation method is used in combination with the Monte Carlo simulations to estimate the CFA models. Essentially, this research utilizes, in order to deal with the small sample size problem, Monte Carlo based DWLS parameter estimation, utilizing shrinkage estimators for the observed covariance matrix, referred to here as MC-SDWLS.

Marcoulides & Sanders (2006) discuss the use of Monte Carlo analysis in two different ways: proactive and reactive. The former, while identified as preferable and more likely to produce valid confidence intervals, is not always easy to conduct since the information about the entire population may not be known. Instead, this study looks at using Monte Carlo in a reactive way. By conducting a thousand iteration runs of varying sample sizes, we can see at what level of sample size we begin to assess meaningful information about the sampling error, and as such be able to better compare the fit indices of the models with more confidence.

Marcoulides proposed doing this through the use of a t-test to assess the significance of one statistics between models (2006).

Testing the MC-SDWLS Approach Using a Simulated Population Model

To further validate the MC-SDWLS method utilized, a Monte Carlo simulation was performed with the McCay-Peet model as the known true model that generates the population data, given that the observed shrunken covariance matrix (with n=23) was generated from the McCay-Peet population model. The simulation was accomplished using the function simulateData within the R package lavaan. Specifically, the shrunken observed covariance matrix, based on the imputed data set of N=23, was used in conjunction with the unconstrained McCay-Peet model as the true population model. Thus, creating a 10,000-record dataset to represent the population under the McCay-Peet model.

Using these population data, the MC-SDWLS simulation was performed to estimate winsorized mean point estimates; winsorized mean fit statistics; and standard errors for these point estimates and fit statistics; for both the McCay-Peet model and the IF-SKD model. A two-sample t-test was performed between the McCay-Peet and the IF-SKD model, using the mean Fmins, across Monte Carlo replications, and standard errors obtained from these Monte Carlo replications (using 1000 Monte Carlo replications). As expected, the t-test statistically significantly favored the McCay-Peet model which was actually the true generating model, when compared with the IF-SKD model.

The *Fmin* is the objective function that is minimized during optimization of the lavaan CFA model. When the data are drawn from a multivariate normal distribution, minimizing the

Fmin (the difference between the observed and implied covariance matrix) also minimizes the so-called Kullback-Liebler divergence. Wang and Jo (2013) explained that the Kullback-Liebler divergence “can be viewed as a measure of the information loss in the fitted model relative to that in the reference model” (p. 409). This fact motivates the use of the Fmin statistic as a way of discriminating the relative differences in goodness of fit between each respective model assumed generating population model. It is worth noting that they do not need to be the same generating population model. Consequently, a t-test test of statistical significance between the Fmin of two different can determine which model is better at approximating their respective reference models.

In summary, in the situation of the Monte Carlo simulation with a known population structure, having the statistical test on the Fmin objective function values favor the McCay-Peet model fit over the IF-SKD fit, when the true generating model was the McCay-Peet model, gives us some confidence that the MC-SDWLS methodology developed here can work to select a best approximating model, in a relative sense (as opposed to an absolute goodness of fit.

CFA Technique for all Models

In this section the technique, measurements and evaluation criteria to be used for each model are presented along with justification for these approaches as outlined in the literature based upon the research instrument and the stated goals of the research.

Before discussing the fit statistics and interpretation guidance criteria for this study, the estimation method used to conduct the CFA must be addressed. There is an array of different estimation methods for conducting a CFA. For example, Brown (2015) points out several estimation methods exist, with Maximum Likelihood (ML) being a common and effective one focused on the analysis of continuous data, though this is influenced by low sample sizes.

Other estimation methods, include: 1) Generalized Least Squares (GLS); 2) Weighted Least Squares (WLS); 3) Diagonally Weighted Least Squares (DWLS), sometimes also referred to by the acronym WLSMV; 4) Unweighted Least Squares (ULS); 4) variants, including robust ML and ML with different standard error reporting. Maximum Likelihood (ML), which is considered one of the better CFA estimator methods, suffers from small sample sizes (Brown, 2015).

Of all the estimation approaches available, the diagonally weighted least squares (DWLS) was chosen. Li (2016), in a recent study, utilized a Monte Carlo approach to evaluate DWLS, ULS and Robust ML under a variety of different ordinal data conditions and distributional shapes. Li (2016) showed that DWLS performed best, especially in accounting for the factor loading and in producing "more accurate interfactor correlation estimates" (p.369). Using a diagonally weighted matrix, as opposed to an inverse matrix, in computing fit statistics, DWLS allows for easier comparison for small sample sizes and handles well with nonnormal data (Rhemtulla, Brosseau-Laird, & Savalei, 2012). Marsh and Grayson (1995) summarized the decision to choose an approach well, stating that "a general approach is to establish that the model is identified, that the iterative estimation procedure converges, that all parameter estimates are within the range of permissible values, and that the standard errors of the parameter estimates have reasonable size" (p. 198). In selecting DWLS and evaluating the models relative to each other, while also looking at the corrected fit indices, allows for rich analysis and comparison on a variety of different fronts, which is a goal for this type of analysis.

Table 4-4

Summary of Fit Statistics

Cutoff Criteria for Several Fit Indexes

Indexes	Shorthand	General rule for acceptable fit if data are continuous	Categorical data
Absolute/predictive fit			
Chi-square	χ^2	Ratio of χ^2 to $df \leq$ 2 or 3, useful for nested models/model trimming	
Akaike information criterion	AIC	Smaller the better; good for model comparison (nonnested), not a single model	
Browne–Cudeck criterion	BCC	Smaller the better; good for model comparison, not a single model	
Bayes information criterion	BIC	Smaller the better; good for model comparison (nonnested), not a single model	
Consistent AIC	CAIC	Smaller the better; good for model comparison (nonnested), not a single model	
Expected cross-validation index	ECVI	Smaller the better; good for model comparison (nonnested), not a single model	
Comparative fit		Comparison to a baseline (independence) or other model	
Normed fit index	NFI	≥ .95 for acceptance	
Incremental fit index	IFI	≥ .95 for acceptance	
Tucker–Lewis index	TLI	≥ .95 can be 0 > TLI > 1 for acceptance	0.96
Comparative fit index	CFI	≥ .95 for acceptance	0.95
Relative noncentrality fit index	RNI	≥ .95, similar to CFI but can be negative, therefore CFI better choice	
Parsimonious fit			
Parsimony-adjusted NFI	PNFI	Very sensitive to model size	
Parsimony-adjusted CFI	PCFI	Sensitive to model size	
Parsimony-adjusted GFI	PGFI	Closer to 1 the better, though typically lower than other indexes and sensitive to model size	
Other			
Goodness-of-fit index	GFI	≥ .95 Not generally recommended	
Adjusted GFI	AGFI	≥ .95 Performance poor in simulation studies	
Hoelter .05 index		Critical *N* largest sample size for accepting that model is correct	
Hoelter .01 index		Hoelter suggestion, *N* = 200, better for satisfactory fit	
Root mean square residual	RMR	Smaller, the better; 0 indicates perfect fit	
Standardized RMR	SRMR	≤ .08	
Weighted root mean residual	WRMR	< .90	< .90
Root mean square error of approximation	RMSEA	< .06 to .08 with confidence interval	< .06

Table 4-4 presents a summary of the fit statistics, compiled by work from Schreiber, Stage, King, Nora and Barlow (2006), that are used to guide the interpretation of the results from the study. In addition, the fit statistics are grouped into categories, type of fit statistic, highlighting their value in interpreting the findings in this study, as well as areas where they are impacted by limitations of the study. While there are specific cutoffs listed, the approach taken in this analysis is to evaluate each model with one another in addition to looking at its overall score on certain indices. This allows for the evaluation of the second null hypothesis asking whether the IF-SKD model reflects physicians' serendipitous knowledge discovery in a clinical setting, while considering its score in comparison to other proposed models. It also allows for a more generalizable interpretation that support calls for future research.

To improve the understanding of the significance of these estimates for each specific model and also between the models, the confidence intervals, point estimates, standardized point estimates and percentiles (2.5% and 97.5%) are calculated. The calculations are performed across all the samples and averaged to provide information about the 1000 simulations for each model. Tofighi and MacKinnon (2015) found the Monte Carlo approach to evaluating results was an effective way to draw on the law of large numbers to evaluate these statistics, further finding that the Monte Carlo approach was as effective as bootstrapping or alternative methods. The reason the percentiles are evaluated is to determine if the distribution of the data is non-normal, which helps provide a better conservative indication of the upper and lower bounds of likely values for any specific model. This helps demonstrate what type of fit is represented by the numbers that are 2.5% and 97.5% underneath the distribution curve. This is significant because, as Tofighi and MacKinnon (2015) further pointed out, that there is a limitation to bootstrapping in evaluating confidence limits "for sample sizes smaller than 100 due to substantial variability in the confidence limits across bootstrap samples" (p. 197).

Assumptions

It is important to discuss the assumptions that are made in this research. Because the sample size is small, there is an assumption that the implied covariance matrix derived from the imputed data is representative of the actual population being studied. That is to say that if there were actually 500 or 1000 respondents to the survey that the covariance matrix for that would be similar to the one implied by the approach taken here. For that reason, this study evaluates the models not only under different specification conditions, but with different Monte Carlo sub-sample size draws (N=23, N=100, N=200, N=500, N=1000) from the estimated population to demonstrate that any observed values in the different models are not attributed to sampling fluctuation. Conceptually, this should also limit the distances between the percentiles for the goodness of fit statistics as sample size increases.

Perception of Serendipity Analysis

This section examines the responses from the Perception of Serendipity questions. Frequency analysis and an Omega reliability coefficient are used to examine the scales internal reliability and to demonstrate the percent of responses submitted by participants with respect to whether Spark contributed to their serendipitous knowledge discovery. McCay-Peet (2013) showed these data, along with two other question sets not employed in this study, as explaining a portion of the variance of the latent variables on the SDE questionnaire. The means of the items on the questionnaire were used in multiple regression to demonstrate this. However, for the purpose of this study, the use of this research instrument was limited to the interpretation of Spark specifically.

The frequency analysis summary and frequency analysis per questions for the data on the Perception of Serendipity Scale are presented in Tables 4-5 through 4-9. This information provides a breakdown on the responses, along with the cumulative percent and valid percent, the percentage with no missing data, reported.

Additionally, reliability statistics, using Omega was used to capture the reliability of the scale amongst the items being measured. At a 0.9 value, the scale is deemed reliable, or very good, according to guidelines for interpreting these statistics (DeVellis, 2003, p. 98-96).

Table 4-5

Summary of Frequency Statistics

		Spark - Impact on Everyday Life	Spark - Impact on my Work	Spark - Encounter Useful Information Not Looking For	Spark - Experience Insight Leading to Unanticipated Outcomes
N	Valid	14	14	18	17
	Missing	11	11	7	8
Median		3.0000	3.0000	4.0000	4.0000
Mode		2.00[a]	3.00[a]	5.00	4.00
Minimum		1.00	2.00	2.00	2.00
Maximum		5.00	5.00	5.00	5.00

a. Multiple modes exist. The smallest value is shown

Table 4-6

Summary of Frequency Statistics - Impact on Everyday Life

		Frequency	Percent	Valid Percent	Cum %
Valid	1.00	1	4.0	7.1	7.1
	2.00	5	20.0	35.7	42.9
	3.00	5	20.0	35.7	78.6
	4.00	1	4.0	7.1	85.7
	5.00	2	8.0	14.3	100.0
	Total	14	56.0	100.0	
Missing	System	11	44.0		
Total		25	100.0		

Table 4-7

Summary of Frequency Statistics - Impact on My Work

		Frequency	Percent	Valid Percent	Cum %
Valid	2.00	3	12.0	21.4	21.4
	3.00	5	20.0	35.7	57.1
	4.00	5	20.0	35.7	92.9
	5.00	1	4.0	7.1	100.0
	Total	14	56.0	100.0	
Missing	System	11	44.0		
Total		25	100.0		

Table 4-8

Summary of Frequency Statistics - Useful Information

		Frequency	Percent	Valid Percent	Cum %
Valid	2.00	1	4.0	5.6	5.6
	3.00	6	24.0	33.3	38.9
	4.00	4	16.0	22.2	61.1
	5.00	7	28.0	38.9	100.0
	Total	18	72.0	100.0	
Missing	System	7	28.0		
Total		25	100.0		

Table 4-9

Summary of Frequency Statistics - Unanticipated Outcomes

		Frequency	Percent	Valid Percent	Cum %
Valid	2.00	1	4.0	5.9	5.9
	3.00	5	20.0	29.4	35.3
	4.00	9	36.0	52.9	88.2
	5.00	2	8.0	11.8	100.0
	Total	17	68.0	100.0	
Missing	System	8	32.0		
Total		25	100.0		

The following figures represent the data distributions reported on the Perception of Serendipity Questionnaire in bar charts.

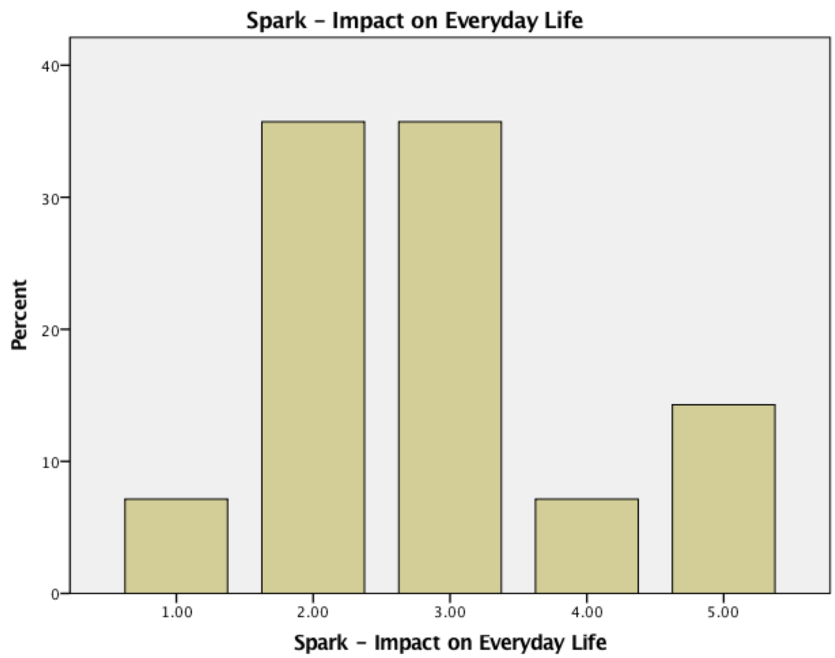

Figure 4-7. Spark bar chart – impact on everyday life.

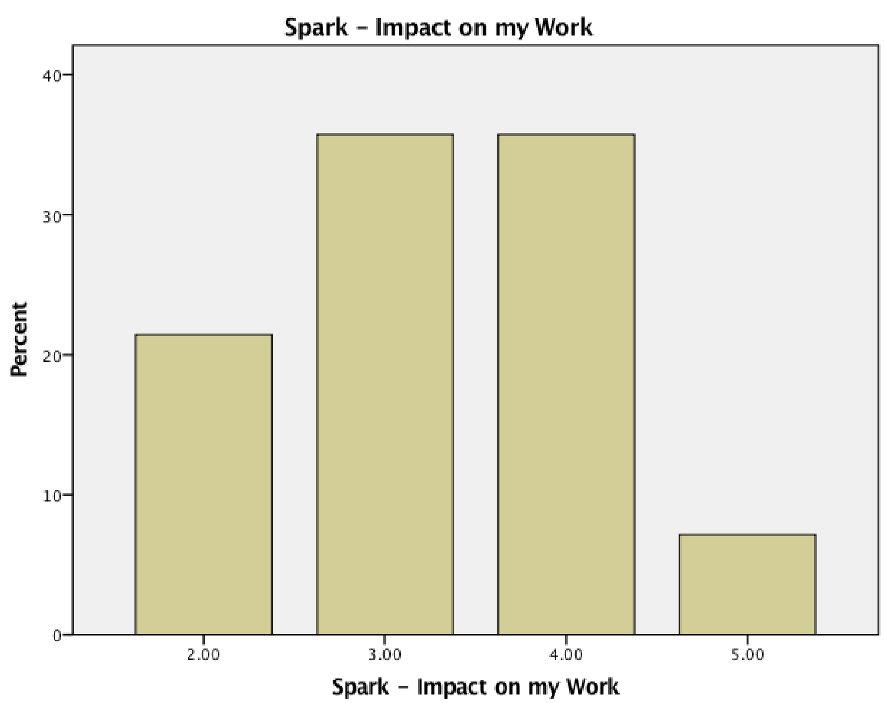

Figure 4-8. Spark bar chart – impact on my work.

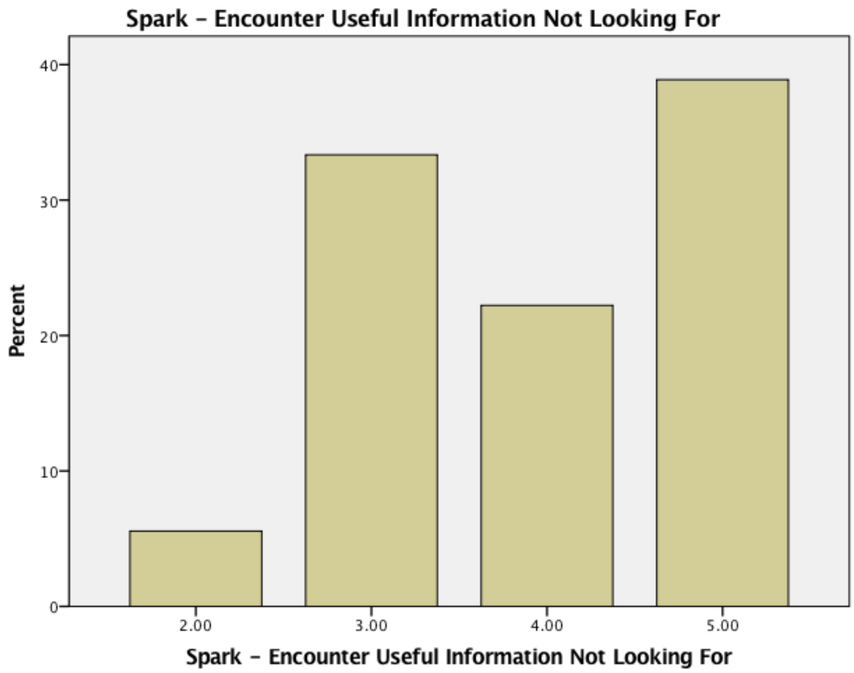

Figure 4-9. Spark bar chart – useful information.

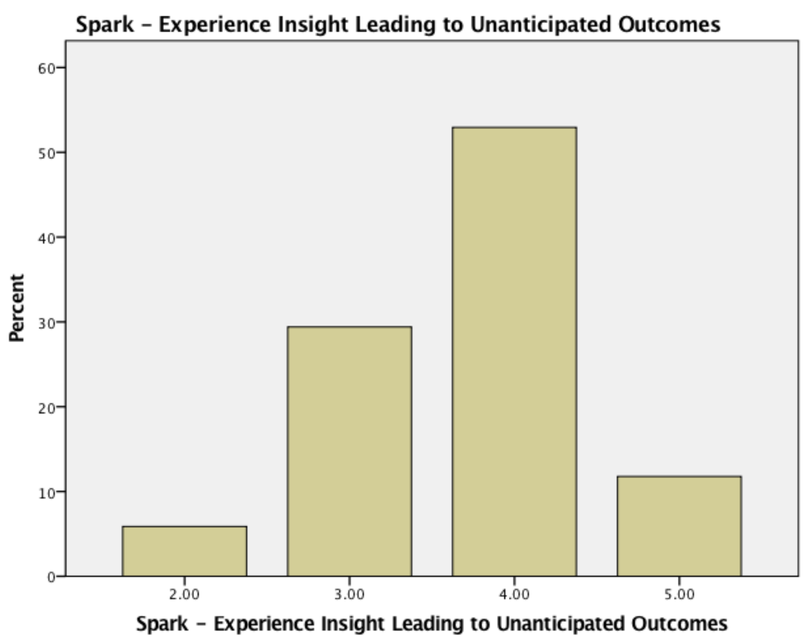

Figure 4-10. Spark bar chart – unanticipated outcomes.

Looking at the frequency output, it is possible to see that for all Perception of Serendipity questions, > 50% of respondents reported Spark contributing to their experience of serendipity either sometimes, frequently or very frequently. For some questions, this was even higher.

Table 4-10 shows a summary of the reporting of Spark contributing to serendipity **A)** either frequently or very frequently, and **B)** sometimes, frequently, or very frequently.

Table 4-10

Breakdown of Select Frequency Statistics

Question	% Reporting Spark contributing to serendipity frequently, or very frequently	% Reporting Spark contributing to serendipity sometimes, frequently, or very frequently
In Spark, I experience serendipity that has an impact on my everyday life.	21.4%	57.1%
In Spark, I experience serendipity that has an impact on my work.	42.8%	78.5%
I encounter useful information, ideas, or resources that I am not looking for when I use Spark.	61.1%	94.4%
In Spark, I experience mixes of unexpectedness and insight that lead to valuable, unexpected outcomes.	64.7%	84.1%

From these results, it is clear that Spark is perceived as contributing to serendipitous knowledge discovery. If limited only to the “frequently” and “very frequently” options, there is still a majority, >50%, frequency, supporting the contention with respect to Spark contributing to participants encountering information they were not seeking which led to unexpected valuable outcomes.

CFA Model Specifications

A path diagram for each model is presented in the next sections. These diagrams highlight the latent variables being evaluated and their corresponding predicted relationship to the indicators used in the Serendipitous Digital Environment (SDE) questionnaire.

The arrows from each latent variable show the proposed relationship that exists between each indicator and its latent variable. The double-ended arrows between latent variables shows that, as part of the CFA analysis, the correlation between latent variables is evaluated.

IF-SKD Model

Figure 4-11 depicts the standardized indicator to latent variables mappings and item errors for the information flow-serendipitous knowledge discovery (IF-SKD) model and the Serendipitous Digital Environment questionnaire at N=23 as captured using the Monte Carlo simulation. Figure 4-12 and Figure 4-13 show the variations on the same model tested to evaluate fit statistics under different model latent variable and indicator fixed conditions. In Figure 4-12, the model is the same, with the exception being the between factor correlations are fixed at zero. Figure 4-13 shows the same model as the freely estimated model Figure 4-11 except that the indicator item errors are fixed at 0.3.

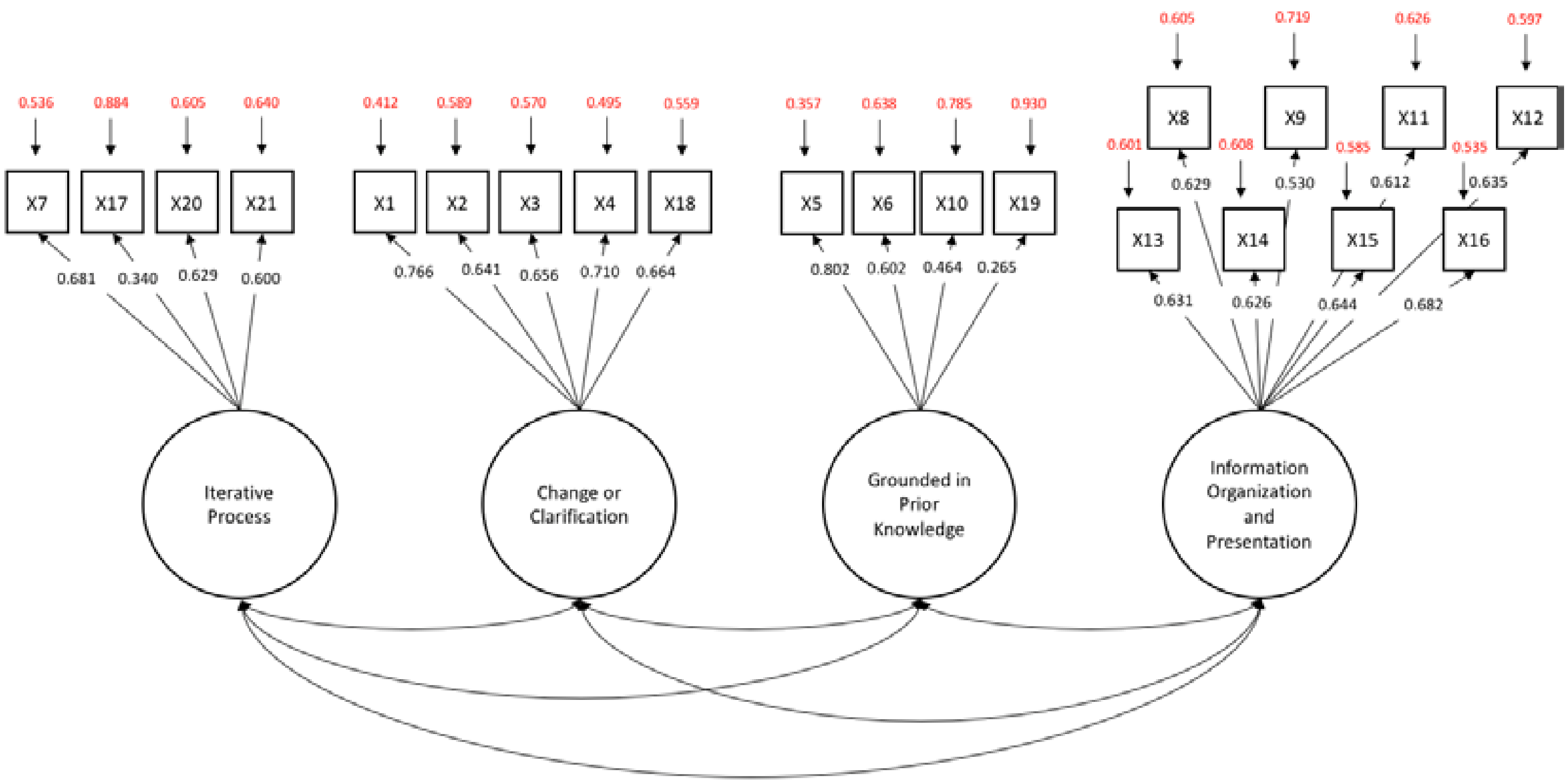

Figure 4-11. IF-SKD model representation.

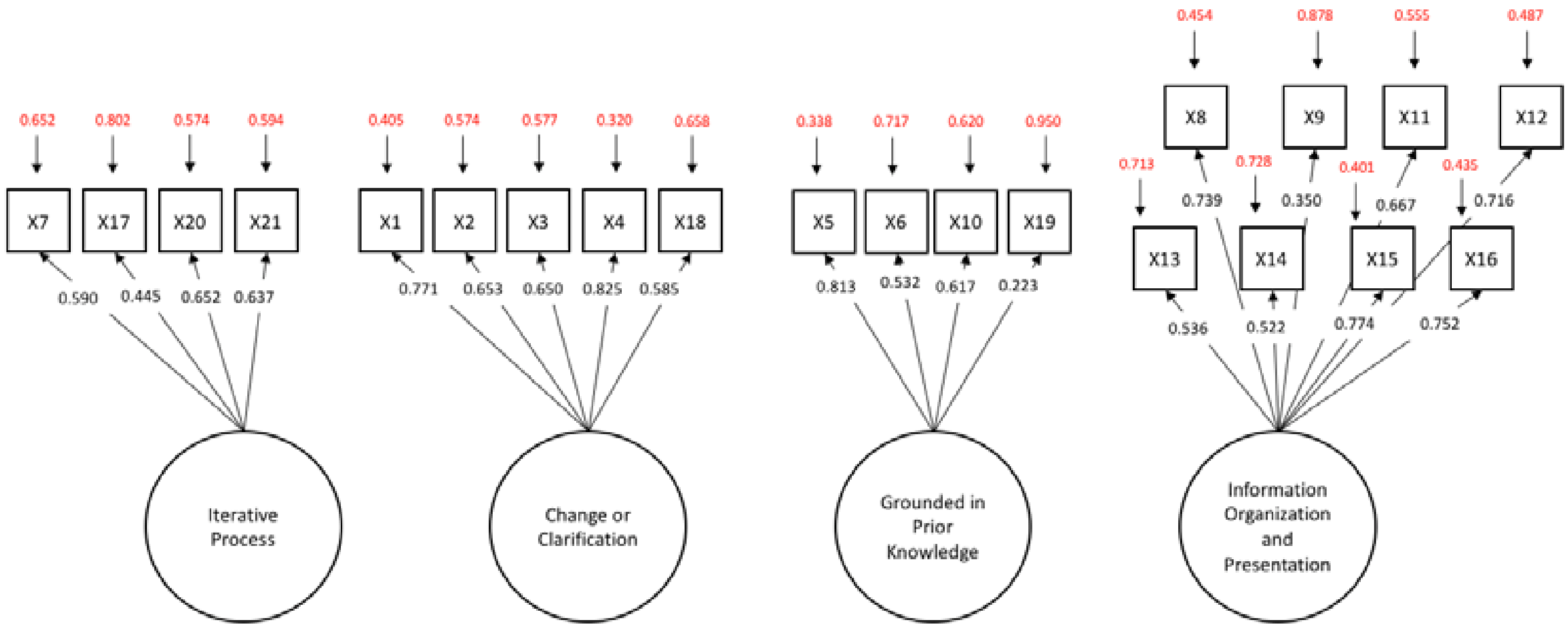

Figure 4-12. IF-SKD model representation – fixed zero correlation between factors.

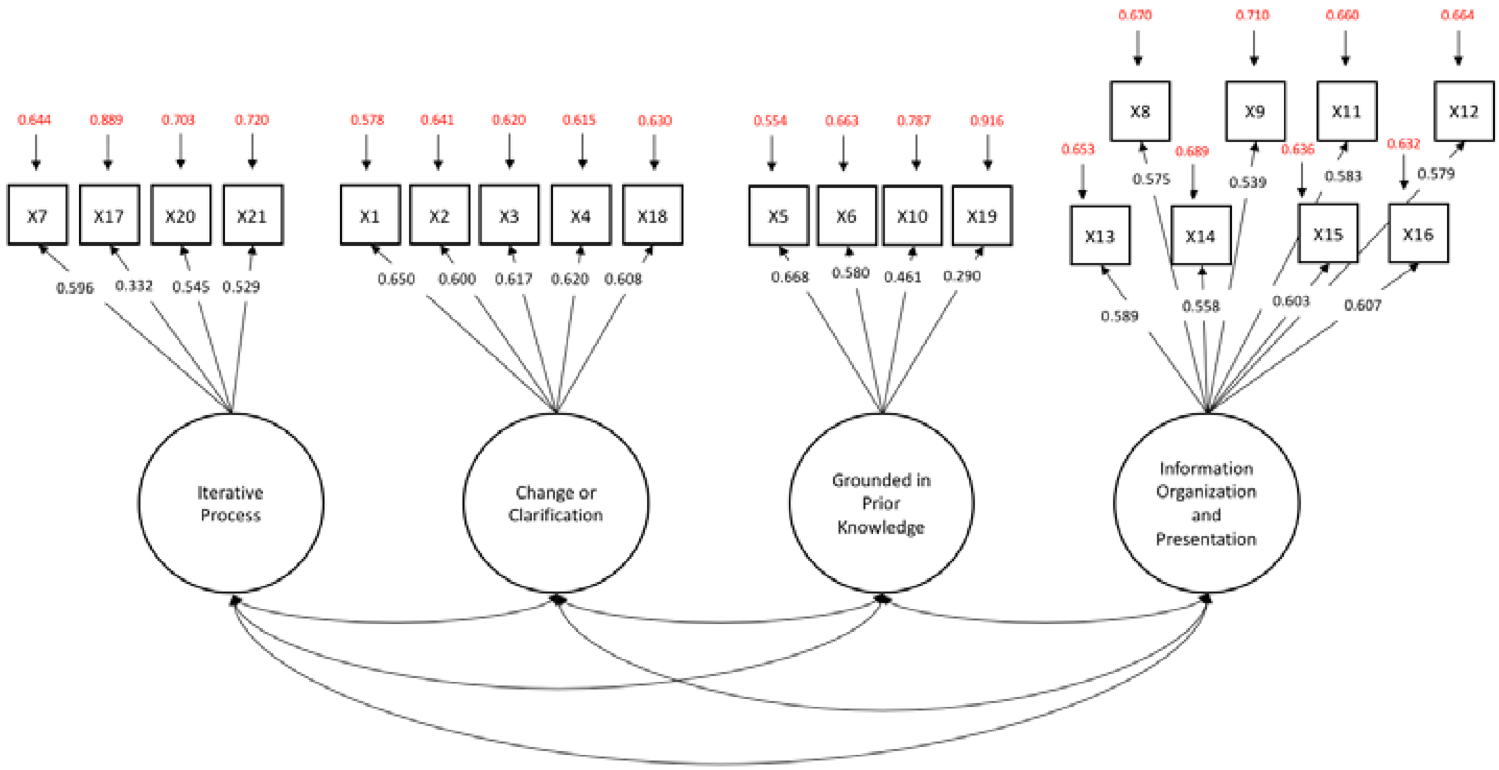

Figure 4-13. IF-SKD model representation – fixed indicator errors.

McCay-Peet Model

Figure 4-14 depicts the standardized indicator to latent variables mappings and item errors for the McCay-Peet model and the Serendipitous Digital Environment questionnaire at N=23 as captured using the Monte Carlo simulation. Figure 4-15 and Figure 4-16 show the variations on the same model tested to evaluate fit statistics under different model latent variable and indicator fixed conditions. In Figure 4-15, the model is the same, though the between factor correlations are fixed at zero. Figure 4-16 shows the same model as the freely estimated model, Figure 4-14, except that the indicator item errors are fixed at 0.3.

Single Factor Model

Figure 4-17 represents the standardized indicator to latent variables mappings and item errors for the Single model and the Serendipitous Digital Environment questionnaire at N=23 as captured using the Monte Carlo simulation. Like the other two models, similarly Figure 4-18 shows the same model as the freely estimated model, Figure 4-17, with the indicator item errors fixed at 0.3.

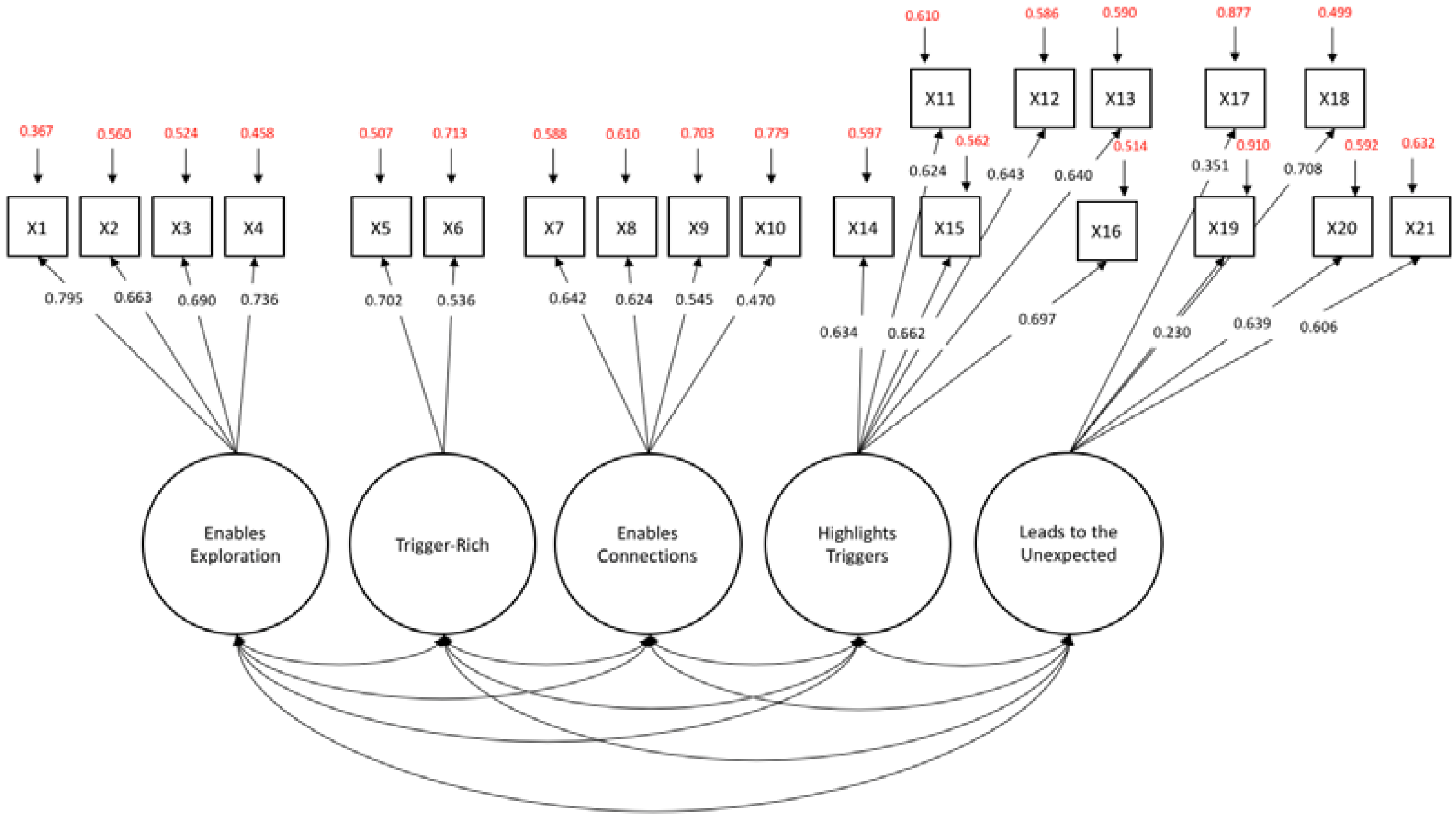

Figure 4-14. McCay-Peet model representation.

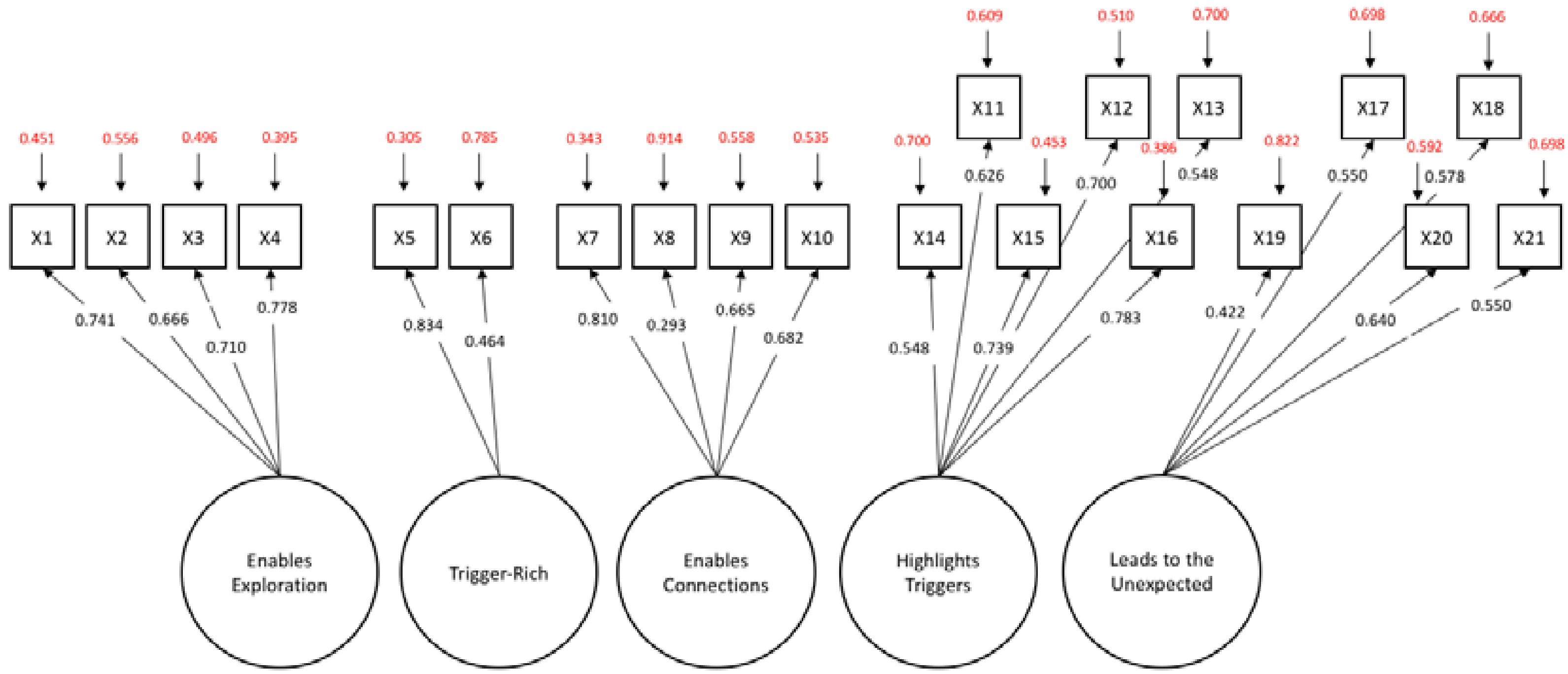

Figure 4-15. McCay-Peet model representation – fixed zero correlations between factors.

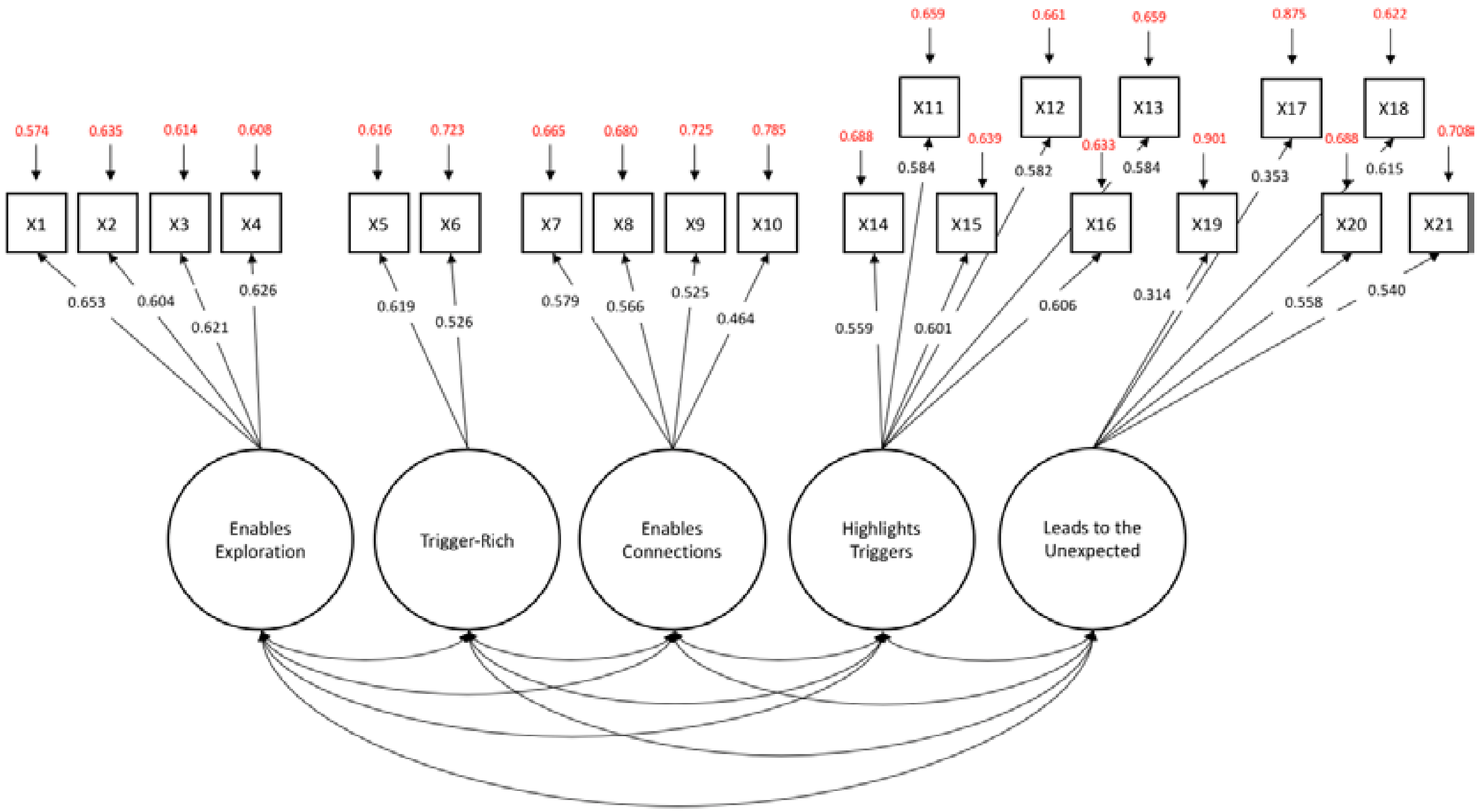

Figure 4-16. McCay-Peet model representation – fixed indicator errors.

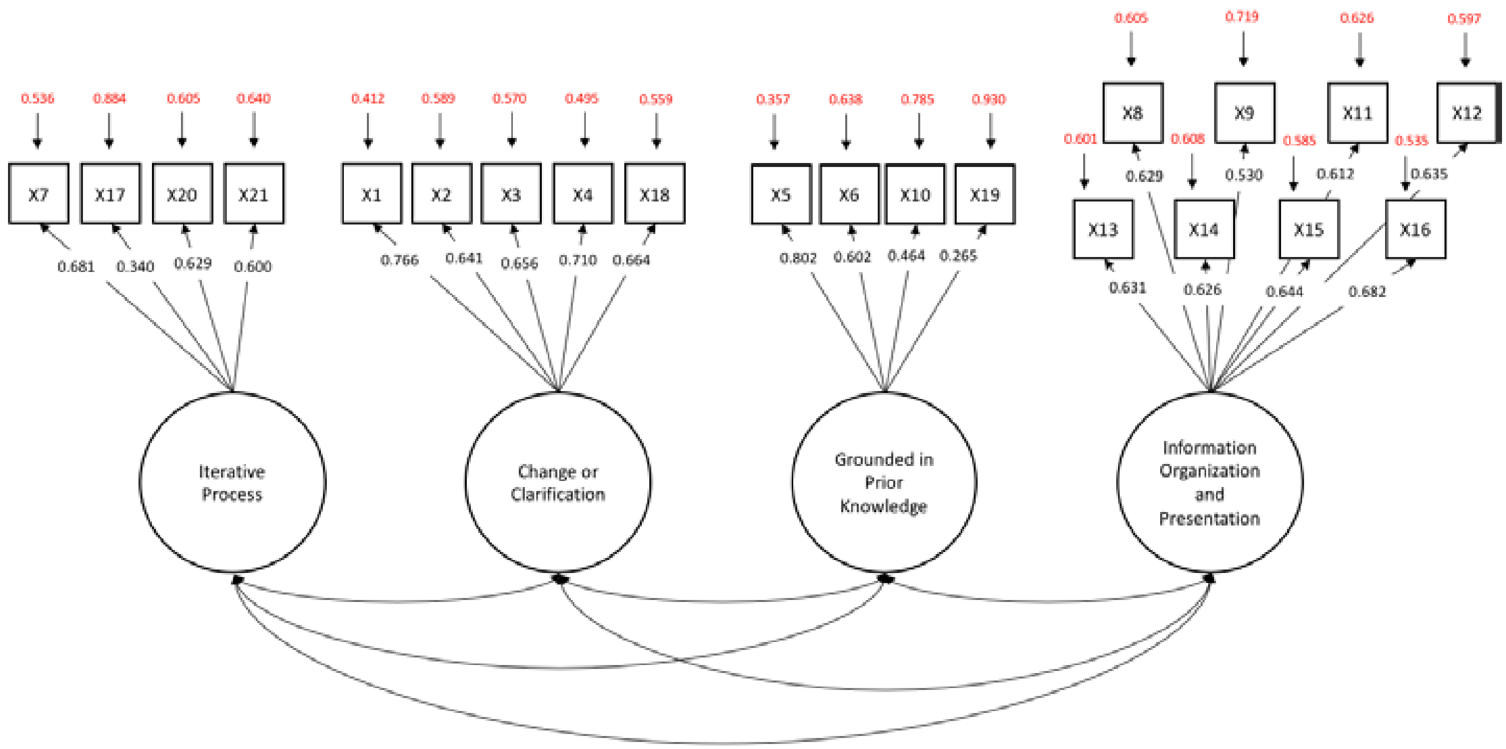

Figure 4-17. Single factor model representation.

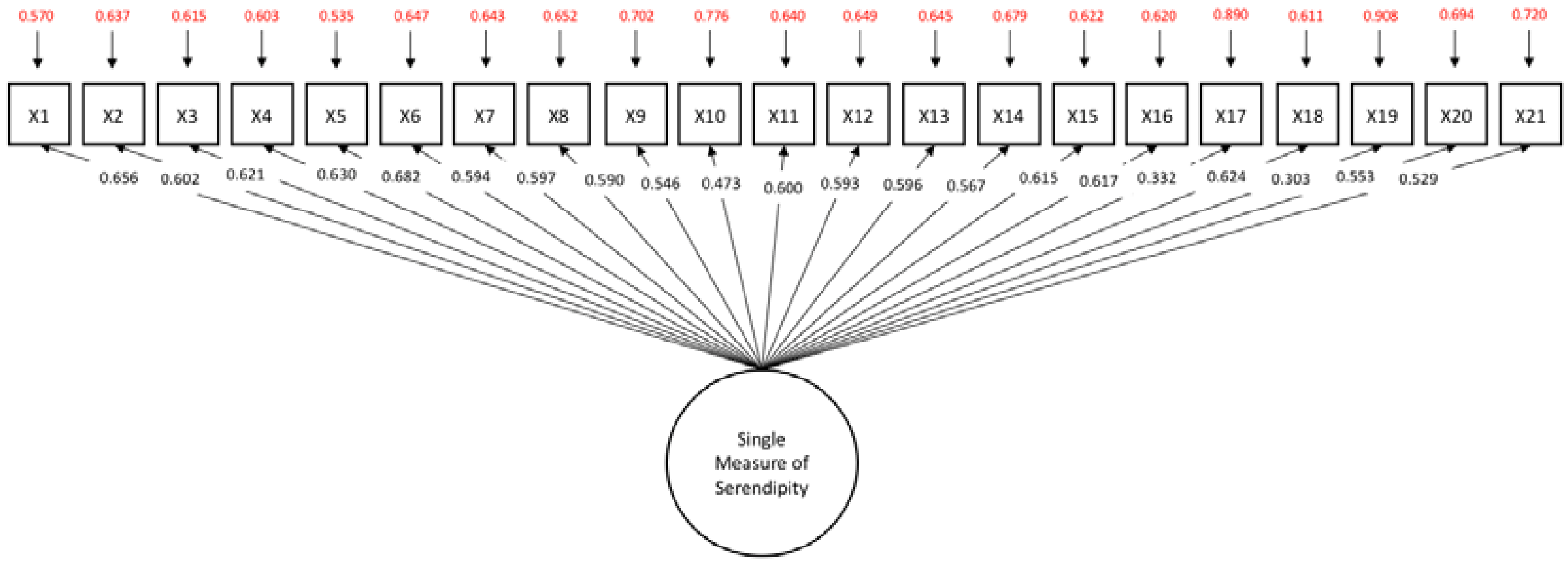

Figure 4-18. Single factor model representation – fixed indicator errors.

IF-SKD Model CFA Analysis

Using varying sized sub-samples of the estimated population created from the imputed data set using the MASS R package, the following fit statistics were obtained for each sub-sample and evaluated using the lavaan CFA function for the IF-SKD model.

Fit Statistics

Tables 4-11 through 4-13 show the mean fit statistics and standard errors for the IF-SKD model specified in its:

- Freely estimated form
- Fixed latent variables to have zero correlation between factors
- Fixed indicator errors (0.3) form

Table 4-11

IF-SKD Model Fit Statistics

IF-SKD.Model.Free		IF-SKD.Model.Free		IF-SKD.Model.Free		IF-SKD.Model.Free		IF-SKD.Model.Free	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
fmin	2.0911	fmin	0.9395	fmin	0.8105	fmin	0.7420	fmin	0.7181
se	0.5980	se	0.2202	se	0.1301	se	0.0774	se	0.0527
chisq	96.1907	chisq	187.9048	chisq	324.2199	chisq	742.0415	chisq	1436.1342
se	27.5067	se	44.0478	se	52.0489	se	77.4004	se	105.4172
pvalue	0.9933	pvalue	0.4857	pvalue	0.0008	pvalue	0.0000	pvalue	0.0000
se	0.0255	se	0.4021	se	0.0037	se	0.0000	se	0.0000
cfi	1.0000	cfi	0.9951	cfi	0.9840	cfi	0.9749	cfi	0.9720
se	0.0000	se	0.0074	se	0.0067	se	0.0043	se	0.0030
tli	1.1623	tli	0.9980	tli	0.9817	tli	0.9712	tli	0.9678
se	0.0827	se	0.0118	se	0.0077	se	0.0049	se	0.0034
nnfi	1.1623	nnfi	0.9980	nnfi	0.9817	nnfi	0.9712	nnfi	0.9678
se	0.0827	se	0.0118	se	0.0077	se	0.0049	se	0.0034
rfi	0.8596	rfi	0.9533	rfi	0.9593	rfi	0.9622	rfi	0.9633

IF-SKD.Model.Free		IF-SKD.Model.Free		IF-SKD.Model.Free		IF-SKD.Model.Free		IF-SKD.Model.Free	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
se	0.0776	se	0.0153	se	0.0090	se	0.0053	se	0.0035
nfi	0.8776	nfi	0.9593	nfi	0.9645	nfi	0.9671	nfi	0.9680
se	0.0676	se	0.0133	se	0.0079	se	0.0046	se	0.0031
pnfi	0.7648	pnfi	0.8360	pnfi	0.8405	pnfi	0.8427	pnfi	0.8436
se	0.0589	se	0.0116	se	0.0069	se	0.0040	se	0.0027
ifi	1.1343	ifi	0.9983	ifi	0.9841	ifi	0.9750	ifi	0.9720
se	0.0652	se	0.0102	se	0.0067	se	0.0043	se	0.0030
rni	1.1415	rni	0.9983	rni	0.9840	rni	0.9749	rni	0.9720
se	0.0720	se	0.0103	se	0.0067	se	0.0043	se	0.0030
rmsea	0.0000	rmsea	0.0218	rmsea	0.0611	rmsea	0.0780	rmsea	0.0827
se	0.0000	se	0.0254	se	0.0118	se	0.0054	se	0.0035
rmr	0.0546	rmr	0.0453	rmr	0.0430	rmr	0.0418	rmr	0.0413
se	0.0082	se	0.0040	se	0.0028	se	0.0018	se	0.0012
srmr	0.1228	srmr	0.0942	srmr	0.0891	srmr	0.0864	srmr	0.0853
se	0.0160	se	0.0095	se	0.0062	se	0.0039	se	0.0027
gfi	0.9950	gfi	0.9977	gfi	0.9980	gfi	0.9981	gfi	0.9982
se	0.0011	se	0.0005	se	0.0003	se	0.0002	se	0.0001
agfi	0.9932	agfi	0.9968	agfi	0.9972	agfi	0.9974	agfi	0.9975
se	0.0015	se	0.0007	se	0.0004	se	0.0002	se	0.0002
pgfi	0.7226	pgfi	0.7245	pgfi	0.7247	pgfi	0.7248	pgfi	0.7249
se	0.0008	se	0.0003	se	0.0002	se	0.0001	se	0.0001
mfi	8.5142	mfi	0.9994	mfi	0.7073	mfi	0.5729	mfi	0.5349
se	4.4814	se	0.2138	se	0.0908	se	0.0443	se	0.0281

Table 4-12

IF-SKD Model – Fixed Zero Correlations between Factors

IF-SKD.Model.0		IF-SKD.Model.0		IF-SKD.Model.0		IF-SKD.Model.0		IF-SKD.Model.0	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
fmin	14.9844	fmin	18.3300	fmin	17.7041	fmin	17.2282	fmin	17.0724
se	4.3182	se	2.6731	se	1.7816	se	1.1198	se	0.7293
chisq	689.2813	chisq	3665.9970	chisq	7081.6334	chisq	17228.2406	chisq	34144.7293
se	198.6378	se	534.6127	se	712.6399	se	1119.8441	se	1458.5758
pvalue	0.0000	pvalue	0.0000	pvalue	0.0000	pvalue	0.0000	pvalue	0.0000
se	0.0000	se	0.0000	se	0.0000	se	0.0000	se	0.0000

IF-SKD.Model.0		IF-SKD.Model.0		IF-SKD.Model.0		IF-SKD.Model.0		IF-SKD.Model.0	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
cfi	0.2944	cfi	0.2473	cfi	0.2449	cfi	0.2431	cfi	0.2427
se	0.0336	se	0.0086	se	0.0061	se	0.0038	se	0.0025
tli	0.2160	tli	0.1636	tli	0.1610	tli	0.1590	tli	0.1585
se	0.0373	se	0.0095	se	0.0068	se	0.0042	se	0.0028
nnfi	0.2160	nnfi	0.1636	nnfi	0.1610	nnfi	0.1590	nnfi	0.1585
se	0.0373	se	0.0095	se	0.0068	se	0.0042	se	0.0028
rfi	0.1601	rfi	0.1564	rfi	0.1573	rfi	0.1576	rfi	0.1578
se	0.0246	se	0.0092	se	0.0066	se	0.0042	se	0.0028
nfi	0.2441	nfi	0.2407	nfi	0.2416	nfi	0.2418	nfi	0.2420
se	0.0221	se	0.0083	se	0.0060	se	0.0038	se	0.0025
pnfi	0.2197	pnfi	0.2166	pnfi	0.2174	pnfi	0.2176	pnfi	0.2178
se	0.0199	se	0.0074	se	0.0054	se	0.0034	se	0.0022
ifi	0.3181	ifi	0.2507	ifi	0.2466	ifi	0.2438	ifi	0.2430
se	0.0379	se	0.0085	se	0.0061	se	0.0038	se	0.0025
rni	0.2944	rni	0.2473	rni	0.2449	rni	0.2431	rni	0.2427
se	0.0336	se	0.0086	se	0.0061	se	0.0038	se	0.0025
rmsea	0.3396	rmsea	0.4298	rmsea	0.4275	rmsea	0.4248	rmsea	0.4240
se	0.0704	se	0.0331	se	0.0221	se	0.0140	se	0.0091
rmr	0.1590	rmr	0.2040	rmr	0.2023	rmr	0.2009	rmr	0.2007
se	0.0525	se	0.0272	se	0.0180	v	0.0112	se	0.0080
srmr	0.3319	srmr	0.4188	srmr	0.4183	srmr	0.4163	srmr	0.4164
se	0.0513	se	0.0260	se	0.0177	se	0.0112	se	0.0079
gfi	0.9615	gfi	0.9539	gfi	0.9553	gfi	0.9563	gfi	0.9566
se	0.0165	se	0.0088	se	0.0058	se	0.0036	se	0.0024
agfi	0.9487	agfi	0.9385	agfi	0.9404	agfi	0.9417	agfi	0.9422
se	0.0220	se	0.0117	se	0.0077	se	0.0048	se	0.0032
pgfi	0.7211	pgfi	0.7154	pgfi	0.7165	pgfi	0.7172	pgfi	0.7175
se	0.0124	se	0.0066	se	0.0044	se	0.0027	se	0.0018
mfi	0.0021	mfi	0.0000	mfi	0.0000	mfi	0.0000	mfi	0.0000
se	0.0079	se	0.0000	se	0.0000	se	0.0000	se	0.0000

Table 4-13

IF-SKD Model – Fixed Indicator (Item) Errors Fit Statistics

IF-SKD.Model.Items.Fixed		IF-SKD.Model.Items.Fixed		IF-SKD.Model.Items.Fixed		IF-SKD.Model.Items.Fixed		IF-SKD.Model.Items.Fixed	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
fmin	2.3403	fmin	1.1776	fmin	0.9971	fmin	0.9101	fmin	0.8806
se	0.6874	se	0.2432	se	0.1507	se	0.0875	se	0.0576
chisq	107.6559	chisq	235.5173	chisq	398.8501	chisq	910.0889	chisq	1761.2041
se	31.6221	se	48.6309	se	60.2710	se	87.4591	se	115.1540
pvalue	0.9887	pvalue	0.2929	pvalue	0.0000	pvalue	0.0000	pvalue	0.0000
se	0.0505	se	0.3566	se	0.0000	se	0.0000	se	0.0000
cfi	1.0000	cfi	0.9909	cfi	0.9780	cfi	0.9684	cfi	0.9652
se	0.0000	se	0.0100	se	0.0079	se	0.0051	se	0.0035
tli	1.1530	tli	0.9921	tli	0.9773	tli	0.9675	tli	0.9641
se	0.0700	se	0.0119	se	0.0081	se	0.0053	se	0.0036
nnfi	1.1530	nnfi	0.9921	nnfi	0.9773	nnfi	0.9675	nnfi	0.9641
se	0.0700	se	0.0119	se	0.0081	se	0.0053	se	0.0036
rfi	0.8631	rfi	0.9480	rfi	0.9549	rfi	0.9585	rfi	0.9597
se	0.0752	se	0.0156	se	0.0094	se	0.0057	se	0.0037
nfi	0.8670	nfi	0.9494	nfi	0.9562	nfi	0.9597	nfi	0.9608
se	0.0731	se	0.0151	se	0.0092	se	0.0055	se	0.0036
pnfi	0.8423	pnfi	0.9223	pnfi	0.9289	pnfi	0.9323	pnfi	0.9334
se	0.0710	se	0.0147	se	0.0089	se	0.0053	se	0.0035
ifi	1.1470	ifi	0.9923	ifi	0.9780	ifi	0.9684	ifi	0.9652
se	0.0666	se	0.0116	se	0.0079	se	0.0051	se	0.0035
rni	1.1486	rni	0.9923	rni	0.9780	rni	0.9684	rni	0.9652
se	0.0680	se	0.0116	se	0.0079	se	0.0051	se	0.0035
rmsea	0.0000	rmsea	0.0342	rmsea	0.0684	rmsea	0.0831	rmsea	0.0873
se	0.0000	se	0.0270	se	0.0110	se	0.0051	se	0.0032
rmr	0.0601	rmr	0.0519	rmr	0.0493	rmr	0.0478	rmr	0.0473
se	0.0080	se	0.0031	se	0.0023	se	0.0014	se	0.0009
srmr	0.1331	srmr	0.1098	srmr	0.1032	srmr	0.0995	srmr	0.0983
se	0.0181	se	0.0096	se	0.0064	se	0.0040	se	0.0027
gfi	0.9943	gfi	0.9971	gfi	0.9975	gfi	0.9977	gfi	0.9978
se	0.0011	se	0.0005	se	0.0003	se	0.0002	se	0.0001
agfi	0.9930	agfi	0.9964	agfi	0.9969	agfi	0.9972	agfi	0.9972
se	0.0014	se	0.0006	se	0.0004	se	0.0002	se	0.0002

IF-SKD.Model.Items.Fixed		IF-SKD.Model.Items.Fixed		IF-SKD.Model.Items.Fixed		IF-SKD.Model.Items.Fixed		IF-SKD.Model.Items.Fixed	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
pgfi	0.8049	pgfi	0.8072	pgfi	0.8075	pgfi	0.8077	pgfi	0.8077
se	0.0009	se	0.0004	se	0.0003	se	0.0002	se	0.0001
mfi	11.0101	mfi	0.8784	mfi	0.6200	mfi	0.4948	mfi	0.4595
se	6.1556	se	0.2079	se	0.0929	se	0.0426	se	0.0264

Information Metrics

Table 4-14

IF-SKD Model – Information Metrics

IF-SKD Model Information Metrics				
Indicator	Unstandardized Factor Loading	(Unstandardized Factor Loading)^2	Variance Estimate	Information Metric (UFL)^2/Variance Est.
X1	1	1	0.116	8.62068966
X2	1.081	1.168561	0.334	3.49868563
X3	1.006	1.012036	0.153	6.61461438
X4	1.17	1.3689	0.105	13.0371429
X5	1	1	0.143	6.99300699
X6	1.153	1.329409	0.444	2.99416441
X7	1	1	0.182	5.49450549
X8	1	1	0.129	7.75193798
X9	0.947	0.896809	0.31	2.89293226
X10	0.566	0.320356	0.154	2.08023377
X11	0.738	0.544644	0.112	4.86289286
X12	0.676	0.456976	0.139	3.28759712
X13	0.919	0.844561	0.102	8.2800098
X14	0.722	0.521284	0.175	2.97876571
X15	0.948	0.898704	0.318	2.82611321
X16	0.928	0.861184	0.302	2.85160265
X17	0.828	0.685584	0.232	2.95510345
X18	0.925	0.855625	0.166	5.15436747
X19	0.843	0.710649	0.205	3.46658049
X20	0.954	0.910116	0.188	4.84104255
X21	0.78	0.6084	0.098	6.20816327

Discussion

Examining the fit statistics across the three variations of the model demonstrates that the model that fixed the between factor loadings to zero did not effectively improve any of the overall fit statistics. In fact, it reduced the fit of most statistics evaluated. The F_{min} statistic jumped to over 17, for example. Moreover, the SRMR (0.416), RMSEA (0.4329), NFI (0.2419), IFI (0.2430) and the PNFI (0.2177) were all considerably further away from the cutoff threshold to be considered a good fit overall or in comparison to the other two models.

The fixed item error IF-SKD model and the freely estimated IF-SKD model demonstrated little overall differences. The F_{min} statistic was less significant (0.8806) compared to the freely estimated model (0.7180), while the comparative fit statistics (CFI, IFI, etc.) were all very close to one another. The PNFI was slightly better on the fixed indicator model, though this statistic penalizes a model for complexity and by fixing the item error loadings, the overall model was therefore less complex. Due to the overall F_{min} performing better for the freely estimated model, the comparison of these models was not overall substantial. However, Brown (2015) points out that, unless justified by theory, refinements to a model in the CFA process are not likely to produce a better solution and are advised against.

McCay-Peet 5 Factor CFA Analysis

Using varying sized sub-samples of the estimated population created from the imputed data set using the MASS R package, the following fit statistics were obtained for each sub-sample and evaluated using the lavaan CFA function for the McCay-Peet model.

Fit Statistics

Tables 4-15 through 4-17 show the mean fit statistics for the McCay-Peet model specified in its:

- Freely estimated form
- Fixed latent variables to have zero correlation between factors
- Fixed indicator errors (0.3) form

Table 4-15

McCay-Peet Model Fit Statistics

McCay.Peet.Model.Free		McCay.Peet.Model.Free		McCay.Peet.Model.Free		McCay.Peet.Model.Free		McCay.Peet.Model.Free	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
fmin	2.0520	fmin	0.9051	fmin	0.7787	fmin	0.7158	fmin	0.6923
se	0.6127	se	0.2092	se	0.1345	se	0.0782	se	0.0532
chisq	94.3899	chisq	181.0270	chisq	311.4823	chisq	715.8090	chisq	1384.6723
se	28.1843	se	41.8316	se	53.8049	se	78.1842	se	106.3733
pvalue	0.9856	pvalue	0.5171	pvalue	0.0024	pvalue	0.0000	pvalue	0.0000
se	0.0629	se	0.3962	se	0.0101	se	0.0000	se	0.0000
cfi	1.0000	cfi	0.9957	cfi	0.9851	cfi	0.9760	cfi	0.9730
se	0.0000	se	0.0068	se	0.0067	se	0.0043	se	0.0030
tli	1.1585	tli	0.9989	tli	0.9825	tli	0.9719	tli	0.9684
se	0.0814	se	0.0114	se	0.0079	se	0.0051	se	0.0035
nnfi	1.1585	nnfi	0.9989	nnfi	0.9825	nnfi	0.9719	nnfi	0.9684
se	0.0814	se	0.0114	se	0.0079	se	0.0051	se	0.0035
rfi	0.8610	rfi	0.9543	rfi	0.9601	rfi	0.9629	rfi	0.9638
se	0.0772	se	0.0143	se	0.0091	se	0.0054	se	0.0036
nfi	0.8815	nfi	0.9611	nfi	0.9660	nfi	0.9683	nfi	0.9692
se	0.0658	se	0.0122	se	0.0078	se	0.0046	se	0.0031
pnfi	0.7514	pnfi	0.8192	pnfi	0.8234	pnfi	0.8254	pnfi	0.8261
se	0.0561	se	0.0104	se	0.0066	se	0.0040	se	0.0026
ifi	1.1274	ifi	0.9991	ifi	0.9851	ifi	0.9760	ifi	0.9731
se	0.0621	se	0.0096	se	0.0067	se	0.0043	se	0.0030
rni	1.1351	rni	0.9991	rni	0.9851	rni	0.9760	rni	0.9730

McCay.Peet.Model.Free		McCay.Peet.Model.Free		McCay.Peet.Model.Free		McCay.Peet.Model.Free		McCay.Peet.Model.Free	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
se	0.0694	se	0.0097	se	0.0067	se	0.0043	se	0.0030
rmsea	0.0000	rmsea	0.0200	rmsea	0.0596	rmsea	0.0773	rmsea	0.0820
se	0.0000	se	0.0247	se	0.0129	se	0.0057	se	0.0036
rmr	0.0546	rmr	0.0447	rmr	0.0424	rmr	0.0411	rmr	0.0406
se	0.0081	se	0.0040	se	0.0028	se	0.0018	se	0.0013
srmr	0.1215	srmr	0.0928	srmr	0.0876	srmr	0.0850	srmr	0.0840
se	0.0163	se	0.0092	se	0.0064	se	0.0042	se	0.0028
gfi	0.9951	gfi	0.9978	gfi	0.9980	gfi	0.9982	gfi	0.9982
se	0.0011	se	0.0005	se	0.0003	se	0.0002	se	0.0001
agfi	0.9931	agfi	0.9969	agfi	0.9972	agfi	0.9974	agfi	0.9975
se	0.0015	se	0.0006	se	0.0004	se	0.0003	se	0.0002
pgfi	0.7068	pgfi	0.7087	pgfi	0.7089	pgfi	0.7090	pgfi	0.7091
se	0.0008	se	0.0003	se	0.0002	v	0.0001	se	0.0001
mfi	8.1178	mfi	1.0116	mfi	0.7235	mfi	0.5859	mfi	0.5478
se	4.1942	se	0.2046	se	0.0964	se	0.0457	se	0.0290

Table 4-16

McCay-Peet Model – Fixed Zero Correlations between Factors

McCay.Peet.Model.0		McCay.Peet.Model.0		McCay.Peet.Model.0		McCay.Peet.Model.0		McCay.Peet.Model.0	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
fmin	15.9546	fmin	19.6321	fmin	18.9733	fmin	18.4547	fmin	18.2871
se	4.6623	se	2.7321	se	1.9566	se	1.2471	se	0.8215
chisq	733.9104	chisq	3926.4120	chisq	7589.3211	chisq	18454.7120	chisq	36574.2140
se	214.4680	se	546.4154	se	782.6573	se	1247.0672	se	1643.0297
pvalue	0.0000	pvalue	0.0000	pvalue	0.0000	pvalue	0.0000	pvalue	0.0000
se	0.0000	se	0.0000	se	0.0000	se	0.0000	se	0.0000
cfi	0.2229	cfi	0.1925	cfi	0.1911	cfi	0.1900	cfi	0.1895
se	0.0274	se	0.0074	se	0.0051	se	0.0032	se	0.0022
tli	0.1366	tli	0.1027	tli	0.1012	tli	0.1000	tli	0.0994
se	0.0304	se	0.0082	se	0.0056	se	0.0036	se	0.0024
nnfi	0.1366	nnfi	0.1027	nnfi	0.1012	nnfi	0.1000	nnfi	0.0994
se	0.0304	se	0.0082	se	0.0056	se	0.0036	se	0.0024
rfi	0.1005	rfi	0.0982	rfi	0.0989	rfi	0.0991	rfi	0.0990

McCay.Peet.Model.0		McCay.Peet.Model.0		McCay.Peet.Model.0		McCay.Peet.Model.0		McCay.Peet.Model.0	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
se	0.0175	se	0.0078	se	0.0055	se	0.0035	se	0.0024
nfi	0.1905	nfi	0.1884	nfi	0.1890	nfi	0.1892	nfi	0.1891
se	0.0158	se	0.0070	se	0.0049	se	0.0032	se	0.0022
pnfi	0.1714	pnfi	0.1695	pnfi	0.1701	pnfi	0.1702	pnfi	0.1702
se	0.0142	se	0.0063	se	0.0044	se	0.0028	se	0.0020
ifi	0.2491	ifi	0.1962	ifi	0.1929	ifi	0.1908	ifi	0.1899
se	0.0337	se	0.0075	se	0.0051	se	0.0032	se	0.0022
rni	0.2229	rni	0.1925	rni	0.1911	rni	0.1900	rni	0.1895
se	0.0274	se	0.0074	se	0.0051	se	0.0032	se	0.0022
rmsea	0.3549	rmsea	0.4457	rmsea	0.4429	rmsea	0.4398	rmsea	0.4389
se	0.0716	se	0.0326	se	0.0234	se	0.0150	se	0.0099
rmr	0.1643	rmr	0.2116	rmr	0.2094	rmr	0.2084	rmr	0.2080
se	0.0529	se	0.0287	se	0.0198	se	0.0123	se	0.0085
srmr	0.3442	srmr	0.4338	srmr	0.4321	srmr	0.4310	srmr	0.4309
se	0.0529	se	0.0278	se	0.0194	se	0.0121	se	0.0084
gfi	0.9591	gfi	0.9506	gfi	0.9520	gfi	0.9532	gfi	0.9536
se	0.0176	se	0.0093	se	0.0064	se	0.0041	se	0.0027
agfi	0.9455	agfi	0.9342	agfi	0.9360	agfi	0.9376	agfi	0.9381
se	0.0235	se	0.0124	se	0.0086	se	0.0055	se	0.0036
pgfi	0.7193	pgfi	0.7130	pgfi	0.7140	pgfi	0.7149	pgfi	0.7152
se	0.0132	se	0.0070	se	0.0048	se	0.0031	se	0.0020
mfi	0.0009	mfi	0.0000	mfi	0.0000	mfi	0.0000	mfi	0.0000
se	0.0031	se	0.0000	se	0.0000	se	0.0000	se	0.0000

Table 4-17

McCay-Peet Model – Fixed Indicator (Item) Errors Fit Statistics

McCay.Peet.Model.Items.Fixed		McCay.Peet.Model.Items.Fixed		McCay.Peet.Model.Items.Fixed		McCay.Peet.Model.Items.Fixed		McCay.Peet.Model.Items.Fixed	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
fmin	2.3371	fmin	1.1400	fmin	0.9832	fmin	0.8857	fmin	0.8614
se	0.7397	se	0.2208	se	0.1455	se	0.0801	se	0.0562
chisq	107.5069	chisq	228.0045	chisq	393.2861	chisq	885.7446	chisq	1722.8798
se	34.0272	se	44.1507	se	58.1985	se	80.1063	se	112.4051
pvalue	0.9770	pvalue	0.2982	pvalue	0.0000	pvalue	0.0000	pvalue	0.0000
se	0.0956	se	0.3471	se	0.0000	se	0.0000	se	0.0000
cfi	0.9998	cfi	0.9918	cfi	0.9783	cfi	0.9693	cfi	0.9659
se	0.0013	se	0.0093	se	0.0078	se	0.0046	se	0.0033
tli	1.1542	tli	0.9928	tli	0.9772	tli	0.9678	tli	0.9642
se	0.0819	se	0.0112	se	0.0082	se	0.0048	se	0.0035
nnfi	1.1542	nnfi	0.9928	nnfi	0.9772	nnfi	0.9678	nnfi	0.9642
se	0.0819	se	0.0112	se	0.0082	se	0.0048	se	0.0035
rfi	0.8577	rfi	0.9485	rfi	0.9549	rfi	0.9588	rfi	0.9597
se	0.0816	se	0.0150	se	0.0097	se	0.0051	se	0.0036
nfi	0.8645	nfi	0.9510	nfi	0.9570	nfi	0.9607	nfi	0.9616
se	0.0777	se	0.0143	se	0.0092	se	0.0049	se	0.0034
pnfi	0.8233	pnfi	0.9057	pnfi	0.9114	pnfi	0.9150	pnfi	0.9158
se	0.0740	se	0.0136	se	0.0088	se	0.0047	se	0.0033
ifi	1.1439	ifi	0.9932	ifi	0.9783	ifi	0.9693	ifi	0.9659
se	0.0749	se	0.0107	se	0.0078	se	0.0046	se	0.0033
rni	1.1469	rni	0.9931	rni	0.9783	rni	0.9693	rni	0.9659
se	0.0780	se	0.0107	se	0.0078	se	0.0046	se	0.0033

McCay.Peet.Model.Items.Fixed		McCay.Peet.Model.Items.Fixed		McCay.Peet.Model.Items.Fixed		McCay.Peet.Model.Items.Fixed		McCay.Peet.Model.Items.Fixed	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
rmsea	0.0006	rmsea	0.0326	rmsea	0.0689	rmsea	0.0827	rmsea	0.0872
se	0.0041	v	0.0258	se	0.0106	se	0.0048	se	0.0032
rmr	0.0592	rmr	0.0512	rmr	0.0489	rmr	0.0473	rmr	0.0468
se	0.0078	se	0.0031	se	0.0022	se	0.0013	se	0.0009
srmr	0.1329	srmr	0.1084	srmr	0.1024	srmr	0.0986	srmr	0.0975
se	0.0198	se	0.0091	se	0.0064	se	0.0037	se	0.0026
gfi	0.9945	gfi	0.9972	gfi	0.9975	gfi	0.9978	gfi	0.9978
se	0.0011	se	0.0004	se	0.0003	se	0.0002	se	0.0001
agfi	0.9930	agfi	0.9965	agfi	0.9969	agfi	0.9972	agfi	0.9972
se	0.0014	se	0.0006	se	0.0004	se	0.0002	se	0.0002
pgfi	0.7893	pgfi	0.7914	pgfi	0.7917	pgfi	0.7919	pgfi	0.7919
se	0.0009	se	0.0003	se	0.0002	se	0.0001	se	0.0001
mfi	10.3377	mfi	0.8895	mfi	0.6219	mfi	0.5047	mfi	0.4675
se	5.9594	se	0.1908	se	0.0889	se	0.0403	se	0.0262

Information Metrics

Table 4-18

McCay-Peet Model – Information Metrics

Indicator	Unstandardized Factor Loading	(Unstandardized Factor Loading)^2	Variance Estimate	Information Metric (UFL)^2/Variance Est.
X1	1	1	0.111	9.00900901
X2	1.082	1.170724	0.328	3.56928049
X3	1.008	1.016064	0.148	6.8652973
X4	1.168	1.364224	0.099	13.7800404
X5	1	1	0.18	5.55555556
X6	1.144	1.308736	0.5	2.617472
X7	1	1	0.184	5.43478261
X8	1.132	1.281424	0.113	11.3400354
X9	1.092	1.192464	0.281	4.24364413
X10	0.599	0.358801	0.15	2.39200667
X11	1	1	0.106	9.43396226
X12	0.912	0.831744	0.136	6.11576471
X13	1.239	1.535121	0.098	15.6645
X14	0.976	0.952576	0.171	5.57061988
X15	1.281	1.640961	0.311	5.27640193
X16	1.257	1.580049	0.294	5.37431633
X17	1	1	0.239	4.18410042
X18	0.973	0.946729	0.174	5.44097126
X19	1.078	1.162084	0.2	5.81042
X20	1.153	1.329409	0.196	6.78269898
X21	0.95	0.9025	0.1	9.025

Discussion

Similar to the IF-SKD model, the McCay-Peet model exhibited differences between the model with between factor correlations set to zero, such as an excessively high F_{min}, in comparison to the freely estimated and fixed indicator model. Again, the PNFI and penalizing fit indices performed better on the fixed item error model, however, the absolute fit statistics on the McCay-Peet models were better on the freely estimate model.

Single Factor CFA Analysis

Using varying sized sub-samples of the estimated population created from the imputed data set using the MASS R package, the following fit statistics were obtained for each sub-sample and evaluated using the lavaan CFA function for the Single model.

Fit Statistics

Tables 4-19 and 4-20 show the mean fit statistics for the Single Factor model specified in its:

- Freely estimated form
- Fixed indicator errors (0.3) form

Table 4-19

Single Factor Model Fit Statistics

Single.Factor.Model		Single.Factor.Model		Single.Factor.Model		Single.Factor.Model		Single.Factor.Model	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
fmin	2.1504	fmin	2.0190	fmin	1.4585	fmin	1.1312	fmin	0.9279
se	0.6227	se	0.3946	se	0.2348	se	0.1159	se	0.0665
chisq	98.9187	chisq	403.8090	chisq	583.3891	chisq	1131.2462	chisq	1855.7074
se	28.6465	se	78.9170	se	93.9175	se	115.8649	se	132.9285
pvalue	0.9865	pvalue	0.0000	pvalue	0.0000	pvalue	0.0000	pvalue	0.0000
se	0.0647	se	0.0000	se	0.0000	se	0.0000	se	0.0000
cfi	1.0000	cfi	0.9389	cfi	0.9463	cfi	0.9554	cfi	0.9602
se	0.0000	se	0.0271	se	0.0158	se	0.0073	se	0.0042
tli	1.1560	tli	0.9321	tli	0.9403	tli	0.9504	tli	0.9558
se	0.0746	v	0.0302	se	0.0176	se	0.0082	se	0.0047

Single.Factor.Model		Single.Factor.Model		Single.Factor.Model		Single.Factor.Model		Single.Factor.Model	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
nnfi	1.1560	nnfi	0.9321	nnfi	0.9403	nnfi	0.9504	nnfi	0.9558
se	0.0746	se	0.0302	se	0.0176	se	0.0082	se	0.0047
rfi	0.8633	rfi	0.8805	rfi	0.9145	rfi	0.9411	rfi	0.9510
se	0.0740	se	0.0338	se	0.0187	se	0.0085	se	0.0048
nfi	0.8770	nfi	0.8925	nfi	0.9231	nfi	0.9470	nfi	0.9559
se	0.0666	se	0.0304	se	0.0168	se	0.0077	se	0.0043
pnfi	0.7893	pnfi	0.8032	pnfi	0.8308	pnfi	0.8523	pnfi	0.8603
se	0.0599	se	0.0274	se	0.0151	se	0.0069	se	0.0039
ifi	1.1351	ifi	0.9392	ifi	0.9465	ifi	0.9554	ifi	0.9602
se	0.0623	se	0.0269	se	0.0157	se	0.0073	se	0.0042
rni	1.1404	rni	0.9389	rni	0.9463	rni	0.9554	rni	0.9602
se	0.0672	se	0.0271	se	0.0158	se	0.0073	se	0.0042
rmsea	0.0000	rmsea	0.1053	rmsea	0.1017	rmsea	0.0998	rmsea	0.0939
se	0.0000	se	0.0197	se	0.0121	se	0.0061	se	0.0038
rmr	0.0559	rmr	0.0591	rmr	0.0520	rmr	0.0493	rmr	0.0457
se	0.0085	se	0.0061	se	0.0036	se	0.0024	se	0.0015
srmr	0.1239	srmr	0.1348	srmr	0.1177	srmr	0.1043	srmr	0.0961
se	0.0162	se	0.0124	se	0.0085	se	0.0050	se	0.0031
gfi	0.9948	gfi	0.9955	gfi	0.9966	gfi	0.9972	gfi	0.9977
se	0.0011	se	0.0008	se	0.0005	se	0.0003	se	0.0002
agfi	0.9931	agfi	0.9941	agfi	0.9955	agfi	0.9963	agfi	0.9969
se	0.0015	se	0.0011	se	0.0006	se	0.0004	se	0.0002
pgfi	0.7461	pgfi	0.7467	pgfi	0.7475	pgfi	0.7479	pgfi	0.7483
se	0.0008	se	0.0006	se	0.0004	se	0.0002	se	0.0001
mfi	9.2100	mfi	0.3640	mfi	0.3815	mfi	0.3917	mfi	0.4353
se	4.7285	se	0.1330	se	0.0862	se	0.0447	se	0.0289

Table 4-20

Single Factor Model – Fixed Indicator (Item) Error Fit Statistics

Single.Factor.Model.Items.Fixed		Single.Factor.Model.Items.Fixed		Single.Factor.Model.Items.Fixed		Single.Factor.Model.Items.Fixed		Single.Factor.Model.Items.Fixed	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
fmin	2.4523	fmin	1.2247	fmin	1.0531	fmin	0.9545	fmin	0.9255
se	0.7782	se	0.2459	se	0.1536	se	0.0887	se	0.0606
chisq	112.8038	chisq	244.9411	chisq	421.2346	chisq	954.4683	chisq	1850.9550
se	35.7966	se	49.1815	se	61.4354	se	88.7462	se	121.1322
pvalue	0.9785	pvalue	0.2751	pvalue	0.0000	pvalue	0.0000	pvalue	0.0000
se	0.0879	se	0.3501	se	0.0000	se	0.0000	se	0.0000
cfi	1.0000	cfi	0.9903	cfi	0.9763	cfi	0.9668	cfi	0.9633
se	0.0000	se	0.0103	se	0.0081	se	0.0051	se	0.0036
tli	1.1539	tli	0.9916	tli	0.9763	tli	0.9668	tli	0.9633
se	0.0767	se	0.0118	se	0.0081	se	0.0051	se	0.0036
nnfi	1.1539	nnfi	0.9916	nnfi	0.9763	nnfi	0.9668	nnfi	0.9633
se	0.0767	se	0.0118	se	0.0081	se	0.0051	se	0.0036
rfi	0.8568	rfi	0.9478	rfi	0.9541	rfi	0.9579	rfi	0.9588
se	0.0805	se	0.0153	se	0.0094	se	0.0055	se	0.0038
nfi	0.8568	nfi	0.9478	nfi	0.9541	nfi	0.9579	nfi	0.9588
se	0.0805	se	0.0153	se	0.0094	se	0.0055	se	0.0038
pnfi	0.8568	pnfi	0.9478	pnfi	0.9541	pnfi	0.9579	pnfi	0.9588
se	0.0805	se	0.0153	se	0.0094	se	0.0055	se	0.0038
ifi	1.1539	ifi	0.9916	ifi	0.9763	ifi	0.9668	ifi	0.9633
se	0.0767	se	0.0118	se	0.0081	se	0.0051	se	0.0036
rni	1.1539	rni	0.9916	rni	0.9763	rni	0.9668	rni	0.9633

Single.Factor.Model.Items.Fixed		Single.Factor.Model.Items.Fixed		Single.Factor.Model.Items.Fixed		Single.Factor.Model.Items.Fixed		Single.Factor.Model.Items.Fixed	
mean.fit (N=23)		mean.fit (N=100)		mean.fit (N=200)		mean.fit (N=500)		mean.fit (N=1000)	
se	0.0767	se	0.0118	se	0.0081	se	0.0051	se	0.0036
rmsea	0.0000	rmsea	0.0357	rmsea	0.0703	rmsea	0.0841	rmsea	0.0884
se	0.0000	se	0.0267	se	0.0103	se	0.0050	se	0.0033
rmr	0.0605	rmr	0.0532	rmr	0.0510	rmr	0.0493	rmr	0.0488
se	0.0076	se	0.0028	se	0.0020	se	0.0013	se	0.0009
srmr	0.1364	srmr	0.1128	srmr	0.1067	srmr	0.1028	srmr	0.1017
se	0.0203	se	0.0097	se	0.0063	se	0.0039	se	0.0027
gfi	0.9942	gfi	0.9970	gfi	0.9974	gfi	0.9976	gfi	0.9977
se	0.0012	se	0.0005	se	0.0003	se	0.0002	se	0.0001
agfi	0.9930	agfi	0.9964	agfi	0.9968	agfi	0.9971	agfi	0.9972
se	0.0014	se	0.0006	se	0.0004	se	0.0002	se	0.0002
pgfi	0.8285	pgfi	0.8308	pgfi	0.8311	pgfi	0.8313	pgfi	0.8314
se	0.0010	se	0.0004	se	0.0003	se	0.0002	se	0.0001
mfi	11.8045	mfi	0.8641	mfi	0.5952	mfi	0.4762	mfi	0.4408
se	7.2555	se	0.2085	se	0.0892	se	0.0420	se	0.0265

Information Metrics

Table 4-21

Single Factor Model – Information Metrics

Indicator	Unstandardized Factor Loading	(Unstandardized Factor Loading)^2	Variance Estimate	Information Metric (UFL)^2/Variance Est.
X1	1	1	0.192	5.20833333
X2	0.862	0.743044	0.302	2.4604106
X3	0.966	0.933156	0.216	4.32016667
X4	1.069	1.142761	0.171	6.68281287
X5	1.05	1.1025	0.189	5.83333333
X6	0.964	0.929296	0.305	3.04687213
X7	0.979	0.958441	0.228	4.2036886
X8	1.099	1.207801	0.17	7.10471176
X9	1.001	1.002001	0.245	4.0898
X10	0.76	0.5776	0.28	2.06285714
X11	1.004	1.008016	0.184	5.47834783
X12	0.941	0.885481	0.213	4.15718779
X13	1.108	1.227664	0.149	8.2393557
X14	0.926	0.857476	0.231	3.71201732
X15	0.958	0.917764	0.268	3.42449254
X16	0.963	0.927369	0.26	3.56680385
X17	0.886	0.784996	0.262	2.99616794
X18	0.904	0.817216	0.24	3.40506667
X19	0.925	0.855625	0.243	3.52109053
X20	0.967	0.935089	0.229	4.08335808
X21	0.978	0.956484	0.191	5.00776963

Discussion

Only two single factor models required testing. Since there was only a single proposed latent variable, it was not necessary to have a model that fixed the correlation between factors to zero. Both models had very similar results across all fit indices. For example, the F_{min}

between the freely estimated model and the model with item errors fixed was 0.9278537 and 0.9254775 respectively. Standard errors were also very close between the models at N=1000.

Limitations

In many respects, while the case is made that the IF-SKD model is applicable in representing the serendipitous knowledge discovery information behavior of clinicians in a clinical setting, the across model comparisons show that there are confounding aspects within the make-up of these models and the questions that require further, more in-depth review. These are be discussed in the next chapter; however, it is worth noting that there are likely more aspects of these models that could be evaluated using different methods, refinements to the current method, or even possibly qualitative follow-up in order to better understand what is required to better distinguish between the models.

Generalizability and Reliability

Before presenting a summary of the findings for the CFA models evaluated as part of this study, it is important to discuss, especially given the methods employed as part of this study, how the results can be viewed from a generalizability and reliability standpoint. The Omega coefficient, a multi-dimensional measure of both generalizability and reliability, is one way to demonstrate this. It shows factor saturation, which means the results would generalize even if theoretically similar items were used. Trizano-Hermosilla and Alvarado (2016) have pointed out, using Monte Carlo analysis comparisons between the traditional Cronbach Alpha and Omega Coefficient, that the Omega Coefficient is better at demonstrated reliability and generalizability, even in unidimensional factor situations.

The Omega Coefficients for all three factor size models can be found in Table 4-22.

Table 4-22

Omega Coefficients for Models

Model	Omega Coefficient Alpha	Omega H Asymptotic
4-Factor	0.97	0.78
5-Factor	0.97	0.79
1-Factor	0.97	N/A with 1 Factor

The values from the Omega Coefficient highlight that the overall relationship between indicators and factors is good, or well saturated. This demonstrates that even with a small sample size, the overall relationship between indicators posed on the questionnaire and a set of given factors is both generalizable and reliable.

Summary of All Model CFA Findings

Brown (2015) discusses a three-pronged approach to evaluate a model in confirmatory factory analysis. First, ascertain whether a combination of different fit type indices (absolute, comparative, parsimonious) yield a good overall fit. Second, if the fit is found not to be good, then it is important to look for "localized areas of strain", or points in the indicators or latent variables where the model is not fitting well (Brown, 2015 p.96). Finally, looking at the statistical significance of the models' parameter estimates can be valuable. Brown (2015) notes that there are many circumstances where a model may appear to fit well in an overall sense, however at a specific indicator or variable level, there may be issues.

The summary of absolute and comparative fit statistics provides a set of evidence that allows for a few broad findings. First, they also performed well independently on some of the absolute fit statistics. For example, both the freely estimated McCay-Peet and IF-SKD models have the lowest F_{min} statistics, which is one of the absolute fit statistics most indicative of the sample closely representing the population, or more attenuated to minimizing the Kullback-Liebler divergence (KLD). However, while all three models met or exceeded some of the cut-off criteria on the comparative fit indices, there were statistically significant differences between them, especially between the McCay-Peet and IF-SKD models compared to the single factor model.

Maydeu-Olivares (2017) pointed out that "the size of the misfit in a covariance structure model cannot be captured by a single effect size parameter because of the multivariate nature of the data" (p. 540). This is consistent with some of the fit statistics between the models capturing fit better in an absolute and comparative sense, even if not in a statistically significant way.

The process for generating the data and the use of the Monte Carlo sampling technique provided for a set of statistics and standard errors that allowed for t-tests to be performed to further evaluate these models statistically from one another. Table 4-23 provides a summary of the t-test p-value significance findings between the IF-SKD and McCay-Peet models on several fit statistics.

Table 4-23

t-*Test Statistics' Summary: McCay-Peet and Single Factor Model*

Model	Fit Statistic	T.Obs	Significance
IF-SKD – McCay-Peet	F_{min}	0.3436296	0.3655805
IF-SKD – McCay-Peet	RMSEA	0.1371495	0.4454632
IF-SKD – McCay-Peet	SRMR	0.3207723	0.3742083
IF-SKD – McCay-Peet	RMR	0.3953301	0.3463207
IF-SKD – McCay-Peet	PNFI	4.650645	**0.000001762796**
IF-SKD – McCay-Peet	PGFI	127.0802	**0**
IF-SKD – McCay-Peet	IFI	-0.2512246	0.5991669
IF-SKD – McCay-Peet	CFI	-0.2505904	0.5989217
IF-SKD – McCay-Peet	TLI	-0.1063003	0.5423226
IF-SKD – McCay-Peet	Chi-Sq	0.3436294	0.3655806

Table 4-24

t-*Test Statistics' Summary: IF-SKD and McCay-Peet*

Model	Fit Statistic	T.Obs	Significance
McCay-Peet – Single Factor	F_{min}	-2.767261	**0.9971475**
McCay-Peet – Single Factor	RMSEA	-0.1936043	0.5767473
McCay-Peet – Single Factor	SRMR	-2.879442	**0.9979869**
McCay-Peet – Single Factor	RMR	-2.585583	**0.9951042**
McCay-Peet – Single Factor	PNFI	-7.305453	**1**
McCay-Peet – Single Factor	PGFI	-269.3283	**1**
McCay-Peet – Single Factor	IFI	2.501573	**0.006221862**
McCay-Peet – Single Factor	CFI	2.496135	**0.006317764**
McCay-Peet – Single Factor	TLI	2.158462	**0.01550533**
McCay-Peet – Single Factor	Chi-Sq	-2.766716	**0.9971427**

Table 4-25

t-*Test Statistics' Summary: IF-SKD and Single Factor Model*

Model	Fit Statistic	T.Obs	Significance
IF-SKD– Single Factor	F_{min}	-2.473646	**0.9932716**
IF-SKD– Single Factor	RMSEA	-2.188251	**0.9856166**
IF-SKD– Single Factor	SRMR	-2.614356	**0.9954966**
IF-SKD– Single Factor	RMR	-2.269956	**0.9883418**
IF-SKD– Single Factor	PNFI	-3.53999	**0.9997954**
IF-SKD– Single Factor	PGFI	-160.6095	**1**
IF-SKD– Single Factor	IFI	2.291769	**0.01101128**
IF-SKD– Single Factor	CFI	2.286917	**0.0111524**
IF-SKD– Single Factor	TLI	2.081595	**0.01875325**
IF-SKD– Single Factor	Chi-Sq	-2.473099	**0.9932613**

The findings indicate, that in all but the case of the PNFI and PGFI fit statistics, the IF-SKD model, while representing better fit statistics overall, is not able to be determined as statistically significantly different than the McCay-Peet model. However, both the IF-SKD model and the McCay-Peet model were statistically significantly better than the single factor on all fit indices examined, with the exception of the RMSEA between the McCay-Peet and single factor model, and the PNFI and PGFI, which favored the single factor model compared to both the IF-SKD and McCay-Peet models.

A last notable mention of the analysis is surrounding the information metric tables for each of the models. McDonald (1999) pointed out that the unstandardized squared factor loadings to the "unique variance of the items" is a "measure of the amount of information about the attribute given by each item. The larger this information measure, the greater is the extent to which the item reduces the error of measurement of the attribute" (p. 22). Evaluating these information metrics provides an initial understanding into the aspects of certain questions, which may or may not be hindering a better interpretation of the model. It may point to questions that need refinement, or it may highlight questions that are no longer needed in the questionnaire. For instance, an initial examination of these metrics, when viewed across all models, shows that item X6 is small across all the models and could be a potential candidate for removal.

Finally, an evaluation of the standardized path diagrams reinforces item X6 as a candidate for removal given its high item error. Item X19 is also an additional candidate to consider for removal, as it has one of the highest item error values consistently across the models.

Summary

The evaluation of the IF-SKD model in reflecting physicians' serendipitous knowledge discovery information behaviour in a clinical setting was analyzed using confirmatory factor analysis. The sample size utilized for this study (N=23) required a series of steps be performed and validated before use, to ensure that the estimated population, and subsequent summary statistics, were accurate and meaningful.

Initial findings suggest that the IF-SKD model should not be ruled out as a model reflective of SDK information behaviour. However, comparative fit statistics, in particular, those that penalize a model for complexity, do not meet the minimum threshold to be considered acceptable, despite being statistically significantly different in comparison between the models.

A t-test evaluation of IF-SKD model to McCay-Peet model showed that on select fit statistics, the IF-SKD models performs statistically significantly better than the McCay-Peet model, yet for most absolute fit statistics, the McCay-Peet model performed slightly better than the IFKSD or single factor model, even though those differences were not statistically significant. Both models were statistically significantly better than the single factor model.

CHAPTER 5

DISCUSSIONS AND CONCLUSIONS

Introduction

This research has reviewed and evaluated the concept of serendipity, and how it relates to the field of information science. Over the past decade, the topic of serendipity has been increasingly studied (Erdelez et al. 2014). As discussed in Chapter 2, the research into serendipity has long roots in information science. Most recently, work by Workman, Fiszman, Rindflesch and Nahl (2014) and McCay-Peet (2013, 2016) have begun efforts to operationalize the concept of serendipity and, specifically, its application to the study of online information systems using quantitative research instruments.

Furthermore, this research has extended the current knowledge on the use of research instruments to quantitatively capture and evaluate serendipity. Existing work by McCay-Peet (2013, 2016) laid groundwork on the application of a quantitative tool (SDE Questionnaire) to capture the essence of serendipity in the study of online systems and the factors, or models, that underlie these items. Through exploratory factor analysis (EFA), McCay-Peet (2013) was able to capture the potential of a four-factor model as capturing the attributes of a serendipitous digital environment. Because exploratory factor methods do not specifically indicate item-to-factor relationships, the original posited five-factor model proposed by McCay-Peet (2013) was used for comparison with the IF-SKD model.

In this study, the application of the SDE questionnaire and the evaluation of underlying models were pursued using confirmatory factor analysis (CFA). Two current models were explored, the McCay-Peet (2013) five-factor model as well as Workman, Fiszman, Rindflesch and Nahl (2014) IF-SKD model, along with a single model, to determine which model was a better fit and how the models compared to one another. The results from this analysis helped shape and extend the current knowledge on the topic and provide opportunities for further research.

Additionally, this study was able to extend an argument for an effective way to analyze a small sample size in the evaluation of a confirmatory factor model through the use of data imputation, covariance shrinkage, estimated population generation and Monte Carlo analysis to evaluate fit statistics for all of the models.

This study was able to demonstrate that the models analyzed were somewhat adequate at reflecting the underlying distribution reflective of a serendipitous digital environment. In some cases, such as with the IF-SKD model and the McCay-Peet model, these differences met or exceeded the threshold required to be considered a good fit. This was not, however, the case for all fit indices, in particular the absolute fit statistics such as the F_{min}, RMSEA, and SRMR. Without a range of fit indices expressing good fit, it is not advisable to assume good fit because some fit indices do meet or exceed a recommended threshold (Brown, 2015).

Overall, the findings showed no model to be, independently, statistically significant in accounting for the sample population. However, in comparison to the other models, the IF-SKD model proved statistically significantly over the McCay-Peet model on select fit statistics, such as the PNFI ($p = 0.000001762796$) and PGFI ($p = 0.0$) and on all fit statistics compared to the single factor model.

Box and Draper (1987) famously noted that “all models are wrong, but some of them are useful” (p. 424). This is an important consideration in this type of study. While the current research provides strong evidence for the underlying concepts explored in the models, it is important to note that in certain circumstances, it is reasonable to reject the null hypothesis if the "null hypothesis is a substantively uninteresting hypothesis, since by definition all models (that is, approximations) are wrong" (Maydeu-Olivares, 2017 p.527). That is not to say that one model stands out as an adequate replacement to the null hypothesis, but rather, that the way in which models are reviewed and evaluated is a process.

The analysis conducted here is able to help shape an improved understanding, not of how the models do not fit, but rather in what ways they do. This explanation by comparison allows for better understanding in how to operationalize the topic of serendipity, how to begin to better understand how we measure it through research instruments, and finally how to begin to improve and hone the tools needed to further explore and refine the models and additional aspects that affect this type of information behavior.

Research Questions' Summary

R1: Does Spark successfully contribute to physicians' serendipitous knowledge discovery?

It was determined through frequency analysis that Spark does successfully contribute to physicians' serendipitous knowledge discovery.

R2: Does the IF-SKD model reflect physician serendipitous knowledge discovery information behavior in the clinical setting?

Using confirmatory factor analysis, it was demonstrated that the IF-SKD model was able to reflect physicians' serendipitous knowledge discovery on several fit statistics; however, not on all. Further research is warranted to better understand the relationship between this model and this type of information behaviour.

Significance of the Study

This study was significant in two principle ways. First, this research presented the second application of a new research instrument, the SDE questionnaire. Additionally, this was the first time that this instrument had been assessed using confirmatory factor analysis. In addition, the study of multiple models has helped provide broader context regarding the application of the indicators to the proposed models' structures.

The second major contribution of this study is the focus on utilizing small sample size to conduct confirmatory factor analysis. The methods and analysis undertaken to successfully analyze these data presented meaningful statistical metrics to compare one model to another, which offers insight into how future analysis can be conducted when small samples are encountered. This is especially useful in the study of serendipity and the application of a research instrument such as the SDE questionnaire which is relatively lengthy, depending on the audience to which it is posed. Moreover, the use of the Omega Coefficient to demonstrate generalizability and reliability of the SDE questionnaire on a four and five factor model is an important finding in this study, particularly given the small sample size. This shows the significance of the research instruments in being able to measure what they intended, namely a multi-dimensional factor model.

Spark is a new way to engage users and promote serendipitous knowledge discovery. Through a frequency analysis of Likert questions designed to measure the direct Perception of serendipity with a specific system, this study showed that Spark did contribute to participants' SKD, or perception of serendipity. Additionally, Spark, being designed based on the IF-SKD model is significant. The relationship between the model and the system is critical to address. As a reflective model of serendipitous knowledge discovery, the IF-SKD model can serve as a springboard for the future development of systems in other domains beyond the medical field, and future studies can further test the reliability of the model in being able to support serendipity. The application of the model to system design is significant and an area that warrants future research and discussion.

As Agarwal (2015) noted, there are competing facets of information seeking behavior that influence serendipity. The social aspects discussed by Cunha (2005) as well as user specific considerations, such as a user's predisposition, have been touched on as well (Erdelez, 1997). Others, such as Nahl (2014), spoke of the motivation, or enthusiasm of the user. This research has helped further delve into the study of serendipity and its existing literature and helps drive future research in consideration of these points.

It is clear from the analysis that two reasonably sound, theoretically driven models, do not fully represent the underlying data from the SDE questionnaire. Refinements to the existing models may be required, or additional models proposed. Moreover, there may be unique aspects of serendipitous knowledge discovery, not studied here, that may be relevant, such as those mentioned in Chapter 2.

Recommendations for Future Research

A significant area for future research would be the application of the SDE questionnaire to another unique population, which would allow for the current models to be examined further. Additionally, while the primary goal of this research was to evaluate the SDE questionnaire and the models posed, further research, especially with a new unique population, could help refine the current knowledge regarding the dimensions that underlie or contribute to the perception of serendipity. For instance, the following hypothetical studies could be beneficial in extending the use of this research instrument:

- Using confirmatory factor analysis, study nurses (RNs), advanced practice nurses (APRNs), and physician's assistants (PAs) to evaluate if the models proposed in this study produce similar results.

- An additional, multi-site study focused on physicians in different clinical settings, such as immediate patient care, a family medicine setting situation and specialty care to evaluate if any differences in task, could account for differences in perception of serendipity.

- A treatment/control study evaluating physicians using a similar fixed patient scenario, in which one group represents the specialty for which the scenario is based and a control group which is not made up of that specialty. Differences in perception of serendipity could be evaluated to determine whether prior knowledge has a substantive impact.

- Another scenario might examine other domains of knowledge to assess differences in serendipity to determine if more specialized fields, such as science, technology, engineering and math (STEM) differ in perception of serendipity compared to the humanities, social sciences, arts, etc. In such a scenario, different systems would have to be utilized.

- Focus groups, or small sample qualitative research studying physicians as a compliment to the study conducted in this research could help elucidate aspects of Spark and the questionnaires that posed challenges to respondents. It could also offer additional insight into areas not well accounted for in the study design. Since this research demonstrated that no specific model, by itself, would be considered an overall good fit, there might be further information that could be found using qualitative or mixed methods research in the future.

- Further examination of the current SDE questionnaire along with a focus on individual user, or organizational characteristics, that may play a role in an individuals' perception of serendipity, and therefore possibly their inclination to find attributes of an online environment as contributing to serendipity, would be worth exploring. One of the models examined as part of this study may remain a good representation of serendipity and its factors in online systems, but may require the addition of a larger shared factor accounting for user specific, social, or work specific aspects that would improve the interpretation and application of the currently proposed models in the literature.

- A review of information metrics in future studies, as demonstrated in this research, would also provide for further interesting findings related to the use of the SDE questionnaire and could pave the way for improvements to its overall construction.

- Also important to future research will be determining improved ways to engage participants, especially physicians. While this study made many efforts to reach out, such as through targeted emails and word of mouth, future studies might utilize additional methods to improve participation, such as snowball sampling.

These types of studies could help with the ongoing refinement of the questionnaire in ways that would allow it to be more effective in future confirmatory factor studies and in its practical use in multiple populations. Additionally, future research will contribute to a better understanding of serendipitous knowledge discovery. Within the medical field in particular, this will allow for the continued development of Spark and other potential solutions. This is important in meeting the challenges presented in Chapter 1 that highlight the gaps in accessing meaningful information in an ever-expanding biomedical information space. Future findings could help shape the way information is presented to medical professionals and their use of this information could help improve the acquisition of new knowledge that could translate to improved patient outcomes.

Summary

The study of serendipity in information science, while not new, is budding in its formal development and application. The work to date has helped set the stage for the wide-scale application of quantitative research instruments to study serendipity in online environments. Current models are rich in their theoretical origins and the application of these models, along with the continued refinement of research instruments, will lead to increasingly effective and efficient ways to evaluate how well an online system is able to contribute to serendipitous knowledge discovery.

APPENDIX A

RESEARCH INSTRUMENT

UNT UNIVERSITY OF NORTH TEXAS A green light to greatne

Informed Consent Notice

University of North Texas Institutional Review Board
Informed Consent Notice

Before agreeing to participate in this research study, it is important that you read and understand the following explanation of the purpose, benefits and risks of the study and how it will be conducted.

Title of Study: A Study of Physicians' Serendipitous Knowledge Discovery in a Clinical Care Setting.

Student Investigator: Mark Hopkins, University of North Texas (UNT) Department of Information Science. **Supervising Investigator**: Dr. Oksana Zavalina.

Purpose of the Study: You are being asked to participate in a research study which will focus on the use of a system called Spark, which is designed to assist physicians with information discovery of the biomedical literature. The goal of this research is to improve understanding of the factors contributing to physicians' serendipitous knowledge discovery in online systems.

Study Procedures: You will be asked to watch a short video of an online information resource called Spark then respond to a series of questions asking about your perception of Spark, in addition to completing a few demographic questions that will take approximately 10-15 minutes of your time.

Foreseeable Risks: No foreseeable risks are involved in this study.

Benefits to the Subjects or Others: This research may not provide an immediate benefit to individual participants, but will significantly aid in the broader understanding of this important type of information behaviour.

Compensation for Participants: None. However, for each completed survey, a donation will be made to the INTEGRIS research passion.

Procedures for Maintaining Confidentiality of Research Records: Data records and informed consent forms, upon collection will be provided to the supervising investigators office for storage on the UNT campus. The confidentiality of your individual information will be maintained in any publications or presentations regarding this study. Confidentiality will be maintained to the degree possible given the technology and practices used by the online survey company. Your participation in this online survey involves risks to confidentiality similar to a person's everyday use of the internet.

Questions about the Study: If you have any questions about the study, you may contact Mark Hopkins at MarkHopkins@my.unt.edu or Oksana Zavalina at Oksana.Zavalina@unt.edu.

Review for the Protection of Participants: This research study has been reviewed and approved by the UNT Institutional Review Board (IRB). The UNT IRB can be contacted at (940) 565-4643 with any questions regarding the rights of research subjects.

This research study has been reviewed and approved by the INTEGRIS Health, Inc. Institutional Review Board (IH IRB). For questions about your rights as a research participant, please contact the IH IRB at 405.949.4184 or irb@integrisok.com

Research Participant's Rights: Your participation in the survey confirms that you have read all of the above and that you agree to all of the following:

- Mark Hopkins has explained the study to you and you have had an opportunity to contact him with any questions about the study. You have been informed of the possible benefits and the potential risks of the study.
- You understand that you do not have to take part in this study, and your refusal to participate or your decision to withdraw will involve no penalty or loss of rights or benefits. The study personnel may choose to stop your participation at any time.
- You understand why the study is being conducted and how it will be performed.
- You understand your rights as a research participant and you voluntarily consent to participate in this study.
- You understand you may print a copy of this form for your records.

Spark Video Overview

Demographic Questions

What is your gender?

Male
Female
Prefer not to say

What is your age group?

- 18-20
- 21-25
- 26-30
- 31-35
- 36-40
- 41-45
- 46-50
- 51-55
- 56-60
- 61-65

Specialty

- Allergy & Immunology
- Anesthesiology
- Cardiology
- Dermatology
- Emergency Medicine
- Family Medicine
- Gastroenterology
- Hematology
- Internal Medicine
- Internal Medicine-Pediatrics

Video Overview Question

Did you watch the overview video?

- ○ Yes
- ○ No

Direct Measure of Serendipity Question Set

Thinking of your perception of Spark from watching the video, please indicate your level of agreement with each of the following statements. If there is a statement that is unclear to you, please select "I Don't Know"

NOTE: You may find these statements repetitive. This is intentional and will help develop a better set of questions.

Survey p. **1 of 5**

	Never	Rarely	Sometimes	Frequently	Very frequently	I Don't Know
In Spark, I experience mixes of unexpectedness and insight that lead to valuable, unanticipated outcomes	○	○	○	○	○	○
In Spark, I experience serendipity that has an impact on my everyday life	○	○	○	○	○	○
In Spark, I experience serendipity that has an impact on my work	○	○	○	○	○	○
I encounter useful information, ideas, or resources that I am not looking for when I use Spark	○	○	○	○	○	○

SDE Scale Question Set 1

Thinking of your perception of Spark from watching the video, please indicate your level of agreement

with each of the following statements. If there is a statement that is unclear to you, please select "I Don't Know"

NOTE: You may find these statements repetitive. This is intentional and will help develop a better set of questions.

Survey p. **2 of 5**

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	I Don't Know
I am surprised by what I find in Spark	O	O	O	O	O	O
My interactions in Spark are unexpectedly valuable	O	O	O	O	O	O
Spark is full of information useful to me	O	O	O	O	O	O
I encounter the unexpected in Spark	O	O	O	O	O	O
I notice content I wouldn't normally pay attention to in Spark	O	O	O	O	O	O

SDE Scale Question Set 2

Thinking of your perception of Spark from watching the video, please indicate your level of agreement

with each of the following statements. If there is a statement that is unclear to you, please select "I Don't Know"

NOTE: You may find these statements repetitive. This is intentional and will help develop a better set of questions.

Survey p. 3 **of 5**

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree	I Don't Know
The way that Spark presents content captures my attention.	O	O	O	O	O	O
Spark is a tool for exploration	O	O	O	O	O	O
I am pointed toward content in Spark	O	O	O	O	O	O
I am alerted to information in Spark that helps me	O	O	O	O	O	O
Spark enables me to make connections between ideas	O	O	O	O	O	O
Spark is an instrument for discovery	O	O	O	O	O	O

SDE Scale Question Set 3

Thinking of your perception of Spark from watching the video, please indicate your level of agreement with each of the following statements. If there is a statement that is unclear to you, please select "I Don't Know"

NOTE: You may find these statements repetitive. This is intentional and will help develop a better set of questions.

Survey p. **4 of 5**

	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly Agree	I Don't Know
I come to understand relationships between ideas in Spark	○	○	○	○	○	○
Information that interests me is highlighted in Spark	○	○	○	○	○	○
I bump into unexpected content in Spark	○	○	○	○	○	○
Spark invites examination of its content	○	○	○	○	○	○
Spark has features that ensure that my attention is drawn to useful information	○	○	○	○	○	○

SDE Scale Question Set 4

Thinking of your perception of Spark from watching the video, please indicate your level of agreement

with each of the following statements. If there is a statement that is unclear to you, please select "I Don't Know"

NOTE: You may find these statements repetitive. This is intentional and will help develop a better set of questions.

Survey p. **5 of 5**

	Strongly disagree	Disagree	Neither agree nor disagreee	Agree	Strongly agree	I Don't Know
I find information of value to me in Spark	○	○	○	○	○	○
There are many ways to explore information in Spark	○	○	○	○	○	○
I stumble upon information in Spark	○	○	○	○	○	○
It is easy to see links between information in Spark	○	○	○	○	○	○
I can see connections between topics in Spark	○	○	○	○	○	○

Thank You

APPENDIX B

UNIVERSITY OF NORTH TEXAS IRB APPROVAL LETTER

UNT
EST. 1890

THE OFFICE OF RESEARCH AND INNOVATION
Research and Economic Development

November 29, 2017

Dr. Oksana Zavalina
Student Investigator: Mark Hopkins
Department of Information Science
University of North Texas

RE: Human Subjects Application No. 17-305

Dear Dr. Zavalina:

In accordance with 45 CFR Part 46 Section 46.101, your study titled "A Study of Physicians' Serendipitous Knowledge Discovery in a Clinical Care Setting" has been determined to qualify for an exemption from further review by the UNT Institutional Review Board (IRB).

Enclosed are the consent documents with stamped IRB approval. Since you are conducting an online study, **please copy the approved language and paste onto the first page of your online survey. You may also use the enclosed stamped document as the first page of your online survey.**

No changes may be made to your study's procedures or forms without prior written approval from the UNT IRB. Please contact The Office of Research Integrity and Compliance at 940-565-4643 if you wish to make any such changes. Any changes to your procedures or forms after 3 years will require completion of a new IRB application.

We wish you success with your study.

Sincerely,

CT

Chad Trulson, Ph.D.
Professor
Chair, Institutional Review Board

CT:jm

APPENDIX C

INTEGRIS IRB APPROVAL LETTER

INTEGRIS

November 28, 2017

Mark Hopkins
3533 NW 116 Terrace
Oklahoma City, OK 73120

RE: Your application dated October 23, 2017 regarding study number 17-041: A Study of Physicians' Serendipitous Knowledge of Discovery in a Clinical Care Setting (Unfunded Study)

Dear Mr. Hopkins:

The INTEGRIS Health Institutional Review Board has reviewed your request for expedited approval of the new study listed above. This type of study qualifies for expedited review under FDA and DHHS (OHRP) Category 7 regulations.

The following have been approved by the IRB:

- Protocol – as submitted with the initial application
- Informed Consent as displayed on the first page of the online survey tool
- Online survey tool
- Recruitment e-mail to physicians
- Accrual of up to 200 subjects

You may conduct your study as described in your application effective immediately. The study is subject to continuing review on or before November 28, 2018, unless closed before that date. If the study is closed prior to the continuing review date, notification regarding the closure and a final report must be submitted to the Board.

Please note that any changes to the study as approved must be promptly reported and approved. Some changes may be approved by expedited review; others require full board review. If you have any questions or require further information, please contact the IRB Coordinator at 405.949.4184 or via e-mail at irb@integrisok.com.

Sincerely,

RC Brown

R.C. Brown, M.D., Chairman
INTEGRIS Health, Inc. Institutional Review Board

APPENDIX D

UNIVERSITY OF NORTH TEXAS IRB MODIFICATION APPROVAL

UNT
EST. 1890

THE OFFICE OF RESEARCH AND INNOVATION
Research and Economic Development

February 8, 2018

Dr. Oksana Zavalina
Student Investigator: Mark Hopkins
Department of Information Science
University of North Texas

Institutional Review Board for the Protection of Human Subjects in Research (IRB)
RE: Human Subject Application #17-305

Dear Dr. Zavalina:

The UNT IRB has received your request to modify your study titled "A Study of Physician's Serendipitous Knowledge Discovery in a Clinical Care Setting." As required by federal law and regulations governing the use of human subjects in research projects, the UNT IRB has examined the request to revise the data collection instrument by amending five multiple choice questions as follows: "Thinking of your perception of Spark from watching the video, please indicate your level of agreement with each of the following statements. If there is a statement that is unclear to you, please select, "I don't know." While this information is present in the first page of the survey, it will help with the interpretation of the five questions being asked. The modification to this study is hereby approved for use with human subjects.

Please contact The Office of Research Integrity and Compliance at (940) 565-4643, if you wish to make changes or need additional information.

Sincerely,

Chad Trulson, Ph.D.
Professor
Chair, Institutional Review Board

CT:jm

APPENDIX E

INTEGRIS IRB MODIFICATION APPROVAL

INTEGRIS

February 27, 2018

Mark Hopkins
3533 NW 116 Terrace
Oklahoma City, OK 73120

RE: Your application regarding the following study: A Study of Physicians' Serendipitous Knowledge of Discovery in a Clinical Care Setting (Unfunded Study) (17-041)

Dear Mr. Hopkins:

The IRB received on February 14, 2018 and reviewed on February 27, 2018 your application for revision of the study listed above. This type of revision qualifies for expedited review under FDA and DHHS (OHRP) regulations and *IRB Review Procedures and Administrative Operations Policy* Section 3.2.2.3(a).

The IRB approves this amendment which includes the following changes:

- Minor wording change to first statement following the video in the online survey tool. The words "from watching the video" will be added to the statement.

You may continue to conduct your study as revised effective immediately. The date for continuing review remains unchanged at November 28, 2018, unless closed before that date.

Please note that any further changes to the study must be promptly reported and approved. If you have any questions or require further information, please contact the IRB Coordinator at 405.949.4184 or via e-mail at irb@integrisok.com.

Sincerely,

R.C. Brown, M.D., Chairman
INTEGRIS Health, Inc. Institutional Review Board

APPENDIX F

RESEARCH STUDY EMAIL NOTIFICATION TEMPLATE

A Study of Physicians' Serendipitous Knowledge Discovery in a Clinical Care Setting

I am a current employee for INTEGRIS Health and this research is part of my dissertation. I would greatly appreciate your time taking the following survey (approximately 10-15 minutes) to support this research. You can access the survey at: https://unt.az1.qualtrics.com/jfe/form/SV_87jlYeqivjG7Htb

Overview: This study aims to improve on the understanding of physicians' information behavior searching the biomedical literature. The role of serendipity specifically, and a new online tool (called Spark) is the focus of this research.

Thank you! Mark Hopkins

APPENDIX G

RStudio R SCRIPT CODE

```
################################################################################
#
#
#   Modified on 3/24/2018
#
#
################################################################################
#
options(scipen=9999)
library(lavaan)
library(missForest)
library(foreign)
library(psych)
library(MASS)
library(corpcor)

# Note - from "cfa" help:  The cfa function is a wrapper for
#                     the more general lavaan function,
#                     using the following default arguments:
#
#                       int.ov.free = TRUE, int.lv.free = FALSE,
#                       auto.fix.first = TRUE (unless std.lv = TRUE),
```

```
#                auto.fix.single = TRUE, auto.var = TRUE,

#                auto.cov.lv.x = TRUE, auto.th = TRUE,

#                auto.delta = TRUE, and auto.cov.y = TRUE.

#

#



ifskd.model.free <- '


          itertv =~ X7 + X17 + X20 + X21

          change =~ X1 + X2 + X3 + X4 + X18

          knwldg =~ X5 + X6 + X10 + X19

          org_vs =~ X8 + X9 + X11 + X12 + X13 + X14 + X15 + X16

          '



model.to.fit<-ifskd.model.free



ifskd.model.0 <- '

          itertv =~ X7 + X17 + X20 + X21

          change =~ X1 + X2 + X3 + X4 + X18
```

```
knwldg =~ X5 + X6 + X10 + X19
org_vs =~ X8 + X9 + X11 + X12 + X13 + X14 + X15 + X16

itertv ~~ itertv
change ~~ change
knwldg ~~ knwldg
org_vs ~~ org_vs

itertv ~~ 0*change
itertv ~~ 0*knwldg
itertv ~~ 0*org_vs

change ~~ 0*knwldg
change ~~ 0*org_vs

knwldg ~~ 0*org_vs
'

model.to.fit<-ifskd.model.0


ifskd.model.items.fixed <- '
```

```
itertv =~ X7 + X17 + X20 + X21

change =~ X1 + X2 + X3 + X4 + X18

knwldg =~ X5 + X6 + X10 + X19

org_vs =~ X8 + X9 + X11 + X12 + X13 + X14 + X15 + X16


itertv ~~ itertv

change ~~ change

knwldg ~~ knwldg

org_vs ~~ org_vs


itertv ~~ change

itertv ~~ knwldg

itertv ~~ org_vs


change ~~ knwldg

change ~~ org_vs


knwldg ~~ org_vs


X1~~.3*X1

X2~~.3*X2
```

```
X3~~.3*X3

X4~~.3*X4

X5~~.3*X5

X6~~.3*X6

X7~~.3*X7

X8~~.3*X8

X9~~.3*X9

X10~~.3*X10

X11~~.3*X11

X12~~.3*X12

X13~~.3*X13

X14~~.3*X14

X15~~.3*X15

X16~~.3*X16

X17~~.3*X17

X18~~.3*X18

X19~~.3*X19

X20~~.3*X20

X21~~.3*X21

'

model.to.fit<-ifskd.model.items.fixed
```

3 McCay-Peet Models: A) Freely Estimated   B) LV Loadings Fixed to Zero   C) Indicator Loadings Fixed to 0.3

```
mccay.peet.model.free <- '
      EnablesExploration =~ X1 + X2 + X3 + X4
      TriggerRich =~ X5 + X6
      EnablesConnections =~ X7 + X8 + X9 + X10
      HighlightsTriggers =~ X11 + X12 + X13 + X14 + X15 + X16
      LeadsToUnexpected =~ X17 + X18 + X19 + X20 + X21
      '
model.to.fit<-mccay.peet.model.free


mccay.peet.model.0 <- '
     EnablesExploration =~ X1 + X2 + X3 + X4
     TriggerRich =~ X5 + X6
     EnablesConnections =~ X7 + X8 + X9 + X10
     HighlightsTriggers =~ X11 + X12 + X13 + X14 + X15 + X16
     LeadsToUnexpected =~ X17 + X18 + X19 + X20 + X21
```

```
EnablesExploration~~EnablesExploration
TriggerRich~~TriggerRich
EnablesConnections~~EnablesConnections
HighlightsTriggers~~HighlightsTriggers
LeadsToUnexpected~~LeadsToUnexpected

EnablesExploration~~0*TriggerRich
EnablesExploration~~0*EnablesConnections
EnablesExploration~~0*HighlightsTriggers
EnablesExploration~~0*LeadsToUnexpected

TriggerRich~~0*EnablesConnections
TriggerRich~~0*HighlightsTriggers
TriggerRich~~0*LeadsToUnexpected

EnablesConnections~~0*HighlightsTriggers
EnablesConnections~~0*LeadsToUnexpected

HighlightsTriggers~~0*LeadsToUnexpected
'
```

```
model.to.fit<-mccay.peet.model.0



mccay.peet.model.items.fixed <- '
            EnablesExploration =~ X1 + X2 + X3 + X4
            TriggerRich =~ X5 + X6
            EnablesConnections =~ X7 + X8 + X9 + X10
            HighlightsTriggers =~ X11 + X12 + X13 + X14 + X15 + X16
            LeadsToUnexpected =~ X17 + X18 + X19 + X20 + X21

            EnablesExploration~~EnablesExploration
            TriggerRich~~TriggerRich
            EnablesConnections~~EnablesConnections
            HighlightsTriggers~~HighlightsTriggers
            LeadsToUnexpected~~LeadsToUnexpected

            EnablesExploration~~TriggerRich
            EnablesExploration~~EnablesConnections
            EnablesExploration~~HighlightsTriggers
            EnablesExploration~~LeadsToUnexpected

            TriggerRich~~EnablesConnections
```

TriggerRich~~HighlightsTriggers

TriggerRich~~LeadsToUnexpected

EnablesConnections~~HighlightsTriggers

EnablesConnections~~LeadsToUnexpected

HighlightsTriggers~~LeadsToUnexpected

X1~~.3*X1

X2~~.3*X2

X3~~.3*X3

X4~~.3*X4

X5~~.3*X5

X6~~.3*X6

X7~~.3*X7

X8~~.3*X8

X9~~.3*X9

X10~~.3*X10

X11~~.3*X11

X12~~.3*X12

X13~~.3*X13

X14~~.3*X14

```
X15~~.3*X15

X16~~.3*X16

X17~~.3*X17

X18~~.3*X18

X19~~.3*X19

X20~~.3*X20

X21~~.3*X21

'

model.to.fit<-mccay.peet.model.items.fixed

###### 2 Single Models:  A) Freely Estimated    C) Indicator Loadings Fixed to 0.3

# Two different methods for setting scale of items;  Function "cfa"

# automatically sets these for us with defaults; However, interestingly setting the

# manifest variable versus the latent variable can produce different results

single.factor.model <- '

serendipity =~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 +

X10 + X11 + X12 + X13 + X14 + X15 + X16 +
```

```
                X17 + X18 + X19 + X20 + X21
                '

model.to.fit<-single.factor.model


#
# serendipity ~~ serendipity

single.factor.model.items.fixed <- '
                        serendipity =~ X1 + X2 + X3 + X4 + X5 + X6 + X7 + X8 + X9 +
                        X10 + X11 + X12 + X13 + X14 + X15 + X16 +
                        X17 + X18 + X19 + X20 + X21
                        X1~~.3*X1
                        X2~~.3*X2
                        X3~~.3*X3
                        X4~~.3*X4
                        X5~~.3*X5
                        X6~~.3*X6
                        X7~~.3*X7
                        X8~~.3*X8
                        X9~~.3*X9
```

```
        X10~~.3*X10

        X11~~.3*X11

        X12~~.3*X12

        X13~~.3*X13

        X14~~.3*X14

        X15~~.3*X15

        X16~~.3*X16

        X17~~.3*X17

        X18~~.3*X18

        X19~~.3*X19

        X20~~.3*X20

        X21~~.3*X21

        '


model.to.fit<-single.factor.model.items.fixed


#################################################################

#

# Change model being tested here:

#

# IFSKD model:

#
```

```
# model.to.fit<-ifskd.model.free

 model.to.fit<-ifskd.model.0

# model.to.fit<-ifskd.model.items.fixed


# McCay-Peet model:

#

# model.to.fit<-mccay.peet.model.free

# model.to.fit<-mccay.peet.model.0

# model.to.fit<-mccay.peet.model.items.fixed


# Single-Factor model:


# model.to.fit<-single.factor.model

# model.to.fit<-single.factor.model.items.fixed




##########################################################################

#

# Create population data with the exact shrunken correlation matrix

# returned from "cov.shrink"

#
```

```
#
# This is the n=23 data set:
#
# > nrow(hopkins.2.df)
# [1] 23
#
#
#
 cov.shrink.est<-cov.shrink(hopkins.2.df)
#
# pop.size<-10000
# hopkins.sim<-mvrnorm(n=pop.size, mu=apply(hopkins.2.df,2,mean),
#                 Sigma=cov.shrink.est, empirical=TRUE)
#
#
# Can run the model on the original data or the simulated data:
#
# hopkins.2.df (original n=23 data);
# hopkins.sim (n=10000) simulated data

model.fit.out<- cfa(model.to.fit,
            data=hopkins.sim, # <----- can change this to hopkins.2.df
```

```
        estimator="DWLS",       # to examine the fit of the model to
        test="default", ridge=.00001,    # the original data
        se="robust")

# Examine eigen-values of the latent variables covariance matrix
eigen(inspect(model.fit.out, "cov.lv"))$values


summary(model.fit.out, fit.measures=TRUE)
fitMeasures(model.fit.out)
standardizedSolution(model.fit.out)



# Sub-sample size - here the original size is used
sub.samp.size<-1000

# Set number of Monte Carlo replications
monte.rep<-1000

# Initialize collection matrices
fit.out.collect<-matrix(NA, ncol=length(fitMeasures(model.fit.out)), nrow=monte.rep)
#coefs.out.collect<-matrix(NA, ncol=length(standardizedSolution(model.fit.out)[,4]),
nrow=monte.rep)
```

```
coefs.out.collect<-matrix(NA, ncol=length(coef(model.fit.out)), nrow=monte.rep)


# Name columns of collection matrices

colnames(fit.out.collect)<-rownames(data.frame(fitMeasures(model.fit.out)))

colnames(coefs.out.collect)<-rownames(data.frame(coef(model.fit.out)))

# colnames(coefs.out.collect)<-rownames(paste(standardizedSolution(model.fit.out)[,1],

#                                   standardizedSolution(model.fit.out)[,2],

#                                   standardizedSolution(model.fit.out)[,3]))


# Change to estimate or not estimate

# the SUB-SAMPLE cov shrinkage estimator;

# This would only be needed for n=23 simulated

# and sub-sampled data; A non-positive definte

# matrix on the observed data returns an ERROR


estimate.cov.shrink.sub.sample<-FALSE


for(i in 1:monte.rep)

{

  hopkins.sub.samp.index<-sample(1:sub.samp.size,  sub.samp.size,

                          replace=TRUE)

  hopkins.sub.df<-hopkins.sim[hopkins.sub.samp.index, ]
```

```
# Subsample estimate of covariance shrinkage estimator

if(estimate.cov.shrink.sub.sample==TRUE)
 {
  cov.sub.shrink.est<-cov.shrink(hopkins.sub.df)
  hopkins.sub.df<-mvrnorm(n=sub.samp.size, mu=apply(hopkins.sub.df,2,mean),
                     Sigma=cov.sub.shrink.est, empirical=TRUE)
  print(max(abs(cov.shrink.est - cov.sub.shrink.est)))
  }

# When fitting the SEM model to the subsampled data here are some possible
# lavaan errors that can occur:
#
#       Non-positive definite observed covariance matrix:  show-stopper
#
# Non-positive definite latent variable covariance matrix:  can proceed, but standard
#                                               errors can't be trusted and loadings COULD be
#                                               suspect; This is why we want to use the Monte-Carlo
#                                               approach - we can get the standard errors
#
```

```
 #              Sample size too small to compute  Gamma:  can proceed, but standard errors can't
be trusted
 #
 #                              Negative variances:  can proceed, but the specific loadings (with
neg.var)
 #                                            and their corresponding standard errors can't be trusted
 #
 #
#   model.fit.out <- cfa(model.to.fit,
#                   data=hopkins.sub.df, estimator="DWLS",
#                   test="default", ridge=.0001,
#                   se="robust")
#
#   eigen(inspect(model.fit.out, "cov.lv"))$values
#
#   summary(model.fit.out, fit.measures=TRUE)
#   fitMeasures(model.fit.out)
#   standardizedSolution(model.fit.out)


  try(model.fit.out <- cfa(model.to.fit,
                   data=hopkins.sub.df, estimator="DWLS",
```

```
                        test="default", ridge=.0000001,
                        se="robust"), silent=TRUE)



  cat("\n")
  cat(i,"\n")
  cat("\n")

  try(fit.out.collect[i, ]<-data.frame(fitMeasures(model.fit.out))[,1], silent=TRUE)
  try(coefs.out.collect[i, ]<-coef(model.fit.out), silent=TRUE)
  #try(coefs.out.collect[i, ]<-standardizedSolution(model.fit.out)[,4])
}

# Note:  make sure to remove NA's created by "non-posdef" samples

summary(fit.out.collect)
summary(coefs.out.collect)

head(fit.out.collect)
tail(fit.out.collect)

head(coefs.out.collect)
```

```
tail(coefs.out.collect)

# Winsorize at 2% levels before calculating summary statistics and quantiles;
# This is to downweight the values of poorly fitting models

fit.winsor<-apply(fit.out.collect, 2, winsor, trim=.02, na.rm=TRUE)
coefs.winsor<-apply(coefs.out.collect, 2, winsor, trim=.02, na.rm=TRUE)

median.fit<-apply(fit.winsor, 2, median, na.rm=TRUE)
median.coefs<-apply(coefs.winsor, 2, median, na.rm=TRUE)

mean.fit<-apply(fit.winsor, 2, mean, na.rm=TRUE)
mean.coefs<-apply(coefs.winsor, 2, mean, na.rm=TRUE)

# Looking at the histogram of the Monte Carlo replications of the coefficients
# helps determine if the sampling distributions are symmetric, unimodal, etc.
# Here we look at the indicator for an item in column 1 across 1000 Monte Carlo reps

hist(coefs.winsor[,1])
```

```
sd.fit<-apply(fit.winsor, 2, sd, na.rm=TRUE)

sd.coefs<-apply(coefs.winsor, 2, sd, na.rm=TRUE)


# Calculate Finite Population Correction (from sampling theory);

# This is not large only 1 part in 1000 (roughly)



FPC<-sqrt((pop.size-sub.samp.size)/(pop.size-1))


sd.coefs.corrected<-sd.coefs * FPC

sd.fit.corrected<-sd.fit * FPC


mean.fit

sd.fit.corrected


mean.coefs

sd.coefs.corrected


mean.fit

sd.fit.corrected
```

```
# Simple percentile based 95% CI;

# Note if the interval contains 0,  then the interval is not significant


CI.95.coefs<-apply(coefs.winsor,  2, quantile, prob=c(.025, .975), na.rm=TRUE)

CI.95.coefs


CI.95.fit<-apply(fit.winsor,  2, quantile, prob=c(.025, .975), na.rm=TRUE)

CI.95.fit




sink(file = "McCay-Peet-Item-Errors-Fixed-23-Unstandardized.html")

library(knitr)

library(kableExtra)

options(knitr.table.format = "html")

(kable(median.fit,caption = "median.fit"))

(kable(mean.fit,caption = "mean.fit"))

(kable(sd.fit.corrected,caption = "sd.fit.corrected"))

(kable(median.coefs,caption = "median.coefs"))

(kable(mean.coefs,caption = "mean.coefs"))

(kable(sd.coefs.corrected,caption = "sd.coefs.corrected"))
```

```
(kable(CI.95.coefs,caption = "CI.95.coefs"))

(kable(CI.95.fit,caption = "CI.95.fit"))

sink()



# Information measures


unstd.col<-data.frame(coefs.winsor)

rownames(unstd.col)<-1:nrow(unstd.col)

unstd.col


# Information metric for choosing items to constrain or remove


info.f1<-unstd.col[c(1:21),1]/unstd.col[c(1:21),1]

info.f1
```

APPENDIX H

OBSERVED COVARIANCE MATRIX FROM IMPUTED DATA

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	0.3694	0.2245	0.2766	0.2779	0.2726	0.2994	0.2787	0.3295	0.3267	0.1884
X2	0.2245	0.6301	0.3664	0.3083	0.4047	0.2322	0.2418	0.3413	0.3017	0.1344
X3	0.2766	0.3664	0.4105	0.3261	0.3212	0.2789	0.2212	0.3833	0.2168	0.1276
X4	0.2779	0.3083	0.3261	0.4520	0.3591	0.4517	0.2952	0.4026	0.3582	0.1652
X5	0.2726	0.4047	0.3212	0.3591	0.4979	0.3636	0.4184	0.3553	0.3801	0.2634
X6	0.2994	0.2322	0.2789	0.4517	0.3636	0.9158	0.4163	0.4723	0.5024	0.2463
X7	0.2787	0.2418	0.2212	0.2952	0.4184	0.4163	0.5143	0.2713	0.4903	0.2971
X8	0.3295	0.3413	0.3833	0.4026	0.3553	0.4723	0.2713	0.5369	0.3060	0.1767
X9	0.3267	0.3017	0.2168	0.3582	0.3801	0.5024	0.4903	0.3060	0.6759	0.3093
X10	0.1884	0.1344	0.1276	0.1652	0.2634	0.2463	0.2971	0.1767	0.3093	0.2682
X11	0.2344	0.2418	0.2674	0.2968	0.2366	0.3346	0.1963	0.3240	0.2634	0.1377
X12	0.2020	0.2847	0.2754	0.2873	0.2870	0.2755	0.2149	0.3243	0.1969	0.1279
X13	0.3214	0.2799	0.2703	0.3296	0.3685	0.3656	0.3936	0.3438	0.4273	0.2286
X14	0.2304	0.2092	0.1927	0.2614	0.2635	0.3486	0.3054	0.2918	0.3209	0.1347
X15	0.2896	0.4404	0.4706	0.3730	0.3146	0.4562	0.2159	0.5216	0.2390	0.1068
X16	0.3375	0.3733	0.3378	0.2680	0.2541	0.3889	0.2512	0.3895	0.3279	0.1312
X17	0.1403	0.2251	0.1374	0.2306	0.2716	0.3920	0.3230	0.2718	0.3689	0.1868
X18	0.2359	0.1584	0.2180	0.3062	0.2414	0.3721	0.2254	0.2999	0.2799	0.1602
X19	0.2193	0.2249	0.2225	0.2327	0.2209	0.3619	0.3251	0.2498	0.4285	0.1781
X20	0.3160	0.2720	0.2267	0.3208	0.3165	0.3165	0.3506	0.3258	0.3851	0.1429
X21	0.2251	0.3521	0.2320	0.2131	0.2652	0.2425	0.2074	0.2849	0.2713	0.1257

	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21
X1	0.2344	0.2020	0.3214	0.2304	0.2896	0.3375	0.1403	0.2359	0.2193	0.3160	0.2251
X2	0.2418	0.2847	0.2799	0.2092	0.4404	0.3733	0.2251	0.1584	0.2249	0.2720	0.3521
X3	0.2674	0.2754	0.2703	0.1927	0.4706	0.3378	0.1374	0.2180	0.2225	0.2267	0.2320
X4	0.2968	0.2873	0.3296	0.2614	0.3730	0.2680	0.2306	0.3062	0.2327	0.3208	0.2131
X5	0.2366	0.2870	0.3685	0.2635	0.3146	0.2541	0.2716	0.2414	0.2209	0.3165	0.2652
X6	0.3346	0.2755	0.3656	0.3486	0.4562	0.3889	0.3920	0.3721	0.3619	0.3165	0.2425

X7	0.1963	0.2149	0.3936	0.3054	0.2159	0.2512	0.3230	0.2254	0.3251	0.3506	0.2074
X8	0.3240	0.3243	0.3438	0.2918	0.5216	0.3895	0.2718	0.2999	0.2498	0.3258	0.2849
X9	0.2634	0.1969	0.4273	0.3209	0.2390	0.3279	0.3689	0.2799	0.4285	0.3851	0.2713
X10	0.1377	0.1279	0.2286	0.1347	0.1068	0.1312	0.1868	0.1602	0.1781	0.1429	0.1257
X11	0.3337	0.2427	0.2574	0.1722	0.3540	0.3038	0.2331	0.3166	0.2486	0.2192	0.2076
X12	0.2427	0.3248	0.2525	0.1455	0.3632	0.2913	0.2037	0.2037	0.1620	0.2020	0.1919
X13	0.2574	0.2525	0.4470	0.2884	0.2900	0.3498	0.2747	0.2492	0.2920	0.4005	0.2397
X14	0.1722	0.1455	0.2884	0.3879	0.2293	0.2949	0.3197	0.1876	0.2867	0.3679	0.2021
X15	0.3540	0.3632	0.2900	0.2293	0.6847	0.4968	0.2380	0.2755	0.3030	0.2412	0.3007
X16	0.3038	0.2913	0.3498	0.2949	0.4968	0.6536	0.2939	0.2429	0.3669	0.3313	0.3234
X17	0.2331	0.2037	0.2747	0.3197	0.2380	0.2939	0.4599	0.2340	0.3550	0.2690	0.2268
X18	0.3166	0.2037	0.2492	0.1876	0.2755	0.2429	0.2340	0.3829	0.2352	0.2147	0.1749
X19	0.2486	0.1620	0.2920	0.2867	0.3030	0.3669	0.3550	0.2352	0.4574	0.2490	0.2343
X20	0.2192	0.2020	0.4005	0.3679	0.2412	0.3313	0.2690	0.2147	0.2490	0.4894	0.2293
X21	0.2076	0.1919	0.2397	0.2021	0.3007	0.3234	0.2268	0.1749	0.2343	0.2293	0.2997

APPENDIX I

SHRINKAGE BASED COVARIANCE MATRIX FROM IMPUTED DATA

	X1	X2	X3	X4	X5	X6	X7	X8	X9	X10
X1	0.4420	0.1859	0.2723	0.2628	0.2478	0.2159	0.2501	0.2906	0.2633	0.2230
X2	0.1859	0.4876	0.2901	0.2345	0.2958	0.1346	0.1745	0.2420	0.1956	0.1280
X3	0.2723	0.2901	0.4492	0.2950	0.2792	0.1923	0.1898	0.3233	0.1671	0.1444
X4	0.2628	0.2345	0.2950	0.4564	0.2999	0.2992	0.2433	0.3262	0.2652	0.1797
X5	0.2478	0.2958	0.2792	0.2999	0.4645	0.2315	0.3315	0.2766	0.2705	0.2753
X6	0.2159	0.1346	0.1923	0.2992	0.2315	0.5375	0.2616	0.2917	0.2836	0.2042
X7	0.2501	0.1745	0.1898	0.2433	0.3315	0.2616	0.4673	0.2085	0.3443	0.3065
X8	0.2906	0.2420	0.3233	0.3262	0.2766	0.2917	0.2085	0.4713	0.2112	0.1792
X9	0.2633	0.1956	0.1671	0.2652	0.2705	0.2836	0.3443	0.2112	0.4956	0.2867
X10	0.2230	0.1280	0.1444	0.1797	0.2753	0.2042	0.3065	0.1792	0.2867	0.4243
X11	0.2521	0.2092	0.2750	0.2932	0.2247	0.2520	0.1840	0.2985	0.2218	0.1703
X12	0.2198	0.2492	0.2866	0.2873	0.2757	0.2100	0.2039	0.3023	0.1678	0.1601
X13	0.3054	0.2139	0.2456	0.2877	0.3092	0.2433	0.3259	0.2798	0.3178	0.2497
X14	0.2323	0.1696	0.1858	0.2422	0.2346	0.2462	0.2683	0.2521	0.2533	0.1562
X15	0.2322	0.2840	0.3609	0.2748	0.2228	0.2562	0.1509	0.3582	0.1500	0.0985
X16	0.2755	0.2450	0.2637	0.2010	0.1832	0.2223	0.1787	0.2724	0.2096	0.1232
X17	0.1318	0.1700	0.1234	0.1989	0.2252	0.2578	0.2644	0.2186	0.2712	0.2017
X18	0.2392	0.1292	0.2114	0.2852	0.2161	0.2642	0.1992	0.2604	0.2222	0.1867
X19	0.2064	0.1702	0.2003	0.2012	0.1836	0.2386	0.2667	0.2014	0.3157	0.1927
X20	0.2893	0.2002	0.1984	0.2698	0.2558	0.2029	0.2797	0.2555	0.2759	0.1505
X21	0.2538	0.3192	0.2502	0.2207	0.2640	0.1915	0.2037	0.2751	0.2394	0.1629

	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21
X1	0.2521	0.2198	0.3054	0.2323	0.2322	0.2755	0.1318	0.2392	0.2064	0.2893	0.2538
X2	0.2092	0.2492	0.2139	0.1696	0.2840	0.2450	0.1700	0.1292	0.1702	0.2002	0.3192
X3	0.2750	0.2866	0.2456	0.1858	0.3609	0.2637	0.1234	0.2114	0.2003	0.1984	0.2502
X4	0.2932	0.2873	0.2877	0.2422	0.2748	0.2010	0.1989	0.2852	0.2012	0.2698	0.2207
X5	0.2247	0.2757	0.3092	0.2346	0.2228	0.1832	0.2252	0.2161	0.1836	0.2558	0.2640
X6	0.2520	0.2100	0.2433	0.2462	0.2562	0.2223	0.2578	0.2642	0.2386	0.2029	0.1915

X7	0.1840	0.2039	0.3259	0.2683	0.1509	0.1787	0.2644	0.1992	0.2667	0.2797	0.2037
X8	0.2985	0.3023	0.2798	0.2521	0.3582	0.2724	0.2186	0.2604	0.2014	0.2555	0.2751
X9	0.2218	0.1678	0.3178	0.2533	0.1500	0.2096	0.2712	0.2222	0.3157	0.2759	0.2394
X10	0.1703	0.1601	0.2497	0.1562	0.0985	0.1232	0.2017	0.1867	0.1927	0.1505	0.1629
X11	0.4358	0.2759	0.2555	0.1814	0.2966	0.2591	0.2287	0.3353	0.2444	0.2096	0.2445
X12	0.2759	0.4342	0.2536	0.1551	0.3079	0.2514	0.2022	0.2183	0.1612	0.1955	0.2287
X13	0.2555	0.2536	0.4556	0.2684	0.2147	0.2635	0.2381	0.2332	0.2536	0.3384	0.2493
X14	0.1814	0.1551	0.2684	0.4452	0.1801	0.2358	0.2940	0.1863	0.2643	0.3298	0.2231
X15	0.2966	0.3079	0.2147	0.1801	0.4971	0.3159	0.1741	0.2176	0.2221	0.1720	0.2641
X16	0.2591	0.2514	0.2635	0.2358	0.3159	0.4917	0.2188	0.1952	0.2738	0.2404	0.2891
X17	0.2287	0.2022	0.2381	0.2940	0.1741	0.2188	0.4578	0.2164	0.3048	0.2247	0.2332
X18	0.3353	0.2183	0.2332	0.1863	0.2176	0.1952	0.2164	0.4444	0.2180	0.1935	0.1941
X19	0.2444	0.1612	0.2536	0.2643	0.2221	0.2738	0.3048	0.2180	0.4574	0.2084	0.2415
X20	0.2096	0.1955	0.3384	0.3298	0.1720	0.2404	0.2247	0.1935	0.2084	0.4630	0.2298
X21	0.2445	0.2287	0.2493	0.2231	0.2641	0.2891	0.2332	0.1941	0.2415	0.2298	0.4298

APPENDIX J

CONFIRMATORY FACTOR ANALYSIS (CFA) ADDITIONAL SUMMARY STATISTICS

McCay.Peet.Model.Free		McCay.Peet.Model.Free		McCay.Peet.Model.Free		McCay.Peet.Model.Free		McCay.Peet.Model.Free	
median.fit (N=23)		median.fit (N=100)		median.fit (N=200)		median.fit (N=500)		median.fit (N=1000)	
fmin	1.914	fmin	0.880	fmin	0.771	fmin	0.711	fmin	0.691
chisq	88.043	chisq	176.007	chisq	308.531	chisq	710.704	chisq	1381.461
pvalue	1.000	pvalue	0.549	pvalue	0.000	pvalue	0.000	pvalue	0.000
cfi	1.000	cfi	1.000	cfi	0.986	cfi	0.976	cfi	0.973
tli	1.143	tli	1.001	tli	0.983	tli	0.972	tli	0.969
nnfi	1.143	nnfi	1.001	nnfi	0.983	nnfi	0.972	nnfi	0.969
rfi	0.883	rfi	0.957	rfi	0.961	rfi	0.963	rfi	0.964
nfi	0.900	nfi	0.963	nfi	0.967	nfi	0.969	nfi	0.969
pnfi	0.768	pnfi	0.821	pnfi	0.824	pnfi	0.826	pnfi	0.826
ifi	1.118	ifi	1.001	ifi	0.986	ifi	0.976	ifi	0.973
rni	1.122	rni	1.001	rni	0.986	rni	0.976	rni	0.973
rmsea	0.000	rmsea	0.000	rmsea	0.060	rmsea	0.077	rmsea	0.082
rmr	0.054	rmr	0.044	rmr	0.042	rmr	0.041	rmr	0.041
srmr	0.119	srmr	0.093	srmr	0.088	srmr	0.085	srmr	0.084
gfi	0.995	gfi	0.998	gfi	0.998	gfi	0.998	gfi	0.998
agfi	0.993	agfi	0.997	agfi	0.997	agfi	0.997	agfi	0.998
pgfi	0.707	pgfi	0.709	pgfi	0.709	pgfi	0.709	pgfi	0.709
mfi	7.903	mfi	1.015	mfi	0.722	mfi	0.587	mfi	0.548

IF-SKD.Model.Free		IF-SKD.Model.Free		IF-SKD.Model.Free		IF-SKD.Model.Free		IF-SKD.Model.Free	
median.fit (N=23)		median.fit (N=100)		median.fit (N=200)		median.fit (N=500)		median.fit (N=1000)	
fmin	1.967	fmin	0.914	fmin	0.803	fmin	0.742	fmin	0.716
chisq	90.490	chisq	182.776	chisq	321.343	chisq	742.308	chisq	1431.184
pvalue	1.000	pvalue	0.491	pvalue	0.000	pvalue	0.000	pvalue	0.000
cfi	1.000	cfi	1.000	cfi	0.985	cfi	0.975	cfi	0.972
tli	1.146	tli	1.000	tli	0.982	tli	0.972	tli	0.968
nnfi	1.146	nnfi	1.000	nnfi	0.982	nnfi	0.972	nnfi	0.968
rfi	0.881	rfi	0.956	rfi	0.960	rfi	0.962	rfi	0.964
nfi	0.896	nfi	0.961	nfi	0.965	nfi	0.967	nfi	0.968
pnfi	0.781	pnfi	0.838	pnfi	0.841	pnfi	0.843	pnfi	0.844
ifi	1.122	ifi	1.000	ifi	0.985	ifi	0.975	ifi	0.972
rni	1.127	rni	1.000	rni	0.985	rni	0.975	rni	0.972
rmsea	0.000	rmsea	0.000	rmsea	0.062	rmsea	0.078	rmsea	0.083
rmr	0.054	rmr	0.045	rmr	0.043	rmr	0.042	rmr	0.041
srmr	0.120	srmr	0.094	srmr	0.089	srmr	0.086	srmr	0.085
gfi	0.995	gfi	0.998	gfi	0.998	gfi	0.998	gfi	0.998
agfi	0.993	agfi	0.997	agfi	0.997	agfi	0.997	agfi	0.997
pgfi	0.723	pgfi	0.725	pgfi	0.725	pgfi	0.725	pgfi	0.725

mfi	8.187	mfi	1.001	mfi	0.706	mfi		mfi	0.535

Single.Factor.Model		Single.Factor.Model		Single.Factor.Model		Single.Factor.Model		Single.Factor.Model	
median.fit (N=23)		median.fit (N=100)		median.fit (N=200)		median.fit (N=500)		median.fit (N=1000)	
fmin	2.011	fmin	1.980	fmin	1.430	fmin	1.123	fmin	0.927
chisq	92.493	chisq	395.998	chisq	572.141	chisq	1122.836	chisq	1854.277
pvalue	1.000	pvalue	0.000	pvalue	0.000	pvalue	0.000	pvalue	0.000
cfi	1.000	cfi	0.943	cfi	0.949	cfi	0.956	cfi	0.960
tli	1.142	tli	0.937	tli	0.943	tli	0.951	tli	0.956
nnfi	1.142	nnfi	0.937	nnfi	0.943	nnfi	0.951	nnfi	0.956
rfi	0.884	rfi	0.886	rfi	0.917	rfi	0.942	rfi	0.951
nfi	0.896	nfi	0.897	nfi	0.925	nfi	0.948	nfi	0.956
pnfi	0.806	pnfi	0.807	pnfi	0.833	pnfi	0.853	pnfi	0.860
ifi	1.124	ifi	0.943	ifi	0.949	ifi	0.956	ifi	0.960
rni	1.128	rni	0.943	rni	0.949	rni	0.956	rni	0.960
rmsea	0.000	rmsea	0.105	rmsea	0.101	rmsea	0.100	rmsea	0.094
rmr	0.055	rmr	0.059	rmr	0.052	rmr	0.049	rmr	0.046
srmr	0.121	srmr	0.134	srmr	0.117	srmr	0.104	srmr	0.096
gfi	0.995	gfi	0.996	gfi	0.997	gfi	0.997	gfi	0.998
agfi	0.993	agfi	0.994	agfi	0.996	agfi	0.996	agfi	0.997
pgfi	0.746	pgfi	0.747	pgfi	0.748	pgfi	0.748	pgfi	0.748
mfi	8.965	mfi	0.352	mfi	0.382	mfi	0.392	mfi	0.435

McCay.Peet.Model.Items.Fixed		McCay.Peet.Model.Items.Fixed		McCay.Peet.Model.Items.Fixed		McCay.Peet.Model.Items.Fixed		McCay.Peet.Model.Items.Fixed	
median.fit (N=23)		median.fit (N=100)		median.fit (N=200)		median.fit (N=500)		median.fit (N=1000)	
fmin	2.130	fmin	1.122	fmin	0.971	fmin	0.881	fmin	0.859
chisq	97.960	chisq	224.345	chisq	388.323	chisq	880.854	chisq	1717.722
pvalue	1.000	pvalue	0.114	pvalue	0.000	pvalue	0.000	pvalue	0.000
cfi	1.000	cfi	0.995	cfi	0.979	cfi	0.970	cfi	0.966
tli	1.140	tli	0.994	tli	0.978	tli	0.968	tli	0.964
nnfi	1.140	nnfi	0.994	nnfi	0.978	nnfi	0.968	nnfi	0.964
rfi	0.883	rfi	0.951	rfi	0.956	rfi	0.959	rfi	0.960
nfi	0.888	nfi	0.953	nfi	0.958	nfi	0.961	nfi	0.962
pnfi	0.846	pnfi	0.908	pnfi	0.913	pnfi	0.915	pnfi	0.916
ifi	1.131	ifi	0.995	ifi	0.979	ifi	0.970	ifi	0.966
rni	1.133	rni	0.995	rni	0.979	rni	0.970	rni	0.966
rmsea	0.000	rmsea	0.035	rmsea	0.069	rmsea	0.083	rmsea	0.087
rmr	0.059	rmr	0.051	rmr	0.049	rmr	0.047	rmr	0.047

srmr	0.127	srmr	0.108	srmr	0.102	srmr	0.099	srmr	0.097
gfi	0.995	gfi	0.997	gfi	0.998	gfi	0.998	gfi	0.998
agfi	0.993	agfi	0.997	agfi	0.997	agfi	0.997	agfi	0.997
pgfi	0.789	pgfi	0.791	pgfi	0.792	pgfi	0.792	pgfi	0.792
mfi	10.166	mfi	0.884	mfi	0.623	mfi	0.505	mfi	0.468
IF-SKD.Model.Items.Fixed		IF-SKD.Model.Items.Fixed		IF-SKD.Model.Items.Fixed		IF-SKD.Model.Items.Fixed		IF-SKD.Model.Items.Fixed	
median.fit (N=23)		median.fit (N=100)		median.fit (N=200)		median.fit (N=500)		median.fit (N=1000)	
fmin	2.176	fmin	1.161	fmin	0.998	fmin	0.904	fmin	0.879
chisq	100.107	chisq	232.180	chisq	399.241	chisq	903.937	chisq	1758.990
pvalue	1.000	pvalue	0.086	pvalue	0.000	pvalue	0.000	pvalue	0.000
cfi	1.000	cfi	0.994	cfi	0.979	cfi	0.969	cfi	0.965
tli	1.142	tli	0.994	tli	0.978	tli	0.968	tli	0.964
nnfi	1.142	nnfi	0.994	nnfi	0.978	nnfi	0.968	nnfi	0.964
rfi	0.887	rfi	0.950	rfi	0.956	rfi	0.959	rfi	0.960
nfi	0.890	nfi	0.952	nfi	0.957	nfi	0.960	nfi	0.961
pnfi	0.865	pnfi	0.924	pnfi	0.930	pnfi	0.933	pnfi	0.934
ifi	1.137	ifi	0.994	ifi	0.979	ifi	0.969	ifi	0.965
rni	1.138	rni	0.994	rni	0.979	rni	0.969	rni	0.965
rmsea	0.000	rmsea	0.037	rmsea	0.069	rmsea	0.083	rmsea	0.087
rmr	0.060	rmr	0.052	rmr	0.049	rmr	0.048	rmr	0.047
srmr	0.129	srmr	0.110	srmr	0.103	srmr	0.099	srmr	0.098
gfi	0.994	gfi	0.997	gfi	0.997	gfi	0.998	gfi	0.998
agfi	0.993	agfi	0.996	agfi	0.997	agfi	0.997	agfi	0.997
pgfi	0.805	pgfi	0.807	pgfi	0.807	pgfi	0.808	pgfi	0.808
mfi	10.604	mfi	0.867	mfi	0.612	mfi	0.496	mfi	0.459
Single.Factor.Model.Items.Fixed		Single.Factor.Model.Items.Fixed		Single.Factor.Model.Items.Fixed		Single.Factor.Model.Items.Fixed			
median.fit (N=23)		median.fit (N=100)		median.fit (N=200)		median.fit (N=500)			
fmin	2.263	fmin	1.203	fmin	1.037	fmin	0.953		
chisq	104.113	chisq	240.686	chisq	414.631	chisq	953.377		
pvalue	1.000	pvalue	0.072	pvalue	0.000	pvalue	0.000		
cfi	1.000	cfi	0.993	cfi	0.977	cfi	0.967		
tli	1.143	tli	0.993	tli	0.977	tli	0.967		
nnfi	1.143	nnfi	0.993	nnfi	0.977	nnfi	0.967		
rfi	0.881	rfi	0.950	rfi	0.955	rfi	0.958		
nfi	0.881	nfi	0.950	nfi	0.955	nfi	0.958		
pnfi	0.881	pnfi	0.950	pnfi	0.955	pnfi	0.958		
ifi	1.143	ifi	0.993	ifi	0.977	ifi	0.967		
rni	1.143	rni	0.993	rni	0.977	rni	0.967		

rmsea	0.000	rmsea	0.038	rmsea	0.070	rmsea	0.084
rmr	0.060	rmr	0.053	rmr	0.051	rmr	0.049
srmr	0.131	srmr	0.113	srmr	0.106	srmr	0.103
gfi	0.994	gfi	0.997	gfi	0.997	gfi	0.998
agfi	0.993	agfi	0.996	agfi	0.997	agfi	0.997
pgfi	0.829	pgfi	0.831	pgfi	0.831	pgfi	0.831
mfi	11.095	mfi	0.856	mfi	0.598	mfi	0.475

McCay.Peet.Model.Free		McCay.Peet.Model.Free		McCay.Peet.Model.Free		McCay.Peet.Model.Free		McCay.Peet.Model.Free	
sd.fit.corrected (N=23)		sd.fit.corrected (N=100)		sd.fit.corrected (N=200)		sd.fit.corrected (N=500)		sd.fit.corrected (N=1000)	
fmin	0.613	fmin	0.209	fmin	0.135	fmin	0.078	fmin	0.053
chisq	28.184	chisq	41.832	chisq	53.805	chisq	78.184	chisq	106.373
pvalue	0.063	pvalue	0.396	pvalue	0.010	pvalue	0.000	pvalue	0.000
cfi	0.000	cfi	0.007	cfi	0.007	cfi	0.004	cfi	0.003
tli	0.081	tli	0.011	tli	0.008	tli	0.005	tli	0.003
nnfi	0.081	nnfi	0.011	nnfi	0.008	nnfi	0.005	nnfi	0.003
rfi	0.077	rfi	0.014	rfi	0.009	rfi	0.005	rfi	0.004
nfi	0.066	nfi	0.012	nfi	0.008	nfi	0.005	nfi	0.003
pnfi	0.056	pnfi	0.010	pnfi	0.007	pnfi	0.004	pnfi	0.003
ifi	0.062	ifi	0.010	ifi	0.007	ifi	0.004	ifi	0.003
rni	0.069	rni	0.010	rni	0.007	rni	0.004	rni	0.003
rmsea	0.000	rmsea	0.025	rmsea	0.013	rmsea	0.006	rmsea	0.004
rmr	0.008	rmr	0.004	rmr	0.003	rmr	0.002	rmr	0.001
srmr	0.016	srmr	0.009	srmr	0.006	srmr	0.004	srmr	0.003
gfi	0.001	gfi	0.000	gfi	0.000	gfi	0.000	gfi	0.000
agfi	0.002	agfi	0.001	agfi	0.000	agfi	0.000	agfi	0.000
pgfi	0.001	pgfi	0.000	pgfi	0.000	pgfi	0.000	pgfi	0.000
mfi	4.194	mfi	0.205	mfi	0.096	mfi	0.046	mfi	0.029
IF-SKD.Model.Free		IF-SKD.Model.Free		IF-SKD.Model.Free		IF-SKD.Model.Free		IF-SKD.Model.Free	
sd.fit.corrected (N=23)		sd.fit.corrected (N=100)		sd.fit.corrected (N=200)		sd.fit.corrected (N=500)		sd.fit.corrected (N=1000)	
fmin	0.598	fmin	0.220	fmin	0.130	fmin	0.077	fmin	0.053
chisq	27.507	chisq	44.048	chisq	52.049	chisq	77.400	chisq	105.417
pvalue	0.026	pvalue	0.402	pvalue	0.004	pvalue	0.000	pvalue	0.000
cfi	0.000	cfi	0.007	cfi	0.007	cfi	0.004	cfi	0.003
tli	0.083	tli	0.012	tli	0.008	tli	0.005	tli	0.003

nnfi	0.083	nnfi	0.012	nnfi	0.008	nnfi	0.005	nnfi	0.003
rfi	0.078	rfi	0.015	rfi	0.009	rfi	0.005	rfi	0.004
nfi	0.068	nfi	0.013	nfi	0.008	nfi	0.005	nfi	0.003
pnfi	0.059	pnfi	0.012	pnfi	0.007	pnfi	0.004	pnfi	0.003
ifi	0.065	ifi	0.010	ifi	0.007	ifi	0.004	ifi	0.003
rni	0.072	rni	0.010	rni	0.007	rni	0.004	rni	0.003
rmsea	0.000	rmsea	0.025	rmsea	0.012	rmsea	0.005	rmsea	0.003
rmr	0.008	rmr	0.004	rmr	0.003	rmr	0.002	rmr	0.001
srmr	0.016	srmr	0.010	srmr	0.006	srmr	0.004	srmr	0.003
gfi	0.001	gfi	0.000	gfi	0.000	gfi	0.000	gfi	0.000
agfi	0.002	agfi	0.001	agfi	0.000	agfi	0.000	agfi	0.000
pgfi	0.001	pgfi	0.000	pgfi	0.000	pgfi	0.000	pgfi	0.000
mfi	4.481	mfi	0.214	mfi	0.091	mfi	0.044	mfi	0.028
Single.Factor.Model		Single.Factor.Model		Single.Factor.Model		Single.Factor.Model		Single.Factor.Model	
sd.fit.corrected (N=23)		sd.fit.corrected (N=100)		sd.fit.corrected (N=200)		sd.fit.corrected (N=500)		sd.fit.corrected (N=1000)	
fmin	0.623	fmin	0.395	fmin	0.235	fmin	0.116	fmin	0.066
chisq	28.646	chisq	78.917	chisq	93.917	chisq	115.865	chisq	132.928
pvalue	0.065	pvalue	0.000	pvalue	0.000	pvalue	0.000	pvalue	0.000
cfi	0.000	cfi	0.027	cfi	0.016	cfi	0.007	cfi	0.004
tli	0.075	tli	0.030	tli	0.018	tli	0.008	tli	0.005
nnfi	0.075	nnfi	0.030	nnfi	0.018	nnfi	0.008	nnfi	0.005
rfi	0.074	rfi	0.034	rfi	0.019	rfi	0.009	rfi	0.005
nfi	0.067	nfi	0.030	nfi	0.017	nfi	0.008	nfi	0.004
pnfi	0.060	pnfi	0.027	pnfi	0.015	pnfi	0.007	pnfi	0.004
ifi	0.062	ifi	0.027	ifi	0.016	ifi	0.007	ifi	0.004
rni	0.067	rni	0.027	rni	0.016	rni	0.007	rni	0.004
rmsea	0.000	rmsea	0.020	rmsea	0.012	rmsea	0.006	rmsea	0.004
rmr	0.008	rmr	0.006	rmr	0.004	rmr	0.002	rmr	0.001
srmr	0.016	srmr	0.012	srmr	0.009	srmr	0.005	srmr	0.003
gfi	0.001	gfi	0.001	gfi	0.000	gfi	0.000	gfi	0.000
agfi	0.001	agfi	0.001	agfi	0.001	agfi	0.000	agfi	0.000
pgfi	0.001	pgfi	0.001	pgfi	0.000	pgfi	0.000	pgfi	0.000
mfi	4.728	mfi	0.133	mfi	0.086	mfi	0.045	mfi	0.029

McCay.Peet.Model.0		McCay.Peet.Model.0		McCay.Peet.Model.0		McCay.Peet.Model.0		McCay.Peet.Model.0	
median.fit (N=23)		median.fit (N=100)		median.fit (N=200)		median.fit (N=500)		median.fit (N=1000)	
fmin	15.341	fmin	19.481	fmin	18.950	fmin	18.462	fmin	18.293
chisq	705.688	chisq	3896.271	chisq	7579.874	chisq	18462.106	chisq	36586.098
pvalue	0.000	pvalue	0.000	pvalue	0.000	pvalue	0.000	pvalue	0.000
cfi	0.219	cfi	0.192	cfi	0.191	cfi	0.190	cfi	0.189
tli	0.132	tli	0.103	tli	0.101	tli	0.100	tli	0.099
nnfi	0.132	nnfi	0.103	nnfi	0.101	nnfi	0.100	nnfi	0.099
rfi	0.100	rfi	0.098	rfi	0.099	rfi	0.099	rfi	0.099
nfi	0.190	nfi	0.188	nfi	0.189	nfi	0.189	nfi	0.189
pnfi	0.171	pnfi	0.170	pnfi	0.170	pnfi	0.170	pnfi	0.170
ifi	0.244	ifi	0.196	ifi	0.193	ifi	0.191	ifi	0.190
rni	0.219	rni	0.192	rni	0.191	rni	0.190	rni	0.189
rmsea	0.353	rmsea	0.445	rmsea	0.443	rmsea	0.440	rmsea	0.439
rmr	0.156	rmr	0.210	rmr	0.208	rmr	0.208	rmr	0.208
srmr	0.345	srmr	0.434	srmr	0.432	srmr	0.431	srmr	0.431
gfi	0.963	gfi	0.951	gfi	0.952	gfi	0.953	gfi	0.954
agfi	0.950	agfi	0.935	agfi	0.937	agfi	0.938	agfi	0.938
pgfi	0.722	pgfi	0.713	pgfi	0.714	pgfi	0.715	pgfi	0.715
mfi	0.000	mfi	0.000	mfi	0.000	mfi	0.000	mfi	0.000
IF-SKD.Model.0		IF-SKD.Model.0		IF-SKD.Model.0		IF-SKD.Model.0		IF-SKD.Model.0	
median.fit (N=23)		median.fit (N=100)		median.fit (N=200)		median.fit (N=500)		median.fit (N=1000)	
fmin	14.734	fmin	18.194	fmin	17.649	fmin	17.189	fmin	17.068
chisq	677.753	chisq	3638.709	chisq	7059.600	chisq	17189.276	chisq	34136.100
pvalue	0.000	pvalue	0.000	pvalue	0.000	pvalue	0.000	pvalue	0.000
cfi	0.292	cfi	0.247	cfi	0.245	cfi	0.243	cfi	0.243
tli	0.213	tli	0.164	tli	0.161	tli	0.159	tli	0.159
nnfi	0.213	nnfi	0.164	nnfi	0.161	nnfi	0.159	nnfi	0.159
rfi	0.162	rfi	0.156	rfi	0.157	rfi	0.157	rfi	0.158
nfi	0.245	nfi	0.241	nfi	0.242	nfi	0.242	nfi	0.242
pnfi	0.221	pnfi	0.217	pnfi	0.217	pnfi	0.217	pnfi	0.218
ifi	0.312	ifi	0.251	ifi	0.246	ifi	0.244	ifi	0.243
rni	0.292	rni	0.247	rni	0.245	rni	0.243	rni	0.243
rmsea	0.343	rmsea	0.429	rmsea	0.427	rmsea	0.425	rmsea	0.424
rmr	0.153	rmr	0.201	rmr	0.202	rmr	0.201	rmr	0.201
srmr	0.333	srmr	0.418	srmr	0.419	srmr	0.416	srmr	0.416
gfi	0.964	gfi	0.955	gfi	0.956	gfi	0.956	gfi	0.957
agfi	0.951	agfi	0.939	agfi	0.941	agfi	0.942	agfi	0.942
pgfi	0.723	pgfi	0.716	pgfi	0.717	pgfi	0.717	pgfi	0.717
mfi	0.000	mfi	0.000	mfi	0.000	mfi	0.000	mfi	0.000

McCay.Peet.Model.0		McCay.Peet.Model.0		McCay.Peet.Model.0		McCay.Peet.Model.0		McCay.Peet.Model.0	
sd.fit.corrected (N=23)		sd.fit.corrected (N=100)		sd.fit.corrected (N=200)		sd.fit.corrected (N=500)		sd.fit.corrected (N=1000)	
fmin	4.662	fmin	2.732	fmin	1.957	fmin	1.247	fmin	0.822
chisq	214.468	chisq	546.415	chisq	782.657	chisq	1247.067	chisq	1643.030
pvalue	0.000	pvalue	0.000	pvalue	0.000	pvalue	0.000	pvalue	0.000
cfi	0.027	cfi	0.007	cfi	0.005	cfi	0.003	cfi	0.002
tli	0.030	tli	0.008	tli	0.006	tli	0.004	tli	0.002
nnfi	0.030	nnfi	0.008	nnfi	0.006	nnfi	0.004	nnfi	0.002
rfi	0.018	rfi	0.008	rfi	0.005	rfi	0.004	rfi	0.002
nfi	0.016	nfi	0.007	nfi	0.005	nfi	0.003	nfi	0.002
pnfi	0.014	pnfi	0.006	pnfi	0.004	pnfi	0.003	pnfi	0.002
ifi	0.034	ifi	0.007	ifi	0.005	ifi	0.003	ifi	0.002
rni	0.027	rni	0.007	rni	0.005	rni	0.003	rni	0.002
rmsea	0.072	rmsea	0.033	rmsea	0.023	rmsea	0.015	rmsea	0.010
rmr	0.053	rmr	0.029	rmr	0.020	rmr	0.012	rmr	0.009
srmr	0.053	srmr	0.028	srmr	0.019	srmr	0.012	srmr	0.008
gfi	0.018	gfi	0.009	gfi	0.006	gfi	0.004	gfi	0.003
agfi	0.023	agfi	0.012	agfi	0.009	agfi	0.005	agfi	0.004
pgfi	0.013	pgfi	0.007	pgfi	0.005	pgfi	0.003	pgfi	0.002
mfi	0.003	mfi	0.000	mfi	0.000	mfi	0.000	mfi	0.000

IF-SKD.Model.0		IF-SKD.Model.0		IF-SKD.Model.0		IF-SKD.Model.0		IF-SKD.Model.0	
sd.fit.corrected (N=23)		sd.fit.corrected (N=100)		sd.fit.corrected (N=200)		sd.fit.corrected (N=500)		sd.fit.corrected (N=1000)	
fmin	4.318	fmin	2.673	fmin	1.782	fmin	1.120	fmin	0.729
chisq	198.638	chisq	534.613	chisq	712.640	chisq	1119.844	chisq	1458.576
pvalue	0.000	pvalue	0.000	pvalue	0.000	pvalue	0.000	pvalue	0.000
cfi	0.034	cfi	0.009	cfi	0.006	cfi	0.004	cfi	0.003
tli	0.037	tli	0.010	tli	0.007	tli	0.004	tli	0.003
nnfi	0.037	nnfi	0.010	nnfi	0.007	nnfi	0.004	nnfi	0.003
rfi	0.025	rfi	0.009	rfi	0.007	rfi	0.004	rfi	0.003
nfi	0.022	nfi	0.008	nfi	0.006	nfi	0.004	nfi	0.002
pnfi	0.020	pnfi	0.007	pnfi	0.005	pnfi	0.003	pnfi	0.002
ifi	0.038	ifi	0.009	ifi	0.006	ifi	0.004	ifi	0.003
rni	0.034	rni	0.009	rni	0.006	rni	0.004	rni	0.003
rmsea	0.070	rmsea	0.033	rmsea	0.022	rmsea	0.014	rmsea	0.009
rmr	0.052	rmr	0.027	rmr	0.018	rmr	0.011	rmr	0.008
srmr	0.051	srmr	0.026	srmr	0.018	srmr	0.011	srmr	0.008
gfi	0.017	gfi	0.009	gfi	0.006	gfi	0.004	gfi	0.002
agfi	0.022	agfi	0.012	agfi	0.008	agfi	0.005	agfi	0.003
pgfi	0.012	pgfi	0.007	pgfi	0.004	pgfi	0.003	pgfi	0.002
mfi	0.008	mfi	0.000	mfi	0.000	mfi	0.000	mfi	0.000

McCay.Peet.Model.Items.Fixed (N=23)		McCay.Peet.Model.Items.Fixed (N=100)		McCay.Peet.Model.Items.Fixed (N=200)		McCay.Peet.Model.Items.Fixed (N=500)		McCay.Peet.Model.Items.Fixed (N=500)	
sd.fit.corrected (N=23)		sd.fit.corrected (N=100)		sd.fit.corrected (N=200)		sd.fit.corrected (N=500)		sd.fit.corrected (N=1000)	
fmin	0.740	fmin	0.221	fmin	0.145	fmin	0.080	fmin	0.056
chisq	34.027	chisq	44.151	chisq	58.199	chisq	80.106	chisq	112.405
pvalue	0.096	pvalue	0.347	pvalue	0.000	pvalue	0.000	pvalue	0.000
cfi	0.001	cfi	0.009	cfi	0.008	cfi	0.005	cfi	0.003
tli	0.082	tli	0.011	tli	0.008	tli	0.005	tli	0.003
nnfi	0.082	nnfi	0.011	nnfi	0.008	nnfi	0.005	nnfi	0.003
rfi	0.082	rfi	0.015	rfi	0.010	rfi	0.005	rfi	0.004
nfi	0.078	nfi	0.014	nfi	0.009	nfi	0.005	nfi	0.003
pnfi	0.074	pnfi	0.014	pnfi	0.009	pnfi	0.005	pnfi	0.003
ifi	0.075	ifi	0.011	ifi	0.008	ifi	0.005	ifi	0.003
rni	0.078	rni	0.011	rni	0.008	rni	0.005	rni	0.003
rmsea	0.004	rmsea	0.026	rmsea	0.011	rmsea	0.005	rmsea	0.003
rmr	0.008	rmr	0.003	rmr	0.002	rmr	0.001	rmr	0.001
srmr	0.020	srmr	0.009	srmr	0.006	srmr	0.004	srmr	0.003
gfi	0.001	gfi	0.000	gfi	0.000	gfi	0.000	gfi	0.000
agfi	0.001	agfi	0.001	agfi	0.000	agfi	0.000	agfi	0.000
pgfi	0.001	pgfi	0.000	pgfi	0.000	pgfi	0.000	pgfi	0.000
mfi	5.959	mfi	0.191	mfi	0.089	mfi	0.040	mfi	0.026

IF-SKD.Model.Items.Fixed (N=23)		IF-SKD.Model.Items.Fixed (N=100)		IF-SKD.Model.Items.Fixed (N=200)		IF-SKD.Model.Items.Fixed (N=500)		IF-SKD.Model.Items.Fixed	
sd.fit.corrected (N=23)		sd.fit.corrected (N=100)		sd.fit.corrected (N=200)		sd.fit.corrected (N=500)		sd.fit.corrected (N=1000)	
fmin	0.687	fmin	0.243	fmin	0.151	fmin	0.087	fmin	0.058
chisq	31.622	chisq	48.631	chisq	60.271	chisq	87.459	chisq	115.154
pvalue	0.051	pvalue	0.357	pvalue	0.000	pvalue	0.000	pvalue	0.000
cfi	0.000	cfi	0.010	cfi	0.008	cfi	0.005	cfi	0.003
tli	0.070	tli	0.012	tli	0.008	tli	0.005	tli	0.004
nnfi	0.070	nnfi	0.012	nnfi	0.008	nnfi	0.005	nnfi	0.004
rfi	0.075	rfi	0.016	rfi	0.009	rfi	0.006	rfi	0.004
nfi	0.073	nfi	0.015	nfi	0.009	nfi	0.006	nfi	0.004
pnfi	0.071	pnfi	0.015	pnfi	0.009	pnfi	0.005	pnfi	0.004
ifi	0.067	ifi	0.012	ifi	0.008	ifi	0.005	ifi	0.003
rni	0.068	rni	0.012	rni	0.008	rni	0.005	rni	0.003
rmsea	0.000	rmsea	0.027	rmsea	0.011	rmsea	0.005	rmsea	0.003
rmr	0.008	rmr	0.003	rmr	0.002	rmr	0.001	rmr	0.001
srmr	0.018	srmr	0.010	srmr	0.006	srmr	0.004	srmr	0.003

gfi	0.001	gfi	0.000	gfi	0.000	gfi	0.000	gfi	0.000
agfi	0.001	agfi	0.001	agfi	0.000	agfi	0.000	agfi	0.000
pgfi	0.001	pgfi	0.000	pgfi	0.000	pgfi	0.000	pgfi	0.000
mfi	6.156	mfi	0.208	mfi	0.093	mfi	0.043	mfi	0.026
Single.Factor.Model.Items.Fixed (N=23)		Single.Factor.Model.Items.Fixed (N=100)		Single.Factor.Model.Items.Fixed (N=200)		Single.Factor.Model.Items.Fixed (N=500)		Single.Factor.Model.Items.Fixed (N=500)	
sd.fit.corrected (N=23)		sd.fit.corrected (N=100)		sd.fit.corrected (N=200)		sd.fit.corrected (N=500)		sd.fit.corrected (N=1000)	
fmin	0.778	fmin	0.246	fmin	0.154	fmin	0.089	fmin	0.061
chisq	35.797	chisq	49.182	chisq	61.435	chisq	88.746	chisq	121.132
pvalue	0.088	pvalue	0.350	pvalue	0.000	pvalue	0.000	pvalue	0.000
cfi	0.000	cfi	0.010	cfi	0.008	cfi	0.005	cfi	0.004
tli	0.077	tli	0.012	tli	0.008	tli	0.005	tli	0.004
nnfi	0.077	nnfi	0.012	nnfi	0.008	nnfi	0.005	nnfi	0.004
rfi	0.080	rfi	0.015	rfi	0.009	rfi	0.005	rfi	0.004
nfi	0.080	nfi	0.015	nfi	0.009	nfi	0.005	nfi	0.004
pnfi	0.080	pnfi	0.015	pnfi	0.009	pnfi	0.005	pnfi	0.004
ifi	0.077	ifi	0.012	ifi	0.008	ifi	0.005	ifi	0.004
rni	0.077	rni	0.012	rni	0.008	rni	0.005	rni	0.004
rmsea	0.000	rmsea	0.027	rmsea	0.010	rmsea	0.005	rmsea	0.003
rmr	0.008	rmr	0.003	rmr	0.002	rmr	0.001	rmr	0.001
srmr	0.020	srmr	0.010	srmr	0.006	srmr	0.004	srmr	0.003
gfi	0.001	gfi	0.000	gfi	0.000	gfi	0.000	gfi	0.000
agfi	0.001	agfi	0.001	agfi	0.000	agfi	0.000	agfi	0.000
pgfi	0.001	pgfi	0.000	pgfi	0.000	pgfi	0.000	pgfi	0.000
mfi	7.255	mfi	0.208	mfi	0.089	mfi	0.042	mfi	0.027

CI.95.fit McCay.Peet.Free (N=1000)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.5839225	1167.845	0	0.9661845	0.9603281	0.9603281	0.9556230	0.9621739	0.8201387	0.9662085	0.9661845	0.0743627	0.0381757	0.0784858	0.9979723	0.9971454	0.7088771	0.4847509
97.5%	0.8128960	1625.792	0	0.9786383	0.9749588	0.9749588	0.9706398	0.9749740	0.8310492	0.9786519	0.9786383	0.0899485	0.0434320	0.0901105	0.9984866	0.9978694	0.7092425	0.6096212
CI.95.fit McCay.Peet.Free (N=500)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.5688737	568.8737	0	0.9666300	0.9608509	0.9608509	0.9511601	0.9583698	0.8168961	0.9666780	0.9666300	0.0660671	0.0374333	0.0766380	0.9978061	0.9969114	0.7087591	0.4986130
97.5%	0.8735332	873.5332	0	0.9836976	0.9808742	0.9808742	0.9725282	0.9765836	0.8324212	0.9837173	0.9836976	0.0881799	0.0448082	0.0931378	0.9985452	0.9979520	0.7092841	0.6766135
CI.95.fit McCay.Peet.Free (N=200)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.5354759	213.3904	0.0000000	0.9694393	0.9641467	0.9641467	0.9390571	0.9480535	0.8081027	0.9695400	0.9694393	0.0310717	0.0367905	0.0752336	0.9973870	0.9963213	0.7084614	0.5354642
97.5%	1.0689990	427.5996	0.0402837	0.9965203	0.9959177	0.9959177	0.9756373	0.9792337	0.8346801	0.9965332	0.9965203	0.0835405	0.0482338	0.1006407	0.9986066	0.9980384	0.7093277	0.9172200
CI.95.fit McCay.Peet.Free (N=100)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.5437595	108.7519	0.0000026	0.9747153	0.9703366	0.9703366	0.9170604	0.9293039	0.7921209	0.9749205	0.9747153	0.0000000	0.0374058	0.0752790	0.9966906	0.9953409	0.7079667	0.6045119
97.5%	1.3933007	278.6601	0.9999921	1.0000000	1.0178329	1.0178329	0.9761297	0.9796534	0.8350379	1.0150898	1.0152004	0.0749923	0.0533532	0.1126731	0.9985980	0.9980262	0.7093216	1.4258790
CI.95.fit McCay.Peet.Free (N=23)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	1.171109	53.87099	0.6933569	1	1.021116	1.021116	0.6461726	0.6984042	0.5953065	1.016885	1.017999	0	0.0396465	0.0955407	0.9924752	0.9894064	0.7049725	1.256869
97.5%	3.672842	168.95073	1.0000000	1	1.408859	1.408859	0.9518511	0.9589588	0.8173982	1.317730	1.348304	0	0.0726523	0.1596757	0.9968421	0.9955543	0.7080744	17.181638
CI.95.fit IFSKD.Model.Free (N=1000)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.6140313	1228.063	0	0.9653585	0.9602474	0.9602474	0.9553669	0.9611055	0.8375348	0.9653817	0.9653585	0.0756071	0.0389667	0.0796957	0.9978858	0.9970866	0.7246552	0.4724089
97.5%	0.8406604	1681.321	0	0.9777121	0.9744237	0.9744237	0.9701946	0.9740267	0.8487947	0.9777232	0.9777121	0.0905303	0.0440101	0.0912977	0.9984219	0.9978269	0.7250445	0.5927073
CI.95.fit IFSKD.Model.Free (N=500)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.5908662	590.8662	0	0.9649433	0.9597710	0.9597710	0.9500569	0.9564781	0.8335024	0.9649864	0.9649433	0.0668318	0.0383250	0.0779978	0.9977469	0.9968974	0.7245543	0.4824769
97.5%	0.9103647	910.3647	0	0.9829184	0.9803982	0.9803982	0.9721505	0.9757312	0.8502800	0.9829370	0.9829184	0.0892484	0.0455764	0.0944051	0.9984952	0.9979278	0.7250977	0.6645245
CI.95.fit IFSKD.Model.Free (N=200)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.5646992	225.8797	0.0000000	0.9669341	0.9620555	0.9620555	0.9365914	0.9447439	0.8232768	0.9670503	0.9669341	0.0343142	0.0373569	0.0769201	0.9973507	0.9963518	0.7242666	0.5253715
97.5%	1.0979516	439.1726	0.0169557	0.9956109	0.9949633	0.9949633	0.9740679	0.9774020	0.8517360	0.9956241	0.9956109	0.0838715	0.0486927	0.1018060	0.9985355	0.9979833	0.7251270	0.8978629
CI.95.fit IFSKD.Model.Free (N=100)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.5634622	112.6924	0.0000008	0.9723997	0.968357	0.968357	0.9131904	0.9260943	0.8070232	0.9728315	0.9723997	0.0000000	0.0375977	0.0756991	0.9966330	0.9953662	0.7237468	0.3822619
97.5%	1.4504273	290.0855	0.9999895	1.0000000	1.016348	1.016348	0.9769430	0.9799074	0.8539193	1.0141723	1.0142458	0.0768815	0.0540522	0.1139623	0.9985441	0.9979951	0.7251332	1.4263076
CI.95.fit IFSKD.Model.Free (N=23)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	1.171287	53.87921	0.8743569	1	1.038816	1.038816	0.6205263	0.6691415	0.5831090	1.032201	1.033826	0	0.0396639	0.0970329	0.9923937	0.9895258	0.7206669	1.637423
97.5%	3.506577	161.30256	1.0000000	1	1.417573	1.417573	0.9504756	0.9568431	0.8338204	1.323340	1.363885	0	0.0725418	0.1586445	0.9969274	0.9957688	0.7239592	18.815300
CI.95.fit Single.Factor.Model.Free (N=1000)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.7964688	1592.938	0	0.9508632	0.9454036	0.9454036	0.9404497	0.9464048	0.8517643	0.9508884	0.9508632	0.0862304	0.0426278	0.0896374	0.9973811	0.9965081	0.7480358	0.3756411
97.5%	1.0726419	2145.284	0	0.9685705	0.9650783	0.9650783	0.9604736	0.9644262	0.8679836	0.9685850	0.9685705	0.1017893	0.0489235	0.1027833	0.9980206	0.9973607	0.7485154	0.4952605
CI.95.fit Single.Factor.Model.Free (N=500)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.9237809	923.7809	0	0.9380477	0.9311641	0.9311641	0.9211577	0.9290419	0.8361377	0.9381107	0.9380477	0.0882669	0.0445068	0.0946220	0.9965823	0.9954430	0.7474367	0.2990606
97.5%	1.3936950	1393.6950	0	0.9682089	0.9646765	0.9646765	0.9559660	0.9603694	0.8643124	0.9682385	0.9682089	0.1130205	0.0545411	0.1154316	0.9977167	0.9969555	0.7482875	0.4789049
CI.95.fit Single.Factor.Model.Free (N=200)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	1.048138	419.2552	0	0.9063550	0.8959500	0.8959500	0.8664059	0.8815653	0.7934088	0.9066275	0.9063550	0.0782433	0.0450671	0.1012306	0.9955377	0.9940503	0.7466533	0.2131297
97.5%	2.010625	804.2500	0	0.9724894	0.9694327	0.9694327	0.9469300	0.9522370	0.8570133	0.9725612	0.9724894	0.1278994	0.0593947	0.1368020	0.9975208	0.9966944	0.7481406	0.3607218
CI.95.fit Single.Factor.Model.Free (N=100)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	1.361686	272.3372	0.0000000	0.8660283	0.8511426	0.8511426	0.7935088	0.8142300	0.7328070	0.8669448	0.8660283	0.0667376	0.0478078	0.1115498	0.9935695	0.9914261	0.7451772	0.1274752
97.5%	2.984235	596.8471	0.0000683	0.9805966	0.9784407	0.9784407	0.9315795	0.9384216	0.8445794	0.9806963	0.9805966	0.1476387	0.0725527	0.1624156	0.9969663	0.9959551	0.7477248	0.6564592
CI.95.fit Single.Factor.Model.Free (N=23)																			
	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	1.271682	58.49738	0.7697153	1	1.033950	1.033950	0.6438849	0.6794964	0.6115467	1.029689	1.030555	0	0.0403697	0.0991426	0.9921835	0.9895780	0.7441376	1.398329
97.5%	3.790764	174.37516	1.0000000	1	1.398901	1.398901	0.9515419	0.9563877	0.8607489	1.324502	1.359011	0	0.0752099	0.1628447	0.9966332	0.9955109	0.7474749	19.413500

CI.95.fit McCay.Peet.zero (N=1000)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	16.57708	33154.17	0	0.1850918	0.0945464	0.0945464	0.0941554	0.1847398	0.1662659	0.1854470	0.1850918	0.4178438	0.1900715	0.4127640	0.9475988	0.9301317	0.7106991	0.0000000
97.5%	20.02267	40045.34	0	0.1944439	0.1049377	0.1049377	0.1044363	0.1939926	0.1745934	0.1948124	0.1944439	0.4594466	0.2266235	0.4498773	0.9592377	0.9456503	0.7194283	0.0000001

CI.95.fit McCay.Peet.zero (N=500)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	16.01605	16016.05	0	0.1836268	0.0929186	0.0929186	0.0920113	0.1828102	0.1645292	0.1843174	0.1836268	0.4096555	0.1845290	0.4067503	0.9446948	0.9262598	0.7085211	0.0000000
97.5%	21.09407	21094.07	0	0.1974915	0.1083238	0.1083238	0.1072139	0.1964925	0.1768433	0.1982036	0.1974915	0.4708088	0.2348956	0.4570199	0.9611190	0.9481587	0.7208393	0.0000001

CI.95.fit McCay.Peet.zero (N=200)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	15.30672	6122.686	0	0.1802936	0.0892151	0.0892151	0.0871492	0.1784343	0.1605909	0.1823073	0.1802936	0.3971960	0.1715192	0.3905750	0.9370479	0.9160638	0.7027859	0.0000000
97.5%	23.48085	9392.339	0	0.2016444	0.1129382	0.1129382	0.1102323	0.1992091	0.1792882	0.2037300	0.2016444	0.4946697	0.2534044	0.4737984	0.9640215	0.9520287	0.7230161	0.0000003

CI.95.fit McCay.Peet.zero (N=100)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	14.62334	2924.668	0	0.1779210	0.0865789	0.0865789	0.0823551	0.1741196	0.1567076	0.1820366	0.1779210	0.3823694	0.1585986	0.3780194	0.9283264	0.9044353	0.6962448	0.000000
97.5%	25.66251	5132.501	0	0.2084885	0.1205428	0.1205428	0.1152131	0.2036918	0.1833226	0.2122886	0.2084885	0.5140067	0.2784647	0.4906196	0.9668698	0.9558264	0.7251524	0.000001

CI.95.fit McCay.Peet.zero (N=23)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	8.272771	380.5475	0	0.1753771	0.0837524	0.0837524	0.0639402	0.1575462	0.1417916	0.1951791	0.1753771	0.2146327	0.0800428	0.2421437	0.9155362	0.8873816	0.6866521	0.0000000
97.5%	26.814994	1233.4897	0	0.2933416	0.2148240	0.2148240	0.1369296	0.2232367	0.2009130	0.3368020	0.2933416	0.5011985	0.2952057	0.4498982	0.9854397	0.9805862	0.7390797	0.0128637

CI.95.fit IFSKD.Model.zero (N=1000)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	15.55100	31102.00	0	0.2372146	0.1524606	0.1524606	0.1517582	0.2365823	0.2129241	0.2375685	0.2372146	0.4046289	0.1838434	0.3993175	0.9511841	0.9349121	0.7133881	0.0000000
97.5%	18.59472	37189.43	0	0.2478533	0.1642815	0.1642815	0.1635141	0.2471627	0.2224464	0.2482081	0.2478533	0.4426798	0.2186224	0.4334722	0.9616412	0.9488549	0.7212309	0.0000002

CI.95.fit IFSKD.Model.zero (N=500)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	14.98559	14985.59	0	0.2334828	0.1305365	0.1305365	0.1491704	0.2342333	0.210828	0.2361914	0.2334828	0.3958273	0.1786469	0.3926730	0.9483339	0.9311119	0.7112504	0.0000000
97.5%	19.64638	19646.38	0	0.2514384	0.1682649	0.1682649	0.1667432	0.2500688	0.223062	0.2520634	0.2514384	0.4542145	0.2255191	0.4399555	0.9634571	0.9512761	0.7223928	0.0000004

CI.95.fit IFSKD.Model.zero (N=200)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	14.18524	3674.097	0	0.2323319	0.1470334	0.1470334	0.1434697	0.2291227	0.2062104	0.2342862	0.2323319	0.3818869	0.1656967	0.3802878	0.9426112	0.9234816	0.7069384	0.000000
97.5%	21.43345	8573.378	0	0.2368837	0.1743152	0.1743152	0.1704750	0.2334273	0.2280847	0.2586518	0.2368837	0.4721478	0.2386548	0.4522032	0.9664960	0.9553280	0.7248720	0.000001

CI.95.fit IFSKD.Model.zero (N=100)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	13.40351	2680.703	0	0.2313235	0.1459150	0.1459150	0.1389231	0.2250308	0.2025277	0.2354004	0.2313235	0.3649216	0.1511815	0.3653251	0.9343670	0.9124894	0.7007753	0.0000000
97.5%	24.08317	4816.634	0	0.2649814	0.1833126	0.1833126	0.1754806	0.2579325	0.2321393	0.2684288	0.2649814	0.4973144	0.2637847	0.4692910	0.9698148	0.9597531	0.7273611	0.0000034

CI.95.fit IFSKD.Model.zero (N=23)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	7.403701	340.5702	0	0.2325271	0.1472523	0.1472523	0.1066143	0.1959528	0.1763576	0.2528429	0.2325271	0.1909246	0.0741061	0.2260703	0.9209792	0.8946389	0.6907344	0.0000000
97.5%	25.017295	1150.7956	0	0.3748565	0.3053961	0.3053961	0.2041732	0.2837559	0.2553803	0.4154950	0.3748565	0.4809487	0.2841529	0.4329132	0.9871026	0.9828034	0.7403269	0.0319214

CI.95.fit McCay.Peet.items.fixed (N=1000)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.7500586	1500.117	0	0.9582391	0.9561511	0.9561511	0.9512337	0.9535559	0.9081485	0.9582494	0.9582391	0.0806666	0.0449505	0.0921211	0.9975317	0.9968899	0.7916918	0.4107022
97.5%	0.9889969	1977.994	0	0.9722601	0.9708731	0.9708731	0.9666698	0.9682570	0.9221495	0.9722657	0.9722601	0.0943338	0.0489750	0.1032061	0.9980610	0.9975569	0.7921119	0.5216756

CI.95.fit McCay.Peet.items.fixed (N=500)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.7338309	733.8309	0	0.9594044	0.9573746	0.9573746	0.9477652	0.9502326	0.9050025	0.9594261	0.9594044	0.0731369	0.0447650	0.0913094	0.9973913	0.9967130	0.7915804	0.4247205
97.5%	1.0546113	1054.6113	0	0.9778400	0.9767320	0.9767320	0.9682621	0.9697734	0.9235937	0.9778494	0.9778400	0.0925578	0.0501171	0.1060713	0.9980821	0.9975835	0.7921287	0.5857275

CI.95.fit McCay.Peet.items.fixed (N=200)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.7165047	286.6019	0.0000000	0.9600445	0.9580467	0.9580467	0.9525558	0.9555769	0.8910257	0.9600963	0.9600445	0.0466468	0.0447052	0.0903962	0.9968407	0.9960192	0.7911434	0.4470410
97.5%	1.3010794	520.4318	0.0000569	0.9912340	0.9907957	0.9907957	0.9709644	0.9723471	0.9260448	0.9912434	0.9912340	0.0897276	0.0538368	0.1166146	0.9980917	0.9975955	0.7921363	0.8044532

CI.95.fit McCay.Peet.items.fixed (N=100)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.7393184	147.8637	0.0000000	0.9653259	0.9635922	0.9635922	0.9094252	0.9137383	0.8702270	0.9654096	0.9653259	0.0000000	0.0459729	0.0905718	0.9961927	0.9952028	0.7906291	0.5162962
97.5%	1.6544641	330.8928	0.9977545	1.0000000	1.0111006	1.0111006	0.9718262	0.9731678	0.9268265	1.0105521	1.0105720	0.0813065	0.0587249	0.1288925	0.9979723	0.9974452	0.7920415	1.3012362

CI.95.fit McCay.Peet.items.fixed (N=23)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	1.275511	58.67352	0.6837902	1	1.019387	1.019387	0.5977667	0.6169207	0.5875435	1.018208	1.018464	0	0.0441371	0.1044141	0.9917443	0.9895979	0.7870987	1.256999
97.5%	4.129063	189.93690	1.0000000	1	1.469383	1.469383	0.9498229	0.9522123	0.9068688	1.431006	1.447032	0	0.0767382	0.1851190	0.9965775	0.9956876	0.7909345	24.827852

CI.95.fit IFSKD.Model.items.fixed (N=1000)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.7636065	1527.213	0	0.9575239	0.9562746	0.9562746	0.9512017	0.9525959	0.9253789	0.9575304	0.9575239	0.0805781	0.0452561	0.0926203	0.9974884	0.9968975	0.8074906	0.4055635
97.5%	1.0035756	2007.151	0	0.9720357	0.9712132	0.9712132	0.9669449	0.9678893	0.9402353	0.9720393	0.9720357	0.0940629	0.0493018	0.1041489	0.9980158	0.9975489	0.8079175	0.5136801

CI.95.fit IFSKD.Model.items.fixed (N=500)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.7504781	750.4781	0	0.9557237	0.9544215	0.9544215	0.9450015	0.9465729	0.9195280	0.9557363	0.9557237	0.0732691	0.0453936	0.0917195	0.9972369	0.9966115	0.8073032	0.3976670
97.5%	1.1242963	1124.2963	0	0.9775229	0.9768618	0.9768618	0.9688162	0.9697072	0.9420013	0.9775286	0.9775229	0.0950820	0.0509243	0.1085151	0.9980197	0.9975537	0.8079207	0.5783516

CI.95.fit IFSKD.Model.items.fixed (N=200)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.7183445	287.3378	0.0000000	0.9599138	0.9587368	0.9587368	0.9524676	0.9543971	0.9077000	0.9599478	0.9599138	0.0453083	0.0451959	0.0902021	0.9967991	0.9960460	0.8069326	0.4436767
97.5%	1.3185959	527.4384	0.0001075	0.9917034	0.9914594	0.9914594	0.9716063	0.9724176	0.9446342	0.9917079	0.9917034	0.0892594	0.0544795	0.1170261	0.9981116	0.9976673	0.8079951	0.8110777

CI.95.fit IFSKD.Model.items.fixed (N=100)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.7369553	147.3911	0.0000000	0.9633608	0.9622832	0.9622832	0.9105805	0.9131354	0.8870458	0.9634118	0.9633608	0.0000000	0.0467007	0.0904856	0.9959772	0.9950307	0.8062673	0.4957338
97.5%	1.7147022	342.9404	0.9989636	1.0000000	1.0113028	1.0113028	0.9729922	0.9737639	0.9459420	1.0109653	1.0109799	0.0829432	0.0599560	0.1304571	0.9979879	0.9975145	0.8078950	1.3309644

CI.95.fit IFSKD.Model.items.fixed (N=23)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	1.374330	63.21918	0.7674051	1	1.030381	1.030381	0.6297632	0.6403414	0.6220459	1.029264	1.029513	0	0.0461560	0.1068288	0.9916401	0.9896731	0.8027563	1.407583
97.5%	4.107822	188.95983	1.0000000	1	1.351446	1.351446	0.9486247	0.9500926	0.9229471	1.331082	1.341405	0	0.0775098	0.1771646	0.9962111	0.9953196	0.8064566	24.521840

CI.95.fit Single.Factor.Model.items.fixed (N=1000)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.8078867	1615.773	0	0.9552823	0.9552823	0.9552823	0.9505798	0.9505798	0.9505798	0.9552823	0.9552823	0.0818588	0.0470382	0.0960122	0.9973656	0.9968364	0.8311364	0.3836679
97.5%	1.0620201	2124.040	0	0.9704645	0.9704645	0.9704645	0.9662079	0.9662079	0.9662079	0.9704645	0.9704645	0.0955175	0.0507868	0.1076968	0.9979097	0.9974917	0.8315915	0.4948056

CI.95.fit Single.Factor.Model.items.fixed (N=500)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.7889514	788.9514	0	0.9546916	0.9546916	0.9546916	0.9448714	0.9448714	0.9448714	0.9546916	0.9546916	0.0743295	0.0469997	0.0949486	0.9971571	0.9965885	0.8309642	0.3885478
97.5%	1.1534483	1153.4483	0	0.9761592	0.9761592	0.9761592	0.9676505	0.9676505	0.9676505	0.9761592	0.9761592	0.0940853	0.0522967	0.1113500	0.9979489	0.9975387	0.8316241	0.5598559

CI.95.fit Single.Factor.Model.items.fixed (N=200)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.7851541	314.0616	0.0000000	0.9571870	0.9571870	0.9571870	0.9522570	0.9522570	0.9522570	0.9571870	0.9571870	0.049901	0.0473300	0.0948628	0.9966210	0.9959452	0.8305175	0.4110860
97.5%	1.4095081	563.8032	0.0000043	0.9895864	0.9895864	0.9895864	0.9699978	0.9699978	0.9699978	0.9895864	0.9895864	0.092012	0.0555352	0.1196028	0.9979412	0.9975294	0.8316176	0.7699256

CI.95.fit Single.Factor.Model.items.fixed (N=100)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	0.7737102	154.7420	0.0000000	0.9637855	0.9637855	0.9637855	0.9110641	0.9110641	0.9110641	0.9637855	0.9637855	0.0000000	0.0483352	0.0935191	0.9959256	0.9951107	0.8299380	0.4918564
97.5%	1.7524729	350.4946	0.9983806	1.0000000	1.0107656	1.0107656	0.9723416	0.9723416	0.9723416	1.0107656	1.0107656	0.0822058	0.0596291	0.1322943	0.9978891	0.9974670	0.8315743	1.3219138

CI.95.fit Single.Factor.Model.items.fixed (N=23)

	fmin	chisq	pvalue	cfi	tli	nnfi	rfi	nfi	pnfi	ifi	rni	rmsea	rmr	srmr	gfi	agfi	pgfi	mfi
2.5%	1.330345	61.19589	0.5588985	1	1.008445	1.008445	0.6349554	0.6349554	0.6349554	1.008445	1.008445	0	0.0474938	0.1077805	0.9913809	0.9896571	0.8261508	1.087329
97.5%	4.485155	206.31711	1.0000000	1	1.376317	1.376317	0.9512627	0.9512627	0.9512627	1.376317	1.376317	0	0.0775527	0.1874711	0.9962202	0.9954642	0.8301835	29.426970

REFERENCES

Agarwal, N. K. (2015). Towards a definition of serendipity in information behaviour. *Information Research, 20*(3), 16.

Arborelius, E., & Timpka, T. (1990). General practitioners' comments on video recorded consultations as an aid to understanding the doctor-patient relationship. *Fam Pract, 7*(2), 84-90.

Bates, M. J. (1989). Design of browsing and berrypicking techniques for online search interface. *Online Review, 13*(5), 407.

Belkin, N. J., Oddy, R. N., & Brooks, H. M. (1982). ASK for information retrieval: Part I. Background and theory. *Journal of Documentation, 38*(2), 61.

Bennett, N. L., Casebeer, L. L., Kristofco, R., & Collins, B. C. (2005). Family physicians' information seeking behaviors: A survey comparison with other specialties. *BMC Med Inform Decis Mak, 5*(1), 9. doi:10.1186/1472-6947-5-9

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychol Bull, 107*(2), 238-246.

Björneborn, L. (2008). Serendipity dimensions and users' information behaviour in the physical library interface. *Information Research, 13*(1).

Box, G. E. P., & Draper, N. R. (1987). *Empirical model-building and response surfaces*. New York: Wiley.

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (Second edition. ed.). New York ; London: The Guilford Press.

Burkell, J., Quan-Haase, A., & Rubin, V. L. (2012). *Promoting serendipity online: recommendations for tool design*. Paper presented at the Proceedings of the 2012 iConference, Toronto, Ontario, Canada. http://dl.acm.org/citation.cfm?doid=2132176.2132274

Case, D. O., & Given, L. M. (2016). *Looking for information: a survey of research on information seeking, needs, and behavior* (Fourth edition. ed.). Bingley, UK: Emerald.

Chen, E. S., Bakken, S., Currie, L. M., Patel, V. L., & Cimino, J. J. (2006). An automated approach to studying health resource and infobutton use. *Stud Health Technol Inform, 122*, 273-278.

Cimino, J. J., Li, J., Bakken, S., & Patel, V. L. (2002). Theoretical, empirical and practical approaches to resolving the unmet information needs of clinical information system users. *Proc AMIA Symp*, 170-174.

Clarke, M. A., Belden, J. L., Koopman, R. J., Steege, L. M., Moore, J. L., Canfield, S. M., & Kim, M. S. (2013). Information needs and information-seeking behaviour analysis of primary care physicians and nurses: a literature review. *Health Information and Libraries Journal, 30*(3), 178-190. doi:10.1111/hir.12036

Costa, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality Disorders. *Journal of Personality Disorders, 6*(4), 343-359. doi:10.1521/pedi.1992.6.4.343

Cunha, Miguel Pina e, Serendipity: Why Some Organizations are Luckier than Others (2005). FEUNL Working Paper No. 472. Available at SSRN: https://ssrn.com/abstract=882782 or http://dx.doi.org/10.2139/ssrn.882782

Currie, L. M., Graham, M., Allen, M., Bakken, S., Patel, V., & Cimino, J. J. (2003). Clinical information needs in context: an observational study of clinicians while using a clinical information system. *AMIA Annu Symp Proc*, 190-194.

Dantonio, L., Makri, S., & Blandford, A. (2012). Coming across academic social media content serendipitously. *Proceedings of the American Society for Information Science and Technology, 49*(1), 1-10. doi:10.1002/meet.14504901002

David, B., Clive, H., & Nigel, C. (1999). Perspectives on information overload. *Aslib Proceedings, 51*(8), 249-255. doi:10.1108/EUM0000000006984

Davies, K., & Harrison, J. (2007). The information-seeking behaviour of doctors: a review of the evidence. *Health Information and Libraries Journal, 24*(2), 78-94. doi:10.1111/j.1471-1842.2007.00713.x

Del Fiol, G., Huser, V., Strasberg, H. R., Maviglia, S. M., Curtis, C., & Cimino, J. J. (2012). Implementations of the HL7 context-aware knowledge retrieval ("infobutton") standard: challenges, strengths, limitations, and uptake. *Journal of biomedical informatics, 45*(4), 726-735. doi:10.1016/j.jbi.2011.12.006

DeVellis, R. F. (2017). *Scale development: Theory and applications* (Fourth edition. ed.). Los Angeles: SAGE.

Erdelez, S. (1997). *Information encountering: a conceptual framework for accidental information discovery*. Paper presented at the Proceedings of an international conference on Information seeking in context, Tampere, Finland.

Erdelez, S. (1999). Information encountering: It's more than just bumping into information. *Bulletin of the American Society for Information Science and Technology, 25*(3), 26-29.

Erdelez, S. (2004). Investigation of information encountering in the controlled research environment. *Information Processing & Management, 40*(6), 1013-1025.

Erdelez, S., Beheshti, J., Heinström, J., Toms, E., Makri, S., Agarwal, N. K., & Björneborn, L. (2016). *Research perspectives on serendipity and information encountering*. Paper presented at the Proceedings of the 79th ASIS&T Annual Meeting: Creating Knowledge, Enhancing Lives through Information & Technology, Copenhagen, Denmark.

Fine, G. A., & Deegan, J. G. (1996). Three principles of Serendip: insight, chance, and discovery in qualitative research. *International Journal of Qualitative Studies in Education, 9*(4), 434-447. doi:10.1080/0951839960090405

Foster, A., & Ford, N. (2003). Serendipity and information seeking: an empirical study. *Journal of Documentation, 59*(3), 321-340.

Goodwin, J. C., Cohen, T., & Rindflesch, T. (2012). *Discovery by scent: Discovery browsing system based on the Information Foraging Theory.* Paper presented at the Bioinformatics and Biomedicine Workshops (BIBMW), 2012 IEEE International Conference.

Gorman, P. (1999). Information seeking of primary care physicians: Conceptual models and empirical studies. In T. D. Wilson & D. K. Allen (Eds.), *Information behaviour: Proceedings of the second international conference on research in information needs, seeking and use in different contexts* (pp. 226-240). Sheefield, UK: Taylor Graham.

Gorman, P. N. (1995). Information needs of physicians. *Journal of the American Society for Information Science, 46*(10), 729-736. doi:10.1002/(SICI)1097-4571(199512)46:10<729::AID-ASI3>3.0.CO;2-2

Gorman, P. N., & Helfand, M. (1995). Information seeking in primary care: how physicians choose which clinical questions to pursue and which to leave unanswered. *Medical Decision Making, 15*(2), 113-119. doi:10.1177/0272989X9501500203

Heeter, C., & Greenberg, B. S. (1985). Profiling the zappers. *Journal of Advertising Research, 25*(2), 15-19.

Heinström, J. (2003). Fast surfers, broad scanners and deep divers as users of information technology – Relating information preferences to personality traits. *Proceedings of the American Society for Information Science and Technology, 40*(1), 247-254. doi:10.1002/meet.1450400131

Heinström, J. (2006). Psychological factors behind incidental information acquisition. *Library & Information Science Research, 28*(4), 579-594. doi:10.1016/j.lisr.2006.03.022

Hider, P. (2006). Search goal revision in models of information retrieval. *Journal of Information Science, 32*(4), 352-361. doi:10.1177/0165551506065811

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1-55. doi:10.1080/10705519909540118

Jiang, T., Zhang, C., Li, Z., Fan, C., & Yang, J. (2018). *Information Encountering on Social Q&A Sites: A Diary Study of the Process.* Paper presented at the 13th International Conference, iConference 2018, Sheffield, UK.

Krinsky, I., & Robb, A. L. (1986). On approximating the statistical properties of elasticities. *The Review of Economics and Statistics, 68*(4), 715-719.

Lawley, J., & Tompkins, P. (2008). *Maximising serendipity: the art of recognising and fostering potential*.

Le, J. V., Pedersen, L. B., Riisgaard, H., Lykkegaard, J., Nexoe, J., Lemmergaard, J., & Sondergaard, J. (2016). Variation in general practitioners' information-seeking behaviour - a cross-sectional study on the influence of gender, age and practice form. *Scand J Prim Health Care, 34*(4), 327-335. doi:10.1080/02813432.2016.1249057

Li, C. H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behav Res Methods, 48*(3), 936-949. doi:10.3758/s13428-015-0619-7

Li, C. H. (2016). The performance of ML, DWLS, and ULS estimation with robust corrections in structural equation models with ordinal variables. *Psychol Methods, 21*(3), 369-387. doi:10.1037/met0000093

Lialiou, P., & Mantas, J. (2016). Online information seeking behaviour by nurses and physicians: A cross-sectional study. *Stud Health Technol Inform, 225*, 33-37.

Makri, S., & Blandford, A. (2012). Coming across information serendipitously - Part 2. *Journal of Documentation, 68*(5), 706-724. doi:10.1108/00220411211256049

Marchionini, G. (1995). *Information seeking in electronic environments* (Vol. 9). Cambridge; New York: Cambridge University Press.

Marcoulides, G. A., & Saunders, C. (2006). Editor's comments: PLS: a silver bullet? *MIS Q., 30*(2), iii-ix.

Marsh, H. W., & Grayson, D. (1995). Latent variable models of multitrait-multimethod data. *Structural equation modeling: Concepts, issues, and applications.* (pp. 177-198). Thousand Oaks, CA, US: Sage Publications, Inc.

Maydeu-Olivares, A. (2017). Assessing the size of model misfit in structural equation models. *Psychometrika*. doi:10.1007/s11336-016-9552-7

McCay-Peet, L. (2013). *Investigating work-related serendipity, what influences it, and how it may be facilitated in digital environments.* (Doctor of Philosophy), Dalhousie University, Halifax, Nova Scotia. Retrieved from https://dalspace.library.dal.ca/handle/10222/42727

McCay-Peet, L., & Toms, E. G. (2010). *The process of serendipity in knowledge work*. Paper presented at the Proceedings of the third symposium on Information interaction in context, New Brunswick, New Jersey, USA. https://dl.acm.org/citation.cfm?doid=1840784.1840842

McCay-Peet, L., & Toms, E. G. (2015). Investigating serendipity: How it unfolds and what may influence it. *Journal of the Association for Information Science and Technology, 66*(7), 1463-1476. doi:10.1002/asi.23273

McCay-Peet, L., Toms, E. G., & Kelloway, K. (2014). Development and assessment of the content validity of a scale to measure how well a digital environment facilitates serendipity. *Information Research, 19*(3), 4-28.

McCay-Peet, L., Toms, E., G, & Kelloway, K. E. (2015). Examination of relationships among serendipity, the environment, and individual differences. *Information Processing & Management, 51*(4), 391-412. doi:10.1016/j.ipm.2015.02.004

McCay-Peet, L., & Wells, P. G. (2017). Serendipity in the Sciences - Exploring the Boundaries. *Proceedings of the Nova Scotian Institute of Science, 49*(1), 97-116.

McDonald, R. P. (1999). *Test theory: a unified treatment*. Mahwah, N.J.: L. Erlbaum Associates.

Mikalef, P., Kourouthanassis, P. E., & Pateli, A. G. (2017). Online information search behaviour of physicians. *Health Information & Libraries Journal, 34*(1), 58-73. doi:10.1111/hir.12170

Miller, C. M., Rindflesch, T. C., Fiszman, M., Hristovski, D., Shin, D., Rosemblat, G., . . . Strohl, K. P. (2012). A closed literature-based discovery technique finds a mechanistic link between hypogonadism and diminished sleep quality in aging men. *Sleep, 35*(2), 279-285. doi:10.5665/sleep.1640

Mooney, C., & Duval, R. (1993). *Bootstrapping.* Retrieved from http://methods.sagepub.com/book/bootstrapping doi:10.4135/9781412983532

Nahl, D. (2004). Measuring the affective information environment of web searchers. *Proceedings of the American Society for Information Science and Technology, 41*(1), 191-197. doi:10.1002/meet.1450410122

Neumaier, A. (1998). Solving ill-conditioned and singular linear systems: a tutorial on regularization. *SIAM Review, 40*(3), 636-666. doi:10.1137/s0036144597321909

Nevitt, J., & Hancock, G. R. (2001). Performance of bootstrapping approaches to model test statistics and parameter standard error estimation in structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal, 8*(3), 353-377. doi:10.1207/S15328007SEM0803_2

Nutefall, J., & Ryder, P. M. (2010). The serendipitous research process. *Journal of Academic Librarianship, 36*(3), 228-234.

Opgen-Rhein, R., & Strimmer, K. (2007). Accurate ranking of differentially expressed genes by a distribution-free shrinkage approach *Statistical Applications in Genetics and Molecular Biology* (Vol. 6).

Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review, 106*(4), 643-675. doi:10.1037/0033-295X.106.4.643

Prasad, S., Gupta, S. C., & Aggarwal, B. B. (2016). Serendipity in cancer drug discovery: Rational or coincidence? *Trends in Pharmacological Sciences, 37*(6), 435-450. doi:10.1016/j.tips.2016.03.004

Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354-373. doi:10.1037/a0029315

Rindflesch, T. C., Kilicoglu, H., Fiszman, M., Rosemblat, G., & Shin, D. (2011). Semantic MEDLINE: An advanced information management application for biomedicine. *Information Services & Use, 31*(1/2), 15-21.

Roberts, R. M. (1989). Serendipity: Accidental discoveries in science. *Serendipity: Accidental Discoveries in Science, by Royston M. Roberts, pp. 288. ISBN 0-471-60203-5. Wiley-VCH, June 1989., 1*.

Schäfer, J., & Strimmer, K. (2005). A shrinkage approach to large-scale covariance matrix estimation and implications for functional genomics. *Statistical Applications in Genetics and Molecular Biology* (Vol. 4).

Solomon, P. (1997). Discovering information behavior in sense making. III. The person. *Journal of the American Society for Information Science, 48*(12), 1127-1138. doi:10.1002/(SICI)1097-4571(199712)48:12<1127::AID-ASI6>3.0.CO;2-W

Spink, A. (2004). Multitasking information behavior and information task switching: an exploratory study. *Journal of Documentation, 60*(4), 336-351.

Stekhoven, D. J., & Buhlmann, P. (2012). MissForest--non-parametric missing value imputation for mixed-type data. *Bioinformatics, 28*(1), 112-118. doi:10.1093/bioinformatics/btr597

Sun, X., Sharples, S., & Makri, S. (2011). A user-centered mobile diary study approach to understanding serendipity in information research. *informationresearch, 16*(3).

Swanson, D. R. (1986). Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspect Biol Med, 30*(1), 7-18.

Swanson, D., R. . (1986). Undiscovered Public Knowledge. *The Library Quarterly, 56*(2), 103-118. doi:10.2307/4307965

Tofighi, D., & MacKinnon, D. P. (2016). Monte carlo confidence intervals for complex functions of indirect effects. *Structural Equation Modeling: A Multidisciplinary Journal, 23*(2), 194-205. doi:10.1080/10705511.2015.1057284

Trizano-Hermosilla, I., & Alvarado, J. M. (2016). Best alternatives to Cronbach's alpha reliability in realistic conditions: Congeneric and Asymmetrical Measurements. *Frontiers in Psychology, 7*(769). doi:10.3389/fpsyg.2016.00769

Wang, C.-P., & Jo, B. (2013). Applications of a Kullback-Leibler divergence for comparing non-nested models. *Statistical Modelling, 13*(5-6), 409-429. doi:10.1177/1471082x13494610

Wilson, P. (1995). Unused relevant information in research and development. *Journal of the American Society for Information Science, 46*(1), 45-51.

Wilson, T. D. (1999). Models in information behaviour research. *Journal of Documentation, 55*(3), 249. doi:10.1108/eum0000000007145

Workman, T. E., Fiszman, M., Cairelli, M. J., Nahl, D., & Rindflesch, T. C. (2016). Spark, an application based on Serendipitous Knowledge Discovery. *Journal of biomedical informatics, 60*, 23-37. doi: 10.1016/j.jbi.2015.12.014

Workman, T. E., Fiszman, M., Rindflesch, T. C., & Nahl, D. (2014). Framing serendipitous information-seeking behavior for facilitating literature-based discovery: A proposed model. *Journal of the Association for Information Science and Technology, 65*(3), 501-512. doi:10.1002/asi.22999

Younger, P. (2010). Internet-based information-seeking behaviour amongst doctors and nurses: a short review of the literature. *Health Information & Libraries Journal, 27*(1), 2-10. doi:10.1111/j.1471-1842.2010.00883.x

Zhang, H., Fiszman, M., Shin, D., Miller, C. M., Rosemblat, G., & Rindflesch, T. C. (2011). Degree centrality for semantic abstraction summarization of therapeutic studies. *Journal of biomedical informatics, 44*(5), 830-838. doi:10.1016/j.jbi.2011.05.001