2016

Five-Factor Model as a Predictor for Spoken Dialog Systems

Teresa G. Carter
Nova Southeastern University, terecart@nova.edu

Five-Factor Model as a Predictor for Spoken Dialog Systems

by
Teresa Carter
terecart@nova.edu

A dissertation report submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in
Information Systems

College of Engineering and Computing
Nova Southeastern University
2016

We hereby certify that this dissertation, submitted by Teresa Carter, conforms to acceptable standards and is fully adequate in scope and quality to fulfill the dissertation requirements for the degree of Doctor of Philosophy.

Maxine S. Cohen, Ph.D. Chairperson of Dissertation Committee	Date
Laurie P. Dringus, Ph.D. Dissertation Committee Member	Date
Steven R. Terrell, Ph.D. Dissertation Committee Member	Date

Approved:

Yong X. Tao, Ph.D., P.E., FASME Dean, College of Engineering and Computing	Date

College of Engineering and Computing

Nova Southeastern University

2016

An Abstract of a Dissertation Report Submitted to Nova Southeastern University
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy

Five-Factor Model as a Predictor for Spoken Dialog Systems

by

Teresa Carter

December 2016

Human behavior varies widely as does the design of spoken dialog systems (SDS). The search for predictors to match a user's preference and efficiency for a specific dialog interface type in an SDS was the focus of this research. By using personality as described by the Five-Factor Method (FFM) and the Wizard of Oz technique for delivering three system initiatives of the SDS, participants interacted with each of the SDS initiatives in scheduling an airline flight. The three system initiatives were constructed as strict system, which did not allow the user control of the interaction; mixed system, which allowed the user some control of the interaction but with a system override; and user system, which allowed the user control of the interaction.

In order to eliminate gender bias in using the FFM as the instrument, participants were matched in gender and age. Participants were 18 years old to 70 years old, passed a hearing test, had no disability that prohibited the use of the SDS, and were native English speakers. Participants completed an adult consent form, a 50-question personality assessment as described by the FFM, and the interaction with the SDS. Participants also completed a system preference indication form at the end of the interaction. Observations for efficiency were recorded on paper by the researcher.

Although the findings did not show a definitive predictor for a SDS due to the small population sample, by using a multinomial regression approach to the statistical analysis, odds ratios of the data helped draw conclusions that support certain personality factors as important roles in a user's preference and efficiency in choosing and using a SDS. This gives an area for future research. Also, the presumption that preference and efficiency always match was not supported by the results from two of the three systems. An additional area for future research was discovered in the gender data. Although not an initial part of the research, the data shows promise in predicting preference and efficiency for certain SDS. Future research is indicated.

Acknowledgments

I cannot express enough thanks to my committee for their continued support and encouragement. Dr. Maxine Cohen, my committee chair, Dr. Laurie Dringus, and Dr. Steven Terrell – to you I offer my sincere appreciation for the learning opportunities that you provided and the sincere interest in my work.

My completion of this dissertation could not have been accomplished without the support of my colleagues who were always there to encourage and advise. To the lab assistants who played the role of wizard, I am truly grateful for all the hours you patiently invested in this study. And to Jammie McCoy, lab and research assistant, thank you for your time and interest in the study but most of all for staying until it was completed.

Finally, to my family who patiently and constantly stayed steadfast while I put in long hours and missed many family functions, I am deeply grateful.

Table of Contents

List of Tables

Tables

List of Figures

Figures

Chapter 1

Introduction

Background

A Spoken Dialog System (SDS) provides an interface between a computer-based application and a user permitting interaction with the application in a natural manner (McTear, 2002). SDS utilizes voice for information transmission for both input and output. Components over and above those of a text dialog system include a speech recognizer and a text to speech module. Most SDS consists of five main components (Wolters, Georgila, Moore, & MacPherson, 2009):

1) Automatic Speech Recognition (ASR) converts audio speech signals into text;

2) Natural Language Understanding (NLU) translates meaning and intention;

3) Dialog Management (DM) controls interaction;

4) Natural Language Generation (NLG) produces text responses; and

5) Text to Speech (TTS) controls speech output.

SDSs are more than just voice dictation systems that only produce a transcript of what the system "interprets" of the user commands but instead, provides an advanced application for speech recognition and analysis as well as processes for language and dialogue (McTear, 2002). With the advent of better technologies, the dream of spoken interaction has gained ground but not to the point of the user fully engaging in a conversation using natural language. Consequently, the interaction has now given way to a problem of higher-level complexity – contextual interpretation of spoken dialog provided by the end user. The spoken words "Yeah right" can have two vastly different meanings depending on the speaker. In the case of an older user, it would convey a confirmation. In contrast, a younger user would consider it a sarcastic negative response (Tepperman, Traum, & Narayanan, 2006).

Developing an interface that matches a user's preference requires a method for identifying user behavior characteristics that guide the SDS design process to a better user experience. Through a Wizard of Oz (WoZ) approach, Wolters et al. (2009) used voice interfaces to study the interaction behavior of users in age ranges of 52 to 84 years for older users and 18 to 29 years of age for younger users. Relevant user groups were delineated by characteristic patterns of behavior (how they interact with the system) based on detailed linguistic analysis of their dialogs.

In this bottom up approach, Wolters, et al. allowed the results to define the user groups. In their analysis of appointment scheduling dialogs, the users broke down into two main groups – social and factual. The "factual" users were able to interact efficiently with the dialog system by rapidly adapting to the interface. The "social" users did not adapt their interaction style because they tended to treat the dialog system like a human. Although "social" users tended to be older, over a third of all older users were in the "factual" group. It appeared that neither cognitive abilities nor gender predicted group membership.

Wolters et al.'s (2009) research concluded that spoken dialog systems should adapt to users based on observed behavior and not on age. In addition, cognitive abilities did not explain any of the remaining variation. However, through Wolters et al.'s (2009) bottom up approach, the only predictor explored was age, and age proved an unreliable predictor of user preference or behavior. This lack of a reliable predictor had been demonstrated by prior researchers as well as Wolters; therefore, leaving the identification and exploration for other differentiators a must. Chickering and Paek (2007) found the need for other differentiators while researching avenues of planning dialog management that accurately takes appropriate actions based on observations and inferred beliefs.

Developing interface designs encompassing user-friendly features, which appeal to a large mass of people, depends on reliable predictions of characteristic patterns of behavior of users in order to design interfaces according to preferences. Determining which characteristic patterns of behavior work best has become an active area of research in the Information Systems field. Wolters et al. (2009) indicated there were three additional candidates for predictors subject to further research - attitude toward technology, personality, and social cognition. Per communication with Maria Wolters (personal communication, March 13, 2012) the question of importance of these three candidates was raised. She explained how these three areas are extremely underrated as potential areas for focused research and expressed great interest in a focused study. Wolters stated that personality as a predictor would be a fertile area for research since little had been done; therefore, with the growing interest in the Five Factor Model (FFM) for personality identification, the use of personality as a predictor would make an interesting study. Macpherson, (personal communication March 23, 2012) singled out personality type as an area that was often overlooked when investigating performance; especially in relationship to cognition and aging. As to SDS, she offered the opinion that personality type would indeed be an area to investigate as it would affect performance especially in the light of a limited number of conducted studies. Whereby setting the stage for research in this area, this study examined one candidate, personality type, as a potential predictor of those characteristics further extending the study of Wolters, Georgila, Moore, and Macpherson (2009).

Problem Statement

The selection of an optimal SDS interface for an individual end user is complex. One size does not fit all. Vendors of human attire from headwear to pants to shoes need to manufacture a wide variety of sizes in order to address variances in the human physical form. A similar problem with “fit” occurs when a human interacts with SDS interfaces. With the use of a SDS, the user can complete a task by simple voice commands either with the device such as a handheld entertainment device, or over distance such as buying an airline ticket over the phone or giving commands via a Bluetooth device. However, voice recognition coupled with technology are proving to be error-prone where system complexity increases and user behavior becomes less predictable (Williams & Young, 2007; M. K. Wolters, V. L. Hanson, & J. D. Moore, 2011). Human behavior varies widely as does the design of SDS interfaces. The initial research problem was - *the selection of an optimal SDS interface for a given individual end user.* The optimal SDS interface for this research was defined as that with the greatest efficiency. Efficiency was defined as minimizing miscues, miscommunications, and other types of errors (McTear, 2002).

Managing the dialog for a SDS took the form of an initiative. The three initiatives used consisted of strict, mixed, and user (Wolters, et al., 2009). Strict initiative was the most common because it was easiest to build. It restricted the user’s control over the input making the user respond only to the options offered by the system. A mixed initiative allowed the user more control while still keeping user responses in a controlled structure. The user initiative operated as its name sounds. It allowed the user as much control over input as could possibly be built into an initiative. With these initiatives, also came confirmations that made up the attributes for an initiative. The confirmations showed how the user responded to the prompts from the system initiative. Naturally, the combination of the attributes, namely initiative and confirmation, that were contained within each SDS interface could vary and were unique. By using three confirmations - explicit, implicit, and none, with three system initiatives – strict initiative, mixed initiative, and user initiative, the resulting nine attributes formed the framework for the dialog types (Table 1).

Table 1

CONFIRMATION AND INITIATIVE			
Confirmation	System initiative dialog	Mixed initiative dialog	User initiative dialog
Explicit	System controlled dialog at each step	User changed dialog direction	System responded to direction of user
Implicit	System controlled dialog at each step	User changed dialog direction	System responded to direction of user
None	System prompted user or ended conversation	System guided user back to original course	System responded to direction of user

As a result, each SDS interface was configured to have multiple “flavors” from the end user perspective. For clarity, the term SDS dialog type was used to describe one of the nine unique flavors (i.e. combination of the three confirmations - explicit, implicit, none, and the three system initiatives - strict, mixed, user) of an SDS interface. Therefore, the scope of the research problem needed to be further narrowed from the concept of a generic SDS interface to that of a more tightly defined and unique SDS dialog type. A more focused research problem could now be stated as – *How does one select an optimal SDS dialog type for a given individual end user?*

Dissertation Goal

As Wolters et al. (2009) indicated, there were three additional candidates for predictors subject to further investigation – attitude to technology, personality, and social cognition. For this research, the candidate predictor of personality was investigated. Personality is defined as a combination of the attributes – behavioral, temperamental, emotional, and mental, that characterized an individual. As an individual's personality was intangible, it was necessary to use an instrument to quantify it.

Several instruments have been designed to classify human personality. The research used the Five-Factor Method (FFM), a taxonomy of personality traits, through the International Personality Item Pool (IPIP) questionnaire in order to measure personality traits (Srivastava, 2012). See Appendix D for more information. The FFM, or sometimes called the Big Five, consisting of a broad range of specific traits derived from statistical factor analyses of co-occurring traits in personality descriptions by the person or of other people making the measurements empirical rather than a theory of personality (Srivastava, 2012). The use of the FFM grew rapidly because of its ability to predict patterns of individual behavior (Costa & McCrae, 1991). The ability to use all or only a portion of the inventory plus the ease of access and free copy right of the instrument makes the FFM an attractive option for research. The FFM also provides a wide range of constructs that can be measured by one or more IPIP questionnaire, giving the researcher the ability to focus on the construct that was most relevant to the research (Goldberg et al., 2006). This coupled with its association of personality traits used in natural language made a good choice for this SDS study.

The goal of this research was – *To determine if personality (as represented by the FFM personality factor) was an effective predictor for selecting the efficient SDS dialog type for a given individual end user.* By allowing the user to experience each of the nine dialog types (three initiatives and three confirmations) when scheduling a plane reservation through the WoZ, an effective predictor for dialog type for individual users was obtained.

Research Questions

Research Question 1 (RQ1): How well did the FFM personality factor serve as a predictor of an individual's preferred SDS dialog type? Preference was investigated by utilizing personality represented by the FFM as a predictor of a user's preference of SDS dialog type.

Research Question 2 (RQ2): How well did the FFM personality factor serve as a predictor of which SDS dialog type an individual used most efficiently? Efficiency was investigated by utilizing personality represented by the FFM as a predictor of a user's efficiency for a SDS dialog type. Efficiency was defined as the amount of miscues, miscommunications, and other types of errors (McTear, 2002).

Research Question 3 (RQ3): Was an individual's preference for an SDS dialog type the same SDS dialog type at which the individual was most efficient in using? The SDS dialog type preferred by a user as compared to efficiency on the SDS was analyzed.

Relevance and Significance

The SDS work was applicable in multiple domains - information provisioning, command and control, tutoring, simulation based training, controlling smart homes, delivering reminders of systems in the health care field in the care of patients with conditions that require monitoring on a continual basis. While research (Czaja & Lee, 2007; Hourcade & Berkel, 2008) took a top down approach which delineated the population by age, Wolters et al. (2009) was significant in this area by taking a bottom up approach which revealed user groups delineated by linguistic analysis instead of by age. Wolters et al. hypothesized that age would be revealed by a unique linguistic difference displayed by definitive age groups. By delineating users into two groups through linguistic analysis, Wolters et al. identified two distinct interactive styles. Interactions were named and defined as (1) factual – where the user treated the interaction as presented with quick adaptation to the situation and (2) social – where the user treated the interaction as another human without adaptation to the situation. These styles show that users can indeed be categorized into groups; that interactive style can be predicted; and that usability of the interface and user interactive style are directly related. While Wolters found the age predictor to be unreliable, the study showed the need for different avenues of delineating users into groups of preferred interactive style. The importance of interaction style of users and how users align with the machine directly impact interface design producing machines that can adapt to the user and producing higher quality interaction in terms of efficiency by reducing miscues, miscommunications, and other types of errors. Providing better predictors of human interaction style will enhance the understanding of how users interact and accept systems.

Barriers and Issues

Two main issues in conducting this type of study were overcoming over-answering, i.e., user providing more information than necessary, and grounding, which is a process where the participants in the group build a consensus of acceptable information, i.e., different users produce different grounding behaviors (Wolters et al. 2009). Barriers were also generated when dialog research approached interface design through dynamic adaptation of user behavior and dialog context (Chickering & Paek, 2007). Using specific user models, although widely accepted as good interface design (Moore et al., 2004) produced limitations where users aligned situation models and adapted their language and speech pattern with the one they were communicating. Although this provided successful conversation and was demonstrated (Niederhoffer & Pennebaker, 2002) through linguistic style in both human-to-human and human-to-computer interaction, it limited true human interaction behavior. Barriers were also produced when users projected emotions and personality onto the computer treating the machine as human (Branigan, Pickering, Pearson, & Mcneil, 2009). All of these models produce limitations in one way or another to developing a more user-friendly interface for SDS.

However, a paradigm of automatically determining dialog strategies derived from data and user simulations helped establish quantifiable research (Georgila, Wolters, Karaiskos, et al., 2008; Lemon, Georgila, Henderson, & Stuttle, 2006). Using these statistical approaches provided for efficient development; automatic optimization; and adaptation of applications already in use for new domains (Wolters, Georgila, Moore, & MacPherson, 2009). Since system input sets were restrictive by nature, user adaptation to the machine was even more important for this study. Using a statistical approach provided a means of measurable results suitable for comparing and contrasting, thereby fostering implications for research and interface design. This approach to simulations, however, has limited research therefore opening the opportunity to expand the body of knowledge in this area.

Assumptions, Limitations and Delimitations

The assumption that participants would answer truthfully in the interaction with the SDS and on the preference form was justified due to the preservation of anonymity and confidentiality by taking strict measures to prevent any personal identification on any of the forms that could be traced back to the participant. The participant was assured that their personal identity would be protected. Also, environmental influences, culture, and biology itself could place limitations on the instrument. A pilot study was conducted in order to assure that the research problem was on target and the design was true to the desired results of data collected. The introduction of SDS into the general population has increased and is expected to continue to be an even larger part of everyday communications.

The limitation of ensuring a random sample is always a consideration. Since the sample was taken from a college campus, the participants represent a global random sampling consisting of faculty, staff, and students from around the world. Therefore, the study results were applied to a larger population which mitigated the limitation.

The choice of problem for this study is in itself a delimitation. In prior research, several other avenues for study were established, however, choosing personality as a predictor narrowed the study and focused the problem statement and goal. In addition, an age range of 18 to 85 years was used, along with requiring the participant to speak native English. The participant had to pass a hearing test to establish an acceptable hearing range with or without hearing aids. The participant also had to physically navigate the building in order to access the lab and use the SDS.

Definitions of Terms

Automatic Speech Recognition (ASR) – One of the five main components in a spoken dialog system. This component converts audio signals of human speech into text strings (Wolters et al., 2009).

Dialog/Dialogue – The conversation that takes place between two human beings, a human being and a device, or two devices (McTear, 2002).

Dialog Interface – This is a way of interacting with your environment by using spoken language (McTear, 2002).

Dialog Management – One of the five main components in a spoken dialog system. This component uses strategies to control dialog in order to achieve the desired results (McTear, 2002).

Factual users – Users who adapt quickly to the system and interact efficiently with them are factual users (Wolters et al., 2009).

Interactive Style – The way in which you interact with a device is your interactive style (Wolters et al., 2009).

Natural Language – A natural language is one that develops naturally rather than a language such as computer code (Sagae, Christian, DeVault, & Traum, 2009).

Natural Language Generation (NLG) – One of the five main components in a spoken dialog system. This component acts as a translator for the computer in order to produce text responses in natural language (McTear, 2002).

Natural Language Understanding (NLU) – One of the five main components in a spoken dialog system, this component determines the meanings and intentions of the recognized utterances (Wolters et al., 2009).

Personality – Characteristics, beliefs, and qualities that make up a person's character is called personality (Srivastava, 2012).

Social Users – Users who interact with a system on human terms and do not adapt their interactive style (Wolters et al., 2009).

Spoken Dialog System (SDS) – This device that takes speech input and produces speech output (Wolters et al., 2009).

Text to Speech Synthesis (TTS) – One of the five main components of a spoken dialog system, this component converts the system utterances into actual speech output (Wolters et al., 2009).

Wizard of Oz – The wizard is used to simulate the spoken dialog system and interact with the participant (Dahlback, Jonsson, & Ahrenberg, 1993).

Summary

A SDS provided an interface between a computer-based application and a user permitting interaction with the application in a natural language. Prior to the start of this study, predictors that enhanced the efficiency of interaction by improving the interface had not been identified. The research investigated if the selection of an optimal SDS interface for a given individual end user could be determined through personality traits. In this research, nine SDS dialog types comprising of three system initiatives with three confirmations each was presented to the user, therefore experiencing all nine-dialog types when scheduling an airline flight using the Wizard of Oz approach. Data was collected on the user's preferences and efficiency on all systems and analyzed to see if there was a predictor for a better interface based on personality traits as described by the FFM.

Chapter 2

Review of the Literature

Introduction

Speech recognition systems have been the center of research for many years in Human Computer Interaction (HCI); however, improvement in the systems will continue to be in the forefront due to the imperfections in human-to-human communication and the intricacies of interpreting contextual cues (Delimarschi, Swartzendruber, & Kagdi, 2014; Pew, 2003). Approach to design, with an emphasis on knowing the user, produces challenges due to lack of research in understanding the user from the viewpoint of experience and all the aspects that make up a more familiar ontology concerning human interaction (Cockton, 2004; Wright & McCarthy, 2008). Maintaining consistency in the design of the interface while addressing a universal usability are the key components in a well-designed interface (Shneiderman, Plaisant, Cohen, Jacobs, Elmqvist, & Diakopoulos, 2017).

SDS Components and Environments

While speech recognition systems are a vital part of HCI research, they are a main component in a SDS, which allows natural spoken language to be used as the interface. Systems that use this type of interface are comprised of several components that allow for functionality (McTear, 2002). McTear described the components as:

- Speech recognition – completes conversion of input speech utterance.
- Language understanding – produces analysis of conversion to produce usable representation for the system.
- Dialogue management – manages and controls the interaction between the user and the system.

- Communication with external system – this is the system, which the user is depending on for information.
- Response generation – this is the actual message that the system will return to the user.
- Speech output – completes text to speech or prerecorded speech used to output the system's message (p. 103-104).

McTear (p. 104-105) explained each component's capacity and associated attributes the component must navigate in order to provide the flexibility needed for successful application of the SDS. As the components have developed, they have come to be more commonly known by the acronyms ASR, NLU, DM, NLG, and TTS and described by Wolters et al. (2009). McTear (2002) also discusses toolkits that are available for developing SDS to support academic research and teaching as well as commercial use. The toolkits mentioned by McTear are in the following list of development environments with a brief description of their purpose:

- The Generic Dialogue System Platform – Denmark – mainly used to support academic research and teaching spoken language technology.
- GULAN – An Integrated System for Teaching Spoken Dialogue Systems Technology – mainly used to support academic research and teaching spoken language technology.
- The CSLU toolkit (Center for Spoken Language Understanding at the Oregon Graduate Institute of Science and Technology) – mainly used to support academic research and teaching spoken language technology.
- CU Communicator system – this is a commercial application.
- Nuance Developer's Toolkit (Nuance Communications) – this is a commercial application.
- SpeechWorks – this is a commercial application.
- Natural Language Speech Assistant (NLSA) (Unisys Corporation) – this is a commercial application.

- SpeechMania: A Dialogue Application Development Toolkit (Philips Speech Processing) – this is a commercial application.
- The REWARD Dialogue platform – this is a commercial application.
- Vocalis SpeechWare – this is a commercial application (p. 155).

Newer toolkits such as the MMDAgent, are open source and open design for the expressed purpose of allowing a variety of different speech interactions for SDS and speech interfaces to be developed using the same toolkit (A. Lee, Oura, & Tokuda, 2013).

Dialog Management

Depending on how dialog is managed, SDS can be categorized in three initiatives: strict, mixed, and user (Wolters et al., 2009). Strict initiative is the most commonly used initiative where the user has no control over the input and must respond based on strictly defined options. This rigidity makes SDS construction easier, but the interaction is unnatural and uncomfortable for many users. Although the technique is in its infancy, user initiative appears to offer the most user control with minimal system prompts or direction (C. Lee, Jung, Kim, Lee, & Lee, 2010). Moreover, the ability to move from a sequential per-turn approach for dialog management to a more activity management approach may be on the horizon (Bohus, Kamar, & Horvitz, 2012).

Spoken sentences contain irregular phenomena such as self-corrections, hesitations, and repetitions (De Mori, Bechet, Hakkani-Tur, & McTear, 2008). Mixed initiatives offer a blend of user control and system control. Each type of initiative produces a variety of errors due to the complexity of user utterances, user assumptions, SDS design assumption, and miscommunications (Wolters et al., 2009). Although strict initiatives may produce improved task completion, mixed initiatives hold promise for a more user-friendly experience especially in identifying specific requests by the user (Quinn & Zaiane, 2014).

SDS Applications

SDS can be found today in many applications, which include:

- Commercial applications for ticket sales (Pieraccini & Huerta, 2008);
- Training people how to train robots (Cakmak & Takayama, 2014);
- Behavior interventions – hypertension management (Giorgino et al., 2005);
- Smart Wheeler – wheelchair for the disabled (Chinaei & Chaib-Draa, 2014);
- Smart-Home controllers (Moller, Krebber, & Smeele, 2006);
- Environmental control systems for in home elder care (Griol, Molina, & Callejas, 2014);
- Automotive and travel information systems (Warnestal & Kronlid, 2014).

SDS will eventually need to be multimodal at the interface level with real-time capability in order to offset the weaknesses of other input or output options on the device (McTear, 2002). In addition, this accommodates the user who is either impaired visually, physically, audibly, or situationally impaired. For example, answering the phone with your hands full; wearing gloves; inappropriate environment such as driving or too much noise. Determining the most appropriate interface for an end user, which is based on the objective of minimizing miscues or miscommunications, strengthens user confidence as exemplified in the use of behavioral and strategy changes by the user in dealing with errors (Shin, Georgiou, & Narayanan, 2013). The use of hand held devices such as the cell phone provides an area where the SDS can greatly improve access to the information on the Web with the ability to spoken interaction where the use of the GUI is not as convenient (Griol, Carbo, & Molina, 2013). Applications of SDS serves in a multitude of domains and is touted as the new frontier in HCI (Lison & Meena, 2014).

Research Areas

The following sections focus on research in user adaptability and interaction, the importance of design accuracy, the search for reliable predictors of user behavior, and prior research which this study was based.

User Adaptability and Interaction

SDS encounters many technical problems that have fueled research offerings especially in the area of user interaction. The different components that make up a dialog system allow the user to receive the input through different methods. Systems dynamically adaptable to user behavior (Chickering & Paek, 2007) and systems that are user specific (Moore, Foster, Lemon, & White, 2004) are the focus for researchers as they appear to foster better efficiency and approval. Statistical optimization approaches where the system can automatically discern dialog strategies using simulations (Georgila, Henderson, & Lemon, 2005; Georgila, Wolters, & Moore, 2008; Lemon et al., 2006) are increasingly popular due to ease of application adaptation and system optimization. Research in SDS also establishes the need for studies that include users of different age groups as well as users who are at different levels of cognitive functioning in their lives; have changes in language habits; and varying degrees of acceptance of new technology. The approach established in current research is unique in both amount of information for each user concerning cognitive abilities and in user satisfaction assessment. Results of studies that establish this type of approach are a key resource in providing data on how characteristics of user behavior such as linguistics, cognitive function, personality, and attitude toward technology can be used to predict interaction with SDS and lays the ground work for future research in determining user groups (Duta, 2014; Georgila, Wolters, Karaiskos, et al., 2008).

Design Accuracy

The importance of design accuracy of the interface has become as essential as predictability when it comes to user performance and use of the interface. The ready acceptance of systems that exhibit a more socially communicative interface as opposed to less interaction has been shown to be the preferred interface for older users (Heerink, Krose, Wielinga, & Evers, 2009) thereby opening research areas to examine user behavior as a predictor. Wolters et al. (2009) demonstrated the effectiveness of using the Wizard of Oz (WoZ) (DiSalvo, Sengers, & Brynjarsdttir, 2010; Li & Bonner, 2014) technique as a tool in simulated conversational capabilities. The WoZ techniques involves using a human (the wizard behind the curtain) to guide the interaction with the user providing appropriate feedback that mimics an actual SDS. By using the WoZ technique, data is easily collected and the user is kept unaware that the interaction is actually managed by another human allowing for an empirical approach for data collection and an interaction beneficial to the user (Ashok, Borodin, Stoyanchev, Puzis, & Ramakrishnan, 2014). Data collected in this fashion can be used in the development of user simulation models allowing real tasks to be administered (Janarthanam & Lemon, 2009). Using the WoZ technique, Wolters et al. (2009) found that by combining technological features with conversational features that the user finds more acceptable and applying both to the interface, the user more readily accepts the device. Although worthwhile, implementing user simulation models to train statistical dialog managers for task oriented domains shows need for adaptability to the user's choice of wording even when the user is on task (Eshky, Allison, & Steedman, 2012).

In exploring the interaction style of a user, research is limited because system design focuses on the needs of the system and not on the needs of the user (Cairns, 2007; Ivanova, 2010). When a user oriented design is used then the approach must be aware of not only the user needs but also the system and the environment in which the interaction takes place (Ivanova, 2010). This elevates the importance of determining how to group preferences of the user in order to design a better user centered interface for the system.

Personality as a Predictor of User Interaction

Personality refers to the individual differences in human behavior. Understanding these differences helps define certain characteristics while examining the person as a whole. When examining possible predictors to define groups of people with the same or similar behavior, personality traits have potential for research, making the FFM a good choice of instrument for determining relationships between personality traits and preferences. Personality can explain a great number of human behaviors while the FFM factors these behaviors into measurable characteristics (Vinciarelli, 2014). The FFM is widely accepted as essentially correct in representation of personality traits (Goldberg, 1992; Goldberg et al., 2006; McCrae & John, 1992). It originates in studies of natural language trait terms (John, Angleitner, & Ostendorf, 1988) using the factored rating scales constructed by Tupes and Christal (1961). Five basic dimensions are used: Extraversion, Agreeableness, Conscientiousness, Neuroticism, and Openness to Experience which have demonstrated factors that have convergent and discriminate validity when used across different instruments, by different researchers, and across years in the human adult (McCrae & John, 1992). With these individual differences among people factored into the five dimensions, personality can be measured. This provides consistent data that can be replicated over time. Since the FFM seeks out personality dimensions in the natural language, the use of this method for determining consistencies in human behavior and preference makes it ideal for use in predicting user preference in SDS (Griol, Callejas, Lopez-Cozar, & Riccardi, 2014). By combining the natural language traits with the lexical information, the level of interest the user has in the SDS can be measured (Jeon, Xia, & Liu, 2014). User preferences that accurately identify certain user interactive styles will facilitate future SDS interface design.

Prior Research

The research is based on the prior research conducted by Wolters et al. (2009). The prior study discussed a bottom up approach to social dialog systems adaptability to users based on observed behavior through a WoZ design (Niederhoffer & Pennebaker, 2002) and cluster analysis. By using the personality factors produced by the FFM, these attributes can be categorized making it possible to perform statistical analysis in order to discover and measure the strength of relationships and norms (McCrae & John, 1992)

Wolters et al. (2009) used a bottom up approach to compare the interaction behavior of older users to younger users. A sampling of 50 participants was asked to schedule appointments through a SDS interaction. A large number of options were presented to the participant with a confirmation of the chosen option in order for the user to retain the information easier. Nine simulated systems were constructed using the WoZ approach where the user interacted with a human wizard while thinking they were interacting with an automated system. This approach not only limited the cost of the experiment but also limited the problems created by the current dialog limitations of ASR and NLU systems. All dialogs used in the WoZ system were identical – participants were asked to book nine appointments with four different health care professionals using four randomized lists of the nine tasks. All task lists were randomly assigned to each participant. All interactions between the participants and the SDS were conducted as a laboratory experiment. All participants were allowed to take breaks in order to combat fatigue. In scoring, the completed task score results and participant recall of the information were recorded. These results showed two main user groups, where one group – the factual group - adapted quickly and the other – the social group - treated the machine as human and showed no change in adaptation style. Wolters et al. (2009) was limited by the variety of users in individual age groups making future research in defining user preferences for all users imperative. Therefore, since the groupings, social and factual, did not support age as a delineator of user's behavior style, further research is needed to determine predictors other than age (Wolters et al., 2009). In examining SDS interfaces in home health care, factors that may or may not prevent user adaptation have been identified emphasizing the need for accurate predictors (M. Wolters, V. Hanson, & J. Moore, 2011). Moreover, to increase the user's ability of recall, the use of spearcons – time compressed speech messages – may help with menu designs in SDS emphasizing the need for better predictors (Wolters, Isaac, & Doherty, 2012).

Summary

This literature review contains the history of SDS in the aspects of the components and environments that presently use the technology and the areas that may benefit from the technology in the future. The need for accurate predictors that will improve user interaction for this technology is documented through a description of research conducted. The search for predictors is a rich research area. Personality traits as described by the FFM are one of those areas that may lead to a predictor for SDS interfaces that provide a richer experience for the user. The use of the Wizard of Oz approach to implementing the interaction for the participant is a proven method for this type research.

Chapter 3

Methodology

Overview of Research Design

The research design type for this study was descriptive research conducted in a laboratory setting. The descriptive research design highlighted potential relationships between the variables in order to identify a predictor for a spoken dialog system. Observations of qualified participants interacting with a SDS as related to how efficient the interaction, were collected by the researcher. Efficiency on a SDS was evaluated by how many attempts were taken to complete the interaction. The higher number of attempts the lower the efficiency for the system. With these observations, the relationship between the participant's efficiency for a SDS dialog type and the participant's personality inventory were identified. The observations provided the necessary data to examine the relationship between efficiency using a SDS dialog type and personality. The participant indicated which SDS was preferred by completing a Participant System Preference Indicator form. With this data, the relationship between the participant's preference for a SDS dialog type and the participant's personality inventory were identified. In addition, relationships between the participant preference and efficiency were examined without personality as a variable in order to identify any correlation between the two.

The actual interaction was conducted using the Wizard of Oz (WoZ) approach where the participants was unaware that they were not interacting with an actual machine. Instead, a lab technician provided each system dialog type and system responses at the correct time rather than an actual SDS. By using a laboratory setting for observations of participants, results were more meaningful (Jackson, 2009). The lab setting was located on the campus of Milligan College in the School of Business and Technology. This allowed for a consistent and reliable interaction by ensuring the participants experienced the same equipment and interaction. In a climate controlled room, the lab setting contained a desk equipped with a headset for the participant to use for the interaction with the SDS and a comfortable chair. The computer and the large print out of the dialog types that the wizard used were located behind a screen and out of sight of the participant. Before the interaction and after being qualified (age, gender, and hearing, etc.,) for the study, the participant completed a Big Five Personality Test. In this study, the participants' personality factors were categorized and quantified using the five factor dimensions of the FFM and scored using the IPIP questionnaire consisting of 50 items.

For the interaction, the participant placed the headset on their head comfortably and listened for the first option to schedule an airline flight from one city in the United States to another. There were ten options in scheduling for each experience with each dialog type. As each option was introduced the participant had the opportunity to speak their choices. The participant experienced three types of systems with their corresponding type of dialog. The systems represented:

1) a strict system limiting the range of interactivity experiences;
2) a mixed system that allowed a more natural interaction but did not allow the user full control of the interaction; and
3) a user system that allowed the user full control of the interaction.

Assumptions included:

- User's preferences might not have matched their most efficient dialog type experience.

- Although randomized, the order of dialog type interaction might have influenced the participant.

- No risk was involved if the participant failed all or any of the tasks presented during the interaction.

- Participants with a higher Agreeableness factor might have been more inclined to volunteer for a research study.

- Larger sample size might have influenced a stronger predictor outcome.

Research Method Employed

Working in the framework of a descriptive research design, the research questions were established and tested. This research investigated personality (as represented by the FFM) as a predictor of user interaction behavior through preference and efficiency. Preference was defined by the participant's selection on the Participant System Preference Indicator form. Efficiency was defined as minimizing miscues, miscommunications, and other types of errors (McTear, 2002).

A presumption of Wolters et al. (2009) was that an end user's preference for a SDS dialog type would match the SDS dialog type for which the user was most efficient at using. Wolters et al. (2009) stated, "Does interaction style affect usability (p. 2:3)? Yes. Interactive style affects efficiency and user satisfaction: "Social" users are less efficient and less satisfied with the system, which is tailored to the "factual" interaction style" (Wolters et al., 2009). The basis for the presumption that efficiency and preference is always matched is questionable as no evidence or prior research was cited to support this claim. Therefore, the scope of this research was expanded to include investigating the validity of the assumption – an end user's preference for a dialog type matches the dialog type at which the user is most efficient. Ergo for each end user, it determined which SDS dialog type they preferred and which SDS dialog type they were most efficient in using.

The SDS dialog types are unique combinations of initiative and confirmation attributes where the initiative takes the value of strict, mixed, or user while the confirmation attribute takes the value of explicit, implicit, or none (refer to Table 1).

First, the issue of preference was investigated. This research study utilized personality (as represented by the FFM) as a predictor for an end user's preference of SDS dialog type. Observations of the interaction were recorded on paper by the researcher and the participant indicated their SDS preference on the Participant System Preference Indicator form.

Research Question 1 (RQ1): How well does the FFM personality factor serve as a predictor of an individual's preferred SDS dialog type?

Using descriptive research methodology, this research replicated and extended the foundational study conducted by Wolters et al. (2009). This type of study solely relied on observations of the variables rather than manipulation of the variables as occurs in a true experiment. In this way, the observed value of the participant's interaction as to preference with the SDS dialog type was matched against the five factors of personality dimension as measured by the FFM. By computing correlations between how much participants preferred a SDS as indicated by the results of the Participant System Preference form, and the participant scores on all five factors of the FFM (Extraversion – E, Agreeableness – A, Conscientiousness – C, Emotional Stability – N, and Intellect/Imagination – O) the strongest indicators of personality dimensions for SDS preference were analyzed in SPSS. In addition, a multinomial regression analysis was run in which the preference for a SDS was the dependent variable and the five personality dimensions were independent variables. Using the regression analysis gave a more accurate result for a predictor of SDS preference based on personality by showing the closest relationship possible (Terrell, 2012). All participants who fell into the same FFM category were expected to have the same behavior and preference. In order to counterbalance the possibility of carryover effects of having multiple dependent variables, each system dialog type was randomly delivered to the participant.

Second, the issue of efficiency was investigated. Efficiency was defined as minimizing miscues, miscommunications, and other types of errors (McTear, 2002). Personality (as represented by the FFM) was used as a predictor in determining the user's efficiency for a SDS dialog type. Again, the researcher observed and recorded on paper the interactions for efficiency.

Research Question 2 (RQ2): How well does the FFM personality factor serve as a predictor of which SDS dialog type an individual used most efficiently?

In this way, the observed value of the participant's interaction as to efficiency with the SDS dialog type was matched against the personality factor as measured by the FFM. Again as in RQ1, by computing correlations between how efficient participants were during the SDS interaction as indicated by the observed miscues, restarts, non-completion of task, and the participant scores on all five factors of the FFM (Extraversion, Agreeableness, Conscientiousness, Emotional Stability, and Intellect/Imagination) the strongest indicators of personality dimensions for SDS efficiency were analyzed in SPSS. In addition, a multinomial regression analysis was run in which the efficiency for a SDS was the dependent variable and the five personality dimensions were independent variables. Using regression analysis gave a more accurate result for a predictor of SDS efficiency based on personality by showing the closest relationship possible (Terrell, 2012). All participants who fell into the same FFM category were expected to have the same behavior and preference. In order to counterbalance the possibility of carryover effects of having multiple dependent variables, each system dialog type was randomly delivered to the participant.

Third, the issue of preference versus efficiency was investigated. The presumption of Wolters et al. (2009) was that preference and efficiency did match. The basis for the presumption that efficiency and preference are always matched was questioned as no evidence or prior research had been found to support this claim.

Research Question 3 (RQ3): Is an individual's preference for an SDS dialog type the same SDS dialog type at which the individual is most efficient in using?

This research analyzed the SDS dialog type preferred by an individual compared to the SDS dialog type, which the individual was most efficient in using. For this research question, a one-way ANOVA was used as the analysis method to compare preference to efficiency. Preference and efficiency were not classified as dependent or independent and could be analyzed in either direction - preference to efficiency or efficiency to preference. This correlation was a good descriptive tool which fit nicely with the research design (Terrell, 2012). In this research question, the FFM five factor scores of the participant were not used for the data analysis. Here, only the preference indicator and efficiency scorings were used.

Procedure

A breakdown of the procedure and time taken for each activity of the research follows:

1. The participant was provided information needed to feel comfortable signing the consent form. Applicant received and completed the consent form (Appendix A). This took an average of 5 minutes. At any time, the participant could refuse to sign the consent form and leave the study. All participants received a copy of the consent form for their files.
2. The applicant received a letter of participation. (Appendix B) The letter was explained in detail. At any time, the participant could refuse to sign the letter and leave the study. All participants received a copy of the letter for their files. This took an average of 2 minutes.
3. The applicant completed the screening form. (Appendix C). This took approximately 2 minutes.
4. Applicants, at this point, were qualified or disqualified. This took approximately 2 minutes.
    a. The following conditions were met to qualify an applicant:
        i. Native English speaker
        ii. Age range 18 to 85
        iii. No hearing problems that could not be corrected by hearing aids
        iv. No physical disability that would prevent mobility in getting to and from the research location or prohibit interaction with the lab equipment or to speak clearly into a microphone.
    b. Applicants that did not meet these requirements were immediately dismissed.

5. With completed consent form, letter of participation, and screening form, the applicant profile was completed. All participants were redacted at this point and any identifying information was removed.
6. The participant took a standard tone and whisper test for hearing performed by a registered nurse. The tuning fork and whisper speech test are routine clinical practice for this purpose. A tuning fork is a metal, two-pronged device that produces a tone when it vibrates. The nurse strikes the tuning fork to make it vibrate and produce a tone. These tests assess how well sounds move through the ear. Sometimes the tuning fork would be placed on the participant's head or behind the ear. Depending on how the participant hears the sound, the nurse could tell if there was a problem with nerves in the ear or with sound getting to the nerves (Rhodes, 2009). In a whispered speech test, the nurse would ask the participant to cover the opening of one ear with their finger. The nurse stood one to two feet behind the participant and whispered a series of words at a soft whisper. The nurse kept saying the words more loudly until the participant can hear them. Each ear was tested separately (Rhodes, 2009). This took approximately 5 minutes. See Appendix D.
    a. Participants completing the hearing test successfully with or without hearing aids were moved on to the next step.
    b. Any participant that did not complete the hearing test successfully with or without hearing aids were immediately dismissed.
7. The participant, now, received the Five-Factor Model Questionnaire (Appendix E) to complete. Comfortable seating was provided at a table in order for the

participant to complete the form. Completion of this part of the process took approximately 10 minutes.

8. At this point, the participant entered the area where the interaction with the SDS took place. Each participant completed a reservation for travel by airline. There were ten options for each system dialog type. The participant listened for options for flights that meet the criteria for the trip and gave their answer. This required using a headset and speaking into a microphone to convey answers to the system options. The participant experienced interaction with three systems, System 1 – strict system, System 2 – mixed system, and System 3 – user system. Systems with their dialog types were given randomly using a random generator to each participant which counterbalanced any effects from multiple dependent variables. In other words, the participant who interacted with System 3 as their first interaction could then interact with System 1 as their second interaction if that was the randomization that the generator produced for the next system for interaction. Comfortable seating was provided and the participant was allowed to take breaks as needed. This part of the interaction took approximately 15 minutes. By utilizing the research technique of Wizard of Oz (WoZ), with a human being acting as the SDS giving the appropriate system responses during participant interaction, the participant was unaware that they were not interacting with an actual machine.
9. At the end of the interaction participants completed a short preference form indicating which SDS system they found most comfortable or most difficult (Appendix F). This took approximately 3 minutes.

10. At this time, the participant was thanked and dismissed. The total time invested by the participant was less than 40 minutes.

The researcher's role was to manage the execution of the interaction with the SDS and the participant. The researcher communicated face-to-face with the participant in the beginning of the interaction to ensure the participant understood the process. When satisfied the participant was ready and understood, the researcher stepped to the side and behind the participant in order to not distract or disturb the participant's concentration while completing the interaction and to record the interaction on paper as observed. The researcher noted the system dialog type that the participant experienced and the interaction that ensued.

The role of the wizard simulated the SDS. The WoZ interacted with the participant by following a script and dialog tree chart of the SDS. The wizard was responsible for randomizing the system dialog types and recording on paper which dialog type each participant received making sure the researcher also recorded the correct dialog type and system for that participant. The wizards consisted of lab assistants who had been carefully chosen for their ability to manipulate the software and their interest in the research. Four lab assistants were chosen, three to actually be wizards and the fourth for a back-up in case one of the assistants was unable to attend a session. A schedule was posted for the hours needed to conduct the experiment. All wizards underwent extensive training by the researcher on the SDS and practiced until comfortable with the interaction. As part of the training process, the wizards interacted with the SDS as both a wizard and then as a participant. Using the console to become familiar with the different dialog types and systems took practice and all wizards had set training hours where they practiced all nine dialog types (three systems times three initiatives) with the researcher as the participant. After the wizard became proficient with the console, they were instructed on how to handle the data correctly and securely. Wizards were responsible for ensuring all data was handled correctly and witnessed the data's placement in the appropriate locking container. When the data was securely locked in the container the wizards' job was finished for that session. Wizards were instructed on the importance of never revealing their part in this research while it was being conducted. Wizards were not allowed to be a participant in the study. Training and practice with the SDS console took approximately two hours. The SDS simulator was a voice synthesizer which spoke the dialog prompts. Using the simulator eliminated variance in speech and reduced the possibility of errors by the wizard. Using a small coded program, each dialog prompt was typed into the computer and appeared on the screen (Appendix G). As the participant completed the tasks, the wizard clicked the appropriate key on the console to trigger the appropriate spoken responses for the participant (Appendix H). Wizards were selected randomly for each participant's interaction.

Dialog Tree and SDS

The dialog tree component of an SDS was the most critical component for the system's usability. It determined what the user was able to request from the system, in which way, and at what time during the dialog session. In order to build a dialog system, it was essential to have a dialog tree that captured information about the dialog partners' interaction patterns and about the task structure of the domain.

A dialog tree (see figure 1 for an example) was composed of the following parts:

- Start symbol – A single point represented by an oval containing the word "Start" and signaling the start of the process.

- End symbols – One or more points represented by ovals containing the word "End" and signaling the end of the process.

- Dialog boxes – The SDS prompted the user to provide information or select commands.

- Connectors – This symbol showed a jump from one point in the process flow to another. Connectors were usually labeled with capital letters (A, B, C) to show matching jump points.

- Jump points – Place in the process flow resuming from a connector.

- Paths – Lines connecting start symbols, end symbols, dialog boxes, connectors, and jump points.

- Utterances – The user's verbal reply to an SDS prompt.

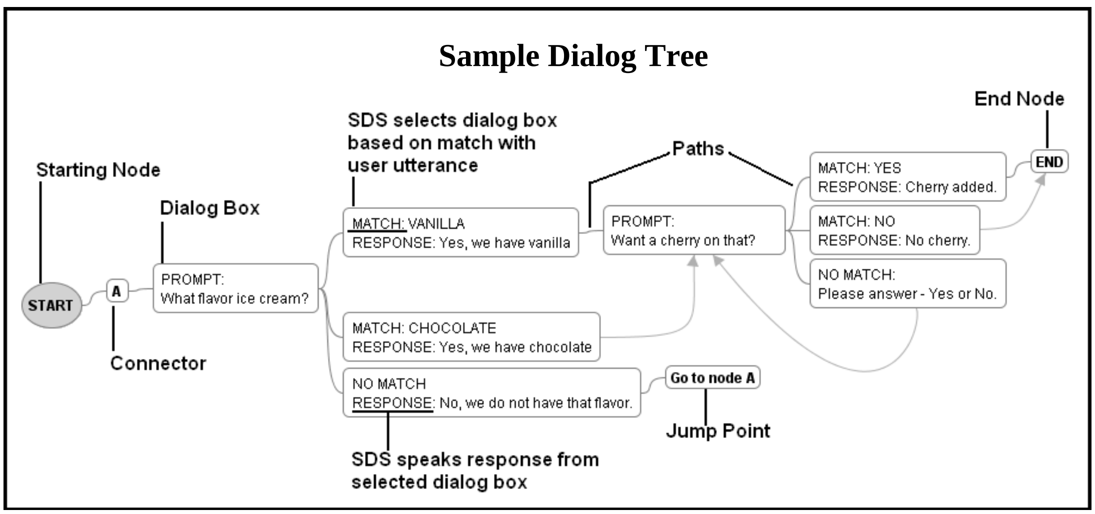


Figure 1. Sample dialog tree showing start and end of decision path.

The researcher defined the SDS dialog type as a unique combination of initiative and confirmation attributes. The initiative takes the value of strict, mixed, or user. The confirmation attribute takes the value of explicit, implicit, or none. Based on a matrix of three initiatives times the three confirmation values yielded nine tested SDS dialog types. All initiatives had to be included to provide a full spectrum of the available initiative types in order to test the supposition that preference and efficiency corresponded with the SDS dialog types. The confirmation attribute was included because it was in the foundation paper of Wolters et al. (2009). This research required a separately created dialog tree for each of the SDS dialog types.

The SDS task was to schedule an airline flight from one United States city to another. The number of vocal responses that it took to complete scheduling ranged from small or large in size and ranged from simple to complex in composition in order to encompass the interaction. In review of Wolters et al. (2009), the data reached a ceiling effect, which limited the range of responses that the participants could produce. After examining the dialog types, the ceiling effect appeared to be the result of a lack of complexity of the SDS tasks per communication with Maria Wolters (personal communication, June 23, 2010). Therefore, reducing the number of attributes and increasing the complexity of choices gave the desired range for responses while allowing for the reduction in the number of SDS dialog types to nine. Participants experienced all nine dialog type attributes on all three systems. All dialog types (Appendix I) were randomly assigned to participants. Appendices G and H are examples that give additional detail for the SDS concept and console.

In order to avoid the ceiling effect experienced by prior researchers, all dialog trees were constructed complex enough to ensure each interaction generated one or more errors during the testing. Complexity was achieved by increasing the difficulty of the choices or options to choose from. An increase in complexity corresponded with more dialog boxes, response options, and paths. Each dialog tree was composed of multiple dialog boxes in order to capture all possible answers that a user might generate for each of the ten options that could be chosen for the particular interaction. An example is where the participant was given the option of days of the week and times available on certain days. In a user system, this option was easy for the participant but in the strict system it was more difficult.

The WoZ technique was used to simulate an actual SDS configuration. The simulation required a person in the role of wizard as described earlier, to determine what action needed to be taken based on current position in the dialog tree and the user's utterance. Due to complexity of the dialog tree, it was necessary to provide the wizard with a visual representation of the process flow. This was accomplished by drawing the process flow for each dialog tree on a large white board. The large size of the white board and the complexity of its contents made any reproduction to standard paper size illegible. Part of the background work performed to test viability of this research was creating a “simple” preliminary dialog tree.

Instrument Development and Validation

The FFM represents the highest hierarchical level of personality trait description (McCrae & John, 1992). By using an IPIP multi-scale inventory questionnaire, individual traits were quantified using the five-factor scales. An existing instrument in the form of a 50-item questionnaire was used in generating data for each of the five dimensions of the FFM. The five dimensions, which are extraversion, agreeableness, conscientiousness, emotional stability, intellect or imagination, were scored by how the participant answered the questions in each of the dimensions represented on the questionnaire. The questionnaire was titled ‘The 50-item IPIP Representation of the Goldberg Markers for the Five-Factor Structure’ in keeping with the structure established by the FFM (Goldberg, 1992). The questionnaire was free for public use and was readily available. The FFM offers a reliable and efficient instrument to measure core attributes of personality and provides an instrument that is easily understood by participants giving a quantifiable assessment (Srivastava, 2012).

Validity is the demonstration that the results of the study are well grounded as proven by internal validity and external validity (Creswell, 2009). Descriptive research cannot describe what caused a situation where one variable affected another; therefore, is not a basis for causal relationship and gives descriptive research a low requirement for internal validity (Jackson, 2009). However, threats to the internal validity were minimized by the researcher through random selection of participants, recruiting a large sample to account for participants who drop out, and by selecting participants that would naturally score higher because of experience (Creswell, 2009). Categories for the participant's personality factors were defined by the FFM negating the possibility that they were subjective.

External validity is the ability of a study to be generalized to a population, in this case participants were chosen from a general population (Adelman, 1991). The researcher minimized threats to the external validity through replicating the research at different times of day to determine if the same results occurred, by conducting the research in different locations, and by limiting participants who had attributes that were easily generalized (Creswell, 2009). The research was conducted in the same location but at different times of the day.

Reliability is achieved when the results are consistent over time, give an accurate description of the population sample that are observed, and can be repeated using the same or similar methodology (Golafshani, 2003). The research design for the study produced results consistent over time with an accurate description of the participants based on the FFM approach to personality factors. This study can be repeated using the same methodology producing reliable results consistently because the methodology and use of WoZ was thoroughly tested in juried research such as Wolters et al., (2009).

Population Sample

The participant population size was statistically calculated using power analysis. Three components are required for a power analysis – significance level (*alpha*), effect size, and sample size. An *alpha* of 0.05 will be used and is considered valid under these circumstances (Bartlett, Kotrlik, & Higgins, 2001). A default effect size of 0.50 will be used based on the researchers Hair, Anderson, Tatham, and Black (1998). With value of .05 assigned to *alpha* and value of .5 assigned to effect size, it is possible to determine the remaining factor - sample size (Hair et al., 1998). The results of the calculation indicated that a minimum sample size of 80 participants was required (See Table 2).

Table 2

CALCULATION OF SAMPLE SIZE				
	Alpha (α) = .05 Effect Size (ES)		Alpha (α) = .01 Effect Size (ES)	
Sample size	**Small (.2)**	**Moderate (.5)**	**Small (.2)**	**Moderate (.5)**
20	.095	.338	.025	.144
40	.143	.598	.045	.349
60	.192	.775	.067	.549
80	.242	.882	.092	.709
100	.290	.940	.120	.823
150	.411	.990	.201	.959
200	.516	.998	.284	.992

It has been noted that the FFM can be skewed by gender and age (Chapman, Duberstein, Sorensen, & Lyness, 2007). In order to have the same distribution of ages and gender, the sample population was created from two pools of 40 males and 40 females. Using the two pools of participants to build the sample, eliminated the bias due to gender and age. Each pool had an equal distribution of age. In addition to the 80 participants needed, extra participants were selected to be alternates for participants who did not show up or failed to complete the experiment. A test pool of 96 (80 X 120%) gave the study the desired 20% additional participants. Therefore, the maximum number of participants required was 96.

The sample of 96 participants were recruited from the population of 2000 on the campus of Milligan College, a liberal arts college in east Tennessee. A campus wide email was sent as an invitation to participate in the study. The population consisted of faculty, staff, and students which met the correct age range, gender distribution, and exposure to the technology needed to participate.

One of the constraints for the study was that the participants had to meet specific requirements. These conditions included falling within the age range, having no physical disabilities, meeting the minimum hearing constraints, and being a native English speaker. The valid age range was 18 to 85 years of age. The instrument for validation was a picture-identification with proof of birth date. Applicants under the age of 18 or over 85 were excluded from the study. Reason for excluding physically disabled applicants was – participants had to be able to enunciate clearly any commands to the SDS and could not be restricted by physical access limitations to the lab. The mechanism for verification was provided as question on the application screening form. The basis for a native English speaker qualification required the participant to have a clear and fluent understanding of colloquial conversational English when interacting with the SDS. Methods for verification included the researcher’s observation and a question on the applicant screening form. The researcher described the process of the experiment to each participant and what information each form was intended to convey. The participant was directed when to complete each form. The participant interacted verbally with the SDS via headset and microphone, which communicated to the person playing the role of WoZ while emulating the SDS. The wizard indicated each system dialog type as either System 1, which is the strict initiative system; System 2, which is the mixed initiative system; or System 3, which is the user initiative system. The system dialog types were randomly assigned to the participant ensuring that the participant interacted with the SDS initiatives without knowing if System 1 was a strict, mixed, or user dialog type, only the wizard and the researcher knew and recorded which system the participant was actually rating. The participant interacted with each system creating a reservation for travel on an airline. They moved through each system dialog type and at the end of each system made their choice on the Participant System Preference Indicator form. Breaks between system interactions were taken if needed and the participant was asked each time if they needed a break. The participant was able to leave the interaction and/or study at any point in time.

Data Collection and Analysis

The researcher observed and collected each participants' interaction and recorded the resulting data on paper. The transcribed data was placed into Excel data files and then used in SPSS for statistical analysis. Data collection from hard copy included the Applicant Screening form, hearing test results, the 50-item IPIP Representation of the Goldberg Markers for the Five-Factor Structure (Big Five Personality Test) results, the Participant System Preference Indicators form, and the observation form for the experiment. The data will be stored for three years in a locked container in a vault on the research site. After such time it will be determined if the data is no longer needed. At determination, all data will be destroyed by security officers in the same manner all secure documents are destroyed on the research site.

The appropriate descriptive statistics was computed and analyzed. Correlation was used as it described how well a statistical model fits a set of observations (Hair et al., 2006). Using correlation analysis helped to determine the probability that the variables were related; however, from there other methods were used to determine more detail (Terrell, 2012). Multinomial regression was used to understand which of the five independent variables (FFM personality factors) could be a predictor for the dependent variable (SDS dialog type preferred) as well as how it was related to the dependent variable of efficiency when using the SDS. By using multinomial regression, the observed values were tested to see how closely the data met the expectation. Pearson's Correlation was used to describe the relationship between efficiency to preference and preference to efficiency.

Formats for Presenting Results

Formats for presenting the results included Excel from Microsoft Office Suite, and IBM SPSS version 23 with advanced add-ons for file output results and statistical analysis. Results were shown in tables and figures.

Resources

The following section provides a listing of all resources used for the successful completion of the research.

- Lab setting containing a standard computer running Windows 10 operating system. No modifications were made except for the modifications required by the SDS to deliver the options in a mechanical voice.
- A maximum of 96 willing and qualified participants

- Nurse to administer a simple hearing test
- Secure container to hold all paperwork
- Forms and questionnaires for Wizard of Oz systems
- Lab research assistants (four) for multiple Wizard of Oz implementations. These assistants are the wizards.
- Designated area at appropriate time for collection and tabulation of questionnaires and forms from observed interaction, the 50-item IPIP Representation of the Goldberg Markers for the Five-Factor Structure (Big Five Personality Test), Participant System Preference Indicator forms, and hearing tests.
- Statistical software – IBM SPSS version 23.

Summary

The objective of the research was to determine if a personality dimension as measured by the FFM is a predictor of a SDS interface that provided the user with an efficient experience. With the proposal approved by the chair and committee and the IRB approval confirmed (Appendix J) the research was conducted. The research produced data using the same WoZ approach as Wolters et al., (2009) but used a personality factor as a dimensional type scheme (specifically the FFM) instead of age as a predictor as in Wolters's study of user characteristic patterns. The end user interactions with multiple SDS dialog types were collected as the recorded data while accomplishing this task. The objective of the research was to determine if a personality dimension was a predictor for SDS dialog type preference and efficiency. Personality was quantified by using the five factor dimensions as scored by the IPIP questionnaire, which was recommended for FFM research (Srivastava, 2012). Efficiency was quantified through observation by the researcher's recording of the participant mistakes, miscues, and incomplete tasks. Secondarily, end user SDS preferences compared to SDS efficiency was quantitatively analyzed to investigate the validity of the presumption that an end user's SDS preference corresponds with his/her efficiency in using a SDS dialog type.

Each qualified participant signed a consent form, completed the IPIP, and then completed the interaction with the SDS. All results from the 50-item IPIP Representation of the Goldberg Markers for the Five-Factor Structure, which is the FFM personality inventory or Big Five Personality Test, Participant System Preference Indicator, and the observation form containing recorded results on paper by the researcher of the interaction were secured as required. Each participant's identity was protected from disclosure and precautions were put in place to ensure results cannot be traced back to the individual. The population sample of 96 for this study was recruited from a college campus of 2000 consisting of faculty, students, and staff. A minimum sample size of 80 was needed. Lab setting and equipment were updated and properly installed to ensure safety and reliability.

The research design provided for not only the efficiency measure of an interface but also for the preference of the user. The research method was implemented using the Wizard of Oz approach to simulate the SDS and gave an understanding of how well the five independent variables of the FFM related to the dependent variables of the SDS dialog type and preference and SDS dialog type and efficiency. To avoid the ceiling effect experienced by prior researchers, all dialog trees were designed with enough complexity to ensure participants would generate errors during the experiment.

Chapter 4

Results

Data Analysis

Analysis was conducted using IBM SPSS version 23. Raw data were first placed in an Excel spreadsheet in order to organize and record the data for each research question and to establish a snapshot of what the participants results looked like as related to the study. Raw data allowed the researcher to see the breakdown of personality dimensions by system preference and efficiency before putting data into the statistical software program. In order to dismiss any gender bias for the FFM, the 80 participants were equally divided by gender with 40 males and 40 females and matched in age. Totals by factor for all personality factors are shown at the bottom of each table. Agreeableness – A has the highest overall total personality factor (2484) for all systems when looking at side-by-side comparison of participant gender and matched age for preference. See Table 3 totals.

Table 3

SIDE-BY-SIDE COMPARISON OF PARTICIPANTS' GENDER AND AGE WITH PREFERENCE AND PERSONALITY FACTORS

	Female Participant	Age	Preferred System	Personality Factors						Male Participant	Age	Preferred System	Personality Factors				
				E	**A**	**C**	**N**	**O**					**E**	**A**	**C**	**N**	**O**
1	1-07	19	Mixed	33	36	29	28	27	**41**	2-04	19	User	21	28	14	32	34
2	1-75	19	User	22	15	25	21	35	**42**	2-40	19	Strict	16	28	29	16	24
3	1-37	20	User	37	33	20	29	18	**43**	2-20	20	User	28	27	29	29	34
4	1-46	20	User	36	28	40	28	31	**44**	2-28	20	Strict	24	37	31	19	34
5	1-48	20	Strict	26	33	21	19	18	**45**	2-37	20	User	21	20	22	31	27
6	1-50	20	User	30	36	32	18	28	**46**	2-38	20	Strict	29	22	29	26	23
7	1-52	20	User	3	29	37	36	28	**47**	2-41	20	User	19	35	19	25	22
8	1-02	21	User	13	29	15	19	20	**48**	2-02	21	Mixed	19	32	35	22	30
9	1-03	21	User	25	23	20	24	20	**49**	2-03	21	User	17	29	29	32	37
10	1-10	21	User	19	29	22	21	34	**50**	2-14	21	User	31	35	36	28	35
11	1-11	21	Mixed	14	30	23	29	31	**51**	2-23	21	User	30	26	29	19	31
12	1-24	21	Strict	12	22	32	19	24	**52**	2-26	21	Mixed	20	26	36	21	34
13	1-42	21	Strict	28	34	33	31	28	**53**	2-39	21	Strict	21	36	29	31	22
14	1-45	21	User	18	40	25	12	18	**54**	2-42	21	Mixed	11	30	30	29	33
15	1-29	22	Mixed	20	33	32	17	29	**55**	2-07	22	User	22	28	32	33	31
16	1-34	22	User	27	35	24	19	25	**56**	2-24	22	Mixed	22	40	29	18	31
17	1-39	22	User	17	39	28	20	33	**57**	2-31	22	User	27	27	19	31	29

Table 3 (continued)

SIDE-BY-SIDE COMPARISON OF PARTICIPANTS' GENDER AND AGE WITH PREFERENCE AND PERSONALITY FACTORS

	Female Participant	Age	Preferred System	Personality Factors						Male Participant	Age	Preferred System	Personality Factors				
				E	A	C	N	O					E	A	C	N	O
18	1-61	22	User	33	39	29	12	32	**58**	2-46	22	User	27	24	36	32	34
19	1-76	22	User	37	32	25	12	27	**59**	2-71	22	User	22	38	22	31	39
20	1-36	23	User	25	25	19	29	20	**60**	2-30	23	Mixed	31	36	24	29	26
21	1-38	23	User	24	33	35	24	22	**61**	2-43	23	User	21	32	31	28	35
22	1-25	24	User	11	30	22	22	10	**62**	2-25	24	Mixed	24	28	30	21	23
23	1-47	25	User	7	30	23	17	21	**63**	2-55	25	Mixed	8	26	34	33	19
24	1-30	28	Mixed	30	38	35	34	36	**64**	2-35	28	User	7	37	26	29	24
25	1-09	35	Strict	13	40	22	17	28	**65**	2-70	35	Mixed	11	35	21	34	34
26	1-67	36	User	30	34	15	8	33	**66**	2-56	36	User	16	25	15	16	40
27	1-23	37	User	23	33	23	26	24	**67**	2-57	37	User	9	27	30	5	27
28	1-54	44	Mixed	24	30	26	31	22	**68**	2-29	44	User	32	39	40	31	23
29	1-60	44	Mixed	31	38	30	31	34	**69**	2-69	44	Strict	6	27	25	23	36
30	1-04	47	User	25	28	33	24	33	**70**	2-27	47	User	10	39	31	37	37
31	1-08	48	User	27	32	37	26	38	**71**	2-05	48	Strict	30	26	33	22	25
32	1-27	53	User	16	32	38	30	16	**72**	2-72	53	User	36	25	24	28	33
33	1-70	57	User	19	37	33	11	31	**73**	2-60	57	User	27	34	25	22	35
34	1-43	58	User	24	40	29	11	27	**74**	2-08	58	Mixed	11	16	28	24	34
35	1-01	61	User	11	32	36	34	22	**75**	2-06	61	User	35	30	21	25	21
36	1-21	62	Mixed	28	32	36	26	33	**76**	2-33	62	User	26	30	15	22	24
37	1-05	63	User	13	39	24	30	30	**77**	2-09	63	Strict	10	28	24	28	33
38	1-41	63	User	13	33	36	20	26	**78**	2-61	63	User	13	22	22	25	26
39	1-06	67	Mixed	19	32	27	31	33	**79**	2-36	67	Strict	35	17	29	13	38
40	1-77	70	User	24	37	39	38	31	**80**	2-32	70	Strict	38	37	34	39	37

Totals: E=1750, A=2484, C=2227, N=1973, O=2290

RQ1: *How well does the FFM personality factor serve as a predictor of an individual's preferred SDS dialog type?*

The raw data breakdown for personality in the predicting of a SDS preference shows that 50 of the 80 participants preferred a user system over the mixed or strict systems. See Table 4 and Figure 2 for user data. This data have a significantly higher average than the 17 of the 80 participants for the mixed system or the 13 of the 80 for the strict system. See Table 5 and Figure 3 for mixed system data. See Table 6 and Figure 4 for strict system data. Since the personality factor for Agreeableness - A was predominant in all preferences, Agreeableness - A was also the predominant factor for the user system.

Factor totals are shown at the bottom of Table 4 where Agreeableness – A has the highest factor total of 1559.

Table 4

USER SYSTEM PREFERENCE WITH PERSONALITY FACTORS																				
	Participant	**Personality Factors**						**Participant**	**Personality Factors**						**Participant**	**Personality Factors**				
		E	**A**	**C**	**N**	**O**			**E**	**A**	**C**	**N**	**O**			**E**	**A**	**C**	**N**	**O**
1	1-75	22	15	25	21	35	**18**	1-10	19	29	22	21	34	**35**	2-60	27	34	25	22	35
2	2-37	21	20	22	31	27	**19**	1-52	3	29	37	36	28	**36**	1-34	27	35	24	19	25
3	2-61	13	22	22	25	26	**20**	2-03	17	29	29	32	37	**37**	2-14	31	35	36	28	35
4	1-03	25	23	20	24	20	**21**	1-25	11	30	22	22	10	**38**	2-41	19	35	19	25	22
5	2-46	27	24	36	32	34	**22**	1-47	7	30	23	17	21	**39**	1-50	30	36	32	18	28
6	1-36	25	25	19	29	20	**23**	2-06	35	30	21	25	21	**40**	1-70	19	37	33	11	31
7	2-56	16	25	15	16	40	**24**	2-33	26	30	15	22	24	**41**	1-77	24	37	39	38	31
8	2-72	36	25	24	28	33	**25**	1-01	11	32	36	34	22	**42**	2-35	7	37	26	29	24
9	2-23	30	26	29	19	31	**26**	1-08	27	32	37	26	38	**43**	2-71	22	38	22	31	39
10	2-20	28	27	29	29	34	**27**	1-27	16	32	38	30	16	**44**	1-05	13	39	24	30	30
11	2-31	27	27	19	31	29	**28**	1-76	37	32	25	12	27	**45**	1-39	17	39	28	20	33
12	2-57	9	27	30	5	27	**29**	2-43	21	32	31	28	35	**46**	1-61	33	39	29	12	32
13	1-04	25	28	33	24	33	**30**	1-23	23	33	23	26	24	**47**	2-27	10	39	31	37	37
14	1-46	36	28	40	28	31	**31**	1-37	37	33	20	29	18	**48**	2-29	32	39	40	31	23
15	2-04	21	28	14	32	34	**32**	1-38	24	33	35	24	22	**49**	1-43	24	40	29	11	27
16	2-07	22	28	32	33	31	**33**	1-41	13	33	36	20	26	**50**	1-45	18	40	25	12	18
17	1-02	13	29	15	19	20	**34**	1-67	30	34	15	8	33							
Totals: E=1106, A=1559, C=1351, N=1212, O=1411																				

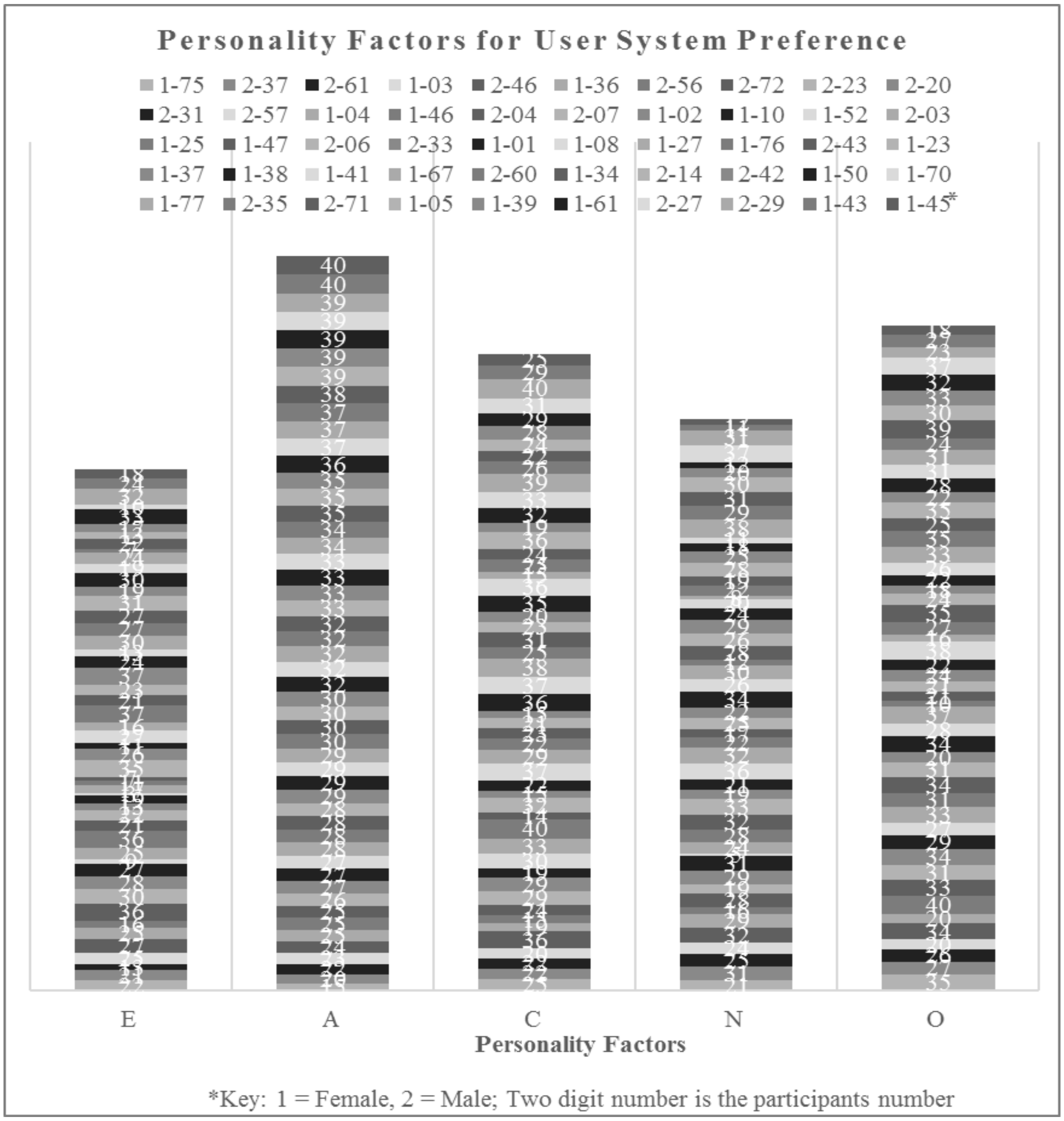


Figure 2. Personality for the user system preference showing Agreeableness – A as the predominant factor.

For the mixed system, 17 of the 80 participants preferred the mixed system initiative. Agreeableness - A was again the dominant personality factor with a total of 538. See Table 5 and Figure 3.

Table 5

	MIXED SYSTEM PREFERENCE WITH PERSONALITY FACTORS					
	Participant	**Personality Factors**				
		E	**A**	**C**	**N**	**O**
1	2-08	11	16	28	24	34
2	2-26	20	26	36	21	34
3	2-55	8	26	34	33	19
4	2-25	24	28	30	21	23
5	1-11	14	30	23	29	31
6	1-54	24	30	26	31	22
7	2-42	11	30	30	29	33
8	1-06	19	32	27	31	33
9	1-21	28	32	36	26	33
10	2-02	19	32	35	22	30
11	1-29	20	33	32	17	29
12	2-70	11	35	21	34	34
13	1-07	33	36	29	28	27
14	2-30	31	36	24	29	26
15	1-30	30	38	35	34	36
16	1-60	31	38	30	31	34
17	2-24	22	40	29	18	31
	Totals: E=356, A=538, C=505, N=458, O=509					

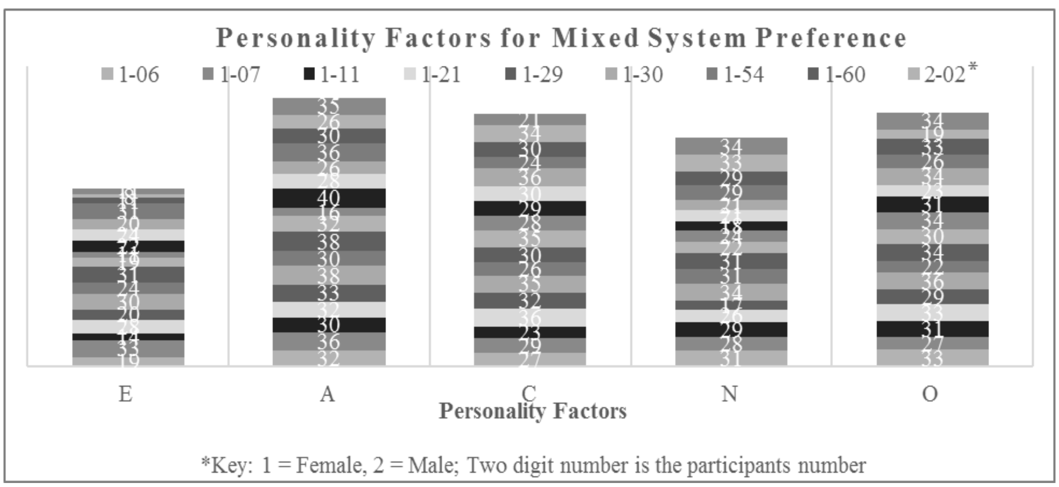


Figure 3. Mixed system data showing Agreeableness – A as the predominent factor.

For the strict system, 13 of the 80 participants preferred the strict system initiative. Again, Agreeableness – A has the highest factor total of 387. Strict system data is represented in Table 6 and Figure 4.

Table 6

	STRICT SYSTEM PREFERENCE WITH PERSONALITY FACTORS					
	Participant	**Personality Factors**				
		E	**A**	**C**	**N**	**O**
1	2-36	35	17	29	13	38
2	1-24	12	22	32	19	24
3	2-38	29	22	29	26	23
4	2-05	30	26	33	22	25
5	2-69	6	27	25	23	36
6	2-09	10	28	24	28	33
7	2-40	16	28	29	16	24
8	1-48	26	33	21	19	18
9	1-42	28	34	33	31	28
10	2-39	21	36	29	31	22
11	2-28	24	37	31	19	34
12	2-32	38	37	34	39	37
13	1-09	13	40	22	17	28
	Totals: E=288, A=387, C=371, N=303, O=370					

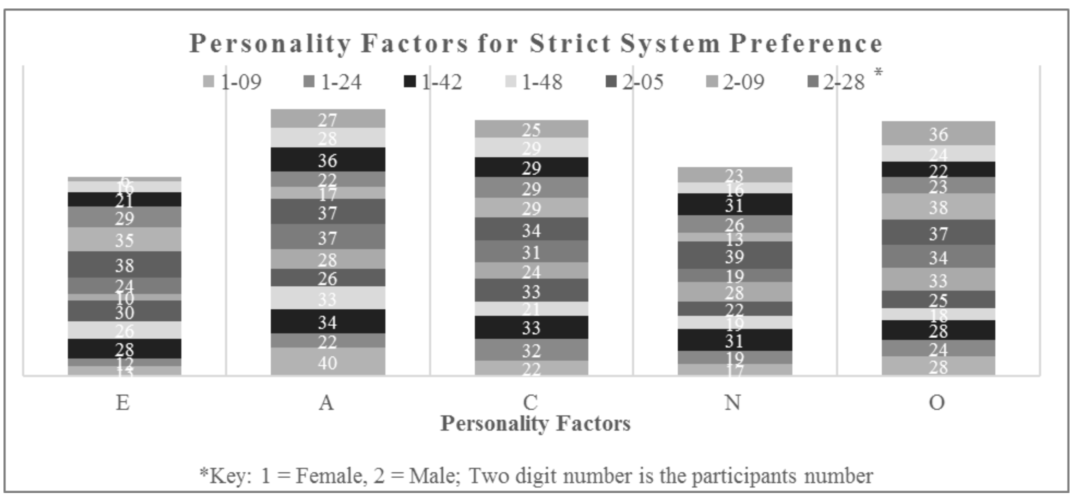


Figure 4. Strict system preference showing Agreeableness – A as the predominant personality factor.

RQ2: *How well does the FFM personality factor serve as a predictor of which SDS dialog type an individual used most efficiently?*

The raw data for personality factor as a predictor of efficiency shows the mistakes made in all systems. The breakdown shows the least mistakes made were in the user system with the most mistakes made in the strict system. In each of these tables a note has been added to explain that the highest factor score is in bold and system efficiency for each dialog type are indicated by the following breakdown of errors: zero (0) = non-errors, one (1) = one (1) errors, two (2) = two (2) errors, three (3) = three (3) errors, and four (4) = four (4) errors. See Table 7.

Table 7

PARTICIPANTS' SYSTEM EFFICIENCY WITH PERSONALITY FACTORS FOR STRICT, MIXED AND USER

Participants	System Efficiency			Personality Factors				
	Strict	Mixed	User	E	A	C	N	O
1-01	1	1	0	11	32	**36**	34	22
1-02	0	0	0	13	**29**	15	19	20
1-03	0	0	0	**25**	23	20	24	20
1-04	1	1	0	25	28	**33**	24	**33**
1-05	1	0	0	13	**39**	24	30	30
1-06	2	0	3	19	32	27	31	**33**
1-07	1	1	0	33	**36**	29	28	27
1-08	1	0	0	27	32	37	26	**38**
1-09	0	1	0	13	**40**	22	17	28
1-10	1	1	1	19	29	22	21	**34**
1-11	1	1	0	14	30	23	29	**31**
1-21	2	1	0	28	32	**36**	26	33
1-23	0	0	2	23	**33**	23	26	24
1-24	0	0	1	12	22	**32**	19	24
1-25	3	2	2	11	**30**	22	22	10
1-27	2	1	2	16	32	**38**	30	16
1-29	1	0	0	20	**33**	32	17	29
1-30	0	0	1	30	**38**	35	34	36
1-34	0	1	0	27	**35**	24	19	25
1-36	4	0	1	25	25	19	**29**	20
1-37	0	0	0	**37**	33	20	29	18
1-38	2	0	1	24	33	**35**	24	22
1-39	0	0	0	17	**39**	28	20	33
1-41	0	1	0	13	33	**36**	20	26
1-42	0	4	1	28	**34**	33	31	28
1-43	0	1	1	24	**40**	29	11	27
1-45	1	0	0	18	**40**	25	12	18
1-46	0	0	1	36	28	**40**	28	31
1-47	0	0	0	7	**30**	23	17	21
1-48	1	0	0	26	**33**	21	19	18
1-50	1	0	1	30	**36**	32	18	28
1-52	1	1	0	3	29	**37**	36	28

Table 7 (continued)

PARTICIPANTS' SYSTEM EFFICIENCY WITH PERSONALITY FACTORS FOR STRICT, MIXED AND USER								
Participants	**System Efficiency**			**Personality Factors**				
	Strict	**Mixed**	**User**	**E**	**A**	**C**	**N**	**O**
1-54	1	1	0	24	30	26	**31**	22
1-60	1	0	0	31	**38**	30	31	34
1-61	0	1	0	33	**39**	29	12	32
1-67	2	0	1	30	**34**	15	8	33
1-70	1	0	0	19	**37**	33	11	31
1-75	0	1	0	22	15	25	21	**35**
1-76	2	1	1	**37**	32	25	12	27
1-77	1	1	2	24	37	**39**	38	31
2-02	0	2	1	19	32	**35**	22	30
2-03	3	1	1	17	29	29	32	**37**
2-04	0	0	0	21	28	14	32	**34**
2-05	3	0	0	30	26	**33**	22	25
2-06	1	1	0	**35**	30	21	25	21
2-07	3	1	1	22	28	32	**33**	31
2-08	0	2	0	11	16	28	24	**34**
2-09	1	0	0	10	28	24	28	**33**
2-14	0	1	0	31	35	**36**	28	35
2-20	1	0	0	28	27	29	29	**34**
2-23	0	1	0	30	26	29	19	**31**
2-24	0	1	1	22	**40**	29	18	31
2-25	0	2	0	24	28	**30**	21	23
2-26	1	0	0	20	26	**36**	21	34
2-27	1	1	0	10	**39**	31	37	37
2-28	1	1	1	24	**37**	31	19	34
2-29	1	1	0	32	39	**40**	31	23
2-30	1	1	1	31	**36**	24	29	26
2-31	1	0	1	27	27	19	31	**29**
2-32	0	2	2	38	37	34	**39**	37
2-33	3	1	0	26	**30**	15	22	24
2-35	0	1	0	7	**37**	26	29	24
2-36	5	3	0	35	17	29	13	**38**
2-37	4	1	0	21	20	22	**31**	27
2-38	1	1	2	**29**	22	**29**	26	23
2-39	0	0	0	21	**36**	29	31	22
2-40	0	1	1	16	28	**29**	16	24
2-41	2	0	0	19	**35**	19	25	22
2-42	1	0	1	11	30	30	29	**33**
2-43	0	2	1	21	32	31	28	**35**
2-46	0	0	0	27	24	**36**	32	34
2-55	1	1	3	8	26	**34**	33	19
2-56	4	4	0	16	25	15	16	**40**
2-57	0	1	1	9	27	**30**	5	27
2-60	1	0	0	27	34	25	22	**35**
2-61	3	1	2	13	22	22	25	**26**
2-69	0	0	0	6	27	25	23	**36**
2-70	0	0	0	11	**35**	21	34	34
2-71	1	1	0	22	38	22	31	**39**
2-72	1	1	1	**36**	25	24	28	33
Totals:	80	60	43	1750	2484	2227	1973	2290

Note. Highest factor score is in bold. Errors made in each system equals efficiency: zero (0) = non-errors, one (1) = one (1) errors, two (2) = two (2) errors, three (3) = three (3) errors, and four (4) = four (4) errors.

Table 8 shows only the user system by efficiency breakdown.

Table 8

SYSTEM EFFICIENCY BASED ON ERRORS MADE IN A USER SYSTEM PER PARTICIPANT						
Participants	**System Efficiency**	**Personality Factors**				
	User	**E**	**A**	**C**	**N**	**O**
1-01	0	11	32	**36**	34	22
1-02	0	13	**29**	15	19	20
1-03	0	**25**	23	20	24	20
1-04	0	25	28	**33**	24	**33**
1-05	0	13	**39**	24	30	30
1-07	0	33	**36**	29	28	27
1-08	0	27	32	37	26	**38**
1-09	0	13	**40**	22	17	28
1-11	0	14	30	23	29	**31**
1-21	0	28	32	**36**	26	33
1-29	0	20	**33**	32	17	29
1-34	0	27	**35**	24	19	25
1-37	0	**37**	33	20	29	18
1-39	0	17	**39**	28	20	33
1-41	0	13	33	**36**	20	26
1-46	0	36	28	**40**	28	31
1-47	0	7	**30**	23	17	21
1-48	0	26	**33**	21	19	18
1-52	0	3	29	**37**	36	28
1-54	0	24	30	26	**31**	22
1-60	0	31	**38**	30	31	34
1-61	0	33	**39**	29	12	32
1-70	0	19	**37**	33	11	31
1-75	0	22	15	25	21	**35**
2-04	0	21	28	14	32	**34**
2-05	0	30	26	**33**	22	25
2-06	0	**35**	30	21	25	21
2-08	0	11	16	28	24	**34**
2-09	0	10	28	24	28	**33**
2-14	0	31	35	**36**	28	35
2-20	0	28	27	29	29	**34**
2-23	0	30	26	29	19	**31**
2-25	0	24	28	**30**	21	23
2-26	0	20	26	**36**	21	34
2-27	0	10	**39**	31	37	37
2-29	0	32	39	**40**	31	23
2-33	0	26	**30**	15	22	24

Table 8 (continued)

SYSTEM EFFICIENCY BASED ON ERRORS MADE IN A USER SYSTEM PER PARTICIPANT						
Participants	**System Efficiency**	**Personality Factors**				
	User	**E**	**A**	**C**	**N**	**O**
2-35	0	7	**37**	26	29	24
2-36	0	35	17	29	13	**38**
2-37	0	21	20	22	**31**	27
2-39	0	21	**36**	29	31	22
2-42	0	11	30	30	29	**33**
2-46	0	27	24	**36**	32	34
2-56	0	16	25	15	16	**40**
2-60	0	27	34	25	22	**35**
2-69	0	6	27	25	23	**36**
2-70	0	11	**35**	21	34	34
2-71	0	22	38	22	31	**39**
1-10	1	19	29	22	21	**34**
1-24	1	12	22	**32**	19	24
1-30	1	30	**38**	35	34	36
1-36	1	25	25	19	**29**	20
1-38	1	24	33	**35**	24	22
1-42	1	28	**34**	33	31	28
1-43	1	24	**40**	29	11	27
1-45	1	18	**40**	25	12	18
1-50	1	30	**36**	32	18	28
1-67	1	30	**34**	15	8	33
1-76	1	**37**	32	25	12	27
2-02	1	19	32	**35**	22	30
2-03	1	17	29	29	32	**37**
2-07	1	22	28	32	**33**	31
2-24	1	22	**40**	29	18	31
2-28	1	24	**37**	31	19	34
2-30	1	31	**36**	24	29	26
2-31	1	27	27	19	31	**29**
2-40	1	16	28	**29**	16	24
2-41	1	19	**35**	19	25	22
2-43	1	21	32	31	28	**35**
2-57	1	9	27	**30**	5	27
2-72	1	**36**	25	24	28	33
1-23	2	23	**33**	23	26	24
1-25	2	11	**30**	22	22	10
1-27	2	16	32	**38**	30	16

Table 8 (continued)

SYSTEM EFFICIENCY BASED ON ERRORS MADE IN A USER SYSTEM PER PARTICIPANT						
Participants	**System Efficiency**	**Personality Factors**				
	User	**E**	**A**	**C**	**N**	**O**
1-77	2	24	37	**39**	38	31
2-32	2	38	37	34	**39**	37
2-38	2	**29**	22	**29**	26	23
2-61	2	13	22	22	25	**26**
1-06	3	19	32	27	31	**33**
2-55	3	8	26	**34**	33	19
Totals:	43	1750	2484	2227	1973	2290

Note. Highest factor score is in bold. Errors made in user system equals efficiency – zero (0) = non-errors, one (1) = one (1) errors, two (2) = two (2) errors, three (3) = three (3) errors, and four (4) = four (4) errors.

Table 9 shows all raw score totals for personality factors. User system efficiency shows 48 non-errors for this system whose personality factor of Agreeableness - A was the highest dimension of personality. These scores pertain only to these 48 non-errors. User system efficiency shows that 23 single errors were made for this system with Agreeableness - A as the dominant personality dimension. More than two errors made for this system shows the personality dimension moves from Agreeableness - A to Emotional Stability – N.

Table 9

USER SYSTEM EFFICIENCY WITH PERSONALITY FACTOR SCORES						
(x) Number Of Errors	**Participants With (x) Number of Errors**	**Personality Factor Totals Based On (x) Number Of Errors**				
		E	**A**	**C**	**N**	**O**
0	48	1029	**1474**	1325	1198	1415
1	23	540	**739**	634	505	656
2	7	154	**213**	207	206	167
3	2	27	58	61	**64**	52
	Totals:	1750	2484	2227	1973	2290

Note. Highest factor score is in bold. The zero means non-errors.

Table 10 is a drill down of Table 9, breaking down the participants' dominant personality factors showing the number of participants in factor. For example, out of the 48 participants with non-errors (0) 17 of the participants' dominant factor was Agreeableness – A. The information in Table 9 shows the overall total scores for personality factors while Table 10 shows the dominate personality factor per participant in a user system. This is discussed later in Findings.

Table 10

DOMINATE PERSONALITY FACTORS IN A USER SYSTEM BASED ON ERRORS MADE PER PARTICIPANT						
(x) Number Of Errors	**Participants With (x) Number of Errors**	**Dominate Personality Factors**				
		E	**A**	**C**	**N**	**O**
0	48	3	17	13	2	13
1	23	2	10	5	2	4
2	7	1	2	2	1	1
3	2	0	0	1	0	1

Note. The zero means non-errors.

Table 11 shows only the mixed system by efficiency breakdown.

Table 11

SYSTEM EFFICIENCY BASED ON ERRORS MADE IN A MIXED SYSTEM PER PARTICIPANT						
Participants	**System Efficiency Mixed**	**Personality Factors**				
		E	**A**	**C**	**N**	**O**
1-02	0	13	**29**	15	19	20
1-03	0	**25**	23	20	24	20
1-05	0	13	**39**	24	30	30
1-06	0	19	32	27	31	**33**
1-08	0	27	32	37	26	**38**
1-23	0	23	**33**	23	26	24
1-24	0	12	22	**32**	19	24
1-29	0	20	**33**	32	17	29
1-30	0	30	**38**	35	34	36
1-36	0	25	25	19	**29**	20
1-37	0	**37**	33	20	29	18
1-38	0	24	33	**35**	24	22
1-39	0	17	**39**	28	20	33
1-43	0	24	**40**	29	11	27
1-46	0	36	28	**40**	28	31
1-47	0	7	**30**	23	17	21
1-48	0	26	**33**	21	19	18
1-50	0	30	**36**	32	18	28
1-60	0	31	**38**	30	31	34
1-67	0	30	**34**	15	8	33

Table 11 (continued)

SYSTEM EFFICIENCY BASED ON ERRORS MADE IN A MIXED SYSTEM PER PARTICIPANT						
Participants	**System Efficiency**	**Personality Factors**				
	Mixed	**E**	**A**	**C**	**N**	**O**
1-70	0	19	**37**	33	11	31
2-04	0	21	28	14	32	**34**
2-05	0	30	26	**33**	22	25
2-09	0	10	28	24	28	**33**
2-20	0	28	27	29	29	**34**
2-26	0	20	26	**36**	21	34
2-31	0	27	27	19	31	**29**
2-39	0	21	**36**	29	31	22
2-42	0	11	30	30	29	**33**
2-41	0	19	**35**	19	25	22
2-46	0	27	24	**36**	32	34
2-60	0	27	34	25	22	**35**
2-69	0	**6**	27	25	23	**36**
2-70	0	11	**35**	21	34	34
1-01	1	11	32	**36**	34	22
1-04	1	25	28	**33**	24	**33**
1-07	1	33	**36**	29	28	27
1-09	1	13	**40**	22	17	28
1-10	1	19	29	22	21	**34**
1-11	1	14	30	23	29	**31**
1-21	1	28	32	**36**	26	33
1-27	1	16	32	**38**	30	16
1-34	1	27	**35**	24	19	25
1-41	1	13	33	**36**	20	26
1-45	1	18	**40**	25	12	18
1-52	1	3	29	**37**	36	28
1-54	1	24	30	26	**31**	22
1-61	1	33	**39**	29	12	32
1-75	1	22	15	25	21	**35**
1-76	1	**37**	32	25	12	27
1-77	1	24	37	**39**	38	31
2-03	1	17	29	29	32	**37**
2-06	1	**35**	30	21	25	21
2-07	1	22	28	32	**33**	31
2-14	1	31	35	**36**	28	35
2-23	1	30	26	29	19	**31**

Table 11 (continued)

SYSTEM EFFICIENCY BASED ON ERRORS MADE IN A MIXED SYSTEM PER PARTICIPANT						
Participants	**System Efficiency**	**Personality Factors**				
	Mixed	**E**	**A**	**C**	**N**	**O**
2-24	1	22	**40**	29	18	31
2-27	1	10	**39**	31	37	37
2-28	1	24	**37**	31	19	34
2-29	1	32	39	**40**	31	23
2-30	1	31	**36**	24	29	26
2-33	1	26	**30**	15	22	24
2-35	1	7	**37**	26	29	24
2-37	1	21	20	22	**31**	27
2-38	1	**29**	22	**29**	26	23
2-40	1	16	28	**29**	16	24
2-55	1	8	26	**34**	33	19
2-57	1	9	27	**30**	5	27
2-61	1	13	22	22	25	**26**
2-71	1	22	38	22	31	**39**
2-72	1	**36**	25	24	28	33
1-25	2	11	**30**	22	22	10
2-02	2	19	32	**35**	22	30
2-08	2	11	16	28	24	**34**
2-25	2	24	28	**30**	21	23
2-32	2	38	37	34	**39**	37
2-43	2	21	32	31	28	**35**
2-36	3	35	17	29	13	**38**
1-42	4	28	**34**	33	31	28
2-56	4	16	25	15	16	**40**
Totals:	60	1750	2484	2227	1973	2290

Note. Highest factor score is in bold. Errors made in user system equals efficiency – zero (0) = non-errors, one (1) = one (1) errors, two (2) = two (2) errors, three (3) = three (3) errors, and four (4) = four (4) errors.

Mixed system efficiency shows 34 non-errors for this system whose personality factor of Agreeableness – A was the highest dimension of personality. These scores pertain only to these 34 non-errors. User system efficiency shows that 37 single errors were made for this system with Agreeableness - A as the dominant personality dimension. Two errors made for this system shows Conscientiousness - C as the dominant personality dimension. Three or more errors shows Intellect/Imagination – O as the dominant personality dimension.

Table 12

MIXED SYSTEM EFFICIENCY WITH PERSONALITY FACTOR SCORES						
(x) Number Of Errors	Participants With (x) Number Of Errors	Personality Factor Totals Based On (x) Number Of Errors				
		E	A	C	N	O
0	34	746	**1070**	910	830	975
1	37	801	**1163**	1060	927	1040
2	6	124	175	**180**	156	169
3	1	35	17	29	13	**38**
4	2	44	59	48	47	**68**
	Totals:	1750	2484	2227	1973	2290

Note. Highest factor score is in bold. Errors made in user system equals efficiency – zero (0) = non-errors, one (1) = one (1) errors, two (2) = two (2) errors, three (3) = three (3) errors, and four (4) = four (4) errors.

Table 13 is a drill down of Table 12, breaking down the participants' dominant personality factors. The information in Table 12 shows the overall total scores for personality factors while Table 13 shows the dominate personality factor per participant in a mixed system. This is discussed later in Findings.

Table 13

DOMINATE PERSONALITY FACTORS IN A MIXED SYSTEM BASED ON ERRORS MADE PER PARTICIPANT						
(x) Number Of Errors	Participants With (x) Number Of Errors	Dominate Personality Factors				
		E	A	C	N	O
0	34	2	16	6	1	9
1	37	4	11	12	3	7
2	6	0	1	2	1	2
3	1	0	0	0	0	1
4	2	0	1	0	0	1

Note. The zero means non-errors.

Table 14 shows only the strict system by efficiency breakdown showing the breakdown of participant by their dominate personality factor.

Table 14

SYSTEM EFFICIENCY BASED ON ERRORS MADE IN A STRICT SYSTEM PER PARTICIPANT

Participants	System Efficiency Strict	Personality Factors				
		E	A	C	N	O
1-02	0	13	**29**	15	19	20
1-03	0	**25**	23	20	24	20
1-09	0	13	**40**	22	17	28
1-23	0	23	**33**	23	26	24
1-24	0	12	22	**32**	19	24
1-30	0	30	**38**	35	34	36
1-34	0	27	**35**	24	19	25
1-37	0	**37**	33	20	29	18
1-39	0	17	**39**	28	20	33
1-41	0	13	33	**36**	20	26
1-42	0	28	**34**	33	31	28
1-43	0	24	**40**	29	11	27
1-45	0	18	**40**	25	12	18
1-46	0	36	28	**40**	28	31
1-47	0	7	**30**	23	17	21
1-61	0	33	**39**	29	12	32
1-75	0	22	15	25	21	**35**
2-02	0	19	32	**35**	22	30
2-04	0	21	28	14	32	**34**
2-08	0	11	16	28	24	**34**
2-14	0	31	35	**36**	28	35
2-23	0	30	26	29	19	**31**
2-24	0	22	**40**	29	18	31
2-25	0	24	28	**30**	21	23
2-32	0	38	37	34	**39**	37
2-35	0	7	**37**	26	29	24
2-39	0	21	**36**	29	31	22
2-40	0	16	28	**29**	16	24
2-43	0	21	32	31	28	**35**
2-46	0	27	24	**36**	32	34
2-57	0	9	27	**30**	5	27
2-69	0	6	27	25	23	**36**
2-70	1	11	**35**	21	34	34
1-01	1	11	32	**36**	34	22
1-04	1	25	28	**33**	24	**33**

Table 14 (continued)

SYSTEM EFFICIENCY BASED ON ERRORS MADE IN A STRICT SYSTEM PER PARTICIPANT						
Participants	**System Efficiency**	**Personality Factors**				
	Strict	**E**	**A**	**C**	**N**	**O**
1-05	1	13	**39**	24	30	30
1-07	1	33	**36**	29	28	27
1-08	1	27	32	37	26	**38**
1-10	1	19	29	22	21	**34**
1-11	1	14	30	23	29	**31**
1-29	1	20	**33**	32	17	29
1-48	1	26	**33**	21	19	18
1-50	1	30	**36**	32	18	28
1-52	1	3	29	**37**	36	28
1-54	1	24	30	26	**31**	22
1-60	1	31	**38**	30	31	34
1-70	1	19	**37**	33	11	31
1-77	1	24	37	**39**	38	31
2-06	1	**35**	30	21	25	21
2-09	1	10	28	24	28	**33**
2-20	1	28	27	29	29	**34**
2-26	1	20	26	**36**	21	34
2-27	1	10	**39**	31	37	37
2-28	1	24	**37**	31	19	34
2-29	1	32	39	**40**	31	23
2-30	1	31	**36**	24	29	26
2-31	1	27	27	19	31	**29**
2-38	1	**29**	22	**29**	26	23
2-41	1	19	**35**	19	25	22
2-55	1	8	26	**34**	33	19
2-60	1	27	34	25	22	**35**
2-71	1	22	38	22	31	**39**
2-72	1	**36**	25	24	28	33
1-06	2	19	32	27	31	**33**
1-21	2	28	32	**36**	26	33
1-27	2	16	32	**38**	30	16
1-38	2	24	33	**35**	24	22
1-67	2	30	**34**	15	8	33
1-76	2	**37**	32	25	12	27
2-42	2	11	30	30	29	**33**
1-25	3	11	**30**	22	22	10

Table 14 (continued)

SYSTEM EFFICIENCY BASED ON ERRORS MADE IN A STRICT SYSTEM PER PARTICIPANT						
Participants	**System Efficiency**	**Personality Factors**				
	Strict	**E**	**A**	**C**	**N**	**O**
2-03	3	17	29	29	32	**37**
2-05	3	30	26	**33**	22	25
2-07	3	22	28	32	**33**	31
2-33	3	26	**30**	15	22	24
2-61	3	13	22	22	25	**26**
1-36	4	25	25	19	**29**	20
2-37	4	21	20	22	**31**	27
2-56	4	16	25	15	16	**40**
2-36	5	35	17	29	13	**38**
Totals:	80	1750	2484	2227	1973	2290

Note. Highest factor score is in bold. Errors made in user system equals efficiency – zero (0) = non-errors, one (1) = one (1) errors, two (2) = two (2) errors, three (3) = three (3) errors, and four (4) = four (4) errors.

The strict system efficiency shows 33 non-errors for this system whose personality factor of Agreeableness - A was the highest dimension of personality. These scores pertain only to these 33 non-errors. Strict system efficiency shows that 30 single errors were made for this system with Agreeableness - A as the dominant personality dimension. More than three errors show Intellect/Imagination – O as the dominant personality dimension. See Table 15.

Table 15

STRICT SYSTEM EFFICIENCY WITH PERSONALITY FACTOR SCORES						
(x) Number of Errors	**Participants with (x) Number of Errors**	**Personality Factor Totals Based on (x) Number of Errors**				
		E	**A**	**C**	**N**	**O**
0	33	692	**1039**	921	760	937
1	30	677	**968**	862	808	878
2	7	165	**225**	206	160	197
3	6	119	**165**	153	156	153
4	3	62	70	56	76	**87**
5	1	35	17	29	13	**38**
	Totals:	1750	2484	2227	1973	2290

Note. Highest factor score is in bold. The zero means non-errors.

Table 16 is a drill down of Table 15 showing the breakdown of participant by their dominate personality factor. The results of this table will be discussed later in Findings.

Table 16

DOMINATE PERSONALITY FACTORS IN A STRICT SYSTEM BASED ON ERRORS MADE PER PARTICIPANT						
(x) Number Of Errors	**Participants With (x) Number Of Errors**	**Dominate Personality Factors**				
		E	**A**	**C**	**N**	**O**
0	33	2	15	9	1	6
1	30	3	11	7	1	8
2	7	1	1	3	0	2
3	6	0	2	1	1	2
4	3	0	0	0	2	1
5	1	0	0	0	0	1

Note. Errors made in user system equals efficiency – zero (0) = non-errors, one (1) = one (1) errors, two (2) = two (2) errors, three (3) = three (3) errors, four (4) = four (4) errors, and five (5) = five (5) errors.

RQ3: *Is an individual's preference for an SDS dialog type the same SDS dialog type at which the individual is most efficient in using?*

Raw data for system preference as to efficiency is shown in Tables 17, 18, and 19 for each system initiative.

Table 17

USER SYSTEM PREFERENCE WITH SYSTEM EFFICIENCY PER PARTICIPANT				
Participant	**System Preference**	**System Efficiency**		
		Strict	**Mixed**	**User**
1-01	User	1	1	0
1-02	User	0	0	0
1-03	User	0	0	0
1-04	User	1	1	0
1-05	User	1	0	0
1-08	User	1	0	0
1-10	User	1	1	1
1-23	User	0	0	2
1-25	User	3	2	2
1-27	User	2	1	2

Table 17 (continued)

USER SYSTEM PREFERENCE WITH SYSTEM EFFICIENCY PER PARTICIPANT				
Participant	**System Preference**	**System Efficiency**		
		Strict	**Mixed**	**User**
1-34	User	0	1	0
1-36	User	4	0	1
1-37	User	0	0	0
1-38	User	2	0	1
1-39	User	0	0	0
1-41	User	0	1	0
1-43	User	0	1	1
1-45	User	1	0	0
1-46	User	0	0	1
1-47	User	0	0	0
1-50	User	1	0	1
1-52	User	1	1	0
1-61	User	0	1	0
1-67	User	2	0	1
1-70	User	1	0	0
1-75	User	0	1	0
1-76	User	2	1	1
1-77	User	1	1	2
2-03	User	3	1	1
2-04	User	0	0	0
2-06	User	1	1	0
2-07	User	3	1	1
2-14	User	0	1	0
2-20	User	1	0	0
2-23	User	0	1	0
2-27	User	1	1	0
2-29	User	1	1	0
2-31	User	1	0	1
2-33	User	3	1	0
2-35	User	0	1	0
2-37	User	4	1	0
2-42	User	1	0	1
2-43	User	0	2	1
2-46	User	0	0	0

Table 17 (continued)

USER SYSTEM PREFERENCE WITH SYSTEM EFFICIENCY PER PARTICIPANT				
Participant	System Preference	System Efficiency		
		Strict	Mixed	User
2-56	User	4	4	0
2-57	User	0	1	1
2-60	User	1	0	0
2-61	User	3	1	2
2-70	User	0	0	0
2-72	User	1	1	1
	Totals:	53	33	25

Note. Errors made in user system equals efficiency – zero (0) = non-errors, one (1) = one (1) errors, two (2) = two (2) errors, three (3) = three (3) errors, and four (4) = four (4) errors.

Table 18

MIXED SYSTEM PREFERENCE WITH SYSTEM EFFICIENCY PER PARTICIPANT				
Participant	System Preference	System Efficiency		
		Strict	Mixed	User
1-06	Mixed	2	0	3
1-07	Mixed	1	1	0
1-11	Mixed	1	1	0
1-21	Mixed	2	1	0
1-29	Mixed	1	0	0
1-30	Mixed	0	0	1
1-54	Mixed	1	1	0
1-60	Mixed	1	0	0
2-02	Mixed	0	2	1
2-08	Mixed	0	2	0
2-24	Mixed	0	1	1
2-25	Mixed	0	2	0
2-26	Mixed	1	0	0
2-30	Mixed	1	1	1
2-41	Mixed	2	0	0
2-55	Mixed	1	1	3
2-71	Mixed	1	1	0
	Totals:	15	14	10

Table 19

STRICT SYSTEM PREFERENCE WITH SYSTEM EFFICIENCY PER PARTICIPANT				
Participant	**System Preference**	**System Efficiency**		
		Strict	**Mixed**	**User**
1-09	Strict	0	1	0
1-24	Strict	0	0	1
1-42	Strict	0	4	1
1-48	Strict	1	0	0
2-05	Strict	3	0	0
2-09	Strict	1	0	0
2-28	Strict	1	1	1
2-32	Strict	0	2	2
2-36	Strict	5	3	0
2-38	Strict	1	1	2
2-39	Strict	0	0	0
2-40	Strict	0	1	1
2-69	Strict	0	0	0
	Totals:	12	13	8

Table 20 shows system preference based on how many errors were made per participant for the user system initiative.

Table 20

USER SYSTEM PREFERENCE COMPARED TO EFFICIENCY ON A SYSTEM BY MISTAKES MADE			
System Efficiency	**User**	**Strict**	**Mixed**
0	30	20	22
1	15	17	25
2	5	4	2
3	0	5	0
4	0	4	1
5	0	0	0
Totals:	50	50	50

Table 21 shows system preference based on number of errors made per participant for the mixed system initiative.

Table 21

MIXED SYSTEM PREFERENCE COMPARED TO EFFICIENCY ON A SYSTEM BY MISTAKES MADE			
System Efficiency	**Mixed**	**Strict**	**User**
0	6	5	11
1	8	9	4
2	3	3	0
3	0	0	2
4	0	0	0
5	0	0	0
Totals:	17	17	17

Table 22 shows system preference based on how many errors were made per participant for the strict system initiative.

Table 22

STRICT SYSTEM PREFERENCE COMPARED TO EFFICIENCY ON A SYSTEM BY MISTAKES MADE			
System Efficiency	**Strict**	**Mixed**	**User**
0	7	6	7
1	4	4	4
2	0	1	2
3	1	1	0
4	0	1	0
5	1	0	0
Totals:	13	13	13

Findings

Findings for the research questions when data was analyzed using IBM SPSS version 23. Results for each research question is shown below.

RQ1: *How well does the FFM personality factor serve as a predictor of an individual's preferred SDS dialog type?*

A multivariate assessment conducted in SPSS, provided analysis of the results by using multinomial logistic regression to predict which initiative participants were likely to end up in based on the personality factors as predictors. It was necessary to run this model twice to obtain a direct multivariate comparison, as each outcome specified a single reference initiative for the outcome (e.g., “user system”), which was compared to the two other categories, strict and mixed. Since significance level could not be obtained due to small sample size for the mixed and strict systems, odds ratios were used to quantify an indicator of a dominate personality factor for the system preference and efficiency. The results given in odds ratios, estimated the change in odds of being in one category versus another for a one-unit change in the predictor in question. The findings focused more on substantive outcomes rather than *p*-values, which were deemed insignificant due to having less than 20 cases in two of the outcome categories: 13 participants who preferred a strict system over mixed or user, and 17 participants who preferred a mixed system over strict or user.

The greater the odds ratio is over one, the more substantive the outcome, with Agreeableness -A and Conscientiousness -C notable when comparing a user system to a strict system based on personality factors. The findings for the user system compared to the strict system found an Agreeableness - A odds ratio of .954. For odds ratios that are less than 1, a conversion had to be made; therefore, 1 / .954 = 1.048; indicating a one-unit increase on the Agreeableness - A scale making a participant 4.8% less likely to choose the strict system compared to the user system. Also of note were the results of the user system compared to the strict system, with a Conscientiousness - C odds ratios of 1.051 to 1, indicating a one-unit increase on the Conscientiousness - C scale making a participant 5.1% more likely to choose the strict system compared to the user system. See Table 23.

Table 23

MULTIVARIATE ASSESSMENT OF MULTINOMIAL LOGISTIC REGRESSION STRICT SYSTEM COMPARED TO USER SYSTEM									
Compared to User System		**Coefficients**	**Std. Error**	**Wald**	**df**	**Sig.**	**Odds Ratio**	**95% Confidence Interval for Odds Ratio**	
								Lower Bound	**Upper Bound**
Strict	Intercept	-.627	2.617	.057	1	.811			
	E	.001	.038	.001	1	.977	1.001	.930	1.078
	A	**-.047**	.053	.776	1	.378	**.954**	.861	1.059
	C	**.049**	.052	.883	1	.347	**1.051**	.948	1.164
	N	-.026	.044	.336	1	.562	.975	.894	1.063
	O	-.003	.052	.003	1	.954	.997	.900	1.105

When comparing a user system to a mixed system based on personality factors, notable factors were Conscientiousness – C, Intellect/Imagination – O and Emotional Stability – N. Therefore, the findings for the user system compared to the mixed system found a Conscientiousness - C odds ratio of 1.058 to 1 indicating a one-unit increase on the Conscientiousness - C scale, making a participant have 5.8% higher odds of being in the mixed category compared to user category. Findings for the user system compared to the mixed system found an Emotional Stability - N odds ratio of 1.038 to 1 indicating a one-unit increase on the Emotional Stability - N scale, making a participant have 3.8% higher odds of being in the mixed category compared to the user category. Findings show the Intellect/Imagination - O odds ratio of 1.039 to 1 indicating a one-unit increase on the Intellect/Imagination scale made a participant 3.9% more likely to choose the mixed system over the user system. See Table 24.

Table 24

MULTIVARIATE ASSESSMENT OF MULTINOMIAL LOGISTIC REGRESSION MIXED SYSTEM COMPARED TO USER SYSTEM									
Compared to User System		**Coefficients**	**Std. Error**	**Wald**	**df**	**Sig.**	**Odds Ratio**	**95% Confidence Interval For Odds Ratio**	
								Lower Bound	**Upper Bound**
Mixed	Intercept	-4.433	2.467	3.229	1	.072			
	E	-.024	.034	.477	1	.490	.977	.913	1.045
	A	.006	.052	.013	1	.908	1.006	.909	1.114
	C	**.057**	.049	1.367	1	.242	**1.058**	.962	1.164
	N	**.037**	.043	.751	1	.386	**1.038**	.954	1.128
	O	**.038**	.049	.608	1	.435	**1.039**	.943	1.145

By running the model with the strict system as the reference category, the outcome as compared to a mixed system determined Emotional Stability - N, Agreeableness - A, and Intellect/Imagination – O as having substantive odds ratios. As a result, findings for a strict system compared to a mixed system found an Emotional Stability - N odds ratio of 1.065 to 1, indicating a one-unit increase on the Emotional Stability - N scale made a participant 6.5% more likely to choose a mixed system compared to a strict system. While in the analysis of the strict system compared to a mixed system, it was found an Agreeableness - A odds ratio of 1.054 to 1, indicated a one-unit increase on the Agreeableness - A scale made a participant 5.4% more likely to choose mixed system compared to a strict system. Also of note, the Intellect/Imagination - O odds ratio of 1.042 to 1, indicating a one-unit increase on the Intellect/Imagination - O scale made a participant 4.2% more likely to choose mixed system compared to a strict system. See Table 25.

Table 25

MULTIVARIATE ASSESSMENT OF MULTINOMIAL LOGISTIC REGRESSION MIXED SYSTEM COMPARED TO STRICT SYSTEM									
Compared To Strict System		**Coefficients**	**Std. Error**	**Wald**	**df**	**Sig.**	**Odds Ratio**	**95% Confidence Interval For Odds Ratio**	
								Lower Bound	**Upper Bound**
Mixed	Intercept	-3.807	3.194	1.420	1	.233			
	E	-.025	.045	.306	1	.580	.975	.893	1.065
	A	**.053**	.065	.652	1	.419	**1.054**	.928	1.198
	C	.007	.063	.014	1	.907	1.007	.890	1.141
	N	**.063**	.054	1.320	1	.251	**1.065**	.957	1.184
	O	**.041**	.064	.421	1	.517	**1.042**	.920	1.182

RQ2: *How well does the FFM personality factor serve as a predictor of which SDS dialog type an individual used most efficiently?*

For the user system initiative, dominate personality factors however show that Agreeableness - A is the most prominent of the factors for the participants in efficiency with forty-eight non-errors. Seventeen participants had Agreeableness - A as their dominant personality factor for the least amount of errors. As a convenience, Table 10 has been added here as a quick reference to the data that is discussed in this paragraph.

Table 10

DOMINATE PERSONALITY FACTORS IN A USER SYSTEM BASED ON ERRORS MADE PER PARTICIPANT						
(x) Number Of Errors	**Participants With (x) Number of Errors**	**Dominate Personality Factors**				
		E	**A**	**C**	**N**	**O**
0	48	3	17	13	2	13
1	23	2	10	5	2	4
2	7	1	2	2	1	1
3	2	0	0	1	0	1

In Tables 26 and 27, dominate personality factors show that Agreeableness - A is the most prominent of the factors for the participants in efficiency in both the mixed and strict system initiatives.

Table 26

DOMINATE PERSONALITY FACTORS IN A MIXED SYSTEM BASED ON ERRORS MADE PER PARTICIPANT						
(x) Number Of Errors	**Participants With (x) Number Of Errors**	**Dominate Personality Factors**				
		E	**A**	**C**	**N**	**O**
0	34	2	16	6	1	9
1	37	4	11	12	3	7
2	6	0	1	2	1	2
3	1	0	0	0	0	1
4	2	0	1	0	0	1

Table 27

DOMINATE PERSONALITY FACTORS IN A STRICT SYSTEM BASED ON ERRORS MADE PER PARTICIPANT						
(x) Number Of Errors	**Participants With (x) Number Of Errors**	**Dominate Personality Factors**				
		E	**A**	**C**	**N**	**O**
0	33	2	15	9	1	6
1	30	3	11	7	1	8
2	7	1	1	3	0	2
3	6	0	2	1	1	2
4	3	0	0	0	2	1
5	1	0	0	0	0	1

Intellect/Imagination – O made fewer mistakes in a user system, and met the .05 threshold for significance with a *p*-value of .031. A one-unit increase in Intellect/Imagination - O resulted in a -.030 decrease in mistakes on a user system. See Table 28.

Table 28

USER SYSTEM EFFICIENCY FOR PERSONALITY FACTORS						
Model		Unstandardized Coefficients*		Ranked Size Of Effect	t	P-Value Sig.
		b	Std. Error	Beta		
1	(Constant)	.824	.694		1.188	.238
	E	7.836E-05	.010	.001	.008	.994
	A	-.001	.015	-.010	-.090	.928
	C	.012	.014	.101	.864	.391
	N	.012	.012	.115	.999	.321
	O	-.030	.014	-.251	-2.204	.031

* Actual Change In The Outcome Of Number Of Mistakes
The Adjusted R Square = .015, Adjustment Of How Much Variance Is Explained - Which Predictors Are Most Effective.

Agreeableness - A participants made fewer mistakes on mixed system and was close to the threshold of .05 with a *p*-value of .141. A one-unit increase in Agreeableness - A resulted in a -.025 decrease in mistakes on a mixed system.

Table 29

MIXED SYSTEM EFFICIENCY FOR PERSONALITY FACTORS						
Model		Unstandardized Coefficients*		Ranked Size Of Effect	T	P-Value Sig.
		b	Std. Error	Beta		
1	(Constant)	.923	.786		1.174	.244
	E	.003	.011	.027	.238	.813
	A	-.025	.017	-.171	-1.488	.141
	C	.011	.016	.084	.709	.481
	N	-.006	.014	-.051	-.433	.666
	O	.013	.016	.097	.840	.403

* Actual change in the outcome of number of mistakes
The adjusted r square = -.018, adjustment of how much variance is explained - which predictors are most effective.

Agreeableness - A participants made fewer mistakes on strict system and met the threshold of .05 with a *p*-value of .004. A one-unit increase in Agreeableness - A resulted in a .065 decrease in mistakes on a strict system.

Table 30

STRICT SYSTEM EFFICIENCY FOR PERSONALITY FACTORS						
Model		**Unstandardized Coefficients***		**Ranked Size Of Effect**	**T**	**P-Value Sig.**
		b	**Std. Error**	**Beta**		
1	(Constant)	2.948	1.017		2.899	.005
	E	.018	.015	.132	1.211	.230
	A	-.065	.022	-.329	-3.015	.004
	C	-.022	.020	-.125	-1.111	.270
	N	.011	.017	.068	.611	.543
	O	.002	.020	.008	.077	.939

* Actual Change In The Outcome Of Number Of Mistakes

The Adjusted R Square = .084, Adjustment Of How Much Variance Is Explained - Which Predictors Are Most Effective.

RQ3: *Is an individual's preference for an SDS dialog type the same SDS dialog type at which the individual is most efficient in using?*

Findings for an individual's preference for an SDS dialog type as compared to efficiency shows that participants who preferred a strict system were more efficient in a mixed system than the system preferred. Participants who preferred a mixed system were more efficient in a user system and those who preferred a user system were most efficient in the user system which was their preferred choice. These finding indicate that all participants were least efficient in the strict system overall. The most efficient system utilized was the user system initiative.

Table 31

DESCRIPTIVES FOR THE ONE-WAY ANOVA MEANS TEST COMPARISON OF SYSTEM EFFICIENCY TO SYSTEM PREFERENCE									
System Preference	System Efficiency	System Preference/ Participant	Efficiency (Mean)	Standard Deviation	Standard Error	95% Confidence Interval (Mean)		Number Of Mistakes	
						Lower Bound	Upper Bound	Min	Max
Strict	Strict	13	0.92	1.498	0.415	0.02	1.83	0	5
	Mixed	17	**0.82**	0.728	0.176	0.45	1.20	0	2
	User	50	1.08	1.192	0.169	0.74	1.42	0	4
	Total	80	1.00	1.158	0.130	0.74	1.26	0	5
Mixed	Strict	13	1.00	1.291	0.358	0.22	1.78	0	4
	Mixed	17	0.76	0.752	0.182	0.38	1.15	0	2
	User	50	**0.68**	0.741	0.105	0.47	0.89	0	4
	Total	80	0.75	0.849	0.095	0.56	0.94	0	4
User	Strict	13	0.62	0.768	0.213	0.15	1.08	0	2
	Mixed	17	0.59	1.004	0.243	0.07	1.10	0	3
	User	50	**0.50**	0.678	0.096	0.31	0.69	0	2
	Total	80	0.54	0.762	0.085	0.37	0.71	0	3

Summary of Results

Results from the SPSS show that although a decisive predictor for a SDS was not found, findings show that Agreeableness - A is the most prominent personality factor and that a logical assumption can be made that the Agreeableness - A factor would be an indicator of preference for a system. Percentages show that of the sample 62.5% preferred the user system initiative, while 21.3% preferred the mixed system initiative, and 16.3% preferred a strict system initiative. All systems show that Agreeableness - A was the dominant personality factor for the population sample. While a participant with Agreeableness - A as a dominant factor in their personality assessment would be more likely to prefer a user system initiative, a participant whose personality assessment contains a higher factor for Conscientiousness - C would be more likely to choose the strict system compared to the user system.

In the case of efficiency, again Agreeableness - A plays a major role in the makeup of the participant when the question of efficiency and SDS system initiative are concerned.

Participants whose personality assessment contain high scores in Agreeableness – A made fewer mistakes in all systems as well as the preferred system. However, although Agreeableness - A was the dominate personality dimension, participants with the Intellect/Imagination – O personality factor made the least errors of those making errors in the user system initiative.

In the presumption of preference and efficiency always matching, the results do not support this presumption. Participants who preferred a strict system initiative were more efficient in a mixed system while participants who preferred a mixed system were more efficient in a user system initiative. However, the participants who preferred a user system initiative were most efficient in the user system. This is the only area where the presumption of preference and efficiency matched.

Chapter 5

Conclusions, Implications, Recommendation, and Summary

Conclusions

Although a definitive predictor for a spoken dialog system was not identified, conclusions can be drawn from the results that would support certain personality factors playing an important role in the user's preference and efficiency in choosing and using a spoken dialog system. Percentages show that of the sample 62.5% (50 out of 80) preferred the user system initiative, while 21.3% (17 out of 80) preferred the mixed system initiative, and 16.3% (13 out of 80) preferred a strict system initiative. Agreeableness - A was the predominate personality factor in both preference and efficiency in all system initiatives. However, when considering efficiency, those participants with a factor of Intellect/Imagination – O made fewer mistakes in the user system initiative even though Agreeableness - A was the dominate personality factor. As for preference and efficiency always matching, the results do not support this presumption in all cases. Participants who preferred a strict system were more efficient in a mixed system while a participant who preferred a mixed system were more efficient in a user system initiative. Where a participant preferred a user system, the results support the presumption that preference and efficiency are matched. Those who preferred a user system were most efficient on their system preference. Although gender was not part of the research questions, the results of gender and preference/efficiency showed interesting results, which are discussed in the Implications section of this chapter, and may provide an area for future research.

Implications

This study enhances the research in this area where eliminating gender bias as assumed in using the FFM personality assessment as the instrument by selecting participants where gender and age match (a male participant and a female participant of the same age). Using the FFM provided a clearer picture of personality dimension through factoring which helped to show a dominating dimension of personality. However, people who volunteer in general show a higher factor for Agreeableness – A (Carlo, Okun, Knight, & de Guzman, 2005). With a population required to complete a personality assessment and system interaction, resulting data would represent a more well distributed personality assessment. This study opens future research in expanding the need to fine tune a predictor through larger population samples and participant involvement. While this study made the dialog types significantly more difficult in an effort to eliminate or decrease the ceiling effect in the foundation paper, future research will be able to utilize different system dialog types in order to improve the results.

Although not a part of the research questions for this study, preference and efficiency were analyzed by gender. The results were interesting in that male participants made more errors in the mixed system initiative than female participants. Out of the 17 participants who preferred the mixed system, nine participants were male while in the strict system out of the 13 participants who preferred this system, nine participants were male. Female participants dominated the user system initiative choice with 28 of the 50 being female. In addition, females overall had higher scores in the Agreeableness - A factor while males had higher factors scores in Emotional Stability – N. In breaking this data down by system, female participants’ dominant personality factors stayed the same

(Agreeableness – A) for all systems. Male participants' dominant personality factors, however, changed from Intellect/Imagination – O in the strict system initiative to Agreeableness – A in the mixed system initiative and back to Intellect/Imagination – O for the user system initiative. One outstanding result was that for the mixed system initiative, male participants held high factor scores in the areas of Agreeableness – A, Conscientiousness – C, and Intellect/Imagination – O with the scores being within five points of each other. This data is reflected in Table 3. Further research is suggested in this area.

Recommendations

Using the findings from this study, a key recommendation would be to recruit not only a larger population sample, but not volunteers. By utilizing a larger sample, a clearer prediction could be made as to exactly which personality factor would play a defined role in preference and efficiency for a SDS. Participants who score higher in Agreeableness - A are more likely to volunteer for a study and should be a consideration in replicating this study. Therefore, a method for having the population sample complete the personality assessment without volunteering to do so would result in a more widespread personality scope which would help to narrow the search for a definitive predictor for an SDS, producing a more accurate snapshot of the population as to personality assessment. An example of this would be to have the personality inventory as part of the log-in to a system where all users must complete the inventory before they are allowed to progress into the system. Complex dialog types were utilized in this study, however, improving on complexity will be an important part of future research keeping in mind the difference in complexity and usability. Also, data analyzed by gender shows interesting results and would be a good research direction for future study.

Summary

The overriding purpose of this study was to determine if personality was a predictor of preference and efficiency in the selection of a spoken dialog system. This study also looked at the presumption that preference and efficiency matched. A literature review was completed where research in this area showed an interest but a lack of focused work in determining predictors for SDS. Using Wolters et al., (2009) as the foundation paper for this research, where age was found not to be a predictor of preference, questions as to what would be a good predictor were formed. Recommendations in the foundation paper suggested that personality would be an area for future research in determining predictors for SDS. Using descriptive research as the research design, choosing an instrument for a personality inventory, what the population sample would look like, and how to construct a SDS for the research began to take shape and the methodology was determined. In the foundation paper, scheduling was used as the task, therefore the task for this research required the participant to schedule an airline flight with departure from one city in the United States and arrival in another city in the United States.

Three system initiatives were used as the SDS systems – 1) a strict system, where the user has limited control of the interaction, 2) a mixed system, where the user has some but not all control of the interaction, and 3) a user system, where the user has control of the interaction, see Table 32. Three confirmation attributes, also used in the foundation paper, were utilized in developing the dialog types. The confirmation attributes consisted of explicit, implicit, and none. These three initiatives and the three confirmations attributes made up the nine dialog types. In order to avoid the ceiling effect experienced by prior researchers, all dialog trees were constructed complex enough to ensure each interaction generated errors during the testing. Complexity was achieved by increasing the difficulty of the choices or options to choose from. An increase in complexity corresponded with more dialog boxes, response options, and paths. Each dialog tree was composed of multiple dialog boxes in order to capture all possible answers that a user might generate for each of the ten options that could be chosen for the particular interaction. Examples: the participant was given the option of days of the week and times available on certain days; or upgrades for seating or hotel reservation discounts. In a user system, this option was easy for the participant but in the strict system it was difficult. The user system was constructed to be difficult by adding complex dialog that required more concentration in maneuvering though the available options.

Table 32

Problem and Goal

To determine if personality (as represented by the FFM personality factor) was an effective predictor for selecting the efficient SDS dialog type for a given individual end user.

3 System Initiatives	**Five-Factor Model**	**Wizard of Oz**
• Strict - user has little control over the interaction. • Mixed - user has some ability to change dialog direction • User - system responds to direction of user (Wolters et al., 2009; Quinn & Zaiane, 2014)	Five Factors of the FFM • Extraversion – E, • Agreeableness – A, • Conscientiousness – C, • Emotional Stability – N, • Intellect/Imagination – O (McCrae & John, 1992)	By using the WoZ technique, the participant was unaware that the SDS was a real person and interacted with the SDS as designed (Li, & Bonner, 2014).

Three research questions were developed as shown below.

RQ1: *How well does the FFM personality factor serve as a predictor of an individual's preferred SDS dialog type?*

RQ2: *How well does the FFM personality factor serve as a predictor of which SDS dialog type an individual used most efficiently?*

RQ3: *Is an individual's preference for an SDS dialog type the same SDS dialog type at which the individual is most efficient in using?*

A population sample size was statistically calculated using power analysis and a minimum population sample was 80 participants with a 20% overage needed to cover participants that might not complete the study making the total recruited sample of 96. This population sample of 96 was recruited from the population of 2000 on the college campus of Milligan College, a liberal arts college in east Tennessee. A campus wide email with a link to register for a time slot to complete the application, was sent as an invitation to participate in the study. Through the link, applicants provided age and gender which made matching age with gender less time consuming. The population consisted of faculty, staff, and students which supplied the correct age range, gender distribution, and exposure to the technology needed to participate. Participants were also required to be 18 years of age at the time of the study and not over the age of 85. They were required to be native English speaking, free of any disability that would restrict them from accessing the facility or any hearing loss that would cause failure of a hearing test with or without hearing aids, or using a headset for the interaction with the SDS. These requirements were verified at the time of application.

The decision to use the FFM, the Big Five Personality Test, as the instrument in determining personality for this study, was based on how the instrument factored out certain dimensions of personality making it a quality tool for a quantitative study (Goldberg, 1992). All five personality factors (Extraversion – E, Agreeableness – A, Conscientiousness – C, Emotional Stability – N, Intellect/Imagination – O) were used to complete the personality assessment, see Table 32. In the construction of the SDS, coding the console to deliver the dialog types was completed in the Java programming language. Once the console was constructed, dialog types in the form of scripts were developed using the initiatives and confirmations matrix, keeping in mind the needed complexity to avoid a ceiling effect. These scripts were loaded into the console, randomly selected for each interaction. Participants interacted with SDS through the use of the Wizard of Oz (WoZ) method. In using the WoZ method, wizards were used to simulate the SDS interacting with the participant following a script and dialog tree chart of the SDS, see Table 32. The wizard was responsible for randomizing the system dialog types and recording on paper which dialog type each participant received making sure the researcher also recorded the correct dialog type and system for that participant. The wizards consisted of four lab assistants who had been carefully chosen for their ability to manipulate the software, agility in using the system, and their interest in the research. All wizards underwent extensive training by the researcher on the SDS and practiced initializing the setup, conducting the interaction, and ending the setup, until comfortable with the entire process.

Recruitment of 96 applicants was completed, participant requirements were validated and redacted. All personal information concerning the participants remained in the initial application. These files were immediately boxed and placed in the vault. The file folder with the forms for the Big Five Personality Test, Participant System Preference Indicators, and the observation form for efficiency used by the researcher was marked only with the gender indicator of the number one for female and the number two for male and the count number of the participant when being verified. The corresponding file number of the gender with matching age were also printed on the folder to validate that the gender had been matched for age. In this way, no personal information was tied to the file. The researcher or wizards had no personal knowledge of any participant that could be used to identify the file to the participant.

The participants interacted with all three system initiatives at this time. Wizards conducted the interaction by loading each dialog type script into the console. When the participants were comfortably seated with the headset adjusted to their liking, the systems were activated and the interaction took place. The wizards were behind a screen out of sight of the participant but in view of the researcher. The researcher made a written record of the interaction marking all miscues, hesitations, miscommunications that occurred during the interaction between the participant and the system. This record was used to determine efficiency. At the end of each system interaction, the participant would record their degree of how easy or how difficult the system. At the completion of all systems, the participant would indicate which system they preferred on the Participant System Preference Indicator. Therein, data was collected through observation, participant preference forms, and personality assessment. This file folder with the collected data was secured for analysis in a locked vault. All data was entered into Excel and analyzed in IBM SPSS version 23.

Findings from this study showed that although a definitive predictor could not be identified, personality factors do play an important role in preference and efficiency. Percentages show that of the sample 62.5% (50 out of 80) preferred the user system initiative, while 21.3% (17 out of 80) preferred the mixed system initiative, and 16.3%

(13 out of 80) preferred a strict system initiative. The personality factor of Agreeableness - A, for example, was the dominate factor in system preference and efficiency in all cases. Examining why certain personality factors play such an important role in preference and efficiency would be a candidate subject for further research.

A multivariate assessment conducted in SPSS, provided analysis of the results by using multinomial logistic regression to predict which initiative participants were likely to end up in based on the personality factors as predictors. It was necessary to run this model twice to obtain a direct multivariate comparison, as each outcome specified a single reference initiative for the outcome (e.g., “user system”), which was compared to the two other categories, strict and mixed. Because the sample was small, a definitive predictor was hard to pinpoint. However, in using odds ratios to quantify how strong an association between a personality trait and a SDS system, the results given in odds ratios, estimated the change in odds of being in one category versus another for a one-unit change in the predictor in question. The findings focused more on substantive outcomes rather than *p*-values, which were deemed insignificant due to having less than 20 cases in two of the outcome categories: 13 participants who preferred a strict system over mixed or user, and 17 participants who preferred a mixed system over strict or user.

The greater the odds ratio is over one, the more substantive the outcome, with Agreeableness - A and Conscientiousness - C notable when comparing a user system to a strict system based on personality factors. The findings for the user system compared to the strict system found an Agreeableness - A odds ratio of .954. For odds ratios that are less than 1, a conversion had to be made; therefore, 1 / .954 = 1.048; indicating a one-unit increase on the Agreeableness - A scale making a participant 4.8% less likely to choose the strict system compared to the user system. Also of note, were the results of the user system compared to the strict system, with a Conscientiousness - C odds ratios of 1.051 to 1, indication that a one-unit increase on the Conscientiousness - C scale made a participant 5.1% more likely to choose the strict system compared to the user system, see Table 33.

Table 33

Research Question 1 (RQ1):
How well does the FFM personality factor serve as a predictor of an individual's preferred SDS dialog type?

User (strict to user)	• Conscientiousness – C: odds ratio of 1.051 indicating a participant is 5.1% more likely to choose a strict system compared to a user system. • Agreeableness – A: odds ratio of 1/.954 = 1.048 indicating a participant is 4.8% less likely to choose a strict system compared to a user system. Since the odds ratio was less than one, a conversion was made in order to represent the ratio correctly.
Mixed (mixed to user)	• Conscientiousness – C: odds ratio of 1.058 indicating a participant is 5.8% more likely to choose a mixed system compared to a user system. • Intellect/Imagination – O: odds ratio of 1.039 indicating a participant is 3.9% more likely to choose a mixed system compared to a user system. • Emotional Stability – N: odds ratio of 1.038 indicating a participant is 3.8% more likely to choose a mixed system compared to a user system.
Strict (mixed to strict)	• Emotional Stability – N: odds ratio of 1.065 indicating a participant is 6.5% more likely to choose a mixed system compared to a strict system. • Agreeableness – A: odds ratio of 1.054 indicating a participant is 5.4% more likely to choose a mixed system compared to a strict system. • Intellect/Imagination – O: odds ratio of 1.042 indicating a participant is 4.2% more likely to choose a mixed system compared to a strict system.

In the case of efficiency, Agreeableness – A was the dominate personality factor in the mixed and the strict systems for the fewest mistakes made. However, for the user system, participants with a dominate personality factor of Intellect/Imagination – N made the fewest mistakes for this system. See Table 34.

Table 34

Research Question 2 (RQ2): How well does the FFM personality factor serve as a predictor of which SDS dialog type an individual used most efficiently?	
User	• Meeting the threshold of .05 with a *p*-value of .031, participants in the Intellect/Imagination - N factor made fewer mistakes in a user system. A one-unit increase in Intellect/Imagination - N resulted in a -.030 decrease in mistakes in the user system.
Mixed	• Close to the threshold of .05 with a *p*-value of .141, participants in the Agreeableness – A factor made fewer mistakes in a mixed system. A one-unit increase in Agreeableness – A resulted in a -.025 decrease in mistakes in the mixed system.
Strict	• Meeting the threshold of .05 with a *p*-value of .004, participants in the Agreeableness – A factor made fewer mistakes in a strict system. A one-unit increase in Agreeableness – A resulted in a .065 decrease in mistakes in the strict system.

As for the presumption that preference and efficiency matched, the findings show that only in the preference of the user system initiative did efficiency match. In all other systems, this presumption could not be supported. In the analysis of the data, it became clear that a larger sample size would have produced a definitive predictor for preference and efficiency. The results in this study show that personality does play a role in a user's choice in the interaction of a SDS. It shows that efficiency also is influenced by personality factors. This study opens future research in the need to fine tune a predictor by using a more varied approach to population samples such as an entire college campus or workplace where students and workers were required to complete a personality assessment. With this large population, participants matching in age with gender could then interact with complex dialog types, see Table 35.

Table 35

Research Question 3 (RQ3):
Is an individual's preference for an SDS dialog type the same as the SDS dialog type at which the individual is most efficient in using?

User	•Participants who preferred a user system were most efficient in the user system which was their preferred choice. •The most efficient system utilized was the user system initiative.
Mixed	• Participants who preferred a mixed system were more efficient in the user system which was not their preferred choice.
Strict	•Participants who preferred a strict system where more efficient in a mixed system than the system preferred. •All participants were least efficient in the strict system overall.

Although not a part of the research questions for this study, preference and efficiency were analyzed by gender. The results were interesting in that male participants made more errors in the mixed system initiative than female participants. Out of the 17 participants who preferred the mixed system, nine participants were male and in the strict system out of the 13 participants who preferred this system, nine participants were also male. Female participants dominated the user system initiative choice with 28 of the 50 being female. In addition, females overall had higher scores in the Agreeableness - A factor while males had higher factors scores in Emotional Stability – N. In breaking this data down by system, female participants' dominant personality factors stayed the same (Agreeableness – A) for all systems. Male participants' dominant personality factors, however, changed from Intellect/Imagination – O in the strict system initiative to Agreeableness – A in the mixed system initiative and back to Intellect/Imagination – O for the user system initiative. One outstanding result was that for the mixed system initiative, male participants held high factor scores in the areas of Agreeableness – A, Conscientiousness – C, and Intellect/Imagination – O with the scores being within five points of each other. The question of whether gender plays a part in the predicting of a SDS is for future research.

In conclusion, although a definitive predictor for a SDS system was not readily apparent, analysis run using odds ratio does show a strong association between certain personality dimensions and SDS dialog types. Further research would show a predictor when using a larger population where participants are required to interact with the SDS, not volunteer. In addition, data collected by gender shows an area of study not yet researched that holds promising results for predicting SDS preference and efficiency.

Appendix A

Adult/General Informed Consent

NSU

Consent Form for Participation in a Research Study Entitled
Five-Factor Model as a Predictor for Spoken Dialog Systems

Funding Source: None

IRB protocol #:

Principal investigator(s)
Teresa A. Carter, MCIS
1224 Centenary Rd
Kingsport, TN 37663
(423) 323-9218

Co-investigator
Maxine Cohen, Ph.D.
Carl DeSantis Building, 4th floor
Nova Southeastern University
3301 College Avenue
Ft. Lauderdale, FL 33314
(954) 262-2072

For questions/concerns about your research rights, contact:
Human Research Oversight Board (Institutional Review Board or IRB)
Nova Southeastern University
(954) 262-5369/Toll Free: 866-499-0790
IRB@nsu.nova.edu

What is the study about?

This study involves research in the area of predictors of user interaction style. The purpose of the study is identifying an effective predictor of characteristic patterns of behavior that may well produce equal or better results for delineating user groups for enhanced interface design for the spoken dialog systems.

Why are you asking me?

Subjects for this study must be a native English speaker and range in age from eighteen to eighty-five. You must be free of any physical disability that would hinder them from accessing the lab location or equipment or any hearing disability that would restrict use of a headset. You are asked to join in this study because you meet the qualifications of the subject described above. Ninety-six participants will be needed for this endeavor.

What will I be doing if I agree to be in the study?

As a qualifying applicant you will required to sign this consent form after reading it in its entirety if you wish to participate. Your participant profile will then be created. All participants will be redacted at this point. All instructions will be read to you before any part of the study is administered. You will be asked to take a standard hearing test comprised of the standard tone and whisper test administered by a registered nurse. If you pass the standard hear test, you will then be required to take and complete a personality test (the Five-Factor Method Personality Inventory). Once completed you will then perform the spoken dialog test where the participant will interact with a voice simulator, which allows the user to schedule appointments. No procedures are experimental, and the expected duration of your participation will be no longer than sixty to ninety minutes. There is no anticipated follow-up. No part of this procedure is likely to cause you stress, pain, or any other unpleasant reaction. The only anticipated circumstances, under which your participation may be terminated by the investigator without regard to your consent, will be failure to pass the standard hearing test.

Is there any audio or video recording?

This research project will not include audio and video recording of the participant's interaction with the simulator or any other part of the participant process.

What are the dangers to me?

Participation in this research involves minimal risk. There will be no harm or physical stress to you as all work will be done in the comfort of a lab environment. Since participation involves minimal risk, there is minimal expectation of any injuries. There is a minimal risk of loss of confidentiality; however, every precaution will be taken to prevent this from happening. All data collected will be stored in a lock container in a locked vault. Any data not used in the study will be immediately incinerated. After three years all data collected until otherwise instructed by the IRB committee, will be incinerated. There is the inconvenience of loss of time, however, all interactions will be scheduled for your convenience, and every consideration will be taken to insure the least amount of time loss. No compensation or medical treatments are expected nor should be needed. If you have further questions or concerns about the risks or benefits of participation in this study, please contact the principal investigator, Teresa Carter at 423-782-6559 or 423-461-8411, or the IRB office at the numbers indicted on page 1.

Are there any benefits for taking part in this research study?

There are no direct benefits.

Will I get paid for being in the study? Will it cost me anything?

There will be no payments made to you for participating in this study. There are no costs to you for your participation in this study.

How will you keep my information private?

Confidentiality is a major concern in any research study containing human subjects. All information obtained in this study is strictly confidential unless disclosure is required by law. The IRB and regulatory agencies may review research records. All data will be stored in a locked container in a locked vault. Any data not used in the research will be immediately incinerated. All data will be stored for three years, after which the data will be incinerated unless otherwise instructed by the IRB committee.

What if I do not want to participate or I want to leave the study?

You have the right to leave this study at any time or refuse to participate. If you do decide to leave or you decide not to participate, you will not experience any penalty or loss of services you have a right to receive. If you choose to withdraw, any information collected about you before the date you leave the study will be destroyed.

Other Considerations:
If significant new information relating to the study becomes available, which may relate to your willingness to continue to participate, this information will be provided to you by the investigators.

Voluntary Consent by Participant:
By signing below, you indicate that

- this study has been explained to you
- you have read this document or it has been read to you
- your questions about this research study have been answered
- you have been told that you may ask the researchers any study related questions in the future or contact them in the event of a research-related injury
- you have been told that you may ask Institutional Review Board (IRB) personnel questions about your study rights
- you are entitled to a copy of this form after you have read and signed it
- you voluntarily agree to participate in the study entitled "" Five-Factor Model as a Predictor for Spoken Dialog Systems"

Participant's Signature: ____________________________ Date: ________________

Participant's Name: _______________________________ Date: ________________

Signature of Person Obtaining Consent: ______________________________

Date: __________________________________

Appendix B
Letter of Participation

Letter of Participation in a Research Study Entitled
Five-Factor Model as a Predictor for Spoken Dialog Systems

Principal investigator(s)
Teresa A. Carter, MCIS
1224 Centenary Rd
Kingsport, TN 37663
(423) 323-9218

Co-investigator
Maxine Cohen, Ph.D.
Carl DeSantis Building, 4th floor
Nova Southeastern University
3301 College Avenue
Ft. Lauderdale, FL 33314
(954) 262-2072

For questions/concerns about your research rights, contact:
Human Research Oversight Board (Institutional Review Board or IRB)
Nova Southeastern University
(954) 262-5369/Toll Free: 866-499-0790
IRB@nsu.nova.edu

Description of Study: Teresa A Carter is a doctoral student at Nova Southeastern University engaged in research for the purpose of satisfying a requirement for a Doctor of Information Systems. The purpose of this study is to determine if personality (as represented by the Five-Factor Model (FFM) personality factor) is an effective predictor for selecting the efficient spoken dialog system (SDS) dialog type for a given individual end user

If you agree to participate, you will be asked to complete the attached forms and questionnaire. The first form is an Adult Consent Form. It will be explained in detail and if at any time you wish to leave or not participant in the study, you may do so without question. The second form is the participant screening form, which establishes that as a participant, you speak English as your native language; that you are between the age of 18 and 85; that you have no disability that would prevent your access to the area where the observation will take place; and that you can pass a standard tone and whisper hearing test with or without hearing aids. The questionnaire is the FFM personality factor form. The data from this questionnaire will be used to categorize all participant personality factors. All personal identification is removed from this document. This data will also be used to establish guidelines for selecting and pairing mentors with beginning teachers. The questionnaire will take approximately fifteen minutes to complete.

Risks/Benefits to the Participant: There may be minimal risk involved in participating in this study. There are no direct benefits to for agreeing to be in this study. The two main risks for you as a participant will be loss of confidentiality and loss of time. All precautions have been made to insure your personal identity is protected. All data is stored in a locked container in a locked vault. In order to insure your time is well spent in this study, scheduling your participation will be at your convenience and every consideration will be taken to make sure you do not spend any more time than required to complete the study. If you have any concerns about the risks/benefits of participating in this study, you can contact the investigators and/or the university's human research oversight board (the Institutional Review Board or IRB) at the numbers listed above.

Cost and Payments to the Participant: There is no cost for participation in this study. Participation is completely voluntary and no payment will be provided.

Confidentiality: Information obtained in this study is strictly confidential unless disclosure is required by law. All data will be secured in a locked container in a locked vault. Your name will not be used in the reporting of information in publications or conference presentations.

Participant's Right to Withdraw from the Study: You have the right to refuse to participate in this study and the right to withdraw from the study at any time without penalty.

I have read this letter and I fully understand the contents of this document and voluntarily consent to participate. All of my questions concerning this research have been answered. If I have any questions in the future about this study they will be answered by the investigator listed above or his/her staff.

I understand that the completion of this questionnaire implies my consent to participate in this study.

Participant's Signature: ______________________________Date:______________

Participant's Name: _________________________________

Initials: _________________ Date: _________________

Appendix C
Applicant Screening Form

Principal investigator(s)
Teresa A. Carter, MCIS
1224 Centenary Rd, Kingsport, TN 37663
(423) 323-9218

Please fill out this questionnaire in its entirety and return to the attendant. Please print when asked for information other than yes or no. The following questions are only for demographics.

Name: ____

Address: ____

In case we need to contact you, would you please supply us with your phone or email address?

Phone: ______________________ **Email: ______________________**

Gender: _______male _________female

Please check yes or no for the following questions. The following questions are for qualification of participants.

Is English your native language? __________Yes __________No

Are you between the age of 18 and 85 years? ___________Yes ___________No

Do you wear any device that aids in hearing? ____________Yes ___________No

Are you free of any physical disability that would hinder you from walking, climbing stairs, sitting at a desk, wearing a headset (earphones) if needed or speaking audibly into a microphone?
_________Yes _______No

Printed Name: ____

Signature: ____

Date: __________________________________

If you are not a qualifying participant this information sheet will be destroyed.
No personal information given on this form will be distributed in any way.

Appendix D
Standard Hearing Test – Tone and Whisper

Tuning Fork

A tuning fork is a metal, two-pronged device that produces a tone when it vibrates. The health professional strikes the tuning fork to make it vibrate and produce a tone. These tests assess how well sounds move through the ear. Sometimes the tuning fork will be placed on the participant's head or behind s/he ear. Depending on how the participant hears the sound, the health professional can tell if there is a problem with the nerves in the ear or with sound getting to nerves (Rhodes, 2009).

Whispered speech test

In a whispered speech test, the health professional will ask the participant to cover the opening of one ear with a finger. The health professional will stand 1ft to 2ft behind the participant and whisper a series of words. The participant will repeat the words that they hear. If they cannot hear the words at a soft whisper, the health professional will keep saying the words more loudly until the participant can hear them. Each ear is tested separately (Rhodes, 2009).

Hearing Test Results

Normal range of hearing results when the participant exhibits the following:

- Participant is able to hear whispered speech accurately
- Participant can hear tones at equal loudness in both ears

Appendix E
The Big Five Personality Test

from personality-testing.info courtesy ipip.ori.org

Introduction

This is a personality test, it will help you understand why you act the way that you do and how your personality is structured. Please follow the instructions below, scoring and results are on the next page.

Instructions

In the table below, for each statement 1-50 mark how much you agree with on the scale 1-5, where 1=disagree, 2=slightly disagree, 3=neutral, 4=slightly agree and 5=agree, in the box to the left of it.

Test

Rating	I....	Rating	I.....
	1. Am the life of the party.		26. Have little to say.
	2. Feel little concern for others.		27. Have a soft heart.
	3. Am always prepared.		28. Often forget to put things back in their proper place.
	4. Get stressed out easily.		29. Get upset easily.
	5. Have a rich vocabulary.		30. Do not have a good imagination.
	6. Don't talk a lot.		31. Talk to a lot of different people at parties.
	7. Am interested in people.		32. Am not really interested in others.
	8. Leave my belongings around.		33. Like order.
	9. Am relaxed most of the time.		34. Change my mood a lot.
	10. Have difficulty understanding abstract ideas.		35. Am quick to understand things.
	11. Feel comfortable around people.		36. Don't like to draw attention to myself.
	12. Insult people.		37. Take time out for others.
	13. Pay attention to details.		38. Shirk my duties.
	14. Worry about things.		39. Have frequent mood swings.

	15. Have a vivid imagination.		40. Use difficult words.
	16. Keep in the background.		41. Don't mind being the center of attention.
	17. Sympathize with others' feelings.		42. Feel others' emotions.
	18. Make a mess of things.		43. Follow a schedule.
	19. Seldom feel blue.		44. Get irritated easily.
	20. Am not interested in abstract ideas.		45. Spend time reflecting on things.
	21. Start conversations.		46. Am quiet around strangers.
	22. Am not interested in other people's problems.		47. Make people feel at ease.
	23. Get chores done right away.		48. Am exacting in my work.
	24. Am easily disturbed.		49. Often feel blue.
	25. Have excellent ideas.		50. Am full of ideas.

- **Extroversion (E)** is the personality trait of seeking fulfillment from sources outside the self or in community. High scorers tend to be very social while low scorers prefer to work on their projects alone.
- **Agreeableness (A)** reflects much individuals adjust their behavior to suit others. High scorers are typically polite and like people. Low scorers tend to 'tell it like it is'.
- **Conscientiousness (C)** is the personality trait of being honest and hardworking. High scorers tend to follow rules and prefer clean homes. Low scorers may be messy and cheat others.
- **Neuroticism (N)** is the personality trait of being emotional.
- **Openness to Experience (O)** is the personality trait of seeking new experience and intellectual pursuits. High scores may day-dream a lot. Low scorers may be very down to earth.

Appendix F
Participant System Preference Indicators

At the end of each system interaction you will be asked to rate the system for difficulty. When you have interacted with all three systems, you will be asked to answer the last question indicating which system of the three you preferred.

System 1 – How difficult was this interaction? Please indicate below by circling one of the choices.

1. Very Difficult   2. Not very difficult   3. Relatively easy   4. Easy, no problem at all

System 2 – How difficult was this interaction? Please indicate below by circling one of the choices.

1. Very Difficult   2. Not very difficult   3. Relatively easy   4. Easy, no problem at all

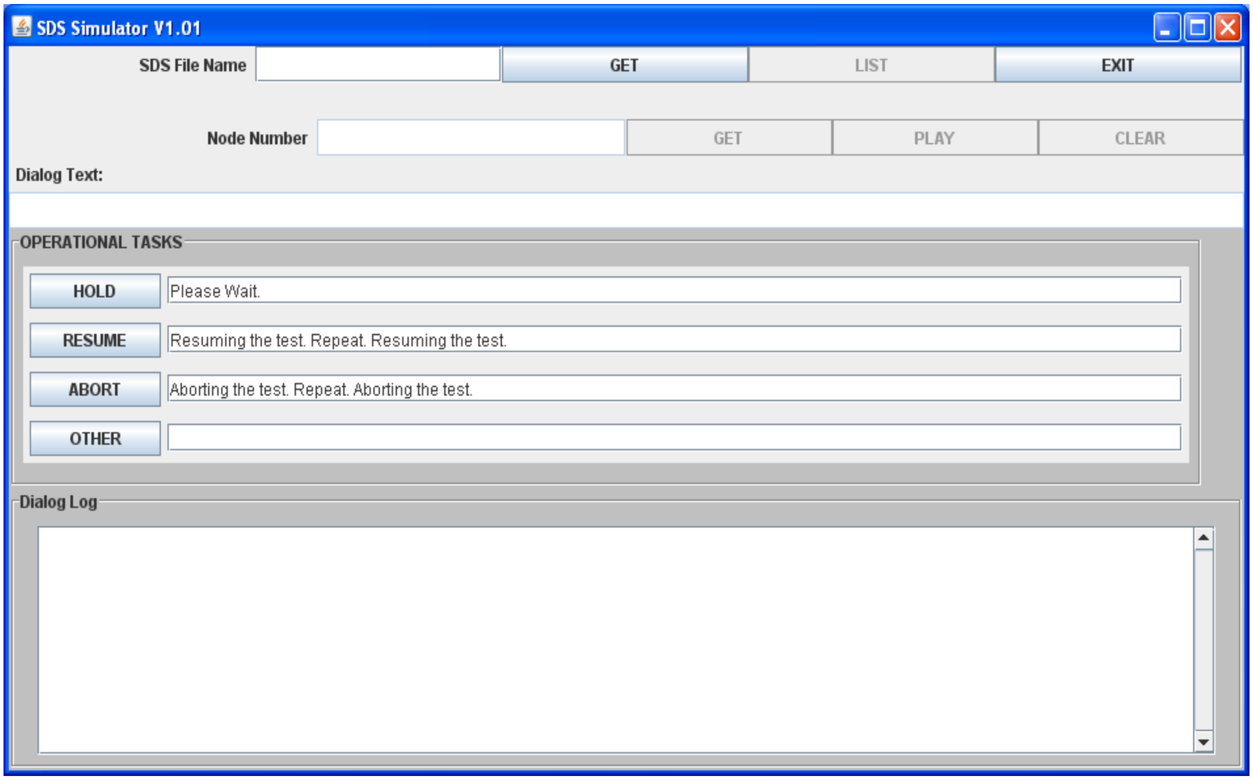
System 3 – How difficult was this interaction? Please indicate below by circling one of the choices.

1. Very Difficult   2. Not very difficult   3. Relatively easy   4. Easy, no problem at all

Now that you have experienced all three systems, which system do you most prefer? Please indicate below by circling one of the choices.

System 1   **System 2**   **System 3**

Appendix G
SDS Simulator Console



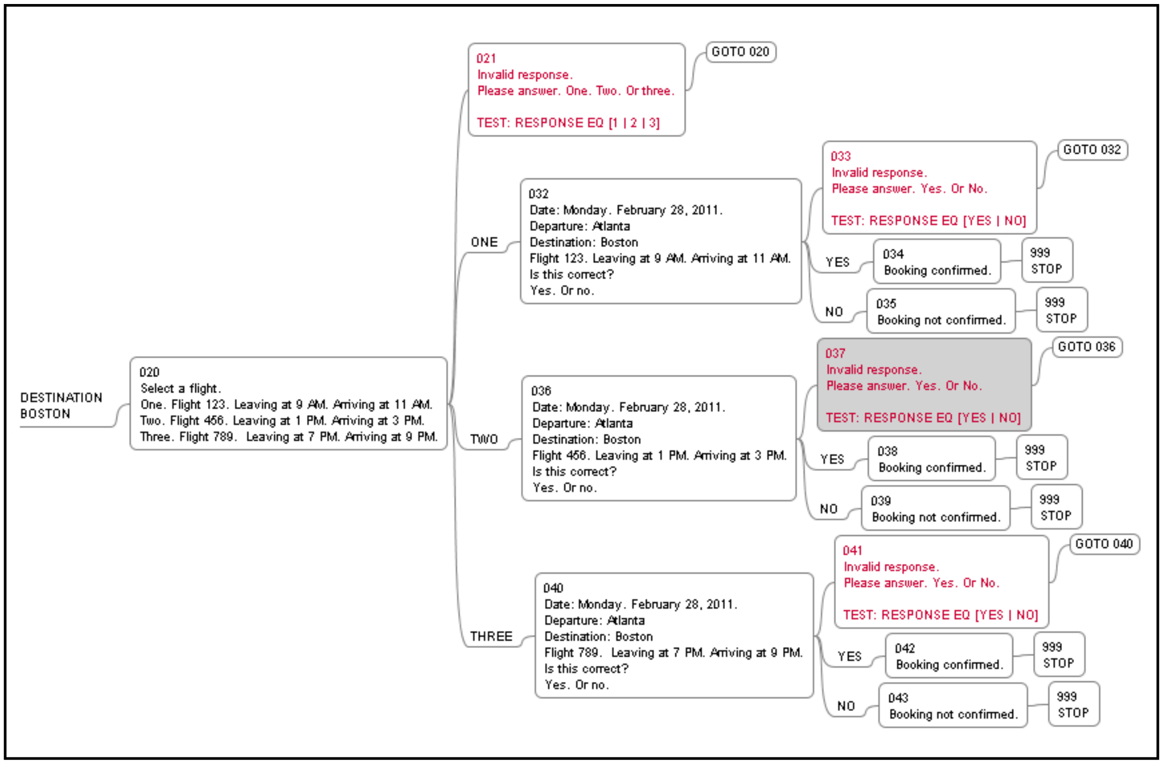
Options	Function
SDS File Name	Holds the file name for the SDS dialog type in the text box.
GET (SDS File Name)	Pressing this button retrieves the file name type in the **SDS File Name** text box.
LIST (SDS File Name)	Pressing this button lists all the node numbers and dialog text for the SDS dialog type specified in the **SDS File Name**.
EXIT	Exit the program.
Node Number	Node number for the SDS dialog type specified in the **SDS File Name**.
GET (Node Number)	Pressing this button retrieves the dialog text of the **Node Number** specified for the **SDS File Name**.
PLAY (Node Number)	Pressing this button plays the dialog text of the **Node Number** specified for the **SDS File Name**.
CLEAR (Node Number)	Pressing this button clears the contents of the **Node Number** text box.
HOLD	Pressing this button plays the hold message found in the adjoining text box. Default contents of the text box can be changed.
RESUME	Pressing this button plays the resume message found in the adjoining text box. Default contents of the text box can be changed.
ABORT	Pressing this button plays the abort message found in the adjoining text box. Default contents of the text box can be changed.
OTHER	Pressing this button plays the other message found in the adjoining text box. Default contents of the text box can be changed.
Dialog Log	Provides a visual history of all the node numbers and dialog text which were played by the master wizard running the program. This field can be scrolled.

Appendix H
SDS Simulator Dialog File

001My name is Otto. I'm here to help you schedule your flight today.
002Begin test. S-D-S-0-2. Say Back To go back to previous menu.
003Ok, let’s get started. Please say month, day, and year of your departure date.
004I'm sorry, please repeat. Please say month, day, and year.
005Ok, sounds good.
006Will you be flying alone or will you need multiple tickets for this flight?
007Great!
008How many adult tickets will you need?
009Ok, sounds good.
010Will you need a child's ticket? Remember, children under the age of two can sit in your lap for free.
011I understand. How many children's tickets will you need?
012Thank you.
013Will you depart from Atlanta, Boston, or Chicago.
014Will you depart from Miami, Houston, or New York?
015I'm sorry, but your entry is incorrect. Say "Back" for previous menu.
016Ok, sounds good.
017In which city will you be arriving? Miami, Houston, or New York.
018In which city will you be arriving? Atlanta, Boston, or Chicago.
019I'm sorry, but your entry is incorrect. Say "Back" for previous menu.

Appendix I
SDS Dialog Type

Using the FreeMind software application, an example of a SDS dialog type was translated into a decision tree. Below is a termination branch which would be part of a SDS dialog type.



Appendix J
IRB Approval
MEMORANDUM

To: Teresa Carter, MS
College of Engineering and Computing

From: Matthew Seamon, Pharm.D., JD
Chair, Institutional Review Board

Date: December 17, 2015

Re: *Five Factor Model as Predictor for Spoken Dialog* – NSU IRB No. 01201515Exp.

I have reviewed the above-referenced research protocol in keeping with Continuing Review requirements by an expedited procedure. On behalf of the Institutional Review Board of Nova Southeastern University, *Five Factor Model as Predictor for Spoken Dialog* is approved. Your study is approved on **January 15, 2016** and is approved until **January 14, 2017**. You are required to submit for continuing review by **December 14, 2016.** As principal investigator, you must adhere to the following requirements:

1) CONSENT: You must use the stamped (dated consent forms) attached when consenting subjects. The consent forms must indicate the approval and its date. The forms must be administered in such a manner that they are clearly understood by the subjects. The subjects must be given a copy of the signed consent document, and a copy must be placed with the subjects' confidential chart/file.

2) ADVERSE EVENTS/UNANTICIPATED PROBLEMS: The principal investigator is required to notify the IRB chair of any adverse reactions that may develop as a result of this study. Approval may be withdrawn if the problem is serious.

3) AMENDMENTS: Any changes in the study (e.g., procedures, consent forms, investigators, etc.) must be approved by the IRB prior to implementation.

4) CONTINUING REVIEWS: A continuing review (progress report) must be submitted by the continuing review date noted above. Please see the IRB web site for continuing review information.

5) FINAL REPORT: You are required to notify the IRB Office within 30 days of the conclusion of the research that the study has ended via the IRB Closing Report form.

The NSU IRB is in compliance with the requirements for the protection of human subjects prescribed in Part 46 of Title 45 of the Code of Federal Regulations (45 CFR 46) revised June 18, 1991.

Cc: Dr. Maxine Cohen
Dr. Ling Wang
Mr. William Smith

NOVA SOUTHEASTERN UNIVERSITY
Institutional Review Board

MEMORANDUM

To: Teresa Carter, M.S.
Graduate School of Computer and Information Sciences

From: Matthew Seamon, Pharm.D., JD RD for MS.
Chair, Institutional Review Board

Date: January 27, 2015

Re: *Five Factor Model as a Predictor for Spoken Dialog* – NSU IRB No. 01201515Exp.

I have reviewed the revisions to the above-referenced research protocol by an expedited procedure. On behalf of the Institutional Review Board of Nova Southeastern University, *Five Factor Model as a Predictor for Spoken Dialog* is approved in keeping with expedited review category #4. Your study is approved on **January 26, 2015** and is approved until **January 25, 2016**. You are required to submit for continuing review by **December 25, 2015**. As principal investigator, you must adhere to the following requirements:

1) CONSENT: You must use the stamped (dated consent forms) attached when consenting subjects. The consent forms must indicate the approval and its date. The forms must be administered in such a manner that they are clearly understood by the subjects. The subjects must be given a copy of the signed consent document, and a copy must be placed with the subjects' confidential chart/file.

2) ADVERSE EVENTS/UNANTICIPATED PROBLEMS: The principal investigator is required to notify the IRB chair of any adverse reactions that may develop as a result of this study. Approval may be withdrawn if the problem is serious.

3) AMENDMENTS: Any changes in the study (e.g., procedures, consent forms, investigators, etc.) must be approved by the IRB prior to implementation.

4) CONTINUING REVIEWS: A continuing review (progress report) must be submitted by the continuing review date noted above. Please see the IRB web site for continuing review information.

5) FINAL REPORT: You are required to notify the IRB Office within 30 days of the conclusion of the research that the study has ended via the IRB Closing Report form.

The NSU IRB is in compliance with the requirements for the protection of human subjects prescribed in Part 46 of Title 45 of the Code of Federal Regulations (45 CFR 46) revised June 18, 1991.

Cc: Dr. Maxine Cohen
Dr. Ling Wang
Mr. William Smith

3301 College Avenue • Fort Lauderdale, Florida 33314-7796
(954) 262-5369 • Fax: (954) 262-3977 • Email: *irb@nova.edu* • Web site: *www.nova.edu/irb*

NOVA SOUTHEASTERN UNIVERSITY
Graduate School of Computer and Information Sciences

NOVA SOUTHEASTERN UNIVERSITY
Institutional Review Board
Approval Date: JAN 2 6 2015
Continuing Review Date: JAN 2 5 2016

Consent Form for Participation in a Research Study Entitled
Five-Factor Model as a Predictor for Spoken Dialog Systems

Funding Source: None

IRB protocol #:

Principal investigator(s)
Teresa A. Carter, MCIS
1224 Centenary Rd
Kingsport, TN 37663
(423) 323-9218

Co-investigator
Maxine Cohen, Ph.D.
Carl DeSantis Building, 4th floor
Nova Southeastern University
3301 College Avenue
Ft. Lauderdale, FL 33314
(954) 262-2072

For questions/concerns about your research rights, contact:
Human Research Oversight Board (Institutional Review Board or IRB)
Nova Southeastern University
(954) 262-5369/Toll Free: 866-499-0790
IRB@nsu.nova.edu

What is the study about?
This study involves research in the area of predictors of user interaction style. The purpose of the study is identifying an effective predictor of characteristic patterns of behavior that may well produce equal or better results for delineating user groups for enhanced interface design for the spoken dialog systems.

Why are you asking me?
Subjects for this study must be a native English speaker and range in age from eighteen to ninety-nine. You must be free of any physical disability that would hinder them from accessing the lab location or equipment or any hearing disability that would restrict use of a headset. You are asked to join in this study because you meet the qualifications of the subject described above. Ninety-six participants will be needed for this endeavor.

What will I be doing if I agree to be in the study?
As a qualifying applicant you will required to sign this consent form after reading it in its entirety if you wish to participate . Your participant profile will then be created. All participants will be redacted at this point. All instructions will be read to you before any part of the study is administered. You will be asked to take a standard hearing test comprised of the standard tone and whisper test administered by a registered nurse. If you pass the standard hear test, you will then be required to take and complete a personality test (the Five-Factor Method Personality Inventory).

Initials: _______________ Date:

3301 College Avenue • Fort Lauderdale, Florida 33314-7796 • (954) 262-2000 • 800-541-6682, ext. 2000
Fax: (954) 262-3915 • Web site: *www.scis.nova.edu*

Once completed you will then perform the spoken dialog test where the participant will interact with a voice simulator, which allows the user to schedule appointments. No procedures are experimental, and the expected duration of your participation will be no longer than sixty to ninety minutes. There is no anticipated follow-up. No part of this procedure is likely to cause you stress, pain, or any other unpleasant reaction.

The only anticipated circumstances, under which your participation may be terminated by the investigator without regard to your consent, will be failure to pass the standard hearing test.

Is there any audio or video recording?
This research project will not include audio and video recording of the participant's interaction with the simulator or any other part of the participant process.

What are the dangers to me?
Participation in this research involves minimal risk. There will be no harm or physical stress to you as all work will be done in the comfort of a lab environment. Since participation involves minimal risk, there is minimal expectation of any injuries. There is a minimal risk of loss of confidentiality; however, every precaution will be taken to prevent this from happening. All data collected will be stored in a lock container in a locked vault. Any data not used in the study will be immediately incinerated. After three years all data collected until otherwise instructed by the IRB committee, will be incinerated. There is the inconvenience of loss of time, however, all interactions will be scheduled for your convenience, and every consideration will be taken to insure the least amount of time loss. No compensation or medical treatments are expected nor should be needed. If you have further questions or concerns about the risks or benefits of participation in this study, please contact the principal investigator, Teresa Carter at 423-782-6559 or 423-461-8411, or the IRB office at the numbers indicted on page 1.

Are there any benefits for taking part in this research study?
There are no direct benefits.

Will I get paid for being in the study? Will it cost me anything?
There will be no payments made to you for participating in this study. There are no costs to you for your participation in this study.

How will you keep my information private?
Confidentiality is a major concern in any research study containing human subjects. All information obtained in this study is strictly confidential unless disclosure is required by law. The IRB and regulatory agencies may review research records. All data will be stored in a locked container in a locked vault. Any data not used in the research will be immediately incinerated. All data will be stored for three years, after which the data will be incinerated unless otherwise instructed by the IRB committee.

What if I do not want to participate or I want to leave the study?
You have the right to leave this study at any time or refuse to participate. If you do decide to leave or you decide not to participate, you will not experience any penalty or loss of services you have a right to receive. If you choose to withdraw, any information collected about you before the date you leave the study will be destroyed.

Initials: _ Date:

Other Considerations:
If significant new information relating to the study becomes available, which may relate to your willingness to continue to participate, this information will be provided to you by the investigators.

Voluntary Consent by Participant:
By signing below, you indicate that

- this study has been explained to you
- you have read this document or it has been read to you
- your questions about this research study have been answered
- you have been told that you may ask the researchers any study related questions in the future or contact them in the event of a research-related injury
- you have been told that you may ask Institutional Review Board (IRB) personnel questions about your study rights
- you are entitled to a copy of this form after you have read and signed it
- you voluntarily agree to participate in the study entitled “”Five-Factor Model as a Predictor for Spoken Dialog Systems”

Participant's Signature: Date:

Participant’s Name: Date: __

Signature of Person Obtaining Consent:

Date:

NOVA SOUTHEASTERN UNIVERSITY
Graduate School of Computer and Information Sciences

NSU

NOVA SOUTHEASTERN UNIVERSITY
Institutional Review Board
Approval Date: JAN 26 2015
Continuing Review Date: JAN 25 2016

Letter of Participation in a Research Study Entitled
Five-Factor Model as a Predictor for Spoken Dialog Systems

Principal investigator(s)
Teresa A. Carter, MCIS
1224 Centenary Rd
Kingsport, TN 37663
(423) 323-9218

Co-investigator
Maxine Cohen, Ph.D.
Carl DeSantis Building, 4th floor
Nova Southeastern University
3301 College Avenue
Ft. Lauderdale, FL 33314
(954) 262-2072

For questions/concerns about your research rights, contact:
Human Research Oversight Board (Institutional Review Board or IRB)
Nova Southeastern University
(954) 262-5369/Toll Free: 866-499-0790
IRB@nsu.nova.edu

Description of Study: Teresa A Carter is a doctoral student at Nova Southeastern University engaged in research for the purpose of satisfying a requirement for a Doctor of Information Systems. The purpose of this study is to determine if personality (as represented by the Five-Factor Model (FFM) personality factor) is an effective predictor for selecting the efficient spoken dialog system (SDS) dialog type for a given individual end user

If you agree to participate, you will be asked to complete the attached forms and questionnaire. The first form is an Adult Consent Form. It will be explained in detail and if at any time you wish to leave or not participant in the study, you may do so without question. The second form is the participant screening form, which establishes that as a participant, you speak English as your native language; that you are between the age of 18 and 85; that you have no disability that would prevent your access to the area where the observation will take place; and that you can pass a standard tone and whisper hearing test with or without hearing aids. The questionnaire is the FFM personality factor form. The data from this questionnaire will be used to categorize all participant personality factors. All personal identification is removed from this document. This data will also be used to establish guidelines for selecting and pairing mentors with beginning teachers. The questionnaire will take approximately fifteen minutes to complete.

Risks/Benefits to the Participant: There may be minimal risk involved in participating in this study. There are no direct benefits to for agreeing to be in this study. The two main risks for you as a participant will be loss of confidentiality and loss of time. All precautions have been made to insure your personal identity is protected. All data is stored in a locked container in a locked vault. In order to insure your time is well spent in this study, scheduling your participation will be at your convenience and every consideration will be taken to make sure you do not spend any more time than required to complete the study. If you have any concerns about the risks/benefits of participating in this study, you can contact the investigators and/or the university's human research oversight board (the Institutional Review Board or IRB) at the numbers listed above.

Cost and Payments to the Participant: There is no cost for participation in this study. Participation is completely voluntary and no payment will be provided.

Confidentiality: Information obtained in this study is strictly confidential unless disclosure is required by law. All data will be secured in a locked container in a locked vault. Your name will not be used in the reporting of information in publications or conference presentations.

Participant's Right to Withdraw from the Study: You have the right to refuse to participate in this study and the right to withdraw from the study at any time without penalty.

I have read this letter and I fully understand the contents of this document and voluntarily consent to participate. All of my questions concerning this research have been answered. If I have any questions in the future about this study they will be answered by the investigator listed above or his/her staff.

I understand that the completion of this questionnaire implies my consent to participate in this study.

Reference List

Adelman, L. (1991). Experiments, Quasi-Experiments, and Case Studies: A review of empirical methods for evaluation decision support systems. *IEEE Transactions on Systems, Man, and Cybernetics, 21*(2), 293-301.

Ashok, V., Borodin, Y., Stoyanchev, S., Puzis, Y., & Ramakrishnan, I. V. (2014). *Wizard-of-Oz evaluation of speech-driven web browsing interface for people with vision impairments*. Paper presented at the Proceedings of the 11th Web for All Conference, Seoul, Korea.

Bartlett, J., Kotrlik, J., & Higgins, C. (2001). Organizational research: Determining appropriate sample size in survey research. *Information Technology, Learning, and Performance Journal*, *19*(1), 43-50.

Bohus, D., Kamar, E., & Horvitz, E. (2012). *Towards situated collaboration*. Paper presented at the NAACL-HLT Workshop on Future Directions and Needs in the Spoken Dialog Community: Tools and Data, Montreal, Canada.

Branigan, H., Pickering, M. J., Pearson, J., & Mcneil, J. F. (2009). Linguistic alignment between humans and computers. *Journal of Pragmatics, 42*(9), 2355-2368.

Cairns, P. (2007). HCI... not as it should be: inferential statistics in HCI research. *Proceedings of the 21st British HCI Group Annual Conference on People and Computers: HCI...but not as we know it* - *Volume 1*, 195-201.

Cakmak, M., & Takayama, L. (2014). *Teaching people how to teach robots: the effect of instructional materials and dialog design*. Paper presented at the Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction, Bielefeld, Germany.

Carlo, G., Okun, M., Knight, G., & de Guzman, M., (2005). The interplay of traits and motives on volunteering: Agreeableness, extraversion and prosocial value motivation. *Personality and Individual Differences, 46*(05), 1293 - 1305.

Chapman, B., Duberstein, P., Sorensen, S., & Lyness, J. (2007). Gender Differences in Five Factor Model Personality Traits in an Elderly Cohort: Extension of Robust and Surprising Findings to an Older Generation. *Personality and Individual Differences*, *43*(06), 1594 - 1603.

Chickering, D., & Paek, T. (2007). Personalizing influence diagrams: applyng online learning strategies to dialog management. *User Model. User-Apapted Interact. Special Issue Statistics Probability. Methods User Model*, *17*(1-2), 71-91.

Chinaei, H. R., & Chaib-Draa, B. (2014). Dialogue POMDP components (part I): learning states and observations. *International Journal of Speech Technology, 17*(4), 309-323. doi: 10.1007/s10772-014-9244-6

Cockton, G. (2004). Value-centred HCI. *Proceedings of the third Nordic conference on Human-computer interaction*, 149-160. doi: http://doi.acm.org/10.1145/1028014.1028038

Costa, P., & McCrae, R. (1991). Revived NEO Personality Inventory (NEO PI-R) and NEO Five Factor Inventory (NEO-FFI). *Odessa, FL: Psychological Assessment Resources, Inc*.

Creswell, J. (2009). *Research Design Qualitative, Quantitative, and Mixed Methods Approaches* (3rd ed.). Thousand Oaks, CA: SAGE Publications.

Czaja, S. J., & Lee, C. C. (2007). The impact of aging on access to technology. *Universal Access Information Society, 5*(4), 341-349. doi: http://dx.doi.org/10.1007/s10209-006-0060-x

Dahlback, N., Jonsson, A., & Ahrenberg, L. (1993). Wizard of Oz studies: why and how. *Proceedings of the 1st international conference on Intelligent user interfaces*, 193-200. doi: http://doi.acm.org/10.1145/169891.169968

De Mori, R., Bechet, F., Hakkani-Tur, D., & McTear, M. (2008). Signal Processing Magazine, . *Institute of Electrical and Electronics Engineers, 25*(3), 50-58.

Delimarschi, D., Swartzendruber, G., & Kagdi, H. (2014). *Enabling integrated development environments with natural user interface interactions*. Paper presented at the Proceedings of the 22nd International Conference on Program Comprehension, Hyderabad, India.

DiSalvo, C., Sengers, P., & Brynjarsdttir, H. (2010). Mapping the landscape of sustainable HCI. *Proceedings of the 28th international conference on human factors in computing systems*, 1975-1984. doi: http://doi.acm.org/10.1145/1753326.1753625

Duta, N. (2014). Natural Language Understanding and Prediction: from Formal Grammars to Large Scale Machine Learning. *Fundamenta Informaticae - Formal Models-Computability, Complexity, Applications, 131*(3-4), 425-440. doi: 10.3233/fi-2014-1023

Eshky, A., Allison, B., & Steedman, M. (2012). Generative goal-driven user simulation for dialog management. *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, 71-81.

Georgila, K., Henderson, J., & Lemon, O. (2005). Learning user smulations for information state update dialog systems. *In Proceedings of Interspeech Conference*.

Georgila, K., Wolters, M., Karaiskos, V., Kronenthal, M., Logie, R., Mayo, N., . . . Watson, M. (2008). A fully annotated corpus for studying the effect of cognitive aging users' interactions with spoken dialog systems. *Proceedings of the 6th International Conference on Language Resources and Evaluation.*

Georgila, K., Wolters, M., & Moore, J. D. (2008). *Simulating the behaviour of older versus younger users when interacting with spoken dialogue systems*. Paper presented at the Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics on Human Language Technologies: Short Papers, Columbus, Ohio.

Giorgino, T., Azzini, I., Rognoni, C., Quaglini, S., Stefanelli, M., Gretter, R., & Falavigna, D. (2005). Automated spoken dialog system for hypertensive patient home management. *International Journal of Medical Informatics, 74*, 159-167.

Golafshani, N. (2003). Understanding reliability and validity in qualitative research. *The Qualitative Report*, *8*(4), 567-606.

Goldberg, L. (1992). The development of markers for the Big-five factor structure. *Psychological Assessment*, *4*(1), 26-42.

Goldberg, L., Johnson, J., Eber, H., Hogan, R., Ashton, M., Cloninger, C., & Gough, H. (2006). The International Personality Item Pool and the future of public-domainpersonality measures. *Journal of Research in Personality, 40,* 84 - 96.

Griol, D., Callejas, Z., Lopez-Cozar, R., & Riccardi, G. (2014). A domain-independent statistical methodology for dialog management in spoken dialog systems. *Computer Speech Language*, *28*(3), 743-768. doi: 10.1016/j.csl.2013.09.002

Griol, D., Carbo, J., & Molina, J. M. (2013). Bringing context-aware access to the web through spoken interaction. *Applied Intelligence*, *38*(4), 620-640. doi: 10.1007/s10489-012-0390-8

Griol, D., Molina, J. M., & Callejas, Z. (2014). Modeling the user state for context-aware spoken interaction in ambient assisted living. *Applied Intelligence, 40*(4), 749-771. doi: 10.1007/s10489-013-0503-z

Hair, J., Anderson, R., Tatham, R., & Black, W. (1998). *Multivariate Data Analysis*: Prentice Hall.

Heerink, M., Krose, B., Wielinga, B., & Evers, V. (2009). Measuring the influence of social abilities on acceptance of an interface robot and a screen agent by elderly users. *Proceedings of the 2009 Bristish Computer Society Conference on Human-Computer Interaction, Cambridge, UK*, 430-439.

Hourcade, J. P., & Berkel, T. R. (2008). Simple pen interaction performance of young and older adults using handheld computers. *Interacting with Computers, 20*(1), 166-183. doi: http://dx.doi.org/10.1016/j.intcom.2007.10.002

Ivanova, A. (2010). Improving the UI of interactive training simulators to address the Y/Z-generations' learning style. *Paper presented at the Proceedings of the 11th International Conference on Computer Systems and Technologies and Workshop for PhD Students in Computing on International Conference on Computer Systems and Technologies*.

Jackson, S. (2009). *Research Methods and Statistics: A critical thinking approach* (3rd ed.). Belmont CA: Wadsworth.

Janarthanam, S., & Lemon, O. (2009). *A wizard-of-oz environment to study referring expression generation in a situated spoken dialogue task*. Paper presented at the Proceedings of the 12th European Workshop on Natural Language Generation, Athens, Greece.

Jeon, J. H., Xia, R., & Liu, X. (2014). Level of interest sensing in spoken dialog using decision-level fusion of acoustic and lexical evidence. *Computer Speech Language, 28*(2), 420-433. doi: 10.1016/j.csl.2013.09.005

John, O., Angleitner, A., & Ostendorf, F. (1988). The lexical approach to personality: A historical review of trait taxonomic research. *European Journal of Personality, 2*, 171-203.

Lee, A., Oura, K., & Tokuda, K. (2013, 26-31 May 2013). *Mmdagent - A fully open-source toolkit for voice interaction systems.* Paper presented at the Acoustics, Speech and Signal Processing (ICASSP), 2013 IEEE International Conference Vancouver, BC.

Lee, C., Jung, S., Kim, K., Lee, D., & Lee, G. (2010). Recent Approaches to Dialog Management for Spoken Dialog Systems. *Journal of Computing Science and Engineering, Vol 4*(No 1), 1 - 22.

Lemon, O., Georgila, K., Henderson, J., & Stuttle, M. (2006). An ISU dialog system exhibiting reinforcement learning of dialog policies: Generic slot-filling in the TALK in-car system. *In Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linquistics (EACL)*.

Li, A. X., & Bonner, J. V. (2014). Using wizard-of-oz method to build multipurpose platform for domestic ambient media research and applications. *Multimedia Tools Applications, 72*(2), 1011-1026. doi: 10.1007/s11042-013-1370-7

Lison, P., & Meena, R. (2014). Spoken dialogue systems: the new frontier in human-computer interaction. *XRDS Crossroads ACM, 21*(1), 46-51. doi: 10.1145/2659891

McCrae, R., & John, O. (1992). An Introduction to the Five-Factor Model and Its Applications. *Journal of Personality, 60*(2), 175-215.

McTear, M. F. (2002). Spoken dialogue technology: enabling the conversational user interface. *ACM Computing Surveys, 34*(1), 90-169. doi: http://doi.acm.org/10.1145/505282.505285

Moller, S., Krebber, J., & Smeele, P. (2006). Evaluating the speech output component of a smart-home system. *Speech Communications*(48), 1-27.

Moore, J. D., Foster, M. E., Lemon, O., & White, M. (2004). Generating tailored, comparative descriptions in spoken dialog. *In Proceedings of the 17th International Florida Artificial Intelligence Research Society Conference*, 917-922.

Niederhoffer, K. G., & Pennebaker, J. W. (2002). Linquistic style matching in social interaction. *Journal of Language and Social Psychology, 21*, 337-360.

Pew, R. W. (2003). Evolution of human-computer interaction: from Memex to Bluetooth and beyond *The human-computer interaction handbook: fundamentals, evolving technologies and emerging applications* (pp. 1-17): L. Erlbaum Associates Inc.

Pieraccini, R., & Huerta, J. (2008). *Where Do We Go from Here? Research and Commerical Spoken Dialogue Systems* (Vol. 39). Netherlands: Springer Netherlands.

Quinn, K., & Zaiane, O. (2014). *Identifying Questions & Requests in Conversation*. Paper presented at the Proceedings of the 2014 International Conference on Computer Science & Software Engineering, Montreal, QC, Canada.

Rhodes, M. (2009, April 22, 2009). Hearing Tests. *WebMD Medical Reference from Healthwise.* Retrieved October 18, 2010, from http://www.webmd.com/a-to-z-guides/hearing-tests?page=6

Sagae, K., Christian, G., DeVault, D., & Traum, D. (2009). Towards Natural Language Understanding of Partial Speech Recognition Results in Dialogue Systems. *Proceedings of North American Chapter of the Association for Computational Linguistics Human Language Technologies 2009: Short Papers*, 53 - 56.

Shin, J., Georgiou, P. G., & Narayanan, S. (2013). Enabling effective design of multimodal interfaces for speech-to-speech translation system: An empirical study of longitudinal user behaviors over time and user strategies for coping with errors. *Computer Speech Language, 27*(2), 554-571. doi: 10.1016/j.csl.2012.02.001

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S., Elmqvist, N., & Diakopoulos, N. (2017). *Designing the User Interface: Strategies for effective human-computer interaction* (6th ed.). Addison-Wesley Longman Publishing Co.

Srivastava, S. (2012). Measuring the Big Five Personality Factors. Retrieved June 4, 2012, from http://psdlab.uoregon.edu/bigfive.html

Tepperman, J., Traum, D., & Narayanan, S. (2006). "Yeah Right": Sarcasm recognition for spoken dialogue systems. *In Proceedings of Interspeech Conference*, 1838-1841.

Terrell, S. R. (2012). *Statistics Translated: A Step by Step Guide to Analyzing and Interpreting Data* (1 ed.): The Guilford Press New York

Vinciarelli, A. (2014). *Personality Computing:How machines can deal with personality traits.* Paper presented at the Proceedings of the 2014 Workshop on Mapping Personality Traits Challenge and Workshop, Istanbul, Turkey.

Warnestal, P., & Kronlid, F. (2014). *Towards a user experience design framework for adaptive spoken dialogue in automotive contexts*. Paper presented at the Proceedings of the 19th international conference on Intelligent User Interfaces, Haifa, Israel.

Williams, J., & Young, S. (2007). Partially obsevable Markov decision processes for spoken dialog systems. *Computer Speech and Language*(21), 393-422.

Wolters, M., Georgila, K., Moore, J. D., & MacPherson, S. E. (2009). Being Old Doesn't Mean Acting Old: How Older Users Interact with Spoken Dialog Systems. *ACM Transactions on Accessible Computing, 2*(1), 1-39. doi: http://doi.acm.org/10.1145/1525840.1525842

Wolters, M., Hanson, V., & Moore, J. (2011). Leveraging large data sets for user requirements analysis. *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, 67-74. doi: 10.1145/2049536.2049550

Wolters, M., Isaac, K., & Doherty, J. (2012). *Hold that thought: are spearcons less disruptive than spoken reminders?* Paper presented at the CHI '12 Extended Abstracts on Human Factors in Computing Systems, Austin, Texas, USA.

Wolters, M. K., Hanson, V. L., & Moore, J. D. (2011). Leveraging large data sets for user requirements analysis. *The proceedings of the 13th international ACM SIGACCESS conference on Computers and accessibility*, 67-74. doi: 10.1145/2049536.2049550

Wright, P., & McCarthy, J. (2008). *Empathy and experience in HCI*. Paper presented at the Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, Florence, Italy.