7-2019

Developing an Automation Locus of Control Scale

Maarten Nelson Devon Edwards

Developing an Automation Locus of Control Scale

by

Maarten Nelson Devon Edwards

A thesis submitted to the College of Aeronautics of
Florida Institute of Technology
in partial fulfillment of the requirements
for the degree of

Master of Science
in
Aviation Human Factors

Melbourne, Florida
July, 2019

We the undersigned committee hereby approve the attached thesis, "Developing an Automation Locus of Control Scale," by Maarten Nelson Devon Edwards.

Brooke Wheeler, Ph.D.
Assistant Professor
College of Aeronautics
Major Advisor

Meredith Carroll, Ph.D.
Associate Professor
College of Aeronautics

Patrick Converse, Ph.D.
Associate Professor
School of Psychology

Korhan Oyman, Ph.D.
Professor and Dean
College of Aeronautics

Abstract

Title: Developing an Automation Locus of Control Scale

Author: Maarten Nelson Devon Edwards

Advisor: Brooke Wheeler, Ph.D.

The industrial and domestic proliferation of automation is such that it has become a core component of the human experience. Both automation design paradigms and human performance must be scrutinized in order to ensure the safety, security, effectiveness and efficiency of man-machine systems across a multitude of domains (Fitts, 1951; Parasuraman, Sheridan, & Wickens, 2000; Rasmussen, 1983). Therefore, the purpose of this study was to develop and validate a measure for the evaluation of control perceptions in the context of human-automation interactions. The scale was developed using a deductive approach to measure development adapting from Rotter (1966) and Levenson (1973) locus of control measures. Results from the solicitation of expert feedback, exploratory factor analyses, and a confirmatory factor analysis supported a three-factor scale structure, and correlational analysis provided preliminary support for the construct validity of the measure. This automation locus of control scale is, therefore,

supported as a novel measure for the evaluation of automation control perceptions. The evaluation of measure generalizability and use of the measure as a means of triangulating automation control perceptions in specific scenarios are recommended.

Table of Contents

List of Figures

List of Tables

Acknowledgement

I would like to take this opportunity to thank my committee chair, Dr. Brooke Wheeler, for her guidance and unwavering dedication throughout this study. Your fervent demand for thorough, quality work will not soon be forgotten. I would also like to thank my committee members, Dr. Meredith Carroll and Dr. Patrick Converse, for their invaluable expertise, constructive feedback, technical guidance, and substantial investments of interest and time.

Similarly, I would like to thank the subject-matter experts who offered much-needed support in the conceptualization and theoretical grounding of the automation locus of control scale. My gratitude is also extended to Dr. Rian Mehta whose incredible support and expertise in the usage of MTurk contributed substantially to the data collection efforts of this study.

Finally, I would like to thank the College of Aeronautics and the Florida Institute of Technology for their administrative, financial, technical, educational, and moral support that enabled my production of this research.

Dedication

To my family and friends,

To say that I could not have done this without you would grossly underestimate your contributions to my success.

Thank you all.

Chapter 1
Introduction

Problem Statement

Automation has become an integral part of the human experience. From the seemingly menial interaction between a person and their virtual assistant to the profound industrial dependency of commercial aviation on complex auto-flight and collision avoidance solutions, automation is undeniably prolific. Consequently, our understanding of how people interact with their ever-increasing ensemble of automated devices is integral to the development and adoption of relevant design philosophies, and operational best practices.

Because the alleviation of human workload is central to the purpose of automation, an examination of the way individuals perceive themselves to be in control of its usage is paramount to the understanding of the effectiveness and efficiency of its usage. As it stands, however, there exists no psychometrically rigorous measure for the examination of such perceptions. To this end, locus of control, a psychological construct rooted in general expectancy theory, is proposed as the basis for a context-specific measure for determining the extent to which an individual perceives their experiences with automation to be the result of their own actions or factors that are external to themselves.

Purpose Statement

The purpose of this study was to develop and validate a valid and reliable empirical measure for locus of control in the context of automation usage for the general population of the United States of America. In order to create and validate an automation locus of control scale, this study consisted of three major phases. Phase 1 consisted of the development of scale items and the preliminary determination of factors into which the items may be assigned. In Phase 2, sample data attained through the administration of the preliminary scale items was tested for internal consistency, and an exploratory factor analysis was conducted. In Phase 3, a confirmatory factor analysis was used to verify the factor loadings from Phase 2, and scale validity was established via correlations with measures of related constructs.

Operational Definitions

The explicit contextual definition of key terms is crucial to the interpretation of both the premise and conclusions of this study. Therefore, the following terms were operationally defined:

Automation

Parasuraman, Sheridan, and Wickens (2000) defined automation as the use of computer hardware or software that can “carry out certain functions that the human operator would normally perform” (p. 286). Their definition and subsequent discussion supported a broad definition of automation that emphasizes system processes rather than physical form. Consequently, their definition of automation was adopted for the purposes of this research.

Locus of Control and Automation Locus of Control

Theoretically rooted in general expectancy theory, locus of control refers to the extent to which an individual perceives their experiences as the result of their own actions (Rotter, 1966). Although classically defined in terms of a unidimensional spectrum of internality versus externality, context-specific measures of locus of control suggest that two dimensions may be insufficient for describing the construct in sufficiently rich detail (Levenson, 1973; Özkan & Lajunen, 2005). This means that the context of use for a given locus of control measure may demand the consideration of a multidimensional locus of control measure.

Given the established influence of context on the dimensional characteristics of locus of control as a general construct, automation locus of control was defined as the extent to which an individual perceives the outcomes of their experiences with automation as a result of their own actions or some other external factor. This was quantified as the aggregate scores for all items on the automation locus of control (A-LOC) scale.

Factor Analysis

Factor analysis is a formal, empirical process for the determination of the number of latent variables being measured by a set of conceptualized scale items (DeVellis, 1991). In the exploratory phase, principle axis factoring with a parallel analysis was used to objectively determine the probable number of latent factors being measured by the initial item set. In the confirmatory phase, a confirmatory factor analysis was used to verify the factor structure and factor loadings of the exploratory factor analysis and quantify the fit of the proposed model.

Validity and Reliability

Validity may be broken down into two major categories, namely internal and external validity (Ary, Jacobs, Sorensen, & Razavieh, 2010). Within the context of this study, internal validity refers to the extent to which the proposed locus of control scale measured the targeted construct based on item content validity as assessed by subject matter experts, construct validity as determined by the establishment of a quantitatively supported nomological network, and quantitatively supported criterion-related validity (Hinkin, 1998).

Reliability refers to the extent to which a measure produces consistent results from one administration to the other within a given ecological setting (Ary, Jacobs, Sorensen, & Razavieh, 2010). In the context of this study, internal scale reliability was quantified using Cronbach's alpha coefficient as a measure of internal consistency (DeVellis, 1991; Hinkin, 1998).

Background

Given the proliferation of automation in contemporary society, researchers have placed considerable effort into investigating human-automation interaction and the influence of user perceptions and cognitive processes on the effectiveness and efficiency with which automation is used (Barg-Walkow & Rogers, 2016; Berberian, Sarrazin, Le Blaye, & Haggard, 2012; Brambilla, et al., 2017). These research efforts are often framed with respect to specific ecological contexts, and attempt to drive the development of design philosophies and operational policies as a means of improving user performance and experiences (Holland, Kochenderfer, & Olson, 2013; Sarter, Woods, & Billings, 1997). Consequently, the availability of valid and reliable measures of human performance, cognitive processes, and perceptions is integral to conducting scholastically rigorous human-automation interaction research.

Locus of control is a psychological construct that seeks to describe the extent to which individuals perceive effective control over their surroundings (Rotter, 1966). Rotter (1966) first proposed the construct of locus of control as antecedent to the concept of general expectancies in social learning theory. Rotter posited that the degree to which an individual expects a given reinforcement following an event or behavior is strengthened by the reinforcement itself. This expectancy is purported to be reduced should subsequent identical behaviors or events fail to be followed by their associated reinforcement.

Since Rotter (1966) published the internal-external scale for generalized expectancies, subsequent researchers have adapted the scale for a variety of settings including aviation safety studies (Hunter, 2002) and behavioral studies for risky driving (Özkan & Lajunen, 2005). Applications of unaltered and modified versions of the instrument have gone on to demonstrate significant correlations between locus of control and changes in general expectancies (Rotter, 1966), risky driving behavior (Özkan & Lajunen, 2005), and career decision-making self-efficacy (Taylor & Pompa, 1990). Consequently, the development of an empirical measure of locus of control in the context of human-automation interaction is a prerequisite for the contextually valid investigation of the relationships between the behavior and cognitive processes of human operators, and their automated systems.

Research Questions and Hypotheses

The primary research questions of this study were:

1. How many latent factors are being measured by the automation locus of control scale?

2. To what extent is the automation locus of control scale internally consistent?

3. To what extent is the automation locus of control scale a valid context-specific measure of locus of control?

The corresponding research hypotheses of this study were:

1. Automation locus of control items will support a three-factor structure.

2. The automation locus of control scale is an internally consistent measure.

3. The automation locus of control scale is a valid context-specific measure for locus of control.

Potential Significance and Generalizability

Potential Generalizability

The intent of this study was to provide a valid and reliable locus of control measure that is specific to the context of human-automation interaction. With respect to population generalizability, the scale was expected to be a reliable measure of automation locus of control most immediately among users of Amazon Mechanical Turk. The scale was further expected to provide valid and reliable automation locus of control data among other homogeneous target populations as well as the general population. With respect to ecological generalizability, the proposed scale should be capable of measuring automation locus of control in the generic context in which it was conceptualized. It should further allow researchers to apply their own context-specific scenarios as required by their own domain-specific investigations. Finally, the scale will serve as a basis for the further generation of specialized automation locus of control measures, such as measures of aircraft cockpit automation locus of control and locus of control for users of self-driving cars.

Rationale, Potential Implications and Applications, and Benefits

This study's theoretical foundation was Rotter's (1966) discussion of general expectancy and locus of control – particularly with respect to its theoretical value as a predictor of human behavior. Preceding context-specific locus of control scales demonstrated their ability to provide richer details regarding the factorial characteristics and psychometric idiosyncrasies of the contexts in which they were developed, and provided a methodological framework for their construction (Hunter, 2002; Jones & Wuebker, 1985; Özkan & Lajunen, 2005). Consequently, this scale will function as an integral tool for social science and human factors researchers by providing a quantitative method for investigating locus of control as either a primary or extraneous variable of interest in their analyses of human interactions with automated tools across a broad scope of ecological settings. Furthermore, the scale provides a base from which researchers may further refine items in order to satisfy their own domain constraints.

Limitations and Delimitations

Limitations of the proposed study included participants' abilities to perceive their experience with automated tools. Because the broadness of the definition of automation included a wide variety of hardware and software, it was possible that participants may have provided responses based on a personal interpretation of the definition of automation that was not commensurate with the contextual definition of the study.

Because the items of the proposed scale were based on those of previously developed locus of control measures, this study was also limited by the item pool from which current measure items were based. Although the preceding locus of control measures have been tested for validity and reliability within their own settings, the possibility existed that the psychometric properties of those items may be limited by the external validity of the studies in which they were conceived. Furthermore, the deductive approach to item generation used by this study prohibited the synthesis of entirely new items.

Finally, the study was limited by using a convenience sample of Amazon Mechanical Turk workers. This accessible population was not necessarily representative of the general population of the United States. Although previous studies supported the use of MTurk as a valid and reliable data source (Buhrmester, Kwang, & Gosling, 2011; Smith, Roster, Golden, & Albaum, 2016; Walter, Seibert, Goering, & O'Boyle Jr, 2018), the use of a convenience sample may have introduced sample biases that make it inappropriate to infer conclusions in the context of the general population of the United States of America.

The delimitations of this study include the decision to utilize preceding locus of control scales as the source for item generation (Levenson, 1973; Rotter, 1966) in order to create the A-LOC scale. The deductive process of item generation has been demonstrated to produce valid and reliable data without requiring word elicitation and word-pairing exercises (Hunter, 2002). Furthermore, the decision to generalize the definition of automation allowed for the application of the measure to a broad scope of human-automation interaction contexts, and for reliability and validity testing with as few constraints applied to participant sampling as practical. Finally, the target population for this study was limited to citizens of the United States of America using Amazon MTurk.

Assumptions

The primary assumption of this study was that participants were able to conceive and apply a general definition of automation for the purpose of providing responses to measure items. Although the measure under investigation may ultimately be used within specific task domains that will provide specific operational definitions or scenarios for automation usage, the development of a general automation scale calls for the testing of the scale by way of a general definition.

Where participants did synthesize specific mental representations of automation, this study assumed that variances in the mental schemas of individual participants will be mediated by the number of participants used in the development sample. Finally, this study assumed that participants were able to comprehend and operate under the paradigm that automation is defined with respect to the functionality a device enables as opposed to the specific form of the system providing that functionality.

Chapter 2
Literature Review

Introduction

The development of an automation locus of control scale is dependent upon the understanding of locus of control as a general construct as well as its adaptation to specific contexts. Similarly, the concept of automation and human-automation interaction must be examined in order to aid in construct definition, and the development of a valid and reliable instrument necessitates a review of scale development and testing strategies. Consequently, this chapter presents a review of established literature on human-automation interaction, locus of control, and its adaptation to specific context domains. Established methodology for the development of psychometric measures is also discussed along with the concepts of instrument reliability and validity. Finally, the use of crowdsource convenience sampling as a method for data collection in a scholarly research setting was discussed.

Human-Automation Interaction

Automation Design Paradigms

Automation may be defined as any software or hardware tool that augments or replaces human agents for the completion of previously human-executed tasks

(Parasuraman & Riley, 1997; Parasuraman, Sheridan, & Wickens, 2000). This definition provides a broad framework for the determination of what specific hardware or software elements within a given content-domain constitute automation based on the consideration of system function rather than form. It follows, therefore, that automation may be categorized with respect to its functionality, and that these categories may be used as the basis for determining the suitability of automated hardware or software solutions for specific applications within specific settings (Parasuraman, Sheridan, & Wickens, 2000).

To this end, Parasuraman, Sheridan, and Wickens (2000) proposed a formal model for the division of automation into a system of four major categories, namely "1) information acquisition; 2) information analysis; 3) decision and action selection; 4) and action implementation" (p. 286). It was further asserted that each major category could be divided into a series of levels in order to describe changing levels in system autonomy. With respect to decision automation, Level 3, for example, describes a design paradigm that is categorized by the provision of a narrowed list of decision alternatives to a human operator. In contrast, Level 10 describes a design paradigm, in which decisions are made and acted upon by the system with no interaction with a human operator (Parasuraman, Sheridan, & Wickens, 2000). This model, however, is limited in that it does not provide guidance on the application of automation levels to any category other than that of decision automation, and it provides no discussion for the choosing of different categories and levels of system automation based on their effect on human performance.

Fitts (1951) offered a model for the allocation of tasks based on the hypothesized strengths and weaknesses of human and machine components within a given system. Fitts' (1951) List, as it is contemporarily known, asserted that humans possess superior sensation and perception capabilities, long-term memory, process flexibility, inductive reasoning, and the application of judgement. Conversely, machines were asserted to be better at rapid task execution, particularly where large forces are required with great precision. Fitts (1951) also argued that machines are able to out-perform humans at task repetition, short-term data storage, deductive and computational reasoning, and process parallelization. Since its inception, Fitts' (1951) List has been both lauded and challenged in research circles (de Winter & Dodou, 2014; Parasuraman, Sheridan, & Wickens, 2000). De Winter and Dodou's (2014) review of function allocation theory under the Fitts' List paradigm discussed the challenges associated with a strict application of the original list to contemporary human-automation systems. This is due in no small portion to the advancement in the capabilities of machines since Fitts' (1951) initial publishing (de Winter & Dodou, 2014). Fitts (1951) is further challenged by contemporary considerations of conflicting alternative models that focus on the complementarity of human-machine interactions, and changes in human behavior as a result of automated systems. Nonetheless, Fitts (1951) has remained a preferred theoretical basis for the examination of human-automation interaction (de Winter & Dodou, 2014; Parasuraman, Sheridan, & Wickens, 2000; Pritchett, Kim, & Feigh, 2014).

Another alternative human-automation design paradigm is that of automation-focused design. Under this paradigm, designers implement automated solutions with the goal of removing human intervention from as many aspects of the system as possible (Parasuraman & Riley, 1997; Parasuraman, Sheridan, & Wickens, 2000; Pritchett, Kim, & Feigh, 2014). This approach is purported to be motivated by the ease and cost effectiveness of automating tasks (Parasuraman, Sheridan, & Wickens, 2000) and is, therefore, blind to changes in user workload and task difficulty (Bainbridge, 1982). To this end, Bainbridge (1982) asserted that the role of system operators has shifted from that of an active participant to that of a system monitor who exercises manual control in the event of a system anomaly. It is further asserted that, despite the execution of tasks by the machine-components of the system, the responsibility for the operation of the system still lies with the operator (Bainbridge, 1982). This approach was reported to constitute considerable risk to the maintenance of system performance. Regarding this risk, Strauch (2018) posited that placing a human operator in a position of oversight and redundancy over a system that was ultimately designed to out-perform them can contribute to considerable declines in overall system performance, given a failure in the automated system.

Human Performance in Human-Automation Interaction

Each of the aforementioned design paradigms are characterized by the focus they placed on human behavior, human performance, and, therefore, system performance. Parasuraman and Manzey (2010) established that it is insufficient to consider automation as a simple replacement for human intervention. Instead, explicit attention must be paid to automation's influence on human activity as this interaction is liable to compromise the performance improvements projections of system designers and implementers. Dekker and Woods (2002) explained that these changes in human activity are the likely result of the fundamental changing of the task itself as a product of the introduction of an automated system. Their discussion, framed with respect to system novelty, illustrated the process of user adaptation that necessarily preceded the optimization of user performance within a new task paradigm under new performance criteria (Dekker & Woods, 2002). Accordingly, Strauch (2018) pointed out the value of a multidimensional approach to human performance considerations and lauded the efforts of Rasmussen (1983) in his examination of human performance in the context of the tasks they must perform.

Rasmussen's (1983) discussion of human performance models introduced a pioneering framework for the multidimensional consideration of human performance in the context of a complex man-machine system, the interpretation of information by system operators under different performance paradigms, the examination of levels of abstraction in the context of user information processing, and the differing roles of qualitative and quantitative models for the evaluation of human performance. With respect to the adoption of a multidimensional approach to operator performance, Rasmussen (1983) offered three performance levels: skill-based, rule-based, and knowledge-based. At the skill-based level, users are purported to operate based on an autonomous, continuous series of sensory-motor patterns that are afforded by an extensive repertoire of prior experiences. At the rule-based level, users are asserted to operate based on the conscious application of explicitly stored rules or procedures. The selection of applicable rules is argued to be based on previous experiences, external instructional sources, or explicit problem-solving efforts (Rasmussen, 1983). Both skill-based and rule-based performance levels are derived from a feedforward approach to action evaluation. The use of feedback to evaluate the effectiveness of control inputs occurs as part of the knowledge-based level of human performance. At this level, it was argued that users employ an explicit expression of goals, and a thorough analysis of the task environment in order to adjust to unfamiliar situations. Rasmussen (1983) further asserted that this level is characterized by the use of explicit mental models as a cognitive representation of the "internal structure of the system," (p. 259). These models are argued to be based on several levels of abstraction that describe a spectrum of user perceptions ranging from the evaluation of the physical form of the system to the general functional purpose of the system. Rasmussen's (1983) discussion of reasons and causes, and levels of abstraction is fundamentally grounded in the assertion that human attentional capabilities are limited. To this end, Rasmussen (1983) stated the following:

> An effective way to counteract limitations of attention seems to be to modify the basis of mental data processing – the mental model of the causal structure – to fit it to the specific task in a way which optimizes the transfer of previous results and minimizes the need for new information. The efficiency of human cognitive processes seems to depend upon an extensive use of model transformations together with a simultaneous updating of the mental models in all categories with new input information, an updating which may be performed below the level of conscious attention and control. (p. 261)

Rasmussen's (1983) multidimensional approach to human performance evaluation and attentional processing provided a theoretical basis for the evaluation of variations of human cognitive processes given changes in system design and operational paradigms. Rasmussen's comments paralleled general expectancy theory by way of the establishment of a relationship between cognitive processes and historic experiences, and the effect of this relationship on operator performance.

With respect information processing, Rasmussen (1983) offered three major categories: signals, signs, and symbols. These categories were characterized by the manner in which information is processed as opposed to the actual form of the information itself. These categories were argued to be commensurate with the performance level in which the user is operating. At the skill-based performance level, information is purported to be interpreted as a series of signals – "continuous quantitative indicators of the time-space behavior of the environment," (p. 206) that are devoid of any meaning, and act only to guide autonomous user processes (Rasmussen, 1983). At the rule-based performance level, information is processed as signs – a series of cues used for the application and modification of rules. It is not until the knowledge-based performance level that information is perceived as a series of symbols – meaningful sets of information that allow for the conscious evaluation of environmental and system characteristics in order to make predictions in unfamiliar situations (Rasmussen, 1983). Therefore, it is reasonable to expect that changes in operator control perceptions could affect information processing, particularly when operating at skill-based or rule-based levels where expectancies based on previous experiences may have a significant influence on operator performance.

Rasmussen (1983) concluded that the effective design of man-machine systems depends heavily upon a design philosophy that emphasizes man-machine system communication with respect to nature of the task. The framing of tasks with respect to user mental processes as opposed to system requirements was also recommended. Finally, Rasmussen (1983) recommended the use of qualitative comparisons to evaluate differences between projected and actual usage strategies of systems in the design-phase. Conversely, quantitative evaluation methods should be employed to “verify the internal consistency” (p. 265) of established cognitive models (Rasmussen, 1983). In the context of the current study, Rasmussen’s (1983) discussion provided an extensive qualitative framework for the evaluation of human-automation interaction. Furthermore, Rasmussen’s discussion on the use of quantitative measures of human performance justified the development of measures for the purpose of verifying observable performance effects based on qualitatively synthesized hypotheses.

Quantitative Evaluations of Human-Automation Interactions

Berberian, Sarrazin, Le Blaye, and Haggard (2012) investigated the relationships between intentional binding and aircraft cockpit automation level, and cockpit automation level and human agency. In the context of their study, intentional binding, an implicit measure of control perceptions, was defined as the variance in an individual's perception of the delay between an action and an outcome based on their perceived control of the action. Increased control perceptions result in perceptions of shorter intervals, whereas decreased control perceptions result in perceptions of increased intervals. Human agency, an explicit measure of perceived control, was described as "a clear feeling that we control our own actions and can thus produce effects in the external environment" (Berberian, Sarrazin, Le Blaye, & Haggard, 2012, p. 1). Agency was measured based on the subjective verbal reports of participants. The study utilized 13 participants, 4 of whom were females, from the Office National d'Etudes et de Recherches Aérospatiales. The mean age was 32 years. Participants were asked to perform a number of simulated collision avoidance tasks for each automation level – referred to as a trial block. Within each block, participants were asked to report the perceived delay between a command engagement point and a resolution confirmation indication. At the end of each block, participants were asked to provide explicit verbal feedback on their perceived level of control within each trial block. With respect to the relationship between intentional binding and automation level, the results of a 4x3 ANOVA produced a significant main effect ($F(3,36) = 26.154$; $p < .01$, $\eta_p^2 = .69$), and post-hoc analysis suggested that, as automation level increased, interval estimates increased as an implicit indicator of a gradual decrease in agency. With respect to the relationship between automation level and explicit judgements of agency, the results of a repeated measures ANOVA suggested a significant effect ($F\ (3,36) = 46.204$; $p < .01$, $\eta_p^2 = .79$), and post-hoc analysis showed that, as automation levels increased, explicitly-stated perceptions of control decreased. Finally, correlations between explicit expressions of agency and intentional binding supported the conclusion that intentional binding can be considered an implicit indicator of human agency ($r = -0.84$, SD = 0.105, $t\ (12) = -28.821$, $p < .001$). The demonstrated relationship between agency and automation level gives credence to the continued development of quantitative measures of operator control perceptions in the context of human-automation interaction. Although Berberian, Sarrazin, Le Blaye, and Haggard (2012) demonstrated the feasibility of intentional binding as a measure of control perceptions, the addition of a robust validated self-reporting measure stands to increase investigators' ability to triangulate their measurements and observations. Such a measure, however, requires a core construct that is grounded in control perception theory, and offers a framework for the development of robust psychological measures.

A Model of Human-Automation Interaction and Control Perceptions

The relatively broad general definition of automation (Parasuraman & Riley, 1997; Parasuraman, Sheridan, & Wickens, 2000) allows researchers to explore automation both as a general technological concept, and as a context-specific phenomenon without the need to stray from or stretch the grounding theory of the construct. The function-centered definition of automation lends itself to a function-centered classification of its many forms (Parasuraman, Sheridan, & Wickens, 2000), the generalized discussion of its design paradigms (Fitts, 1951), and the generalized evaluation and discussion of system-wide performance particularly when a human operator is a part of that system (Parasuraman & Manzey, 2010; Rasmussen, 1983; Strauch, 2018).

The existing literature supported a conceptual model, illustrated in Figure 1, that describes the progression from system goals to desirable system performance via the independent but interacting processes of automation design and operator behavior. The model suggests that, operator behavior is influenced both by the system goals and the automation design, and both automation design and operator behavior influence system performance. The findings of Berberian, Sarrazin, Le Blaye, and Haggard (2012) supported the consideration of a mediating effect of operator control perceptions. In the context of the current study, such an effect is proposed to function as shown in Figure 2. Note now that the diagram implies the supported hypothesis that control perceptions can be calibrated by way of system design as a means of influencing operator performance (Berberian, Sarrazin, Le Blaye, & Haggard, 2012; James & Rotter, 1958; Phares, 1957).

Locus of Control

Foundations of Locus of Control

Locus of control is a psychological construct that describes the extent to which individuals perceive their experiences as primarily influenced by their actions, known as an internal locus of control, or by factors beyond their control, known as an external locus of control (Rotter, 1966). In order to create an instrument that quantitatively measured general locus of control, Rotter's general internal-external (I-E) locus of control scale was the result of factor and item analyses on Phares' (1957) instrument for the measurement of generalized expectancies. The resultant 60-item scale was further refined due to high correlations (r = [-.35, -.40]) of some of the items to measures of social desirability. The final 29-item I-E scale was distributed to a sample (n = 400), and biserial item correlations for the instrument were determined. Internal consistency estimates were found to be stable, 1-month test-retest reliability was found to be consistent between two different samples, and the new correlation to the Marlow-Crowne social desirability scale was significantly reduced (r = [-.07, -.35]). Based on the findings of his and other supporting studies, Rotter (1966) concluded that locus of control varies both between different individuals and between different situations experienced by a single individual.

Rotter’s work on social learning theory, general expectancies and locus of control set the foundation for the widely accepted unidimensional locus of control construct that forms the core of an expanse of control perceptions investigations. Lefcourt’s (1966) discussion of the underlying theory of the internality-externality construct illustrated the substantial contrast between its foundation in expectancy theory and the motivationally-driven constructs that preceded it. In so doing, Lefcourt highlighted the applicability of locus of control to situations well beyond those in which survival or success formed the core of a subject’s cognitive processes. Of note was Lefcourt’s references to preceding investigations of general expectancies that demonstrated the significant effect of task structure on changes in expectancies. These investigations included but were not limited to James and Rotter (1958), and Phares (1957) both of which investigated the effect of stated sources of success, skill verses chance, on changes of expectancies for an otherwise controlled reinforcement. That is to say that, in either case, reinforcement stimuli were fixed by the experimenters. The manipulation of experiment instructions to suggest that either chance or skill determined task outcome resulted in changes in participants’ locus of control. The studies, therefore, supported the ability to alter locus of control based on task parameters, and demonstrated how those alterations in control perceptions elicited changes in participant behavior.

Factor Structure of Locus of Control

Noting the growing popularity of the construct, Rotter (1975) sought to clarify the foundations of the theory, and address trends in literature that did not seem consistent with the core construct. This included a discussion of the growing body of research that debated the factor structure of locus of control. At its inception, the Rotter general internal-external locus of control (I-E) instrument was built around a unidimensional conceptualization of the locus of control construct. Rotter explained that, in the case of the general I-E instrument, the decision to adopt a unidimensional model was made based on the superior explained variance statistic produced by factor analyses performed on sample data (Rotter, 1975). Rotter does not, however, purport a single-factor solution to be the correct answer. Rather, Rotter's discussion cautioned against experimenter attempts to ratify a single unidimensional or multidimensional model. Drawing parallels to the construct of dependency, Rotter supported the subdivision of this construct into latent factors and noted the tendency of factor structures to vary as a result of the sample data from which they were developed.

Having established that sub-dimensional structure for the locus of control construct is subject to the interpretation of the instrument by one's development sample, it remains prudent to examine the factor structures of existing multidimensional measures as a means of targeting a probable number of factors that, although suited the context of use for the measure, may be grounded in wider construct theory. Lefcourt, Von Baeyer, Ware and Cox (1979) and Özkan and Lajunen (2005), for example, illustrated complex multidimensional scale structures that were conceptualized for the purpose of measuring locus of control in markedly different research domains. Where Lefcourt, Von Baeyer, Ware and Cox's factor structure was design to enable the measurements of "goal specific" (p. 288) locus of control across major subscales of achievement and affiliation, Özkan and Lajunen (2005) aimed to measure locus of control in the text of accident causation perceptions. As such, the factor structure of their instrument was divided into subscales for perceptions of causation by the driver, other drivers, the vehicle and environment, and fate. Although both scales share the core thread of internality-externality, the underlying factor structure of the traffic locus of control measure is heavily tailored to its context of use.

By contrast, there are a number of studies that support a common three-factor structure for locus of control measurement across a number of domains including the evaluation of locus of control among adult educators (Kourmousi, Xythali, & Koutras, 2015), in the context of social activism (Levenson & Miller, 1976), and in the context of health perceptions (Ross, Ross, Short, & Cataldo, 2015). In general, the common three-factor structure consists of an Internal subscale that measures the extent to which an individual perceives their experiences to be the result of their own actions, a Powerful Others subscale that measures the extent to which an individual perceives their experiences to be the result of the will or actions of some influential or powerful individual, and a Chance subscale that measures the extent to which an individual perceives their experiences to be the result of random probabilities or fate. Based on the repeated appearance of this subscale structure among established measures, it is reasonable to expect a novel, context-specific measure to adopt a similar factor structure.

Quantitative Investigations of Locus of Control

In furtherance of Rotter's (1966) assertion that locus of control can act as a viable predictor of human behavior, there is a substantial body of knowledge that explores the relationships between locus of control and other variables. These investigations may be general in both context and using a general instrument (Rotter, 1966), specific in context and using a general instrument (Oğuz & Sariçam, 2016; Thompson, 2010), or specific in context and using a context-specific instrument (Chittaro, 2014; Hunter, 2002; Özkan & Lajunen, 2005; Ross, Ross, Short, & Cataldo, 2015). At any rate, empirical investigations of locus of control have considered locus of control as either an independent variable wherein changes in other dependent variables were observed as an effect of changes in locus of control, or as a dependent variable where manipulations of some other independent variable are correlated to a change in reported locus of control.

Oğuz and Sariçam (2016) investigated the relationship between locus of control and critical thinking disposition in pre-service teachers. Although no explicit description of a target population was provided, an appropriate target population can be inferred to be pre-service teachers in Kütahya, Turkey, based on the sample. The accessible population consisted of students of the Dumlupinar University, Faculty of Education. They used a convenience sample of 347 participants, of whom 203 were female. The ages of participants ranged from 17 to 24 with an average age of 20.4. With respect to grade level, the sample consisted of first-years (n = 188) and seniors (n = 159). Respondents were asked to complete the Rotter Internal-External Locus of Control Scale (1966), the Critical Thinking Dispositions Scale (Sosu, 2013), and a personological information form. The relationship between locus control and critical thinking disposition was analyzed via Spearman's correlation. Oğuz and Sariçam (2016) found a negative relationship ($r = -.44$, $p < .01$) between locus of control scores and critical decision-making dispositions. Furthermore, the statistically significant results of the regression model ($B = -.46$, $p < .01$) suggest that locus of control scores may be used to predict critical decision-making disposition scores.

Thompson (2010) investigated the relationship between locus of control and decision-making styles. The target population of the study was business managers of for-profit business in the United States. The accessible population consisted of members of an e-mail marketing database that was used for the solicitation of participants. The study estimated that approximately 200,000 eligible participants were contacted. The study attained 237 participants, who provided responses to the Rotter I-E instrument and the Decision-Making Inventory. Thompson found no significant relationship between decision-making style and locus of control. This conclusion is supported by the lack of statistical significance across hypotheses that attempted to determine the relationship between locus of control and analytical decision style ($r = .025$, $p = .705$, $n = 237$), locus of control and conceptual decision style ($r = -.112$, $p = .085$, $n = 237$) , locus of control and directive decision style ($r = .109$, $p = .095$, $n = 237$), and locus of control and behavioral decision style ($r = -.025$, $p = .705$, $n = 237$). Although the relationship between locus of control and decision-making style was found to be insignificant in this context, the investigation of an updated or contextually specific scale for locus of control was cited as an area of interest for future research. Furthermore, the importance of the understanding of personal preferences and cognitive processes as a means of ensuring optimal decision-making performance was supported.

Özkan and Lajunen (2005) sought to develop and implement a locus of control scale for evaluating factors associated with risky driving behaviors. Having discussed the potential shortcomings of generalized scales for measuring locus of control in specialized contexts, Özkan and Lajunen proposed a multidimensional traffic locus of control scale (T-LOC). The scale consisted of 16 prompts that each represented the cause of an accident. Using a 5-point Likert scale, participants were asked to indicate the degree to which they believed the accident to be possible, given their own driving styles. The scale was distributed as a part of a study questionnaire that also contained the Driver Behavior Questionnaire instrument and a request for demographic data. The scale was distributed to a sample of students from the Middle East Technical University where $n = 348$. Results of the hierarchical regression suggested that the Self subscale of the T-LOC instrument was a predictor of the total number of accidents ($\beta = 0.17$, $p < .001$), the number of active accidents ($\beta = 0.18$, $p < .001$), the total number of offences ($\beta = 0.11$, $p < .05$), aggressive violations($\beta = 0.19$, $p < .001$) , ordinary violations ($\beta = 0.26$, $p < .001$), and errors ($\beta = 0.24$, $p < .001$) . Consequently, Özkan and Lajunen concluded that an internal locus of control orientation predicted reported driver behavior and demonstrated the value of a specialized locus of control scaled for inferring contextually relevant conclusions using the factor structure of a multidimensional instrument.

Hunter (2002) investigated locus of control in the context of aviation safety and developed a scale that measured the internality-externality of pilots based on instrument items that were framed in the context of aviation safety. This scale was a modification of the Jones and Wuebker (1985) Safety LOC scale. Hunter hypothesized that pilots with an internal safety locus control orientation were less likely to be involved in an accident. The investigation solicited participation from visitors of a Federal Aviation Administration website over 6 months. The study received 477 responses. The internal-external subscales were negatively correlated ($r = -.419$, $p < .001$), and a comparison of the mean scores of the subscales supported the assertion that pilots would score significantly higher on the internal subscale than the external subscale ($t = 69.1$, $df = 476$, $p < .001$). Furthermore, correlation of the combined locus of control scores and the Hazardous Events Scale scores support the hypothesis that pilots with more internal locus of control orientations were less likely to have been involved in a hazardous event ($r = -0.162$, $n = 170$, $p < .05$). Consequently, the study supported the use of locus of control as a possible predictor for identifying pilots who are more likely to be involved in a hazardous event.

Where the preceding studies have examined locus control as a predictor of human behavior, Chittaro (2014) examined the effect of persuasive play on locus of control. Although a description of the target population was not provided, it can be inferred that the author intended for the study to be generalizable to passengers of commercial airlines. The study recruited 24 participants, 11 of whom were female. Participant ages ranged from 19 to 55 with a mean age of 30.5. Participants were asked to complete a modified version of Hunter’s (2002) Aviation Safety Locus of

Control Scale, a risk perception instrument, a brace position knowledge test, and a demographics form. Participants were then asked to play a computer game that provided interactive instruction on how to assume the brace position required by emergency landing on commercial aircraft. The knowledge test, locus of control instrument, and risk perception instrument were re-administered following completion of the game. Chittaro (2014) found statistical significance in the differences between pre-intervention and post-intervention measures of internal locus of control scores ($F(1,23) = 17.05$, $p < .001$) and external locus of control scores ($F(1,23) = 7.58$, $p = .01$). The findings suggest that persuasion play can have a significant effect on aviation safety locus of control. The lack of a control group, however, means that there was no way to verify that the treatment did indeed produce the effect. Nonetheless, the findings demonstrate that locus of control is not fixed, and that it can change based on external factors.

Related Constructs

In order to establish the nomological network required to assert the construct validity of the proposed scale, constructs related to locus of control must be examined. To this end, Skinner (1966) established locus of control as one in a substantial collection of established constructs, all of which relate to the concept of control. This collection is purported to include constructs of efficacy, agency, and autonomy. Similarly, Galvin, Rendel, Collins, and Johnson (2018) discussed the social learning and general expectancies theories that gave rise to the generalized locus of control construct, and asserted positive correlations between locus of control and measures of self-esteem and intrinsic task motivation. These assertions were supported by Johnson, Rosen and Levy (2008), and Kourmousi, Xythali and Koutras (2015) who both observed positive relationships between locus of control and self-esteem. Similarly, Ng, Sorensen and Eby (2006) observed a positive relationship between locus of control and intrinsic task motivation. With regard to discriminant constructs, Kalnback and Hinsz (1999) investigated the relationship between locus of control and goal commitment. Based on the statistical insignificance of their results, their findings supported the assertion that locus of control and goal commitment are not related constructs. Similarly, the relationship between decision-making style and locus of control has been investigated on numerous occasions with no support for a significant relationship between the two constructs (Hornaday & Curran, 1987; Thompson, 2010). It is, therefore, reasonable to expect a self-report measure of control perceptions to correlate significantly with self-esteem and intrinsic task motivation. Conversely, it is expected that locus of control will not correlate significantly with measures of goal commitment, and decision-making style.

Per the guidance of DeVellis (1991), the aim of establishing criterion-related validity is to demonstrate an empirical relationship with an established criterion variable. Accordingly, the discussion of the nomological network of locus of control by Galvin et al. (2018) highlighted Lilly and Virick's (2006) demonstration of the relationship between work locus of control and trust perceptions. Lilly and Virick (2006) found a statistically significant positive correlation between locus of control and organizational trust ($b = .43$, $p < .001$). Similarly, the findings of correlations between locus of control scores and interpersonal trust demonstrated strong positive correlations between measures for a cross-sectional sample of husbands and wives as part of a nonverbal communications study (Sabatelli, Buck, & Dreyer, 1983). The findings of these studies support the use of trust as a criterion variable for control perceptions, and suggest that locus of control and trust have a significant, positive relationship such that, as locus of control become more internal, measures of trust should increase.

A Model of the Measurement of Locus of Control

In this section, the locus of control construct was defined around the variability of expectancies based on the relationship between the occurrence of a reinforcement and the perceived proximity of a subject's actions to the cause of that reinforcement. The established malleability of locus of control based on continued experiences across contexts supports the use of locus of control as probable determinant of human behavior that is capable of being calibrated (Lefcourt, 1966; Rotter, 1966). Analysis of the factor structure of locus of control is established to be sensitive to the perceptions of a given sample (Rotter, 1975); however, some consistency has been observed in the development of three-factor general and context-specific measures of the construct (Levenson, 1973; Ross, Ross, Short, & Cataldo, 2015). The numerous empirical studies that explore locus of control as both an independent variable (Hunter, 2002; Oğuz & Sariçam, 2016; Özkan & Lajunen, 2005) and a dependent variable (Chittaro, 2014) supported the development of a conceptual model, illustrated in Figure 3. The model illustrates the potential role of an automation locus of control measure as a means of observing locus of control perceptions that mediate the interaction between automation design and operator behavior. Note the introduction of a feedback loop that would allow for the evaluation, interpretation, recalibration and reevaluation of operator control perceptions in the interest of optimal system performance.

Scale Development

The development of a valid and reliable psychometric measure is dependent on a systematic approach to item generation, refinement and testing. In the initial phases of scale development, Downing (2006) called for the explicit establishment of a content definition as the foundation for all other development tasks. Framed with respect to the development of achievement tests, Downing (2006) stressed the importance of defining the content domain and construct of a given test as necessary prerequisites for the establishment of valid inferences based on test administration. Similarly, DeVellis (1991) and Hinkin (1998) advocated for the establishment of the content domain and thorough understanding of the construct under investigation as integral first steps in the scale development process.

After defining the construct and establishing of the content domain of the proposed measure, there are several recorded methods for item generation illustrated in the literature. Hinkin (1998) described two approaches to item generation, namely the deductive approach and the inductive approach. Hinkin's (1998) deductive approach to item generation is a process whereby the existing body of knowledge regarding the target construct is substantial enough to enable the synthesis of an initial item-pool. This strategy is evident in the development of the Hunter (2002) Aviation Safety Locus of Control Scale, and the Spector (1988) Work Locus of Control Scale. Conversely, the inductive approach to item development was described as an approach whereby the novelty of the proposed measure and its associated construct precluded the development of an initial item-pool based on established theory alone. Instead, researchers are required to solicit the input of a sample in order to form the basis for item-development. This process is evident in Cremer's (2015) development of a Perception of Airport Sustainability Scale, and Jian, Bisantz, Drury and Linas' (2000) development of the Checklist for Trust between People and Automation.

Having developed an initial item pool, Hinkin (1998) called for the evaluation of the content validity of the items. This process was purported to allow for the reduction of the initial item pool based on the relevance of each item to the established construct definition. To accomplish this, Hinkin (1998) proposed several strategies. Firstly, Hinkin (1998) suggested the distribution of the initial item pool to a sample. Respondents would be provided with a set of scale items along with a definition. It would then be the task of the respondents to rate the extent to which each item corresponds to a single definition. The process would be repeated for all items and all definitions. Alternatively, Hinkin (1998) proposed the examination of the "proportion of the respondents who assign an item to its intended construct," and "the degree to which each rater assigned an item to its intended construct" (p. 111). Finally, Hinkin (1998) suggested an item sorting task where respondents would assign items to their associated definitions or to a category indicative of their being unsuitable for any of the provided definitions. Prior to dispersion to a development sample, DeVellis (1991) encouraged the consideration of validation items in order to control for the confounding effect of respondent biases and motivations. For example, DeVellis (1991) suggested the inclusion of a social desirability measure to test for the influence of respondents' motivation to respond in a manner factored by society on their responses to the proposed measure.

The development and construct validation of the initial item-pool is typically followed by distribution of the pool to a development sample. This process involves consideration of the target population, for which the scale is intended, as well as the number of participants required for the acquisition of valid development feedback (DeVellis, 1991). With respect to sample size, DeVellis (1991) suggested that, although the determination of absolute sample minimums has been the subject of debate, a sample size of 300 is regarded as adequate in order to mitigate for confounding variances between subjects. Hinkin (1998) suggested that 10 participant responses per item is considered desirable, particularly considering the effect of sample size on exploratory and confirmatory factor analysis results. However, Hinkin (1998) also noted that samples as small as 150 may be adequate for maintaining exploratory factor analysis accuracy given appropriately strong item intercorrelations.

Factor Analysis

Rooted in the theory of the common factor model, factor analysis is a statistical procedure for the investigation of the relationship between a set of indicators and one or more latent variables (Brown, 2006; DeVellis, 1991). Accordingly, Brown (2006) established that the common factor model asserted that "each indicator is a linear function of one or more common factors and one unique factor," (p. 13) and that factor analysis itself discriminates the common variance and unique variance of each indicator in relation to the latent variable.

Exploratory Factor Analysis is a data-driven strategy that allows for the establishment of factor loadings free of a priori specifications regarding the number of latent factors (Brown, 2006). This is the preferred exploratory method for the initial establishment of the number of latent factors measured by an item-pool, and to quantify the meaningfulness of each item in the context of the latent factors (Brown, 2006; DeVellis, 1991). Prior to the completion of an exploratory factor analysis, Hinkin (1998) suggested an initial analysis of the inter-item correlations and the deletion of any item with a correlation of less than .4. This functioned to ensure the domain commonality of the proposed items prior to the instigation of latent factors and factor loadings (Hinkin, 1998).

Confirmatory factor analysis is a theory-based, benchmark-driven approach to investigating the relationship between indicators and latent factors. Brown (2006) explained that, unlike the exploratory factor analysis approach, the number of factors, factor loading patterns, and factor-indicator independence or covariance parameters may all be determined on an a priori basis in order to produce a sample correlation matrix. Confirmatory factor analyses are, therefore, used in scale development as a means of validating the factor structure of a measure based on preceding development processes (Brown, 2006; DeVellis, 1991; Hinkin, 1998).

Validity and Reliability

Although historical discussions of validity are centered around the assessment of the extent to which an instrument truly measures its targeted construct, Ary, Jacobs, Sorensen and Razavieh (2010) pointed out that recent explanations of validity are more so focused on “the interpretation and meaning of the scores derived from [an] instrument,” as opposed to the instrument itself (p. 225). It was argued, therefore, that the concept of validity in the context of the testing of hypothetical constructs referred to the extent to which the results of a measure that are based on the operational definition of a given construct may be used to make inferences based on the more abstract conceptual definition of the investigated construct (Ary, Jacobs, Sorensen, & Razavieh, 2010). In order to demonstrate the validity of a proposed measure, Hinkin (1998) called for the demonstration of convergent validity, discriminant validity, and criterion-related validity. Convergent validity refers to the extent to which the proposed measure correlates with similar constructs; discriminant validity refers to the extent to which the proposed measure fails to correlate with dissimilar constructs; and criterion-related validity refers to the extent to which the proposed measure correlates with other theoretically correlated variables (Hinkin, 1998).

Ary, Jacobs, Sorensen, and Razavieh (2010) defined reliability as a reference to the consistency of the results produced by a given measure. From a theoretical perspective, this refers to the evaluation and management of random and systematic errors of measurement and functions as a necessary prerequisite to any assertions of measure validity (Ary, Jacobs, Sorensen, & Razavieh, 2010). Quantitative reliability testing may be parsed into three major categories: test-retest, equivalent forms, and internal consistency (Ary, Jacobs, Sorensen, & Razavieh, 2010). Where test-retest and equivalent forms strategies require multiple instrument administrations to a static sample, internal-consistency measures achieve adequately rigorous quantitative results in a single measure administration (Ary, Jacobs, Sorensen, & Razavieh, 2010). In the context of scale development, Hinkin (1998) suggested the use of Cronbach's (1951) coefficient alpha as the preferred measure of internal consistency. A minimum alpha coefficient of .7 was suggested as an indicator of adequate coverage of the construct to which the item is claimed to be related (Churchill, 1979; Hinkin, 1998).

Context-Specific Locus of Control Scale Development

Spector (1988) developed the Work Locus of Control scale in order to quantify the extent to which individuals perceived their experiences in a professional work environment as the results of their actions. The study utilized six independent samples in order to narrow the initial 49-item pool down to a final set of 16 scale items that were tested for consistency and reliability. The initial 49-item pool was created via “a conceptual analysis of the locus of control construct and how it relates to work behavior” (Spector, 1988, p. 336). Based on the results of the distribution of this pool to the first independent sample of participants, the pool was reduced based on the following criteria: “acceptable item-total correlations, lack of correlation with social desirability” (Spector, 1988, p. 336), and the balancing of internal and external subscale items. Table 1 shows the results of sample testing for each of the six testing scenarios. Based on the findings shown in Table 1, Spector (1988) concluded that that the Work Locus of Control Scale was a “viable” (p. 339) context-specific measure of locus of control. Although the exact details of Spector’s (1988) development process were not reported, the use of a deductive approach to item generation could be inferred. Results of the correlational analyses for construct validity demonstrated the ability to develop a context-specific measure for locus of control that more closely correlates to context-specific criterion measures than general measures of locus of control.

Bradley and Sparks (2002) developed a locus of control measure for service situations. In defining their context-specific locus of control construct, Bradley and Sparks (2002) adopted the traditional internal-external dichotomy of the original Rotter (1966) measure. Within this dichotomy, multiple facets were established. To this end, Bradley and Sparks (2002) offered the following:

> Thus, within the domain, we proposed that customers' sense of control could be derived from perceptions of (a) their abilities to manage the service encounter, (b) the interpersonal influence strategies they select, and/or (c) the amount of effort they invest in the encounter. Similarly, within the external domain, we proposed that control over service could be perceived to reside in (a) management personnel, philosophy, and practices; (b) the skills and attitudes of the service staff; and/or (c) luck and chance events. (p. 315)

Bradley and Sparks' (2002) division of locus of control into three facets provided theoretical grounding for the expectation of multiple latent factors in the context of the proposed scale. Scale development began with the initial conceptualization of a 70-item pool based on input from three focus groups and the consideration of existing locus of control measures. Examination of the pool for content distinctiveness and face validity saw the reduction of the pool to 48 items – 24 items for each internal and external, and 8 items per facet. Thereafter, the study was divided into three phases. In the first stage, data was collected from 265 participants and a principle components factor analysis was used to determine the factor loadings of the 24-item pool. Results of the analysis suggested that the scale had three construct dimensions – "internal, powerful others…, and luck or chance" (Bradley & Sparks, 2002, p. 316). Initial evaluations of concurrent validity were enabled via investigations of participant pre-purchase research behavior, responses to positive word-of-mouth, and the tendency for participants to examine multiple service alternatives in search of superior service. Having further refined the item-pool to 14 items, data was collected from an independent set of 302 participants and subjected to a confirmatory factor analysis. Results suggested good discriminant validity among the three Service Locus of Control subscales and demonstrated model superiority to a two-factor or single-factor model. Further investigations of the concurrent validity of the scale were enabled via correlations of subscales to the Search Benefits Scale (Srinivasan & Ratchford, 1991). Based on significant correlations to the Powerful Others ($r = .28$, $p < .01$) and Internal ($r = .23$, $p < .01$) subscales, Bradley and Sparks (2002) concluded that the Service Locus of Control scale could be used as a component measure for the investigation of consumer preservice search behaviors. The third and final part of the study was dedicated to further validation testing of the measure. For this stage, 205 participants provided responses to the Service Locus of Control Scale, the Rotter (1966) I-E Scale, the Busseri, Lefcourt, and Kerton (1998) Consumer Locus of Control Scale, the Lambert (1980) Consumer Powerlessness Scale, the Murray (1991) Non-Search Purpose Tendency Scale, and the Crosby and Stephens (1987) Generalized Satisfaction Scale. Results of this phase supported the factor structure confirmed in the second research phase and demonstrated superior relationships to theoretically related constructs than general locus of control measures. Thus, Bradley and Sparks (2002) demonstrated the feasibility of a multidimensional factor solution for locus of control that is better suited to measurements of control perceptions within the context of their development.

Crowdsourced Convenience Samples

Commonly discussed in the context of the research process at large, "external validity refers to the extent to which the findings of a study can be generalized to other subjects, settings, and treatments" (Ary, Jacobs, Sorensen, & Razavieh, 2010, p. 292). This concept is crucial to the scientific method as it illustrates the process of making conclusions about a comparatively large target population based on an observed treatment effect within a comparatively small representative sample (Ary, Jacobs, Sorensen, & Razavieh, 2010; Ferguson, 2004). In the context of scale development, assurance of the generalizability of the scale to the target population is accomplished via explicit consideration of the representativeness of the samples used throughout its development (DeVellis, 1991; Hinkin, 1998). Consequently, sampling strategy constitutes a significant concern for scale development efforts.

Under ideal circumstances, population-based probability sampling is regarded as the gold standard for sampling strategies based on its theoretical ability to manage the confounding effects of demographic factors (Ary, Jacobs, Sorensen, & Razavieh, 2010; Bornstein, Jager, & Putnick, 2013; Jager, Putnick, & Bornstein, 2017). Population-based probability samples, however, have been noted to be impractical on the bases of extensive cost and resource requirements (Bornstein, Jager, & Putnick, 2013; Landers & Behrend, 2015). Consequently, crowdsourced convenience samples are offered as a practical alternative for the acquisition of sample data (Chandler & Shapiro, 2016; Landers & Behrend, 2015). One of the main resources for the solicitation of crowdsource sample data is Amazon Mechanical Turk (MTurk) – a web-based crowdsource labor platform that has, over the years, been used to enable the collection of substantial quantities of data from workers on the platform (Chandler & Shapiro, 2016).

Although lauded for its ability to provide vast quantities of sample data, the representativeness, integrity, and, therefore, validity of MTurk sample data has been challenged with good reason (Landers & Behrend, 2015). Chandler and Shapiro (2016) illustrated several challenges with regard to the collection of data via Amazon MTurk including potential limitations regarding the representativeness of the sample, character misrepresentation, malingering, and the familiarization of practiced participants with research procedures and established measures. With respect to the mitigation of character misrepresentation, Wessling, Huber and Netzer (2017) called for the use of a pre-screening strategy in order to determine participant eligibility. Chandler and Shapiro (2016), however, illustrated the importance of presentation in the implementation of a pre-screening procedure by depicting the increase in desirable responses once workers were made aware that the pre-screening procedure was, in fact, to determine participant eligibility in a following study. With respect to subject inattentiveness, Fleischer, Mead and Huang (2015) called for the communication of the gravity of the study to the worker as a means of inspiring attentiveness. Chandler and Shapiro (2016), however, suggested the active approach of considering workers' task acceptance ratios as an indicator of attentiveness. This, they postulated, was a superior method to the implementation of attentiveness prompts – a strategy with which experienced workers are assumed to be familiar (Chandler & Shapiro, 2016).

Empirical Investigations of External Generalizability

In order to quantify the extent to which crowdsourced convenience samples can contribute to academic research fields, it is imperative that investigations that compare data collected on crowdsourced platforms with more traditional laboratory and field data be conducted. To this end, Buhrmester, Kwang and Gosling (2011) investigated the relationship between data collected via Amazon MTurk and traditional data collection methods. The results suggested that the MTurk sample produced good coefficient alpha and test-retest reliability coefficients. Consequently, it was concluded that the MTurk sample either met or exceeded psychometric property criteria regardless of compensation amounts, and that MTurk was a valid source of academic sample data (Buhrmester, Kwang, & Gosling, 2011).

Similarly, Walter Seibert, Goering and O'Boyle (2018) evaluated the convergence of data collected from online data panels and conventional data sources. Results of the study were based on 90 independent samples and 32,121 participants. Analysis of the effect sizes of the two major data sources showed no statistically significant difference between the data solicitation methods thereby supporting the notion of converging external validity among the online and conventional sampling strategies (Walter, Seibert, Goering, & O'Boyle Jr, 2018). Sample reliability was established via the comparison of online panel data to a prior reliability generalization study. Results of the comparisons supported adequate internal consistency of the online data sources. Consequently, Walter, Seibert, Goering, and O'Boyle (2018) supported the use of online panel data as a viable data source with demonstrated convergence with conventional sampling strategies.

Finally, Smith, Roster, Golden, and Albaum (2016) investigated the differences in demographics, survey-taking experience and data quality between MTurk sample data, and data from an unspecified general household panel via the Qualtrics online survey platform. Where the aforementioned studies supported the use of online panels and crowdsource sampling as a viable substitute for more traditional laboratory and field data, Smith, Roster, Golden, and Albaum (2016) found significant differences in data convergence in the context of their performance criteria and demographics analyses. The results of the study showed that the exact source of crowdsourced samples can have a significant effect on data generalizability, particularly where the inclusion of non-US respondents – which make up the majority of the MTurk respondent pool – is concerned (Smith, Roster, Golden, & Albaum, 2016).

Conclusion

Research regarding automation design paradigms established a multilevel approach to automation design, and illustrated that automation is defined less with respect to its physical form, and more with respect to its functional capabilities, and the degree to which a given system interacts with the user (Parasuraman, Sheridan, & Wickens, 2000). Regarding the determination of what functions should be automated, the literature offered a user-centered task allocation approach that emphasized the importance of differentiating the specific strengths of humans versus machines (Fitts, 1951), and an automation-centered approach that sought to automate as many tasks as possible (Parasuraman & Riley, 1997; Parasuraman, Sheridan, & Wickens, 2000; Pritchett, Kim, & Feigh, 2014). The latter of these approaches were established to constitute considerable risk to system performance particularly given the failure of an automated system component (Strauch, 2018). With respect to the evaluation of performance in the context of human-automation interaction, multidimensional approaches to human performance, attentional processing, and automation perception are offered (Rasmussen, 1983). Furthermore, the consideration of both qualitative and quantitative approaches to human performance researcher was suggested. Where qualitative approaches were purported to be of value to the synthesis of general concepts and hypotheses, quantitative approaches were offered as a means of validating said hypotheses (Rasmussen, 1983).

Context specific measures of locus of control have been developed in order to investigate the relationship between locus of control and other variables within specific context domains (Bradley & Sparks, 2002; Hunter, 2002; Jones & Wuebker, 1985; Özkan & Lajunen, 2005; Spector, 1988). Context-specific investigations of locus of control also enriched the multidimensional structure of the general locus of control construct, and supported the increased value of context-specific scales as predictors of human behavior within specific settings (Bradley & Sparks, 2002; Özkan & Lajunen, 2005). Where previous studies established context-specific measures of locus of control as viable indicators of human behavior within those contexts, no such measures exist in the context of human-automation interaction. Therefore, the conceptualization and testing of a quantitative measure for automation locus of control is a critical next step that will provide a valid and reliable method for the establishment of automation locus of control as a predictor of human behavior within a human-automation interaction setting.

Chapter 3
Methodology

Population and Sample

Population

The target population for this study was defined as members of the general public in the United States. As of 2018, the United States Census Bureau (2019) estimated the total population of the United States of America to be 327,167,434 people (166,038,755 females). The median age was reported at 38.2 years, 36.9 years for males and 39.5 years for females (United States Census Bureau, 2019). For the purpose of this study, the accessible population was defined as workers of Amazon MTurk who are over the age of 18 years. Although no precise figure is known, the total number of registered MTurk users has been reported as being in excess of 500,000 with a probable US worker-base of approximately 15,000 users as of 2016 (Chandler & Shapiro, 2016).

Sample

For each research phase, the study used a crowdsource convenience sample of workers on Amazon MTurk. Participants were recruited via the MTurk human intelligence tasks (HIT) system, which provides monetary incentives for the completion of tasks requiring a human agent. In order to allow for the assessment of the ecological and population generalizability of the proposed study, demographic and relevant professional information were collected. These parameters included age, gender, ethnicity, country of origin, and highest education achieved.

Procedures

Phase 1: Item Generation

Phase 1 of this study consisted of the deductive synthesis of a preliminary item pool based on a review of established locus of control scales (Hinkin, 1998). Scale items were compiled into a master database along with a record of their study origin, context-of-use, and the dimension and latent factor that they were purported to measure. These items were then reframed such that the context of the item was shifted to that of human-automation interaction while the factor to which each item relates was maintained to the greatest degree possible. This process was consistent with the approach of Hunter (2002) in the development of his aviation safety locus of control measure. Once a preliminary list of items was generated, the full list was reviewed by three subject matter experts for face and content validity (Jones & Wuebker, 1985). Experts were tasked with sorting the preliminary item pool based on a predetermined set of anticipated factor groupings. This task provided expert support for the number of factors to be expected (Hinkin, 1998). Items not assigned to any anticipated factor were placed in their own category and considered for deletion from the development pool.

Phase 2: Reliability and Factor Analysis

Following item generation, inspection and preliminary grouping, the item-pool was administered as a single instrument to a sample of Amazon MTurk workers. Based on the approaches of DeVellis (1991), Hinkin (1998), and Zygmont and Smith (2014), an item-to-respondent ratio of 1:10 was suggested as the ideal standard for determining the appropriate sample size for an exploratory factor analysis for scale development purposes. However, sample sizes as low as 150 participants have been noted to provide reliable factor analysis results given strong inter-item correlations (Hinkin, 1998). Similarly item-to-response ratios as low as 1:4 (Hinkin, 1998) to 1:5 (Zygmont & Smith, 2014) have been suggested as adequate criteria for the determination of sample size. In an effort to ensure the robustness of the analysis and provide a reasonable margin for the removal of outliers and inattentive responses, a sample size of 600 participants was targeted for this phase of scale development. Workers were not required to be MTurk Masters; however, they were required to be in the United States of America, and task visibility was set to “Hidden” to prevent viewing of the task by unqualified candidates.

Following data collection, the data were screened for inattentive and ineligible responses. Screening for participant attentiveness consisted of the removal of duplicate Google Form submissions based on participants' worker IDs. If the ID appeared more than once within the dataset, all responses associated with that worker ID were removed. The data were also screened for excessive yea-saying and nay-saying by checking the number of identical A-LOC measure item responses from an individual participant. If a response consisted of more than 55 identical responses out of 57 items not including the automation experience question, participant inattentiveness was assumed due to their insensitivity to the conceptually opposing items of the measure, and their response was removed. Erroneous responses to demographic data prompts were also considered as a criterion for assumed participant inattentiveness. If, for example, a participant entered their age in the field requesting the participant's country of origin, the response was removed. Removals due to inattentiveness were also processes based on participants' non-adherence to questionnaire instructions. This included any communication by the participant regarding the non-provision of a questionnaire completion code via the MTurk HIT form, the Google Form, or email; the entering of an invalid worker ID as determined via a comparison of the Google Form data and the MTurk results output; or the submission of any Google Form that did not also have a matching MTurk HIT completion record. Checks for participant eligibility were conducted based on reported age and country of origin. If a participant reported either an age of less than 18 years or country of origin other than the United States of America or its territories, their responses were removed from the dataset. Univariate and multivariate outliers were then removed, and descriptive statistics for sample demographics were calculated in SPSS Release 26.0.0.0. Initial scale reliability and item-total correlations were determined using SPSS, and any item with an item-total correlation of less than .4 was excluded from further analysis. Items were also excluded on the basis of expert suggestions, so long as the item suggested for removal had an initial item-total correlation of less than .5. A parallel analysis was then conducted within SPSS using the rawpar.sps program developed by O'Connor (2000). The script was configured for the use of the principle axis factoring method on 5000 parallel datasets based on permutations of the raw dataset. The decision to compute parallel datasets based on permutations of raw data was made in order to produce a model that was more robust against violations of assumptions of sample normality (O'Connor, 2000). The results of this analysis were used to determine the number of factors to be retained during the principle axis factoring procedure.

An exploratory factor analysis using the principle axis factoring method with direct oblimin rotation was then performed. The Kaiser-Meyer-Olkin Measure and Sample Adequacy and Bartlett's Test for Sphericity were calculated in order to determine the dataset's suitability for analysis via factor analysis, and the number of significant factors was cross-checked via an application of Kaiser's eigenvalue rule and an inspection of the scree plot (DeVellis, 1991; Hinkin, 1998; Tabachnick & Fidell, 2013). Items were then removed based on weak loadings (<.4), loading to the incorrect factor, or ambiguous cross-loadings (Hinkin, 1998). Once weak, incorrectly loaded, and cross-loaded items were removed, item removals continued based on the content of the item, and the item's contribution to the explained variance of the subscale and the measure. The factor analysis was re-run after the removal of each item.

Phase 3: Confirmatory Factor Analysis and Validity Testing

Based on the approach of Hinkin (1998), a minimum sample size of 200 was suggested for the performance of a confirmatory factor analysis. In order to determine the appropriate minimum sample size for the correlational analyses required for the establishment of construct validity, an a priori power analysis was conducted using G*Power 3.1.9.2. The test family was set to Exact, the statistical test was set to Correlation: Bivariate normal model, and a priori was selected as the power analysis type. The analysis assumed two-tailed tests throughout the validity testing phase where ρ H_1 was set to .3, α error probability was set at .05, power was set at .9, and ρ H_0 was set to 0. The power analysis determined that a minimum sample size of 112 participants was necessary for correlational analyses. In order to enhance the robustness of the model and account for the removal of inattentive responses, ineligible responses and outliers, a sample size of 400 was targeted for this phase. The MTurk parameters remained the same as in Phase 2; however, an additional qualification was created in order to exclude workers who had completed Phase 2 from partaking in the study again. This assured two independent samples between the two data collection phases.

For this stage of the development process, the reduced item pool if 17 items (Table 10) was distributed along with a collection of published instruments (see Validation Measures below) for use in determining the construct validity of the proposed measure (Hinkin, 1998). Tests for convergent validity were enabled via the distribution of the Rotter (1966) general I-E scale, the Rosenberg (1965) self-esteem scale, and the Dishman and Ickes (1981) self-motivation inventory. Tests for discriminant validity were enabled by comparison to the decision style inventory (Rowe & Mason, 1987), and criterion-related validity was tested using the trust in automation scale (Jian, Bisantz, & Drury, 2000).

Following Phase 3 data collection, the data were screened for inattentive and ineligible responses. Screening for participant attentiveness consisted of the removal of duplicate Google Form submissions based on participants' worker IDs. If the ID appeared more than once within the dataset, all responses associated with that worker ID were removed. Two explicit attention checks were included in this phase. These checks consisted of prompts that direct the participant to select a specific response (e.g. “While completing surveys about automation perceptions, please select strongly agree in response to this question”). If the participant did not respond appropriately to either of these checks, the participant’s response was removed from the dataset. The data were also screened for excessive yea-saying and nay-saying by checking the number of identical A-LOC measure item responses from an individual participant. If a response consisted of more than 16 identical responses out of 18 items not including the automation experience question, participant inattentiveness was assumed due to their insensitivity to the conceptually opposing items of the measure, and their response was removed. Erroneous responses to demographic data prompts were also considered as a criterion for assumed participant inattentiveness. If, for example, a participant entered their age in the field requesting the participant’s country of origin, the response was removed. Removals due to inattentiveness were also processes based on participants’ non-adherence to questionnaire instructions. This included any communication by the participant regarding the non-provision of a questionnaire completion code via the MTurk HIT form, the Google Form, or email; the entering of an invalid worker ID as determined via a comparison of the Google Form data and the MTurk results output; or the submission of any Google Form that did not also have a matching MTurk HIT completion record. Checks for participant eligibility were conducted based on reported age and country of origin. If a participant reported either an age of less than 18 years or country of origin other than the United States of America or its territories, their responses were removed from the dataset. Both univariate and multivariate outliers were then removed and descriptive statistics were calculated in SPSS. Parallel analysis and principle axis factoring procedures were repeated as a means of verifying the factor structure observed in Phase 2. Subsequently, a three-factor model, based on the verified results of the Phase 2 analysis, was developed in AMOS 26 with Internal, Powerful Others, and Chance as the three factors, and a confirmatory factor analysis was performed to assess model fit. Criteria for the determination of adequate model fit were determined based on Hooper, Coughlan and Mullen (2008). Accordingly, an insignificant chi-squared statistic was desirable but not expected. Other measures of fit included in this analysis were root mean squared error of approximation (RMSEA), the comparative fit index (CFI), and the root mean square residual (RMR). Linear correlations between the proposed measure and the validation scales were then tested, and internal consistency was tested using Cronbach's alpha for both the A-LOC measures and its subscales. All correlational and reliability testing was completed in SPSS.

Validation Measures

The following scales were included in the Phase 3 data collection for the purposes of determining A-LOC validity.

Rotter (1966) *General I-E Scale.* Rotter's (1966) general locus of control scale is a 29-item, forced-choice, unidimensional measure for the determination of general locus of control orientation. Each item consisted of a pair of prompts from which the participant was required to choose one based on their agreement with either prompt. Disregarding items that were designed for the purpose of preventing the discovery of the purpose of the scale by participants, it was possible for participants to score zero to 23 points, where higher scores depicted a more external locus of control orientation.

Rosenberg (1965) *Self-Esteem Measure.* The Rosenberg (1965) is a 10-item, unidimensional Guttman scale that functions as a general measure of self-esteem based on the extent to which a respondent agrees or disagrees with a given scale item. Possible scores range from 10 to 40 points, where higher scores indicate higher self-esteem.

Dishman and Ickes (1981) *Self-Motivation Inventory*. The self-motivation inventory (Dishman & Ickes, 1981) is a 40-item multidimensional measure that uses Likert scaling to quantify respondent self-motivation. Based on 19 positively keyed items and 20 negatively keyed items, possible scores range from 40 to 200, where higher scores depict higher self-motivation.

Decision-Style Inventory (Rowe & Mason, 1987). The decision-style inventory is a 20-item measure for the determination of how respondents perceive, understand, and respond to stimuli. These responses, termed decision styles, are divided into four major categories: directive, analytical, conceptual, and behavioral. Respondents with a directive decision style are practically oriented with preferences for data specificity and structure. Respondents with an analytical style prefer intensive data analysis in order to optimize solutions to problems. Respondents with a conceptual style have broad, creative tendencies with reliance on intuition and emotion. Respondents with a behavioral style are primarily socially oriented. A given respondent may have more than one decision style, but the order of dominance is determined by scores assigned to each category. It is, therefore, possible to score a maximum of 160 for a given style, and a minimum of 20, with higher scores indicating increased dominance of that style.

Trust in Automation Scale (Jian, Bisantz, & Drury, 2000). The trust in automation scale is a 12-item, Likert-type scale that quantifies the level of trust that a person has in an automated system. Accounting for the negative coding of five items, possible scores range from 12 to 84, where higher scores indicate higher levels of trust in automation.

Human Subjects Research

Prior to conducting this research, an exemption application was submitted to the Institution Review Board, which systematically reviewed the study so as to ensure that the study was both safe and ethical for human participants. It was anticipated that participants were exposed to no greater risk than that of day-to-day life. Participation was voluntary, and participants were permitted to withdraw from the study at any time. Because participants were Amazon MTurk workers, they had the choice to accept the survey tasks, and they were compensated. Each participant was paid USD $0.25 as compensation for their time and effort. Data ownership resides with the participants, and members of the research team did not collect personally identifying information as a part of the study, thus all participants remained anonymous. Should a participant have decided that he or she no longer wished to be a part of the study, the data collected was excluded from the dataset and destroyed.

Chapter 4
Results

The following chapter provides the results of the analyses conducted for each of the study's three phases. In Phase 1, a formal construct definition was established, and items were generated based on the items of existing locus of control scales. In Phase 2, tests for initial item reliability and factor structure were conducted. The initial item pool was then reduced in order to produce a usable instrument. In Phase 3, the factor structure observed in Phase 2 was verified, a confirmatory factor analysis was used to assess model fit, and construct validation was conducted via correlational analyses.

Phase 1: Construct Definition and Item Generation

The purpose of this phase was to define the construct of automation locus of control, develop an initial hypothesis regarding the structure of its underlying factors, and generate an initial pool based on the items of established locus of control scales. Automation locus of control was defined as the extent to which operators perceive the outcomes of their use of automation as the result of their own actions or of influences external to themselves. This definition was primarily based on Rotter's (1966) general definition of locus of control, which emphasized the relationship between a subject's perception of the cause of a specific experience, and the occurrence of that experience in the context of the actions of the subject. Based on the multi-dimensional structure of existing locus of control scales (Levenson, 1973; Özkan & Lajunen, 2005), a multidimensional factor structure consisting of three major three factors was considered. Each factor would be measured by a subscale such that items grouped on the Internal subscale measured the extent to which an individual perceived their experiences as the result of their own actions. Items grouped to the Powerful Others subscale measured the extent to which an individual perceived their experiences as the result of influences of the system itself or the designers of the system. Items grouped to the Chance subscale measured the extent to which individuals perceived their experiences with automation as the result of chance irrespective of their actions, or the influences of the system or its designers.

A deductive approach to item generation was used in which the items of published locus of control scales were reframed to fit the context of the automation locus of control construct. Origin scales were chosen based on their documented merit as valid and reliable measures of locus of control, the generalizability of their verbiage based on their context of use, and the latent factor structure of each instrument. To this end, the Rotter (1966) general locus of control measure was selected for its notoriety as the foundation of the construct and its measurement, and the Levenson (1973) multidimensional locus of control measure was chosen on the basis of its empirically supported multidimensional factor structure (Presson, Clark, & Benassi, 1997). Initial item generation based on the items of the aforementioned measures resulted in the creation of 57 items that were designed with the intent of capturing a unidimensional measure of internality verses externality (Rotter, 1966), and a multi-factor structure consisting of three distinct latent factors: internal, powerful others, and chance (Levenson, 1973).

Following initial development, scale items were distributed to three subject-matter experts. These experts were all current professors of the Florida Institute of Technology with professional and educational backgrounds in human factors, aerospace engineering, or industrial-organization psychology. Results of the item sorting task produced relatively consistent item groupings with respect to Internal, and Chance items. Some variation was observed among proposed Powerful Others items that was likely due to variations in experts' conceptualization of the multidimensional structure of the locus of control construct. The list of proposed items, and results of the item sorting task are available in Table 2.

Phase 2: Reliability Testing and Exploratory Factor Analysis

The purpose of this phase was to examine the initial reliability of the proposed item-set, establish a preliminary factor structure that is consistent with both established theory and the collected data, and to reduce the size of the initial item-pool to that of a practical scale for use in subsequent analyses.

Data Screening and Demographics

For this initial phase, 600 responses were collected from workers of Amazon MTurk from May 15, 2019 to May 17, 2019. As an assurance of data quality, responses were screened and cleaned prior to any further analyses (Tabachnick & Fidell, 2013). Checks for duplicate responses determined that that 10 participants had submitted the instrument twice. These 20 cases were removed. Subsequent checks for inattentive responding were also conducted resulting in the removal of a further 98 responses. Finally, screening for ineligible participants based on age and country of origin resulted in the removal of 18 responses. Following the removal responses from inattentive and ineligible participants, 464 responses remained.

The handling of outliers followed the guidance of Tabachnick and Fidell (2013) with the removal of both univariate (12) and multivariate (41) outliers. A univariate outlier was defined as any case for which a response to a single scale item differed from the mean response to that item by greater than 3.29 standard deviations. A multivariate outlier was defined as any case for which the combined responses of all scale items produced a significant Mahalanobis Distance (p <

.001). Following the removal of all outliers, a final dataset of 411 responses remained. This produced an item-to-response ratio of 1:7.2. By virtue of total size and item-to-response ratio, the resultant sample was determined to be adequate for factor analysis (Hinkin, 1998; Zygmont & Smith, 2014), and was used for all remaining Phase 2 analyses. Demographic statistics for the sample are available in Table 3. The data consisted of responses from 207 (50.4%) males and 204 (49.6%) females. Respondents' ages ranged from 19 years to 81 years with an average age of 38 years, and a standard deviation of 12.1 years. With respect to race and ethnicity, the majority of participants identified as Caucasian/White ($N = 331$). The remainder of the sample consisted of 25 respondents who identified as African American/Black, 24 who identified as Latin American/Hispanic, 23 who identified as Asian, two who identified as American Indian, five who identified as mixed race, and one participant who specified neither a race nor ethnicity. Regarding the educational background of the sample, 93 respondents indicated having completed their high school diploma/GED, 63 indicated having completed an associate's degree, approximately half of the participants reported having completed their bachelor's degree (N = 194), 49 indicated having completed a master's degree, and seven indicated having completed a philosophical doctorate. The remaining five participants reported partial completion of a college degree, completion of vocational training or trade school, and completion of other professional degree programs.

In order to ensure adequate understanding of the study's definition of automation by participants, respondents were asked to indicate their experience with automation based on a list of automated systems at varying capability levels (see item 1 in the Appendix). Results suggested that the use of autocorrect/predictive text was the most common form of automation experience among respondents ($N = 327$) followed by automated navigation aids ($N = 305$), motion-activated lighting ($N = 301$), low-level kettle/coffee-maker automation ($N = 291$), auto-curated media ($N = 265$), and sunlight-sensitive lighting ($N = 188$). Experience with production robotics and self-driving cars was limited among the sample with system usage being reported by 52 participants and 30 participants respectively. The full statistical output for automation experience is available in Table 3.

Initial Reliability Testing and Item Reduction

As a preliminary measure of internal consistency, Cronbach's alpha was calculated based on responses to all items in the preliminary item pool. The initial Cronbach's alpha for the measure was $\alpha = .94$. Corrected item-total correlations were also determined, and items with item-total correlations, $r < .4$ were removed from the item pool (Clark & Watson, 1995; Hinkin, 1998). Based on this criterion, 21 items were removed from the initial pool (Table 4). Further reduction of the pool was accomplished through the removal of items that were recommended for removal by at least one subject-matter expert, and that also had item-total correlations of $r < .5$, or were suggested for removal by all three subject-matter experts. Based on these criteria a further five items were excluded from the pool. Consequently, 31 items remained to be subjected to factor analysis procedures.

Exploratory Factor Analysis

The exploratory factor analysis determined the number of factors to be retained for the preliminary construction of subscales based on sample data and a priori theory, to determine which of the proposed items loaded to which factor, and to enable further reduction of the item-pool based on the factor loadings. Results of the parallel analysis supported the retention of four factors based on raw data eigenvalues that were greater than the 95th percentile of the random dataset (Table 5 and Figure 4). Following the parallel analysis, principle axis factoring was used to further examine the number of factors to be retained, examine item groupings based on retained factors, and enable decisions on item retention. Results of the Kaiser-Meyer-Olkin measure of sampling adequacy produced a value of .95, and Bartlett's test of sphericity produced $\chi^2 = 8230.4$, $df = 465$, $p < .001$. These results suggested that the current dataset was suitable for factor extraction (Table 6). Results of the initial factor analysis were consistent with the findings of the parallel analysis insofar as the retention of four factors was supported based on Kaiser's eigenvalue rule, and the structure of the scree plot (Table 7, Table 8, Figure 5).

Given the consistent support for the retention of four factors, item groupings based on those factors were determined via the interpretation of the pattern matrix. This initial pattern matrix, provided in Table 9, demonstrated a number of ambiguous and weak item loadings that marginally supported the extraction of the fourth factor, but gave little merit of its retention based on theory. Decisions for item removal were made based on weak item loadings (<.4), or cross loadings for which a single factor failed to load at double the factor loading of any other factor. Items that did not load to the desired latent factor were also removed. The analysis was repeated after the removal of each item, and, once items with weak loadings or ambiguous cross-loadings were removed, three latent factors remained. Item reduction continued based on the content of each item, and the impact of the removal of that item on the total explained variance of the resultant measure. The resultant pattern matrix of the final item-set is available in Table 10. The factor structure of the item-set was supported by the eigenvalues and loadings depicted in Table 11 and Table 12, the scree plot depicted in Figure 6, and the post-reduction parallel analyses matrix and sequence plot depicted in Table 13 and

Figure 7 respectively. Total explained variance for the measure at this stage was 62% with Internal, Powerful Others and Chance subscales producing explained variances of 12%, 10% and 40% respectively.

Phase 3: Confirmatory Factor Analysis and Validity Testing

The purpose of this phase was to examine the reliability of the instrument, verify the factor structure observed in the Phase 2 factor analysis, assess the fit of the exploratory factor model by way of a confirmatory factor analysis, and established the construct validity of the proposed instrument by way of correlations for convergent, divergent and criterion-related validity.

Data Screening and Demographics

For this phase, 431 initial responses were collected from workers of Amazon MTurk between May 30, 2019 and June 2, 2019. Responses were subject, as they were in Phase 2, to screening for respondent ineligibility and inattentiveness (Tabachnick & Fidell, 2013). Checks for duplicate responses determined that that five participants had submitted the instrument twice. These 10 cases were removed. Subsequent checks for inattentive responding were also conducted resulting in the removal of a further 152 responses. Finally, screening for ineligible participants based on country of origin resulted in the removal of 11 responses. The removal of inattentive and ineligible responses resulted in the retention of 258 cases prior to the removal of outliers. Outlier removal followed identical procedures to those of Phase 2. Following the removal of all outliers, a final dataset of 246 responses remained. This produced an item-to-response ratio of 1:14.5. By virtue of total size and item-to-response ratio (Hinkin, 1998; Zygmont & Smith, 2014), the resultant sample was determined to be adequate for factor analysis.

Full demographic data for the sample are provided in Table 14. The data consisted of responses from 102 (41.5%) males and 144 (58.5%) females. Respondents' ages ranged from 19 years to 74 years with an average age of 38 years, and a standard deviation of 12.7 years. With respect to race and ethnicity, the majority of participants identified as Caucasian/White ($N = 196$). The remainder of the sample consisted of 17 respondents who identified as African American/Black, 14 who identified as Latin American/Hispanic, 12 who identified as Asian, one who identified as Eurasian, four who identified as mixed race, and two participants who identified as Middle Eastern. Regarding the educational background of the sample, 54 participants indicated having completed their high school diploma/GED, 35 indicated having completed an associate's degree, 116 reported having completed their bachelor's degree, 32 indicated having completed a master's degree, and six indicated having completed a philosophical doctorate. The remaining three participants reported partial completion of a college degree, and completion of other professional degree programs.

An in Phase 2, respondents were asked to indicate their experience with automation based on a list of automated systems at varying capability levels (Table 14). Results suggested that the use of autocorrect/predictive text was the most common form of automation experience among respondents ($N = 220$) followed by automated navigation aids ($N = 214$), motion-activated lighting ($N = 208$), low-level kettle/coffee-maker automation ($N = 191$), auto-curated media ($N = 189$), and sunlight-sensitive lighting ($N = 110$). Experience with production robotics and self-driving cars was limited among the sample with usage being reported by 20 participants and 16 participants respectively.

Reliability Testing and Exploratory Factor Analysis

As measure of internal consistency, Cronbach's alpha was calculated based on responses to the reduced item pool. The overall Cronbach's alpha for the current measure was $\alpha = .85$. Subscale internal reliability was determined to be $\alpha = .76$, $\alpha = .84$ and $\alpha = .92$ for Internal, Powerful Others, and Chance subscales respectively. Following the reliability analyses, a parallel analysis was conducted to verify the factor structure observed in Phase 2. Results of the analysis (Table 15, Figure 8) supported the retention of three factors based on raw data eigenvalues that were greater than the 95^{th} percentile of the random dataset. Principal axis factoring with a direct oblimin rotation was also repeated to further verify the number of factors to be retained, as well as the factors to which each item had loaded. Results of the Kaiser-Meyer-Olkin measure of sampling adequacy produced a value of .87, and Bartlett's test of sphericity produced $\chi^2 = 2127.3$, $df = 136$, $p < .001$. These results suggested that the current dataset was suitable for factor extraction (Table 16). Results of the initial factor analysis were consistent with the findings of the parallel analysis and the Phase 2 exploratory analysis insofar as the retention of three factors was supported based on Kaiser's eigenvalue rule, and the structure of the scree plot (Table 17, Table 18, Figure 9). The pattern matrix was also consistent with the Phase 2 findings regarding which item loaded to which factor (Table 19).

Confirmatory Factor Analysis

In order to assess the fit of the proposed measure, a confirmatory factor analysis using the maximum likelihood estimation (Tabachnick & Fidell, 2013) was conducted based on the three-factor model established in Phase 2 and verified by the secondary factor analysis performed in Phase 3. This model, depicted in Figure 10, grouped individual scale items to latent factors on an a priori basis in a manner commensurate with the principal axis pattern matrices in Table 13 and Table 19. Results of the confirmatory factor analysis produced $\chi^2 = 251.6$, $df = 116$,

$p < .001$. Results for other metrics of model fit were RMSEA = .07, CFI = .94, RMR = .05.

Validity Testing

A correlational analysis was conducted to determine the relationship between the proposed automation locus of control scale and its subscales, and the battery of validation scales. Accordingly, means, standard deviations, alpha reliabilities and inter-subscale correlations for the automation locus of control scale are provided in Table 21. The analysis showed significant subscale-total correlations, and significant correlations between the Internal subscale, and Powerful Others and Chance subscales. The relationship between Powerful Others and Chance was not significant.

Hypothesis Testing

In order to test the convergent, divergent and discriminant validity of the measure, the following hypotheses were developed:

> $H1_0$: There is no significant relationship between total automation locus of control scores and general locus of control scores.

> $H1_A$: There is a significant relationship between total automation locus of control scores and general locus of control scores.

$H2_0$: There is no significant relationship between total automation locus of control scores and self-esteem scores.

$H2_A$: There is a significant relationship between total automation locus of control scores and self-esteem scores.

$H3_0$: There is no significant relationship between total automation locus of control scores and self-motivation scores.

$H3_A$: There is a significant relationship between total automation locus of control scores and self-motivation scores.

$H4_0$: There is no significant relationship between total automation locus of control scores and directive decision style scores.

$H4_A$: There is a significant relationship between total automation locus of control scores and directive decision style scores.

$H5_0$: There is no significant relationship between total automation locus of control scores and analytical decision style scores.

$H5_A$: There is a significant relationship between total automation locus of control scores and analytical decision style scores.

$H6_0$: There is no significant relationship between total automation locus of control scores and conceptual decision style scores.

$H6_A$: There is a significant relationship between total automation locus of control scores and conceptual decision style scores.

$H7_0$: There is no significant relationship between total automation locus of control scores and behavioral decision style scores.

$H7_A$: There is a significant relationship between total automation locus of control scores and behavioral decision style scores.

$H8_0$: There is no significant relationship between total automation locus of control scores and trust in automation scores.

$H8_A$: There is a significant positive relationship between total automation locus of control scores and trust in automation scores.

With respect to the establishment of convergent validity, Hypothesis 1 tested the relationship between automation locus of control and a measure of general locus of control. Correlational analyses produced a significant relationship between total automation locus of control scores and general locus of control ($r = -.185$, $p = .004$). Because the scores of the general locus of control increase as respondent orientations become more external, the apparent inverse relationship between the two measures was expected. The null hypothesis, $H1_0$, was rejected. Likewise, Hypothesis 2 determined the relationship between automation locus of control and self-esteem. The relationship between the total automation locus of control score and the self-esteem measure was positive and significant ($r = .187$, $p = .003$). Therefore, the null hypothesis, $H2_0$, was rejected. Finally, Hypothesis 3 tested the relationship between automation locus of control and self-motivation. Results of the correlational analysis were positive and significant ($r = .279$ $p < .001$). The null hypothesis, $H3_0$, was rejected.

With respect to discriminant validity, Hypothesis 4 examined the relationship between automation locus of control and directive decision style scores. The results of the analysis were not significant ($r = -.095$, $p = .138$). Therefore, the null hypothesis, $H4_0$, was accepted. Hypothesis 5 examined the relationship between automation locus of control and analytical decision style scores. Results of the correlational analysis were positive and significant ($r = .233$, $p < .001$). Therefore, the null hypothesis, $H5_0$, was rejected. Hypothesis 6 tested the relationship between automation locus of control and conceptual decision style scores. Results of the analysis were insignificant ($r = -.012$, $p = .854$). Consequently, the null hypothesis, $H6_{0,}$ was accepted. Finally, hypothesis 7 examined the relationship between automation locus of control and behavioral decision style scores. Results of the analysis were insignificant (*r* = -.111, *p* = .083). Therefore, the null hypothesis, $H7_0$, was accepted.

Regarding the establishment of criterion-related validity, hypothesis 8 investigated the relationship between automation locus of control scores and trust in automation. Results of the correlational analysis supported a significant positive relationship (*r* =.183, *p* = .004). Therefore, the null hypothesis, $H8_0$ was rejected.

Chapter 5
Conclusion

Overview

Given the extent to which automation has become a normal and, in cases, integral part of personal and professional life, the analysis of human-automation interaction remains a prerequisite for its informed design, and safe, effective and efficient use (Fitts, 1951; Parasuraman & Manzey, 2010; Rasmussen, 1983). This includes the analysis of control perceptions as a probable indicator or determinant of operator performance (Berberian, Sarrazin, Le Blaye, & Haggard, 2012). To this end, locus of control was offered as an established construct for the determination of the outcome expectancies of system users. Rooted in general expectancy theory, locus of control was chosen for its established nomological network, its prolific use as an indicator of human control perceptions, and its hypothesized value as a predictor of human behavior (Lefcourt, 1966; Rotter, 1966; Rotter, 1975). The applicability of locus of control to human-automation interaction evaluation was further bolstered by the demonstrated ability to both inductively and deductively develop and deploy context-specific locus of control measures based on refined construct definitions and factor structures within a specific content domain (Hunter, 2002; Lefcourt, Von Baeyer, Ware, & Cox, 1979; Ross, Ross, Short, & Cataldo, 2015; Özkan & Lajunen, 2005). Therefore, the purpose of this study was to develop and validate a locus of control measure that could quantify operators' locus of control in the context of human-automation interaction. This measure was proposed as an additional method for analyzing control perceptions as an indicator and modifier of human behavior and, thus, system performance.

Discussion and Interpretation

The item pool, derived in similar fashion to the development of Hunter's (2002) aviation safety locus of control scale, was based on the items of the Rotter (1966) general I-E measure and the Levenson (1973) multidimensional locus of control measure. The a priori targeting of three latent factors, based on Levenson's multidimensional measure, resulted in the development of 57 initial items that were designed to conform with established conceptual definitions for Internal, Powerful Others, and External subscales (Table 2). Preliminary support for a three-factor structure was demonstrated via the subject-matter expert feedback. Although there was some variation as to the exact composition of factor groupings for the initial pool, all subject-matter experts made use of all three available factor groups. Regarding the face validity of the items, expert feedback was consistent in the suggested removal of items whose wording targeted comprehension of the system as opposed to outcome expectancies.

The primary objectives of Phase 2 were the examination of the initial reliability of the proposed item-set, the establishment of a preliminary factor structure consistent with both established theory and collected data, and the reduction of the initial item-pool to a practical but theoretically meaningful scale for use in subsequent analyses. The high Cronbach's alpha ($\alpha = .94$) suggested that the initial item pool adequately captured the content domain of the automation locus of control construct (Churchill, 1979), and demonstrated the high internal reliability of the initial item pool. The removal of items with weak item-total correlations ($r < .4$) reduced the size of the item-pool, and eliminated items that did not measure the common construct captured by the remaining items. Similarly, the removal of items based on the suggestions of subject matter experts and low item-total correlations (<.5) eliminated items that did not measure the same core construct as the rest of the item-pool and supported the face validity of the measure. Although the removal criteria were either empirically grounded or based on unanimous expert opinion, it should be noted that these procedures precluded the examination of factor loadings for these items. Nonetheless, their preliminary exclusion was assumed to have a negligible effect on the final factor structure of the instrument. Furthermore, removal criteria that combined individual expert suggestions with relatively low item-total correlations aligned well with the consistency in observed experts' suggested removal of items that targeted system understanding as opposed to outcome expectancies.

The results of the parallel analysis and principal axis factoring procedures demonstrated consistent initial support for the retention of four latent factors; however, the retention of the fourth factor was only marginal across all results. The resultant pattern matrix (Table 9) demonstrated weak and ambiguously loading items that were not consistent with the grounding theory of the construct. The incremental removal of weakly loaded and ambiguously cross-loaded items was, therefore, justified (Clark & Watson, 1995; Hinkin, 1998). Following item reduction, the exploratory analyses supported a three-factor solution that was conceptually consistent with the internal, powerful others, and chance subscales of other multidimensional measures of locus of control (Bradley & Sparks, 2002; Levenson, 1973; Ross, Ross, Short, & Cataldo, 2015). Although the three-factor solution is supported by a number of established scales, it stands to note that the emergence of a factor structure is influenced by the characteristics of the sample. It is, therefore, not appropriate to assert strict model correctness based data from one sample. Rather, the structure is indicative of the differences to which members of the sample are sensitive for any given analysis (Rotter, 1975). Although the observed thee-factor structure is indicative of the perceptual capabilities of the sample, it is plausible that empirical support for a given subscale structure of A-

LOC may change based on the characteristics of the study sample. Pilots, for example, may respond in a pattern that is indicative of perceived differences in the influences of individual system designers, aircraft manufacturers, and system regulators; whereas, a smartphone user may only perceive their phone at the level of the manufacturer branding under which it was sold. This consideration notwithstanding, the existing support for a three-dimensional structure based on perceived influences of oneself, powerful others, and chance supported the adoption of this model as the foundation of the scale.

Total variance explained (62%) was also deemed adequate based on Hinkin's (1998) 60% acceptance criteria for scale development. Explicit effort was required on the part of the researchers to circumvent the insensitivity of purely empirical item removal criteria to the construct definition and content domain of the measure (Nunnally & Bernstein, 1994). As such, some items that negligibly increased explained variance were discarded in favor of items that were more conceptually meaningful to the measurement of the construct. This consideration was instrumental in the retention of scale items that maximized explained variance based on the differences that members of the target population were likely able to perceive. Items that attempted to differentiate between the roles of system designers and system integrators as distinct elements in automation manufacturing were discarded in favor of items that introduced considerations of operator proficiency. Specifically, participants were assumed to be insensitive to the differences between “The outcomes of my use of automation are chiefly controlled by system designers” and “The outcomes of my use of automation are chiefly controlled by system integrators.” Conversely, “The outcomes of my use of automation are chiefly controlled by system designers” and “Although I may be a proficient user, the outcomes of my use of automation are determined by system designers” were determined to be perceivably different based on the introduction of elements of operator proficiency.

The primary objectives of Phase 3 were the determination of the internal reliability of A-LOC, the assessment of the fit of the three-factor model using a confirmatory factor analysis, and the evaluation of convergent, discriminant, and criterion-related validity. By this stage, the A-LOC measure consisted of 17 items scored on a Likert-type scale from Strongly Disagree, scored as -2, to Strongly Agree, scored as +2. Items on the Powerful Others and Chance subscales were reverse scored so that higher overall scale scores indicated more internal automation locus of control. A participant’s overall A-LOC score was the summed score for all items across all subscales. Subscale scores were also calculated as the summed scored for all items within a given subscale.

Regarding scale reliability, results of Cronbach's alpha analyses for both the overall measure and all three subscales supported adequate internal consistency with subscale Cronbach's α scores ranging from .76 to .92 and an overall Cronbach's α of .85. These results exceeded the reliability criteria of .7 (Hinkin, 1998), and provided a good indication that scale items were capturing a similar content domain based on the consistency of participants' responses (Churchill, 1979). Comparisons of the Phase 2 exploratory factor analysis and the results of secondary parallel and principle axis factoring analyses further supported the reliability of the factor structure of the scale. Although testing of the factor structure across different target populations would best support the external generalizability of the three-factor structure, the observed similarities provide good initial support for the robustness the factor structure using two independent samples from a single target population.

The suitability of the three-factor model for A-LOC was evaluated via a confirmatory factor analysis using a maximum likelihood estimation. Results of the analysis produced a significant chi-squared statistic ($\chi^2 = 251.6$, $df = 116$, $p < .001$), which suggested that the sample covariance matrix differed significantly from the estimated population covariance matrix (Tabachnick & Fidell, 2013). However, the robustness of the chi-squared test of model fit has been challenged due to its sensitivity to violations in multivariate normality, and over-rejection of the null hypothesis due to minute differences between the sample and estimated covariance matrices that are inherent in sufficiently large sample sizes (Hooper, Coughlan, & Mullen, 2008; Kline, 2011; Tabachnick & Fidell, 2013). Consequently, the analysis of alternate measures of model fit is recommended, and the examination of the RMSEA, CFI and RMR is supported (Hooper, Coughlan, & Mullen, 2008). Analyses of the other measures of model fit, depicted in Table 20, generally suggested that the model approached good fit (Hooper, Coughlan, & Mullen, 2008). The RMSEA (.07) supported favorable model parsimony and suggested good model fit based on a modification of the chi-squared criterion that enables the determination of model fit based on reasonable imperfections in the comparison of the observed model and the estimated population model (Brown, 2006; Hooper, Coughlan, & Mullen, 2008). Similarly, the CFI (.93) suggested good model fit based on the comparison of the sample covariance matrix and an estimated null matrix that assumes uncorrelated latent factors (Hooper, Coughlan, & Mullen, 2008). Good model fit was also supported by the RMR (.05), which demonstrated that the residuals between the observed covariance matrix and the hypothesized covariance matrix were reasonably similar (Hooper, Coughlan, & Mullen, 2008). The item loadings of the confirmatory factor analysis (Figure 10) supported a strong relationship between individual items within each sub-factor

(Clark & Watson, 1995; Hinkin, 1998), and the low covariances between latent factors demonstrated that the subscales are likely measuring different elements of the overall construct (Nunnally & Bernstein, 1994). Overall, the resultant factor structure of the A-LOC measure was empirically supported via exploratory and confirmatory analyses and aligned well with the grounding theory of the automation locus of control construct.

The construct validity of the A-LOC measure was divided into individual analyses of convergent, divergent and criterion-related validity. With respect to convergent validity, significant correlations between A-LOC and its subscales, and Rotter's (1966) general I-E scale, Rosenberg's (1965) self-esteem measure, and Dishman and Ickes' (1981) self-motivation inventory were expected. Results only partially supported the convergent validity of the measure as total A-LOC scores supported significant but weak positive correlations for all measures (Table 22). However, analysis of the individual subscale scores only supported the convergent validity of the Chance subscale for all measures of the convergent test battery. The Internal subscale had a weak, significant, positive correlation with the self-motivation measure, and the Powerful Others subscale was not correlated significantly with any convergent measure. Regarding the weak but significant correlations between A-LOC subscale scores and convergent validity variables, the findings of these analyses are consistent with the weak but statistically significant validity correlations found by Bradley and Sparks (2002) and Hunter (2002). However, the development and validation of other published locus of control scales demonstrated convergent validity that was supported by Pearson's correlations with absolute values in the range of .4 to .6 (Kourmousi, Xythali, & Koutras, 2015; Lindbloom & Faw, 1982). The relative weakness of observed correlations notwithstanding, the results of the convergent validity tests provided partial support for the convergent validity of the A-LOC measure by virtue of their significance and expected directionality.

Analysis of the A-LOC measure's discriminant validity was, for the most part, supported by the correlational analyses between the overall A-LOC scores, and the subscale scores of the decision style inventory (Rowe & Mason, 1987). Significant correlations were observed between the total A-LOC score and the analytical decision style scores, as well as between directive, analytical and behavioral decision style scores, and the A-LOC Chance subscale (Table 22). The combination of significant relationships between total scores and Chance subscale scores, and the analytical decision style scores are of interest because their positive significant correlations support the conceptual convergence of the two scales. Consideration of the conceptual definition of the analytical decision style supported this relationship because the style denotes an individual who tends to seek extensive amounts of information in order to make decisions. By contrast, the comparatively very weak or insignificant subscale correlations of the remaining decision styles with total, Internal, Powerful Others, and Chance A-LOC scores are indicative of a lack of a conceptual relationship between the two constructs. However, the relationships that are significant may suggest a link between decision styles and locus of control, particularly where considerations of chance are concerned, that is not observable in other domains or outside the paradigm of a multidimensional construct.

Analysis of the criterion-related validity of automation locus of control required both statistical significance and the appropriate directionality of the relationship between A-LOC scores and trust in automation scores for satisfactory support to be considered. With respect to total A-LOC scores and Chance subscale scores, the observation of a weak, significant relationship between A-LOC scores and trust in automation supported the criterion-related validity of the scale. Much like the analysis of convergent validity, however, the relationships between trust in automation, and both Internal and Powerful Others subscale scores were not significant. It is possible that the observed lack of significance among the A-LOC subscales is due to fundamental differences in the perceptions of trust between people, and trust between people and machines (Hoff & Bashir, 2015; Lee & Moray, 1992). The possibility also exists that the unidimensional measures of locus of control used in the reference studies precluded the observation of significant relationships at the level of detail that the A-LOC measure provides.

Limitations

This study was limited by the use of Amazon MTurk for the recruitment of the samples used for data collection. In general, convenience samples are liable introduce unique sample characteristics and biases that limit the generalizability of results to the target population from which the data was collected (Ary, Jacobs, Sorensen, & Razavieh, 2010). For this study, the use of Amazon MTurk added layers of complexity both with respect to assurances of data quality, and the achievement of two independent samples. Although there are studies that suggest comparable quality between MTurk samples and other convenience samples (Buhrmester, Kwang, & Gosling, 2011; Walter, Seibert, Goering, & O'Boyle Jr, 2018), relying on those results as an assurance of data quality is not advised particularly if one’s goal is the generalization of findings beyond the research sample. In order to assure the integrity of the data used for this study, explicit attention checks, tests for yea-saying and nay-saying, checks for erroneous responses to demographic prompts, determinations of non-adherence to survey instructions, and checks for participant ineligibility based on age and country of origin were used.

The use of remote participants also constituted a limitation on the types of tests that could be performed as measures of criterion-related validity as there was no way to administer and monitor performance tasks such as the MATB-II (National Aeronautics and Space Administration, 2016) without sacrificing the ability to directly oversee the task. This was also a limitation of the use of remote participants for the completion of self-report measures; however, self-report measures were determined to have remained within the realm of practicality whereas simulated performance tasks did not.

Normality of the data was another issue in developing the A-LOC scale because there are assumptions of normality for confirmatory factor analysis procedures, and correlational analyses (Table 23). With respect to the confirmatory factor analysis, the maximum likelihood estimation is generally robust against deviations from data normality given a sufficiently large (> 2,500) sample size (Tabachnick & Fidell, 2013). The comparatively small sample size of 246 responses, therefore, means that the criteria for robustness on the basis of sample size was not satisfied.

The use of a deductive approach to item generation limited the number and variety of items in the initial development pool to contextually shifted recreations of existing items. Although this approach was chosen due to its purported ability to reliably generate scale items, the variety of items created is severely limited by the development scales that are both available and suitable for reframing. Rotter's (1966) and Levenson's (1973) scales were chosen as the base from which new items were generated based on their public availability and their framing in sufficiently general context so that reframing in the context of human-automation interaction was feasible. The consideration of more scales that are written in a sufficiently general context would have provided a wider variety of source items from which factor structures could be derived and scale reduction decisions made.

Participants' apparent inability to perceive automation in its day-to-day use was also a potential limitation of the study. The lower-than-expected indications of experience with typing aids such as autocorrect and predictive text from samples of MTurk workers suggested that, although members of that population may engage in regular use of automation, their mental models of the technology do not include an explicit awareness or understanding of its automated functions. Therefore, it is possible that the factor structure and construct validity tests may have been influenced by participants' insensitivity to nuances in contemporary automation design.

Practical Implications

In addition to providing modest preliminary support for the construct validity of the A-LOC measure, the groundwork laid by this study offers several key elements in the development of a valid and reliable control perceptions measure for use in a human-automation interactions context. First, the literature review established the relationship between control perceptions and operator behavior, giving credence to the consideration of locus of control as a variable of interest for user-centered evaluations of human-machine system performance. The propagation of such an idea is critical to both designers and operators particularly in cases where system performance has tangible implications for operational safety and security. Second, the formal establishment of a context-specific construct definition provided a starting point from which either deductive or inductive approaches to scale development may proceed (DeVellis, 1991; Hinkin, 1998; Nunnally & Bernstein, 1994). Without this foundation, it is entirely possible to stray significantly from the core construct as new scales are developed and existing scales are refined or adapted (Rotter, 1975). Third, the establishment of an initial item pool, based on a deductive approach to item generation, both demonstrated the feasibility of adapting existing measures, and provided a repository of items that may be revised, reframed, and revalidated for use in subsequent iterations of this measure. This effort also provided additional support for a general three-factor structure (Internal, Powerful Others and Chance) of the locus of control construct that is generally robust to changes in context of use, and demonstrated how the consideration of subscales can result in the identification of new relationships between latent factors and other variables, particularly those that are context specific, that would have otherwise been lost in the aggregate scores of a unidimensional measure. Finally, the correlational analyses highlighted the potential uniqueness of the relationships between the automation locus of control construct and other variables. Constructs, such as critical decision-making and problem solving, may be considered as alternative measures for the establishment of convergent validity that may be more appropriate than a measure of self-esteem.

Future Research

In light of the findings of the current study, future research for the establishment of the nomological network of automation locus of control is proposed, and the evaluation of criterion-related validity through considerations of alternate criterion variables and data triangulation methods is suggested. Investigations for the establishment of the external generalizability of the measure are also strongly encouraged. Finally, the use of the A-LOC measure as a tool for the evaluation of operator control perceptions is recommended.

Although modest support for the construct validity of the measure was observed, these results were based primarily on weak correlations on the Chance subscale and the total measure scores. The results of the correlational analyses for self-motivation and decision style supported the investigation of measures of problem solving and decision-making as possible alternate measures of convergent validity. Future investigations could also benefit from the targeting of construct validity measures at the subscale level, and via comparisons of the current scale to other multidimensional measures of locus of control.

With respect to criterion related validity, the use of criteria other than psychometrics was supported by the literature (Hunter, 2002; Özkan & Lajunen, 2005). In this respect, the examination of the relationship between A-LOC scores and performance metrics derived from an evaluation tool, such as MATB-II (National Aeronautics and Space Administration, 2016), could better demonstrate the criterion-related validity of the measure. Additional support may be attained via the triangulation of A-LOC scores, MATB-II performance metrics, and intentional binding (Berberian, Sarrazin, Le Blaye, & Haggard, 2012).

The external generalizability of the measure depends on its continued evaluation across different target populations. It is generally regarded that the external validity of any research finding is only as comprehensive as its sample allows (Ary, Jacobs, Sorensen, & Razavieh, 2010), and the effects of sample characteristics have been noted to have substantial effect on the factor structure of locus of control scales (Rotter, 1975). By extension, the potential influence of sample characteristics on the results of tests for factor structure and validity cannot be overlooked. Therefore, the evaluation of the current A-LOC measure is recommended using other generally accepted academic target populations such as college students.

Although the current findings called for further investigation of the construct validity of the A-LOC measure, the findings also supported the factor structure of the scale and provided reasonable preliminary support for the A-LOC scale as a measure of automation control perceptions. Consequently, the use of the A-LOC measure is suggested to investigate the control perceptions of system operators within specific task domains. These include pilot perceptions of cockpit automation, medical practitioner control perceptions of automated medical equipment, and driver control perceptions of self-driving cars. It is projected that, in any of these task domains, the A-LOC measure would serve to support user-centered investigations of system performance, thereby enabling informed system design through detailed considerations of human-automation interaction.

Conclusion

The automation locus of control scale was developed to enable the empirical evaluation of the effect of control perceptions on operator behavior in the context of human-automation interaction. Established theory, expert feedback, and exploratory factor analyses supported a three-factor structure (Internal, Powerful Others, and Chance) to measure control perception orientations. Results of a confirmatory factor analysis demonstrated good model fit, and correlational analyses demonstrated adequate preliminary construct validity for the measure. The unexpected lack of statistical significance for the relationships between Internal and Powerful Others subscales, and measures for convergent and criterion-related validity suggested that the traditional nomological network for locus of control may not be generalizable to considerations of automation control perceptions. The unexpected significance of the relationship between Chance and overall A-LOC scores, and a measure of analytical decision style suggested that measures that assess decision-making and critical thinking may correlate significantly with automation control perceptions. Nonetheless, the resultant A-LOC scale was supported as a novel measure of control perceptions in the context of human-automation interaction, and its use as a means informing automation design through an evaluation of operator behavior was encouraged.

References

Ary, D., Jacobs, L. C., Sorensen, C., & Razavieh, A. (2010). *Introduction to research in education* (8th ed.). Belmont, CA: Wadsworth Cengage Learning.

Bainbridge, L. (1982). Ironies of automation. *IFAC Proceedings, 16*(6), 129-135.

Barg-Walkow, L. H., & Rogers, W. A. (2016). The effect of incorrect reliability information on expectations, perceptions, and use of automation. *Human Factors, 58*, 242-260.

Berberian, B., Sarrazin, J.-C., Le Blaye, P., & Haggard, P. (2012). Automation technology and sense of control: A window on human agency. *Public Library of Science One, 7*(3), 1-6.

Bornstein, M. H., Jager, J., & Putnick, D. L. (2013). Sampling in developmental sciences: Situations, shortcomings, solutions, and standards. *Developmental Review, 33*, 357-270.

Bradley, G. L., & Sparks, B. A. (2002). Service locus of control: Its conceptualization and measurement. *Journal of Service Research, 4*(4), 321-324.

Brambilla, A., Alvi, H., Himanshu, V., Lalanne, D., Jusselme, T., & Andersen, M. (2017). "Our inherent desire for control": A case study of automation's impact on the perception of control. *Energy Procedia, 122*, 925-930.

Brown, T. A. (2006). *Confirmatory Factor Analysis for Applied Research.* New York, NY: The Guilford Press.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science, 6*(1), 3-5.

Busseri, M. A., Lefcourt, H., & Kerton, R. R. (1998). Locus of control for consumer outcomes: Predicting Consumer Behavior. *Journal of Applied Social Psychology, 28*(12), 1067-1087.

Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology, 12*, 53-81.

Chittaro, L. (2014). Changing user's safety locus of control through persuasive play: An application to aviation safety. *Persuasive Technology* (pp. 31-42). Switzerland: Springer International Publishing.

Churchill, G. A. (1979). A paradigm for developing better measures of marketing constructs. *Journal of Marketing Research, 16*(1), 64-73.

Clark, L. A., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment, 7*(3), 309-319.

Cremer, I. (2015). Creating and validating a perception of airport sustainability (PASS) scale (Doctoral dissertation).

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 291-334.

Crosby, L. A., & Stephens, N. (1987). Effects of relationship marketing on satisfaction, retention, and process in the life insurance industry. *Journal of Marketing Research, 24*, 404-411.

de Winter, J. C., & Dodou, D. (2014). Why the Fitts list has persisted throughout the history of function allocation. *Cognition, Technology and Work, 16*(1), 1-11.

Dekker, S. W., & Woods, D. D. (2002). MABA-MABA or Abracadabra? Progress on human-automation co-ordination. *Cognitions, Technology and Work, 4*, 240-244.

DeVellis, R. F. (1991). *Scale Development: Theory and applications* (Vol. 26). Newbury Park, California: Sage.

Dishman, R. K., & Ickes, W. (1981). Self-motivation and adherence to therapeutic exercise. *Journal of Behavioral Medicine, 4*(4), 421-438.

Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing, & T. M. Haladyna, *Handbook of test development* (pp. 3-25). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Ferguson, L. (2004). External validity, generalizability and knowledge utilization. *Journal of Nursing Scholarship, 36*(1), 16-22.

Fitts, P. M. (1951). *Human engineering for an effective air-navigation and traffic-control system.* Oxford, England: National Research Council.

Fleischer, A., Mead, A. D., & Huang, J. (2015). Inattentive responding in MTurk and other online samples. *Industrial and Organizational Psychology, 8*(2), 196-202.

Galvin, B. M., Randel, A. E., Collins, B. J., & Johnson, R. E. (2018). Changing the focus of locus (of control): A targeted review of the locus of control literature and agenda for future research. *Journal of Organizational Behavior, 39*(7), 820-833.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 2*(1), 104-121.

Hoff, K. A., & Bashir, M. (2015). Trust in automation: Integrating empirical evidence on factors that influence trust. *Human Factors, 57*(3), 407-434.

Holland, J. E., Kochenderfer, M. J., & Olson, W. A. (2013). Optimizing the next generation collision avoidance system for safe, suitable and acceptable operational performance. *Air Traffic Control Quarterly, 21*(3), 275-297.

Hooper, D., Coughlan, J., & Mullen, M. (2008). Structural equation modelling: Guidelines for determining model fit. *Electronic Journal of Business Research Methods, 6*(1), 53-60.

Hornaday, R. W., & Curran, K. E. (1987). Decision styles and student simulation performance: A replication. *Developments in Business Simulation and Experiential Exercises, 14*, 99.

Hunter, R. D. (2002). Development of an aviation safety locus of control scale. *Aviation Space and Environmental Medicine, 73*(12), 1184-1188.

Jager, J., Putnick, D. L., & Bornstein, M. H. (2017). More than just convenient: The scientific merits of homogeneous convenience samples. *Monographs of the Society for Research in Child Development, 82*(2), 13-30.

James, W., & Rotter, J. B. (1958). Partial and 100% reinforcement under chance and skill conditions. *Journal of Experimental Psychology, 55*, 397-403.

Jian, J.-Y., Bisantz, A. M., & Drury, C. G. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics, 4*(1), 53-71.

Johnson, R. E., Rosen, C. C., & Levy, P. E. (2008). Getting to the core of core self-evaluation: a review and recommendations. *Journal of Organizational Behavior, 29*, 391-413.

Jones, J. W., & Wuebker, L. (1985). Development and validation of the safety locus of control scale. *perceptual and Motor Skills, 61*(1), 151-161.

Kalnbach, L. R., & Hinsz, V. B. (1999). A conceptualization and test of the influences of individual differences in goal-setting situations. *Journal of Applied Social Psychology, 29*(9), 1854-1878.

Kline, R. B. (2011). *Principles and Practice of Structural Equation Modeling.* New York: The Guilford Press.

Kourmousi, N., Xythali, V., & Koutras, V. (2015). Reliability and validity of the multidimensional locus of control IPC scale in a sample of 3668 Greek educators. *Social Sciences, 4*, 1067-1078.

Lambert, Z. V. (1980). Consumer alienation, general dissatisfaction, and consumerism issues: Conceptual and managerial perspectives. *Journal of Retailing, 56*, 3-24.

Landers, R. N., & Behrend, T. S. (2015). An inconvenient truth: Arbitrary distinctions between organizational, mechanical turk, and other convenience samples. *Industrial and Organizational Psychology, 8*(2), 142-164.

Lee, J. D., & Moray, N. (1992). Trust, control strategies and allocation of function in human machine systems. *Ergonomics, 22*, 671-691.

Lefcourt, H. M. (1966). Internal versus external control of reinforcement: A review. *Psychological Bulletin, 65*(4), 206-220.

Lefcourt, H. M., Von Baeyer, C. L., Ware, E. E., & Cox, D. J. (1979). The multidimensional-multiattributional causality scale: The development of a goal specific locus of control scale. *Canadian Journal of Behavioral Science/Revue canadienne des sciences du comportement, 11*(4), 286-304.

Levenson, H. (1973). Multidimensional locus of control in psychiatric patients. *Journal of Consulting and Clinical Psychology, 41*(3), 397-404.

Levenson, H., & Miller, J. (1976). Multidimensional locus of control in sociopolitical activists of conservative and liberal ideologies. *Journal of Personality and Social Psychology, 33*(2), 199-208.

Lilly, J. D., & Virick, M. (2006). The effect of personality on perceptions of justice. *Journal of Managerial Psychology, 21*(5), 438-458.

Lindbloom, G., & Faw, T. T. (1982). Three measurements of locus of control: What do they measure? *Journal of Personality Assessment, 46*(1), 70-71.

Murray, K. B. (1991). A test of services marketing theory: Consumer information acquisition activities. *Journal of Marketing, 55*, 10-25.

National Aeronautics and Space Administration. (2016, July 13). *MATB-II*. Retrieved from MATB-II: https://matb.larc.nasa.gov/

Ng, T. W., Sorensen, K. L., & Eby, L. T. (2006). Locus of control at work: a meta-analysis. *Journal of Organisational Behavior, 27*, 1057-1087.

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric Theory* (3rd ed.). New York: McGraw-Hill, Inc.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behavior Research Methods, Instrumentation, and Computers, 32*, 396-402.

Oğuz, A., & Sariçam, H. (2016). The relationship between critical thinking disposition and locus of control in pre-service teachers. *Journal of Education and Training Studies, 4*(2), 182-192.

Özkan, T., & Lajunen, T. (2005). Multidimensional traffic locus of control scale (T-LOC): factor structure and relationship to risky driving. *Personality and Individual Differences, 38*, 533-545.

Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors, 52*, 381-410.

Parasuraman, R., & Riley, V. (1997). Humans and automation: use, misuse and abuse. *Human Factors, 39*(2), 230-253.

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, 30*(3), 286-297.

Phares, E. J. (1957). Expectancy changes in skill and chance situations. *The Journal of Abnormal and Social Psychology, 53*(3), 339-342.

Presson, P. K., Clark, S. C., & Benassi, V. A. (1997). The levenson locus of control scales: Confirmatory factor analysis and evaluation. *Social Behavior and Personality, 25*(1), 93-104.

Pritchett, A. R., Kim, S., & Feigh, K. M. (2014). Measuring human-automation function allocation. *Journal of Cognitive Engineering and Decision Making, 8*(1), 52-77.

Rasmussen, J. (1983). Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. *IEEE Transactions on Systems, Man, and Cybernetics, SMC-13*(3), 257-266.

Rosenberg, M. (1965). *Society and the adolescent self-image.* Princeton, NJ: Princeton University Press.

Ross, T. P., Ross, L. T., Short, S. D., & Cataldo, S. (2015). The multidimensional health locus of control scale: Psychometric properties and form equivalence. *Psychological Reports: Mental and Physical Health, 116*(3), 889-913.

Rotter, J. B. (1966). Generalized experiences for internal versus external control of reinforcement. *Psychological Monographs: General and Applied, 80*(1), 1-27.

Rotter, J. B. (1975). Some problems and misconceptions related to the construct of internal versus external control of reinforcement. *Journal of Consulting and Clinical Psychology, 43*(1), 56-57.

Rowe, A., & Mason, R. (1987). *Managing with style: A guide to understanding, assessing, and improving decision making.* San Francisco, CA: Jossey-Bass Publishers.

Sabatelli, R. M., Buck, R., & Dreyer, A. (1983). Locus of control, interpersonal trust, and nonverbal communication accuracy. *Journal of Personality and Social Psychology, 44*(2), 399-409.

Sarter, N. B., Woods, D. D., & Billings, C. E. (1997). Automation Surprises. In *Handbook of Human Factors & Ergonomics* (2nd ed., pp. 1926-1943). Wiley.

Skinner, E. A. (1966). A guide to constructs of control. *Journal of Personality and Social Psychology, 71*(3), 549-570.

Smith, S. M., Roster, C. A., Golden, L. L., & Albaum, G. S. (2016). A multi-group analysis of online survey respondent data quality: Comparing a regular USA consumer panel to MTurk samples. *Journal of Business Research, 69*, 3139-3148.

Sosu, E. M. (2013). The development and psychometric validation of a critical thinking disposition scale. *Thinking Skills and Creativity, 6*, 107-119.

Spector, P. E. (1988). Development of the work locus of control scale. *Journal of Occupational Psychology, 61*, 335-340.

Srinivasan, N., & Ratchford, B. (1991). An empirical test of a model of external search for automobiles. *Journal of Consumer Research, 18*(2), 233-242.

Strauch, B. (2018). Ironies of automation: Still unresolved after all these years. *IEEE Transactions on Human-Machine Systems, 48*(3), 419-433.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). New York: Pearson Education Inc.

Taylor, K. M., & Pompa, J. (1990). An examination of the relationships among career decision making self-efficacy, career salience, locus of control, and vocational indecision. *Journal of Vocational Behaviour*, 17-31.

Thompson, T. S. (2010). *The impact of locus of control on decision making styles of mid-level managers (Order No. 3398345).* (Doctoral dissertation). Retrieved from https://search.proquest.com/docview/219916906?accountid=27313

United States Census Bureau. (2019). *Annual Estimates of the Resident Population by Sex, Age, Race, and Hispanic Origin for the United States and States: April 1, 2010 to July 1, 2018.* Washington, D.C.: U.S. Department of Commerce.

Walter, S. L., Seibert, S. E., Goering, D., & O'Boyle Jr, E. H. (2018). A tale of two sample sources: Do results from online panel data and conventional data converge? *Journal of Business and Psychology*, 1-28.

Wessling, K. S., Huber, J., & Netzer, O. (2017). MTurk character misrepresentation: Assessment and Solutions. *Journal of Consumer Research, 44*(1), 211-230.

Zygmont, C., & Smith, M. R. (2014). Robust factor analysis in the presence of normality violations, missing data, and outliers: Empirical questions and possible solutions. *The Quantitative Methods for Psychology, 10*(1), 40-55.

Tables

Table 1. Means, standard deviations, coefficient alphas, and correlations for Work Locus of Control (WLCS)

	Sample					
	1	2	3	4	5	6
Mean	41.7	36.8	39.2	38.0	39.4	36.9
Standard Deviation	9.6	9.9	11.9	9.0	9.1	9.6
Coefficient alpha	0.85	0.85	0.85	0.75	0.80	0.85
Correlations of WLCS with:						
Job satisfaction	-42* (82)	-54* (35)	-62* (99)	-68* (256)	-	-43* (496)
Commitment	-20 (84)	-26 (39)	-26* (99)	-	-	-
Intention	13 (83)	14 (39)	35* (99)	38* (286)	13 (160)	-
Autonomy	-18 (83)	-	-	-	-	-10* (496)
Influence	-18 (83)	-45 (39)	-47* (98)	-	-	-
Role stress	-	-	-	32* (287)	-	-
Tenure	-	08 (38)	05 (95)	-10 (52)	-	-07 (496)
Consideration	-	-26 (37)	-34* (93)	-34*	-	-
Initiating structure	-	-31 (38)	-35* (95)	(52)	-	-
Social desirability	005 (149)					

General I-E	57* (144)	55* (160)	49* (496)

Note. *p<0.05. Numbers in parentheses are sample sizes. Decimal points omitted in correlations. Reprinted with permission from Spector (1988, p. 338).

Table 2. Phase 1 Initial Item Pool and Subject-Matter Expert Feedback

Origin Scale	Item	Factor Group			
		1	2	3	NF[a]
Rotter (1966)	People's difficulties with automation result from deficiencies in their use of the system.	3			
Rotter (1966)	One of the major reasons automation is ineffective is because operators do not take the time to understand it.	3			
Rotter (1966)	Capable people who fail to effectively use their automation have failed due to an inappropriate application of automation.	2			
Rotter (1966)	People who can't make effective use of automation don't understand how to use it effectively.	3			
Rotter (1966)	Believing in fate has never turned out as well as making decisions and applying a definitive action plan to my automation usage.	1	1		1
Rotter (1966)	Being successful with automation is a matter of your efforts as an operator, luck has little or nothing to do with it.	3			
Rotter (1966)	The average operator maintains influence over automated processes.	2	1		
Rotter (1966)	When I make plans, I am almost certain that I can find a way to have automation conform to those plans.	2	1		
Rotter (1966)	In my case, getting what I want out of an automated system has little or nothing to do with luck.		2	1	1
Rotter (1966)	Getting automation to do what you want depends on operator proficiency, luck has little or nothing to do with it.	2	1		

Rotter (1966)	By taking an active role in automation observation and control, people can better control their outcomes.	2	1			
Rotter (1966)	When it comes to automation, there really is no such thing as luck.		2	1		
Rotter (1966)	It is impossible for me to believe that the outcomes of my use of automation come down to chance or luck.		1	1		1
Rotter (1966)	Automation is ineffective because people do not effectively apply it to satisfy their needs.	2	1			
Rotter (1966)	The outcomes of my use of automation are the result of my own actions.	2	1			
Rotter (1966)	In the long run, operators are responsible for poor automation performance both individually and systematically.	3				
Rotter (1966)	Most misfortunes with automation are the result of a lack of ability, ignorance, or laziness on the part of the operator.	3				
Rotter (1966)	There is a direct connection between my understanding of automation, and my performance with the system.	2	1			
Levenson (1973)	My success with automation depends on my ability as an operator.	2	1			
Levenson (1973)	Whether or not I experience difficulty with automation is dependent on my proficiency with the system.	2	1			
Levenson (1973)	When I make plans, I manipulate the system to conform to my plans.	2	1			
Levenson (1973)	I determine the outcomes of my experiences with automation.	2	1			
Levenson (1973)	The outcomes of my use of automation are determined by my own actions.	2	1			
Levenson (1973)	Whether or not I experience a malfunction with automation is dependent on my proficiency with the system.	2	1			
Rotter (1966)	Automation will always be ineffective no matter how much operators try to understand it.			1		2
Rotter (1966)	No matter how hard you try, some automation will just be ineffective.			1		2

Rotter (1966)	Automation will function as it was designed, and there is little the average operator can do to influence it.		2	1		
Rotter (1966)	It is not always wise to plan too far ahead because automated processes are unpredictable anyhow.		1	1		1
Rotter (1966)	Many times, we might as well leave the automation to do what it will.			1		2
Rotter (1966)	As far as automation oversight is concerned, most operators are merely observers to processes that are beyond their control or understanding.		1	2		
Rotter (1966)	Most people don't realize the extent to which they are controlled by their automation.		2			1
Rotter (1966)	Many times, I feel that I have little influence over the outcomes of my experiences with automation.		1	2		
Rotter (1966)	Sometimes I feel like I don't have enough control over automated systems.	1	1	1		
Rotter (1966)	Most of the time, I can't understand why automation behaves the way that it does.					3
Levenson (1973)	The outcomes of my experiences with automation are chiefly controlled by the system.		2	1		
Levenson (1973)	Getting what I want out of automation requires attempts at conforming to the design of the system.		3			
Levenson (1973)	In order to have my plans work, I make sure that they fit in with the design of the system.	2	1			
Levenson (1973)	I feel like the outcomes of my use of automation are mostly determined by the design of the system.		2	1		
Rotter (1966)	The idea that system designers control operator experiences is nonsense.	1	2			
Rotter (1966)	Most people don't realize the extent to which system designers play a role in the outcomes of people's use of automated systems.		2	1		
Rotter (1966)	There's not much use in trying to control automation. System designers have already decided how it will operate in my use-case.		2			1
Rotter (1966)	Most people don't realize the extent to which they are controlled by automation system designers.		2			1
Levenson (1973)	I feel like the outcomes of my use of automation are mostly determined by system designers.		2	1		

Levenson (1973)	Although I may be a proficient user, the outcomes of my use of automation are determined by system designers.	2	1		
Levenson (1973)	Whether or not I experience a system malfunction depends mostly on the design of the system.	2	1		
Levenson (1973)	Although I may be a proficient user, the outcomes of my use of automation are determined by system implementers.	2	1		
Levenson (1973)	The outcomes of my use of automation are chiefly controlled by system designers.	2	1		
Levenson (1973)	The outcomes of my use of automation are chiefly controlled by system integrators.	2	1		
Rotter (1966)	Many of the difficulties operators face with automation are partly due to bad luck.		3		
Rotter (1966)	Without the right amount of luck, one cannot be an effective system operator.		3		
Rotter (1966)	I have often found that the outcomes of my use of automation are mostly down to chance.		3		
Rotter (1966)	Being successful with automation has a lot to do with being in the right place at the right time.		3		
Levenson (1973)	The outcomes of my use of automation are mostly controlled by accidental happenings.		3		
Levenson (1973)	Often, there is no chance of mitigating the influence of bad luck over the outcomes of my use of automation.		3		
Levenson (1973)	When I have a pleasant outcome with automation, it is usually because I get lucky.		2		1
Levenson (1973)	I have often found that what is going to happen will happen.		2		1
Levenson (1973)	Whether or not I experience an automation malfunction is mostly a matter of luck.		3		

Note: Factor group columns indicate the number of experts who assigned an item to a given factor group. a. NF = No Factor.

Table 3. Phase 2 Demographics and Automation Experience

Variable		Frequency	Percentage
Sex			
	Male	207	50.36%
	Female	204	49.64%
Race/Ethnicity			
	African American/Black	25	6.08%
	American Indian	2	0.49%
	Asian	23	5.60%
	Caucasian/White	331	80.54%
	Latin American/Hispanic	24	5.84%
	Mixed Race	5	1.22%
	Unspecified	1	0.24%
Highest Education Obtained			
	High School Diploma/GED	93	22.63%
	Associate's Degree	63	15.33%
	Bachelor's Degree	194	47.20%
	Master's Degree	49	11.92%
	Philosophical Doctorate	7	1.70%
	Doctor of Jurisprudence	1	0.24%
	Doctor of Medicine	1	0.24%
	Some College	1	0.24%
	Trade School	1	0.24%
	Vocational Training/Licensure	1	0.24%
Automation Experience			
	Kettle/Coffee-Maker with an Automatic Shut-Off	291	70.80%
	Motion-Sensor-Activated Lighting	301	73.24%
	Sunlight-Sensitive Lighting	188	45.74%
	Automatic Route Planning and Navigation (Google Maps/Apple Maps/Waze)	305	74.21%
	Autocorrect/Predictive Text	327	79.56%

Auto-curated Media (Spotify Suggested Music/YouTube Recommendations/Netflix Recommendations)	265	64.48%
Production Robotics	52	12.65%
Self-Driving Cars	30	7.30%

Table 4. Phase 2 Initial Item Reduction

Item	Corrected Item-Total Correlation
Being successful with automation is a matter of your efforts as an operator, luck has little or nothing to do with it.	.53
In my case, getting what I want out of an automated system has little or nothing to do with luck.	.54
Getting automation to do what you want depends on operator proficiency, luck has little or nothing to do with it.	.54
By taking an active role in automation observation and control, people can better control their outcomes.	.40
When it comes to automation, there really is no such thing as luck.	.47
The outcomes of my use of automation are the result of my own actions.	.44
There is a direct connection between my understanding of automation, and my performance with the system.	.42
My success with automation depends on my ability as an operator.	.43
Whether or not I experience difficulty with automation is dependent on my proficiency with the system.	.46
The outcomes of my use of automation are determined by my own actions.	.41
Automation will always be ineffective no matter how much operators try to understand it.	.58
No matter how hard you try, some automation will just be ineffective.	.55
It is not always wise to plan too far ahead because automated processes are unpredictable anyhow.	.62

As far as automation oversight is concerned, most operators are merely observers to processes that are beyond their control or understanding.	.50
Many times, I feel that I have little influence over the outcomes of my experiences with automation.	.61
Sometimes I feel like I don't have enough control over automated systems.	.57
The outcomes of my experiences with automation are chiefly controlled by the system.	.52
I feel like the outcomes of my use of automation are mostly determined by the design of the system.	.40
I feel like the outcomes of my use of automation are mostly determined by system designers.	.53
Although I may be a proficient user, the outcomes of my use of automation are determined by system designers.	.53
Although I may be a proficient user, the outcomes of my use of automation are determined by system implementers.	.49
The outcomes of my use of automation are chiefly controlled by system designers.	.53
The outcomes of my use of automation are chiefly controlled by system integrators.	.51
Many of the difficulties operators face with automation are partly due to bad luck.	.63
Without the right amount of luck, one cannot be an effective system operator.	.63
I have often found that the outcomes of my use of automation are mostly down to chance.	.68
Being successful with automation has a lot to do with being in the right place at the right time.	.67
The outcomes of my use of automation are mostly controlled by accidental happenings.	.65
Often, there is no chance of mitigating the influence of bad luck over the outcomes of my use of automation.	.68
When I have a pleasant outcome with automation, it is usually because I get lucky.	.67
Whether or not I experience an automation malfunction is mostly a matter of luck.	.64
Items Removed Based on Low Item-Total Correlations ($r < .4$)	
People's difficulties with automation result from deficiencies in their use of the system.	.29

One of the major reasons automation is ineffective is because operators do not take the time to understand it.	.40
Capable people who fail to effectively use their automation have failed due to an inappropriate application of automation.	.13
People who can't make effective use of automation don't understand how to use it effectively.	.32
Believing in fate has never turned out as well as making decisions and applying a definitive action plan to my automation usage.	.24
The average operator maintains influence over automated processes.	.39
When I make plans, I am almost certain that I can find a way to have automation conform to those plans.	.23
Automation is ineffective because people do not effectively apply it to satisfy their needs.	.23
In the long run, operators are responsible for poor automation performance both individually and systematically.	.25
Most misfortunes with automation are the result of a lack of ability, ignorance, or laziness on the part of the operator.	.35
When I make plans, I manipulate the system to conform to my plans.	.15
I determine the outcomes of my experiences with automation.	.34
Whether or not I experience a malfunction with automation is dependent on my proficiency with the system.	.22
Automation will function as it was designed, and there is little the average operator can do to influence it.	.37
Most people don't realize the extent to which they are controlled by their automation.	.38
Getting what I want out of automation requires attempts at conforming to the design of the system.	.20
In order to have my plans work, I make sure that they fit in with the design of the system.	-.03
The idea that system designers control operator experiences is nonsense.	.25
Most people don't realize the extent to which system designers play a role in the outcomes of people's use of automated systems.	.12
Whether or not I experience a system malfunction depends mostly on the design of the system.	.38
I have often found that what is going to happen will happen.	.38
Items Removed Based on Expert Suggestions and Low Item-Total Correlations (r < .5)	

It is impossible for me to believe that the outcomes of my use of automation come down to chance or luck.	.41
Many times, we might as well leave the automation to do what it will.	.41
Most of the time, I can't understand why automation behaves the way that it does.	.53
There's not much use in trying to control automation. System designers have already decided how it will operate in my use-case.	.50
Most people don't realize the extent to which they are controlled by automation system designers.	.46

Table 5. Phase 2 Initial Parallel Analysis Matrix Output

Factor	Eigenvalues		
	Observed Data	Generated Data	
		Means	95th Percentile
1	10.67	0.63	0.71
2	3.79	0.56	0.61
3	2.55	0.50	0.55
4	0.67	0.45	0.49

Note: Only retained factors are presented. Factors are retained if the raw data eigenvalue is greater than the 95th percentile eigenvalue for that factor.

Table 6. Phase 2 Initial Kaiser-Meyer-Olkin and Bartlett's Test Results

Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy		0.95
Bartlett's Test of Sphericity	Approx. Chi-Square	8230.4
	df	465
	Sig.	.000

Note: The KMO Measure of Sampling Adequacy and Bartlett's Test of Sphericity indicate the suitability of the sample data for factor extraction. KMO values above .6 and significant ($p < .05$) Bartlett's test results support the use of factor analysis for this dataset.

Table 7. Phase 2 Initial Principal Axis Eigenvalues and Explained Variance

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	11.05	36	36
2	4.25	14	49
3	2.94	9	59
4	1.13	4	62

Note: Only retained factors are presented. Factors are retained if the initial eigenvalue is greater than one. Support for the retention of the fourth factor is marginal.

Table 8. Phase 2 Initial Principal Axis Extraction Sums of Squared Loadings and Explained Variance

Factor	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total
1	10.67	34	34	9.38
2	3.79	12	47	1.23
3	2.55	8	55	6.63
4	0.67	2	57	5.91

Note: a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

Table 9. Phase 2 Initial Principal Axis Pattern Matrix

Item	Factor 1	Factor 2	Factor 3	Factor 4
Internal				
Being successful with automation is a matter of your efforts as an operator, luck has little or nothing to do with it.		-.25		-.64
In my case, getting what I want out of an automated system has little or nothing to do with luck.	.24	-.26		-.57
Getting automation to do what you want depends on operator proficiency, luck has little or nothing to do with it.		-.24		-.69
By taking an active role in automation observation and control, people can better control their outcomes.				-.56
When it comes to automation, there really is no such thing as luck.		-.29		-.51
The outcomes of my use of automation are the result of my own actions.				-.61
There is a direct connection between my understanding of automation, and my performance with the system.				-.61
My success with automation depends on my ability as an operator.				-.79
Whether or not I experience difficulty with automation is dependent on my proficiency with the system.				-.66
The outcomes of my use of automation are determined by my own actions.				-.75
Powerful Others				
Automation will always be ineffective no matter how much operators try to understand it.	.71			
No matter how hard you try, some automation will just be ineffective.		.20	0.31	-.28

It is not always wise to plan too far ahead because automated processes are unpredictable anyhow.	.58	.21	
As far as automation oversight is concerned, most operators are merely observers to processes that are beyond their control or understanding.	.39		.26
Many times, I feel that I have little influence over the outcomes of my experiences with automation.	.31	.37	.33
Sometimes I feel like I don't have enough control over automated systems.	.22	.33	.36
The outcomes of my experiences with automation are chiefly controlled by the system.			.54
I feel like the outcomes of my use of automation are mostly determined by the design of the system.			.63
I feel like the outcomes of my use of automation are mostly determined by system designers.			.81
Although I may be a proficient user, the outcomes of my use of automation are determined by system designers.			.85
Although I may be a proficient user, the outcomes of my use of automation are determined by system implementers.			.77
The outcomes of my use of automation are chiefly controlled by system designers.			.89
The outcomes of my use of automation are chiefly controlled by system integrators.			.79
Chance			
Many of the difficulties operators face with automation are partly due to bad luck.	.84		
Without the right amount of luck, one cannot be an effective system operator.	.89		
I have often found that the outcomes of my use of automation are mostly down to chance.	.85		

Being successful with automation has a lot to do with being in the right place at the right time.	.79
The outcomes of my use of automation are mostly controlled by accidental happenings.	.93
Often, there is no chance of mitigating the influence of bad luck over the outcomes of my use of automation.	.76
When I have a pleasant outcome with automation, it is usually because I get lucky.	.91
Whether or not I experience an automation malfunction is mostly a matter of luck.	.79

Table 10. Phase 2 Post-Reduction Principal Axis Factoring Pattern Matrix

Subscale	Item	Factor		
		1	2	3
Internal				
	Getting automation to do what you want depends on operator proficiency, luck has little or nothing to do with it.	.26	.57	
	There is a direct connection between my understanding of automation, and my performance with the system.		.60	
	My success with automation depends on my ability as an operator.		.83	
	Whether or not I experience difficulty with automation is dependent on my proficiency with the system.		.67	
	The outcomes of my use of automation are determined by my own actions.		.73	
Powerful Others				
	The outcomes of my experiences with automation are chiefly controlled by the system.			.62

	I feel like the outcomes of my use of automation are mostly determined by the design of the system.		.72
	Although I may be a proficient user, the outcomes of my use of automation are determined by system designers.		.77
	The outcomes of my use of automation are chiefly controlled by system designers.		.78
Chance			
	Many of the difficulties operators face with automation are partly due to bad luck.	.84	
	Without the right amount of luck, one cannot be an effective system operator.	.87	
	I have often found that the outcomes of my use of automation are mostly down to chance.	.86	
	Being successful with automation has a lot to do with being in the right place at the right time.	.81	
	The outcomes of my use of automation are mostly controlled by accidental happenings.	.92	
	Often, there is no chance of mitigating the influence of bad luck over the outcomes of my use of automation.	.75	
	When I have a pleasant outcome with automation, it is usually because I get lucky.	.91	
	Whether or not I experience an automation malfunction is mostly a matter of luck.	.79	

Table 11. Phase 2 Post-Reduction Principal Axis Initial Eigenvalues and Explained Variance

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	7.14	42	42
2	2.55	15	57
3	2.05	12	69

Note: Only retained factors are presented. Factors are retained if the initial eigenvalue is greater than one.

Table 12. Phase 2 Post-Reduction Principal Axis Extraction Sums of Squared Loadings and Explained Variance

Factor	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total
1	6.84	40	40	6.54
2	2.08	12	52	3.03
3	1.64	10	62	3.17

Note: a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

Table 13. Phase 2 Post-Reduction Parallel Analysis Matrix

Factor	Eigenvalues		
	Observed Data	Generated Data	
		Means	95^{th} Percentile
1	6.81	0. 41	0.49
2	2.00	0.34	0.40
3	1.56	0.28	0.33

Note: Only retained factors are presented. Factors are retained if the raw data eigenvalue is greater than the 95^{th} percentile eigenvalue for that factor.

Table 14. Phase 3 Demographics and Automation Experience

Variable		Frequency	Percentage
Sex			
	Male	102	41.46%
	Female	144	58.54%
Race/Ethnicity			
	African American/Black	17	6.91%
	Asian	12	4.88%
	Caucasian/White	196	79.67%
	Eurasian	1	0.41%
	Latin American/Hispanic	14	5.69%
	Middle Eastern	2	0.81%
	Mixed Race	4	1.63%
Highest Education Obtained			
	High School Diploma/GED	54	21.95%
	Associate's Degree	35	14.23%
	Bachelor's Degree	116	47.15%
	Master's Degree	32	13.01%
	Philosophical Doctorate	6	2.44%
	Doctor of Chiropractic	1	0.41%
	Doctor of Jurisprudence	1	0.41%
	Some College	1	0.41%
Automation Experience			
	Kettle/Coffee-Maker with an Automatic Shut-Off	191	77.64%
	Motion-Sensor-Activated Lighting	208	84.55%
	Sunlight-Sensitive Lighting	110	44.72%
	Automatic Route Planning and Navigation (Google Maps/Apple Maps/Waze)	214	86.99%
	Autocorrect/Predictive Text	220	89.43%
	Auto-curated Media (Spotify Suggested Music/YouTube Recommendations/Netflix Recommendations)	189	76.83%

Production Robotics	20	8.13%
Self-Driving Cars	16	6.50%

Table 15. Phase 3 Parallel Analysis Matrix

Factor	Eigenvalues		
	Observed Data	Generated Data	
		Means	95^{th} Percentile
1	5.23	0.56	0. 67
2	2.61	0.46	0.53
3	1.43	0.38	0.44

Note: Only retained factors are presented. Factors are retained if the raw data eigenvalue is higher than the 95^{th} percentile eigenvalue for that factor.

Table 16. Phase 3 Kaiser-Meyer-Olkin and Bartlett's Test Results

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.87
Bartlett's Test of Sphericity	Approx. Chi-Square	2127.3
	df	136
	Sig.	.000

Note: The KMO Measure of Sampling Adequacy and Bartlett's Test of Sphericity indicate the suitability of the sample data for factor extraction. KMO values above .6 and significant ($p < .05$) Bartlett's test results support the use of factor analysis for this dataset.

Table 17. Phase 3 Principal Axis Initial Eigenvalues and Explained Variance

Factor	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	5.62	33	33
2	3.10	18	51
3	1.97	12	63

Note: Only retained factors are presented. Factors are retained if the initial eigenvalue is greater than one.

Table 18. Phase 3 Principal Axis Extraction Sums of Squared Loadings and Explained Variance

Factor	Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings[a]
	Total	% of Variance	Cumulative %	Total
1	5.23	31	31	5.10
2	2.67	16	46	2.49
3	1.47	9	55	2.54

Note: a. When factors are correlated, sums of squared loadings cannot be added to obtain a total variance.

Table 19. Phase 3 Principal Axis Pattern Matrix

Source	Item	Factor		
		1	2	3
Internal				
Rotter (1966)	Getting automation to do what you want depends on operator proficiency, luck has little or nothing to do with it.	.25		.45
Rotter (1966)	There is a direct connection between my understanding of automation, and my performance with the system.			.56
Levenson (1973)	My success with automation depends on my ability as an operator.			.78

Levenson (1973)	Whether or not I experience difficulty with automation is dependent on my proficiency with the system.			.66
Levenson (1973)	The outcomes of my use of automation are determined by my own actions.			.63
Powerful Others				
Levenson (1973)	The outcomes of my experiences with automation are chiefly controlled by the system.		.62	
Levenson (1973)	I feel like the outcomes of my use of automation are mostly determined by the design of the system.		.78	
Levenson (1973)	Although I may be a proficient user, the outcomes of my use of automation are determined by system designers.		.81	
Levenson (1973)	The outcomes of my use of automation are chiefly controlled by system designers.		.79	
Chance				
Rotter (1966)	Many of the difficulties operators face with automation are partly due to bad luck.	.80		
Rotter (1966)	Without the right amount of luck, one cannot be an effective system operator.	.85		
Rotter (1966)	I have often found that the outcomes of my use of automation are mostly down to chance.	.84		
Rotter (1966)	Being successful with automation has a lot to do with being in the right place at the right time.	.79		
Levenson (1973)	The outcomes of my use of automation are mostly controlled by accidental happenings.	.78		
Levenson (1973)	Often, there is no chance of mitigating the influence of bad luck over the outcomes of my use of automation.	.63		
Levenson (1973)	When I have a pleasant outcome with automation, it is usually because I get lucky.	.83		

Levenson (1973)	Whether or not I experience an automation malfunction is mostly a matter of luck.	.70

Table 20. Phase 3 Confirmatory Factor Analysis Fit Statistics for the Three-Factor Model of the Automation Locus of Control Measure

Fit Indices		Result
Chi-Squared	χ^2	251.6
	df	116
	p	.000
Root Mean Square Error of Approximation		.07
Comparative Fit Index		.93
Root Mean Square Residual		.05

Table 21. Phase 3 Descriptive Statistics, Reliability Statistics, and Inter-Subscale and Subscale-Total Correlations

		Automation Locus of Control			
Subscale		Internal	Powerful Others	Chance	Total
Descriptive and Reliability Statistics					
	M	4.03	-1.69	7.57	9.92
	SD	2.94	3.05	5.52	7.92
	Cronbach's α	.76	.84	.92	.85
Automation Locus of Control Inter-Subscale Correlations					
Internal		1.000			
Powerful Others		.23**	1.000		
Chance		.24**	0.07	1.000	
Automation Locus of Control Subscale-Total Correlations					
Total		.63**	.52**	.81**	1.000

Note: **. Correlation is significant at the .01 level (2-tailed).

Table 22. Phase 3 Validation Correlations

	Automation Locus of Control			
	Internal	Powerful Others	Chance	Total
General Locus of Control	-.11	.07	-.24**	-.19**
Self-Esteem	.09	-.06	.26**	.19**
Self-Motivation	.16*	-.02	.33**	.28**
Decision-Style Inventory (Directive)	-.004	.003	-.14*	.10
Decision-Style Inventory (Analytical)	.07	-.019	.31**	.23**
Decision-Style Inventory (Conceptual)	-.02	-.09	.04	.01
Decision-Style Inventory (Behavioral)	-.05	.08	-.18**	-.11
Trust in Automation	.08	.07	.18**	.18**

Note: **. Correlation is significant at the .01 level (2-tailed). *. Correlation is significant at the .05 level (2-tailed).

Table 23. Phase 3 Tests for Normality

Measure	Shapiro-Wilk		
	Statistic	df	Sig.
Automation Locus of Control (Internal)	0.96	246	.000
Automation Locus of Control (Powerful Others)	0.96	246	.000
Automation Locus of Control (Chance)	0.93	246	.000
Automation Locus of Control (Total)	0.99	246	.129
General Locus of Control	0.99	246	.083
Self-Esteem	0.98	246	.001
Self-Motivation	0.99	246	.112
Decision Style Inventory (Directive)	0.99	246	.017
Decision Style Inventory (Analytical)	0.99	246	.195
Decision Style Inventory (Conceptual)	0.98	246	.003
Decision Style Inventory (Behavioral)	1.00	246	.560
Trust in Automation	0.98	246	.000

Figures

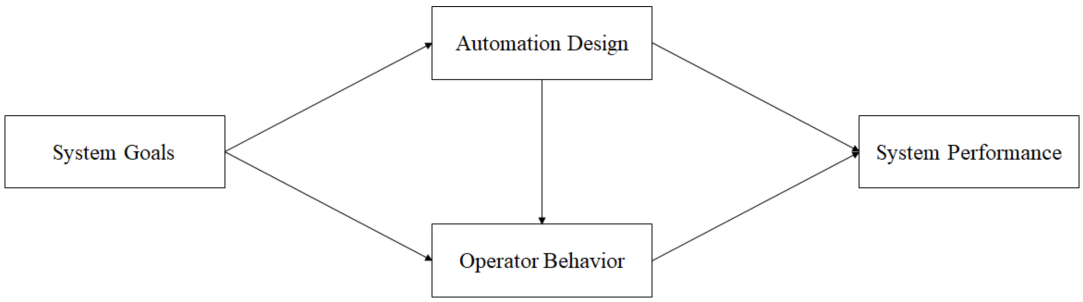

Figure 1. Conceptual Model for Human-Automation Interaction.
The model depicts system performance as the combined output of automation design and operator behavior based on prescribed system goals, and the influence of automation design on operator behavior.

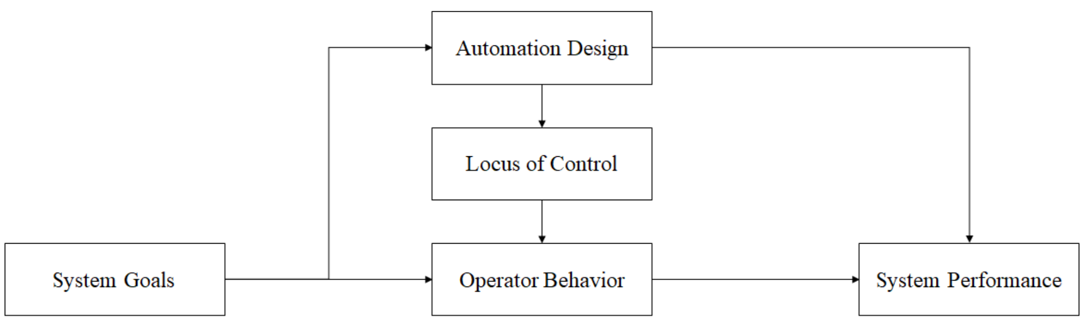

Figure 2. Conceptual Model for the Mediating Effect of Locus of Control on Human-Automation Interaction.
The model depicts system performance as the combined output of automation design and operator behavior based on prescribed system goals and illustrates locus of control as a mediator between automation design and operator behavior.

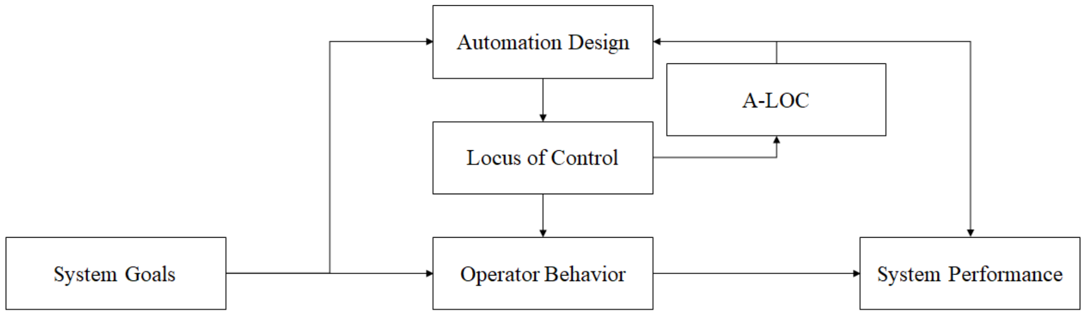

Figure 3. Conceptual Map for the Measurement of Locus of Control in the Context of Human-Automation Interaction.
The conceptual diagram illustrates the relationships among system goals, automation design, operator behavior, and system performance, and shows how an automation locus of control (A-LOC) measure could provide feedback for the adjustment of automation design.

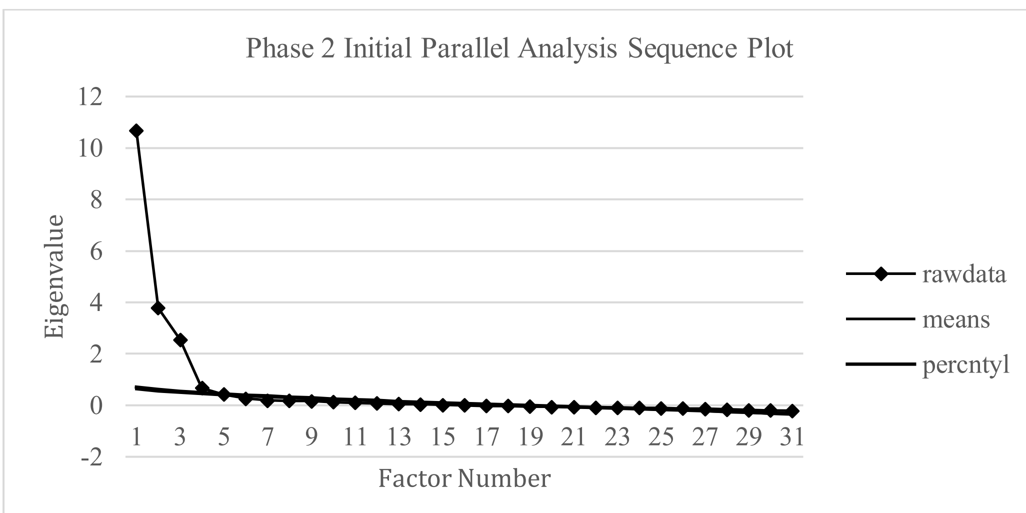

Figure 4. Phase 2 Initial Parallel Analysis Sequence Plot.
The sequence plot marginally supports the retention of four factors based on the number of points above the 95th percentile (percntyl) line.

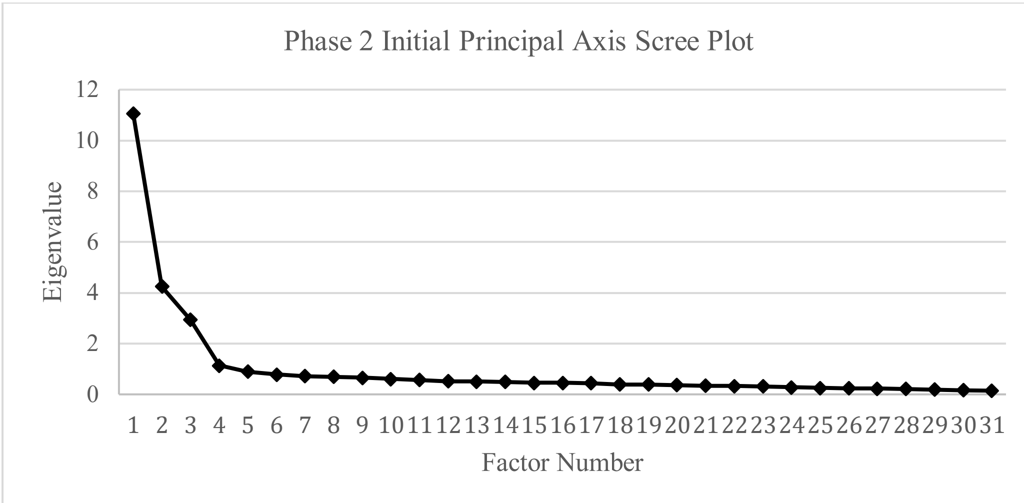

Figure 5. Phase 2 Initial Principal Axis Scree Plot.
This scree plot marginally supports the retention of four factors based on the position of the elbow of the plot.

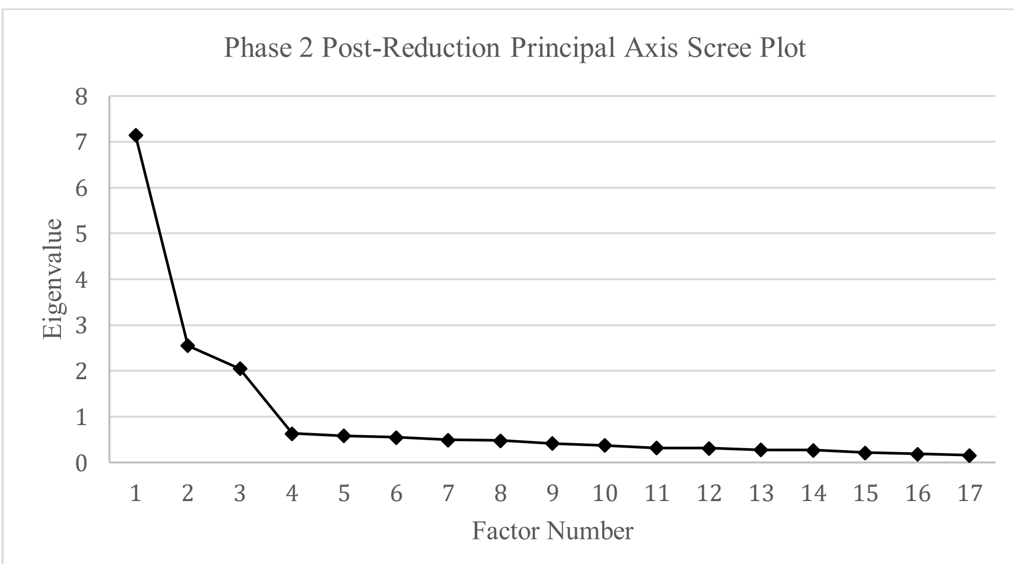

Figure 6. Phase 2 Post-Reduction Principal Axis Scree Plot.
This scree plot supports the retention of three factors based on the position of the elbow of the plot.

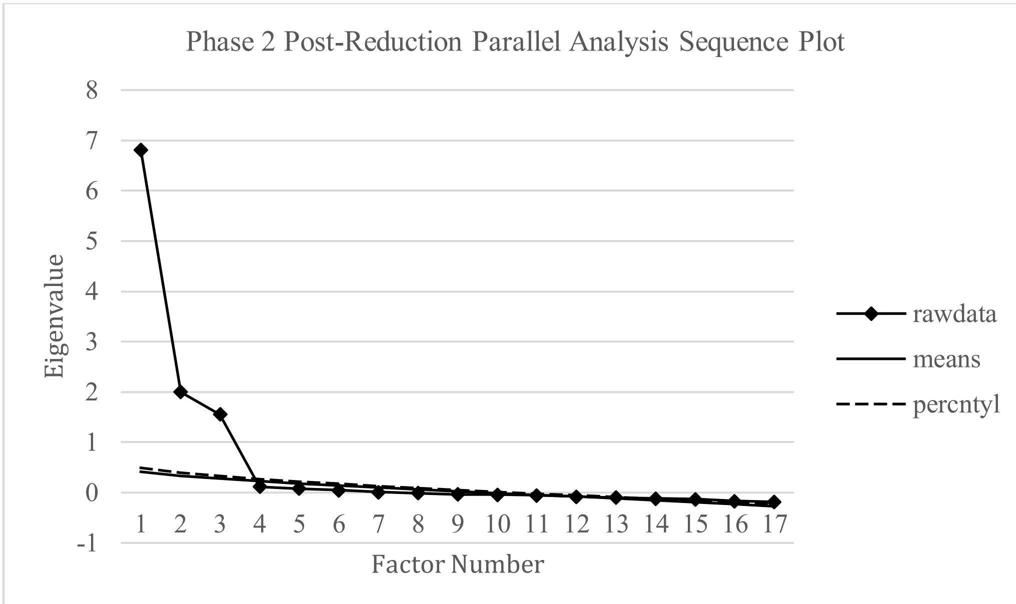

Figure 7. Phase 2 Post-Reduction Parallel Analysis Sequence Plot.
The sequence plot supports the retention of three factors based on the number of points above the 95th percentile (percntyl) line.

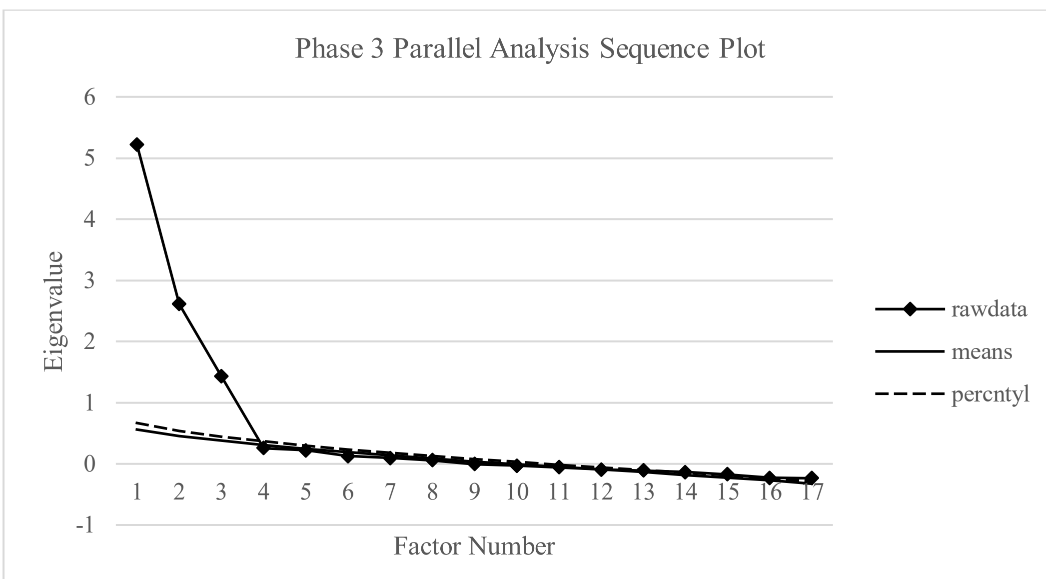

Figure 8. Phase 3 Parallel Analysis Sequence Diagram.

This sequence plot supports the retention of three factors based on the number of points above the 95th percentile (percntyl) line.

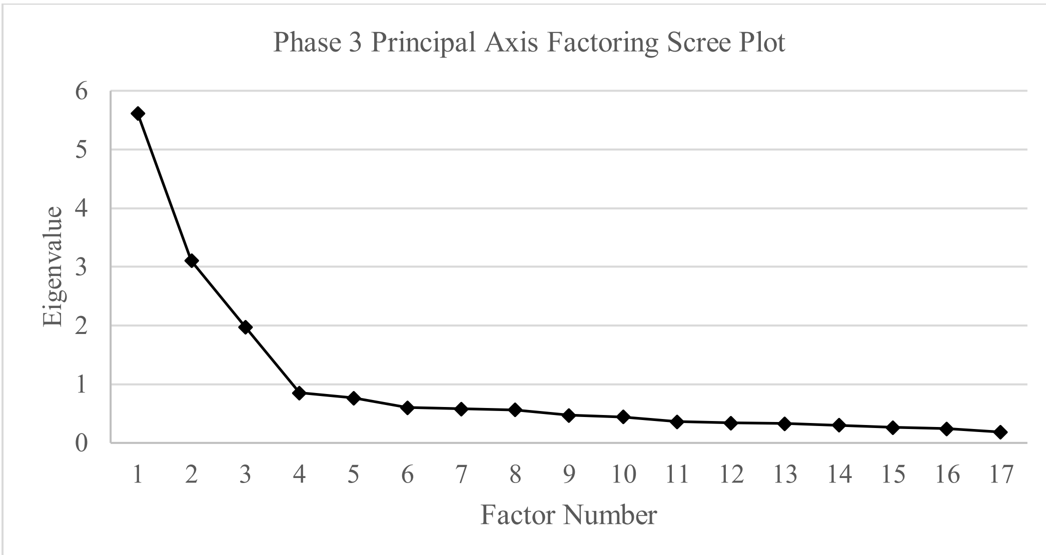

Figure 9. Phase 3 Principal Axis Scree Plot.
This scree plot supports the retention of three factors based on the position of the elbow of the plot.

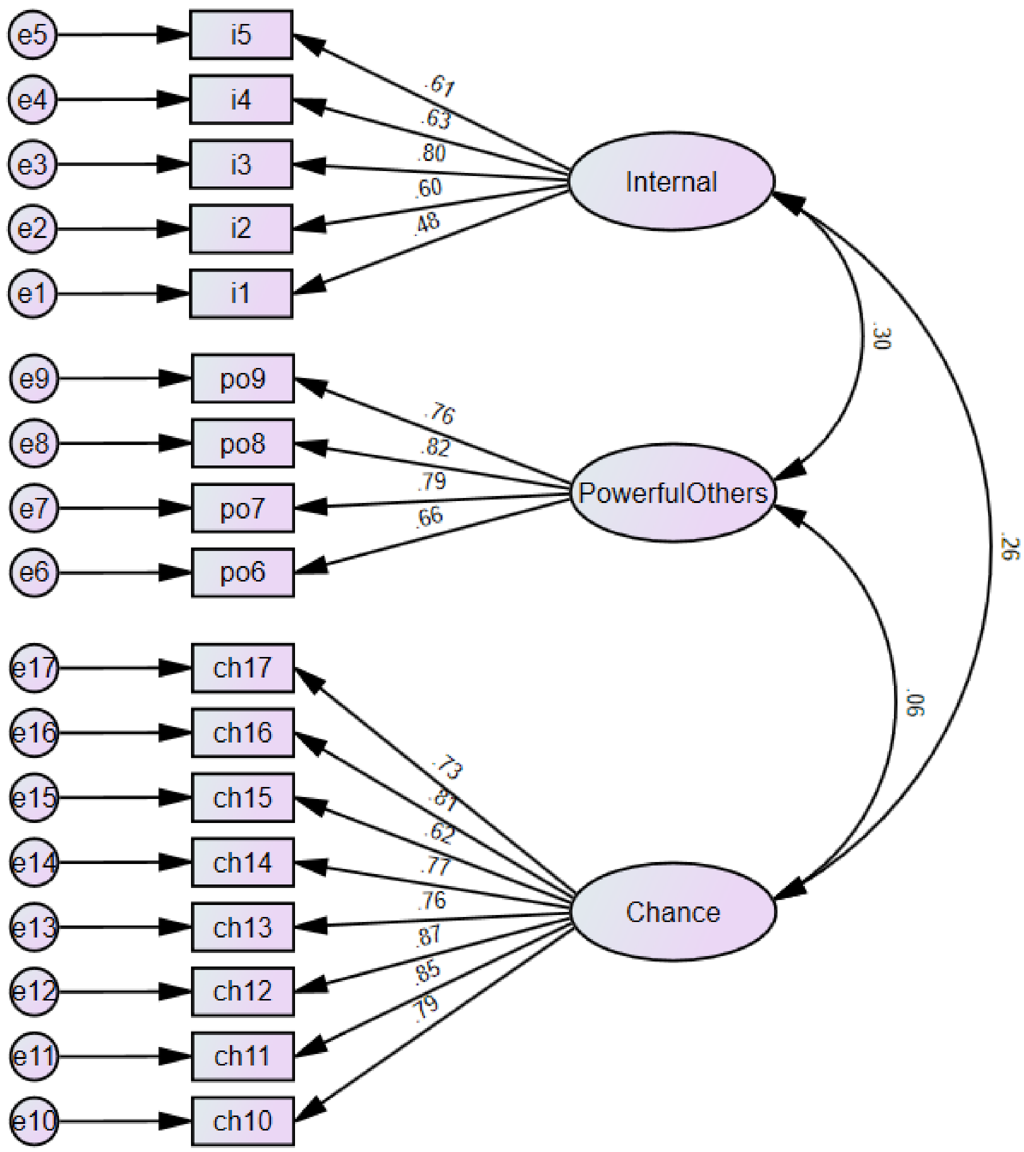

Figure 10. Phase 3 A-LOC Confirmatory Factor Analysis Results.
The diagram depicts the mapping of each item of the automation locus of control (A-LOC) scale, labeled i1 through ch17, to latent factors. The ovals to the right represent the three latent factors (Internal, Powerful Others, and Chance), and the circles on the left illustrates the unique variance of each item. Factor loadings are displayed on the arrows between the items and their latent factors, while the arrows between latent factors represent the covariances between latent factors.

Appendix

This questionnaire attempts to capture control perceptions in the context of human automation interaction. In the context of this questionnaire, automation is a general term that refers to the use of hardware or software to help or replace a human operator for the completion of a task.

An "operator" or "user" refers to an individual who interacts with the system in order to achieve a goal or complete a task. A "system designer" refers to any person involved in the creation and development of a system with respect to what it can do, and how users are meant to interact with it.

Low level examples of automation include the automated shut-off feature of your coffee-maker or kettle, when exterior lighting automatically turns on as the result of the sun setting, or doors that open automatically once a sensor detects movement. Mid-level examples of automation include the use of a GPS-enabled device for route planning and navigation (Google Maps, Waze, Apple Maps), and non-adaptive cruise control. High-level examples of automation include industrial assembly-line robotics, and self-driving cars.

When completing the questions below, try to consider automation in as general a sense as you can, and select the option that best matches the degree to which you agree or disagree with the statement. Your responses should reflect your beliefs on the outcomes of automation usage as it exists today.

There are no wrong answers.

1. Check each of the following automation examples that you have experience using.
   - o Kettle/Coffee-Maker with an Automatic Shut-Off
   - o Motion-Sensor-Activated Lighting
   - o Sunlight-Sensitive Lighting
   - o Automatic Route Planning and Navigation (Google Maps/Apple Maps/Waze)
   - o Autocorrect/Predictive Text Auto-curated Media (Spotify Suggested Music/YouTube Recommendations/Netflix Recommendations)

- Production Robotics
- Self-Driving Cars

2. Getting automation to do what you want depends on operator proficiency, luck has little or nothing to do with it.
   - Strongly Disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

3. There is a direct connection between my understanding of automation, and my performance with the system.
   - Strongly Disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

4. My success with automation depends on my ability as an operator.
   - Strongly Disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

5. Whether or not I experience difficulty with automation is dependent on my proficiency with the system.
   - Strongly Disagree
   - Disagree
   - Neutral
   - Agree
   - Strongly Agree

6. The outcomes of my use of automation are determined by my own actions.
   - Strongly Disagree
   - Disagree
   - Neutral

- Agree
- Strongly Agree

7. The outcomes of my experiences with automation are chiefly controlled by the system.
    - Strongly Disagree
    - Disagree
    - Neutral
    - Agree
    - Strongly Agree

8. I feel like the outcomes of my use of automation are mostly determined by the design of the system.
    - Strongly Disagree
    - Disagree
    - Neutral
    - Agree
    - Strongly Agree

9. Although I may be a proficient user, the outcomes of my use of automation are determined by system designers.
    - Strongly Disagree
    - Disagree
    - Neutral
    - Agree
    - Strongly Agree

10. The outcomes of my use of automation are chiefly controlled by system designers.
    - Strongly Disagree
    - Disagree
    - Neutral
    - Agree
    - Strongly Agree

11. Many of the difficulties operators face with automation are partly due to bad luck.
    - Strongly Disagree

- Disagree
- Neutral
- Agree
- Strongly Agree

12. Without the right amount of luck, one cannot be an effective system operator.
    - Strongly Disagree
    - Disagree
    - Neutral
    - Agree
    - Strongly Agree

13. I have often found that the outcomes of my use of automation are mostly down to chance.
    - Strongly Disagree
    - Disagree
    - Neutral
    - Agree
    - Strongly Agree

14. Being successful with automation has a lot to do with being in the right place at the right time.
    - Strongly Disagree
    - Disagree
    - Neutral
    - Agree
    - Strongly Agree

15. The outcomes of my use of automation are mostly controlled by accidental happenings.
    - Strongly Disagree
    - Disagree
    - Neutral
    - Agree
    - Strongly Agree

16. Often, there is no chance of mitigating the influence of bad luck over the outcomes of my use of automation.

- Strongly Disagree
- Disagree
- Neutral
- Agree
- Strongly Agree

17. When I have a pleasant outcome with automation, it is usually because I get lucky.
    - Strongly Disagree
    - Disagree
    - Neutral
    - Agree
    - Strongly Agree

18. Whether or not I experience an automation malfunction is mostly a matter of luck.
    - Strongly Disagree
    - Disagree
    - Neutral
    - Agree
    - Strongly Agree