2020

Creation of Two Valid Scales: Willingness to Fly in an Aircraft and Willingness to Pilot an Aircraft

Stephen Rice
ERAU, rices15@erau.edu
Scott R. Winter
Embry-Riddle Aeronautical University, scott.winter@mac.com
John Capps
Embry-Riddle Aeronautical University, johncappsdesign@gmail.com
Justin Trombley
Embry-Riddle Aeronautical University, tromblj4@my.erau.edu
John Robbins
Embry-Riddle Aeronautical University, robbinsj@erau.edu
Mattie Milner
Embry-Riddle Aeronautical University, milnerm1@my.erau.edu
Tracy L. Lamb
Embry-Riddle Aeronautical University, LAMBT4@my.erau.edu

Follow this and additional works at: https://commons.erau.edu/ijaaa

Part of the Other Psychology Commons

The massive scope of aviation research covers a plethora of topics, such as accident analysis, controls and displays, crew resource management, etc. (Meister & Gawron, 2010). This has facilitated the creation of several validated metrics within aviation and human performance research. Contingent upon the research design, and available resources, the particular research instruments vary, allowing the researchers to focus on specific data collection methods and/or desired deliverables.

Accurately measuring flight performance is a critical component of aviation research and researchers often utilize mathematical outputs such as root mean square error (RMSE), the mean of error scores (ME), and standard deviation (SD) (Hubbard, Rockway, & Waag, 1989; Meister & Gawron, 2010). In addition, observing participants' response time to certain stimuli provides useful information; however, the reliability of experimenters' observations may be questioned. Regarding nonflight performance metrics, researchers discuss the use of Task Load Index (TLX), a commonly used metric to solicit information on participant workload (Vidulich & Tsang, 1985). Furthermore, researchers may collect physiological data, such as heart-rate, respiration, eye tracking, blink rate, and oxygen consumption.

Research into the challenges associated with willingness in aviation by both passengers and pilots is essential to find solutions to many challenges. A scale to measure willingness may support solving issues that are central to safety, product, and service development. In 2018, an International Air Transport Association report found that the largest number of accidents occurred in 'Generation 3 and 4' jet aircraft (IATA, 2018). This trend is set to increase in 2019, as a result of two Boeing 737-Max accidents (National Transport Safety Board, 2019). Passengers' and pilots' willingness to fly can be affected by their perceptions of safety, services, and products, which in turn effects their emotions and decisions.

For passengers, newer generation aircraft are quieter, faster, operate at lower cabin pressures, and often have attractive interior designs to make the passenger experience more pleasant. For pilots the newer aircraft feature higher levels of advanced technology and automation compared to earlier and simpler aircraft (ICAO, 2013). Pilots' willingness to pilot also involves their willingness to embrace challenges of learning new operating philosophies (e.g., flying a Boeing compared to an Airbus) (EASA, 2013). Both the passengers' and the pilots' willingness to fly/pilot, has foundational parameters that are important for a safe, sustainable and successful future airline industry.

Participants may also provide personal responses through the use of subjective scales, which allow participants to report their own thoughts, feelings and perceptions. Within the aviation and human performance research realms, these measures provide valuable information through the use of interviews or surveys (Meister & Gawron, 2010). The current study focuses on the development of two

concurrent, yet independent valid subjective measures: a willingness to fly scale for use with passengers, and a willingness to pilot scale for use with aircraft pilots.

Justification for Willingness Scales in Aviation

Research has indicated that perceptions, decisions and willingness to use new and emergent technologies such as drones, driverless cars, autonomous aircraft, and even trusting new biofuel (Rains et al., 2017), has a high dependence on emotions (Anania et al., 2018; Cook, Noyes, & Masakowski 2017; Lerner, Lerner, & Tiedens 2001; Lerner, Valdesolo, & Kassam, 2015; Mehta, Rice, Winter, & Oyman, 2014; Rice & Winter, 2019). Embracing these new emergent technologies requires acceptance, trust and a submission or commitment to a change of behavior. Previous studies have shown that emotional responses can include being fearful, angry, happy, sad, disgusted or surprised (Ekman & Friesen (1971). Without valid scales to identify and validate the needs and probable responses of the user groups (passengers and pilots), safety messaging, training programs, system design, outreach and education campaigns will not have a grounding premise for development, which is of paramount importance in a highly complex, dynamic and increasingly technology centric aviation industry (Billings, 1996; EASA, 2013; Wickens, Mavor, & MacGee, 1997).

As it relates to subjective measures, very few valid measures have been developed for use with passengers. A few previously developed scales focused on passengers' trust of commercial airline pilots (Rice, Mehta, Steelman, & Winter, 2014; Rice, Metha, Winter, & Oyman, 2015), affect (Rice & Winter, 2015), and privacy (Mehta, Rice, Winter, Moore, & Oyman, 2015). However, aviation and human performance research has neglected to address the need for a valid scale measuring the construct of the passengers' willingness to fly. Ultimately, commercial aviation is dependent upon providing customers with a service; therefore, it is crucial to understand if passengers are unwilling to fly in certain situations (Ragbir, Baugh, Rice, & Winter, 2018; Rice, Winter, Mehta, & Ragbir, 2019).

Previous subjective measures have focused on pilot-related constructs, such as workload (Gawron, 2008), situational awareness (Endsley, 1988), hazardous attitudes (Hunter, 2005), risk perception (Hunter, 2006), situational judgment (Hunter, 2003), and locus of control (Hunter, 2002). A valid measure of willingness to pilot (WTP) can fill an existing gap in the body of literature as the pilot's willingness to pilot score may provide an indicator as to other potentially dangerous human factors issues such as anxiety, mistrust or misplaced trust in the system (Endsley, 2017), which could lead to unsafe behaviors, reduced situational awareness, and impaired decision making (Flin, O'Connor, & Crichton, 2008).

The Value of a Validated Instrument

Instrument validity is a critical component for conducting appropriate and meaningful research, particularly when humans are involved (Edmonds & Kennedy, 2017; Gawron, 2008). The concept of validity relates to whether an instrument measures what it intends to measure. For example, if a teacher used an instrument intended to measure students' math skills, but the instrument actually measured students' reading comprehension, it would be an invalid instrument. Invalid instruments potentially introduce fatal flaws into the research design causing corrupt data, failed studies, and/or increased exposure to a Type I error. However, validity alone is not sufficient, as an instrument must also demonstrate reliability. Reliable instruments continuously produce the same results, regardless of when, or how many times, they are tested, assuming new information has not been collected. Therefore, useful and meaningful data collection instruments must be both valid and reliable.

Current Study

The purpose of the current study was to produce two valid instruments which could be used by researchers in the aviation field to collect data regarding a passenger's willingness to fly (WTF) in various scenarios, and data regarding a pilot's willingness to pilot (WTP). Therefore, two separate scale building and psychometric assessments were conducted. First, researchers constructed and validated a willingness to fly (WTF) in an aircraft scale. Second, researchers created the willingness to pilot (WTP) in an aircraft scale through replication of the WTF scale constructs, but replaced "flying" with "piloting." The findings demonstrate the steps taken by the researchers to demonstrate the validity of the measures along with their reliability.

Methodology – Willingness to Fly Scale (WTF) For Use with Passengers

Scale Building

The process used to develop the Willingness to Fly Scale involved five stages: 1) item generation, 2) nominal paring of the items, 3) Likert-scale paring, 4) factor analysis and reliability testing, and 5) sensitivity test. Similar methods have been used in the past (Rice, Mehta, Steelman, & Winter, 2014; Rice, Mehta, Winter & Oyman, 2015), and these methods follow Hinkin's (1998) framework for scale development.

Stage 1: Item Generation

In this stage, the goal was to generate items that may eventually comprise the final scale and this was accomplished in three ways. First, we searched the literature for terms that appeared relevant to the concept of 'willingness to fly.'

Second, we solicited terms from aviation experts, and third, we solicited terms from potential passengers. Potential passengers are the end-users of this scale; therefore, to ensure construct validity, we wanted them to provide terms representing willingness to fly in an aircraft.

Participants. Two hundred and forty-two (111 females) participants were recruited via a convenience sample using Amazon's ® Mechanical Turk ® (MTurk). MTurk provides a pool of participants who participate in studies for compensation. Previous research has revealed that MTurk data is reliable and robust (Buhrmester, Kwang, & Gosling, 2011; Germine et al., 2012; Rice, Winter, Doherty, & Milner, 2017). All participants identified as airline consumers, native English speakers, and were located in the United States, for the current and subsequent studies. The mean age was 29.64 ($SD = 8.22$) years. Additional aviation experts were recruited from the community.

Procedure, Materials and Stimuli. Participants first gave electronic consent via SurveyMonkey ® and were presented with the following instructions: "*In the context of being willing to fly in an airplane, please enter 5 words or phrases that you feel are strongly relevant to the concept of being willing to fly. Each answer should include only a one word or one sentence phrase.*" Lastly, participants provided demographics information, were debriefed, paid and dismissed.

Results. Upon completion, the researchers were able to gather 178 unique terms from this exercise (e.g., comfortable, confident, eager, etc.). Another 39 items were generated from the literature and aviation experts. Not surprisingly, the most common word that appeared was "willing." All the terms were reviewed for correct spelling and lowercase font, thus allowing equal saliency in the following stages.

Stage 2: Nominal Paring

The goal of stage 2 was to begin the process of paring down the initial list of items to focus on more relevant terms with a stronger connection to "willingness to fly." During this stage, participants read each item and determined the item's relevance to the construct of "willingness to fly."

Participants. Two hundred and sixty-nine (123 females) participants were recruited from the MTurk community. The mean age was 34.30 ($SD = 11.59$) years.

Procedure, Materials, and Stimuli. Participants first gave electronic consent. All of the items from the first stage were presented to participants. After reading each item, participants were asked to rate if the item was related to 'willingness to fly' or not related to 'willingness to fly;' and they could also choose "I don't know." Lastly, participants provided demographics information, debriefed, paid, and dismissed.

Results. An a priori determination was made that at least 80% of the participants had to agree that a particular item was related to willingness to fly

before it could be included in the next stage. There were 64 items that met these criteria.

Stage 3: Likert-scale Paring

The purpose of this stage was to provide a more sensitive measure of an item's relevancy to the concept of "willingness to fly". Participants read through the 64 items from Stage 2 and rated them on a scale from 0 (Not at all related to willingness to fly) to +3 (Extremely related to willingness to fly). A prior determination was made that a score of at least 2.5 was required for inclusion in the final scale.

Participants. Two hundred and seventy-three (121 females) participants were recruited via a convenience sample using MTurk. The mean age was 31.75 ($SD = 10.73$) years.

Procedure, Materials and Stimuli. Participants first gave electronic consent. Each item generated from Stage 2 was presented to the participants through the following statement request, "*In the context of being willing to fly in an airplane, please rate how strongly each word below is related to willingness to fly.*" Participants provided their responses on a scale from 0 (Not at all related to willingness to fly) to +3 (Extremely related to willingness to fly). Lastly, participants provided demographics information, were debriefed, paid and dismissed.

Results. Seven items met the inclusion criteria for the next stage and these items are present in the final scale, as well (see Appendix A).

Stage 4: Factor Analysis

In this stage, the final scale of seven items was tested for reliability and validity. Participants were presented with two different hypothetical scenarios and used the scale to provide their responses to the scenarios.

Participants. Five hundred and one (234 females) participants were recruited from MTurk for this stage. The mean age was 36.89 ($SD = 11.99$) years.

Procedure, Materials, and Stimuli. Participants first gave electronic consent and were then presented with two hypothetical scenarios. After each scenario, they responded to the Willingness to Fly scale (see Appendix A). The two scenarios were designed to ensure variance in the data, while avoiding skewed or bipolar distributions. The scenarios read as follows:

Scenario One

Imagine that you're a passenger on a 14-hour flight from one country to another. There are 300 passengers on board. You are told that the pilots on board will be allowed to use CRIP during this flight. Controlled Rest in Position (CRIP) allows one pilot to nap while remaining in the cockpit seat during the non-critical stages of flight. In some countries where CRIP is

used, they have a strict policy on how it is implemented. For example, when the pilot wants to take a nap, s/he has to inform the co-pilot and flight attendants. When the pilot wakes, s/he cannot perform any essential tasks until s/he is fully awake (e.g. 20 minutes). However, this does not apply to emergency situations, where the pilot might be expected to operate the aircraft immediately.

Scenario Two

Imagine a situation where you are flying on a commercial airplane from one major city to another. The autopilot is engaged, and it will control the aircraft from the moment it takes off until the moment it lands. There is one human pilot in the cockpit to monitor the autopilot in case of emergency.

Lastly, participants provided demographics information, debriefed, paid, and dismissed.

Results. For the first scenario, a factor analysis using the principal components and varimax rotation was conducted on the data. This analysis revealed that all items loaded strongly on a single factor, with 87.22% of the variance being explained by this model. Figure 1 shows the Screen Plot from the analysis. In order to test internal consistency of the scale, a Cronbach's Alpha test was conducted, which resulted in a value of .975, which indicates extremely high consistency between items. In order to rest reliability, a Guttmann's Split Half test was conducted, which resulted in a value of .949, indicating extremely high reliability.

For the second scenario, a factor analysis using the principal components and varimax rotation was conducted on the data. This analysis revealed that all items loaded strongly on a single factor, with 89.35% of the variance being explained by this model. Figure 2 shows the Screen Plot from the analysis. In order to test internal consistency of the scale, a Cronbach's Alpha test was conducted, which resulted in a value of .980, which indicates extremely high consistency between items. In order to rest reliability, a Guttmann's Split Half test was conducted, which resulted in a value of .963, indicating extremely high reliability.

Stage 5: Sensitivity Test

Data from the previous stage supports evidence for scale reliability and validity; however, the data does not allow researchers to measure the scale's sensitivity. If the scale is unable to distinguish between different conditions, then it does not provide much utility to the researcher. To test sensitivity, we presented participants with two different hypothetical scenarios and used the scale to measure differences in willingness to fly.

Participants. Two hundred and thirty-six (121 females) participants were recruited from MTurk for this stage. The mean age was 36.66 ($SD = 10.61$) years.

Procedure, Materials, and Stimuli. Participants first gave electronic consent. Participants were then presented with one of the two following hypothetical scenarios:

Scenario One

> *Imagine a situation where you need to fly on a commercial airplane flight from one major city to another. The autopilot is engaged, and it will control the aircraft from the moment it takes off until the moment it lands. There is a human pilot in the cockpit to monitor the autopilot in case of emergency.*

Scenario Two

> *Imagine a situation where you need to fly on a commercial airplane flight from one major city to another. The autopilot is engaged, and it will control the aircraft from the moment it takes off until the moment it lands. There is no human pilot in the cockpit to monitor the autopilot in case of emergency.*

Following this, they were asked to rate their willingness to fly using the newly-constructed willingness to fly scale. Lastly, they provided demographics information, debriefed, paid and dismissed.

Results. Prior to analysis, the scores from each scale were averaged for a single willingness to fly score. The scores for the two conditions were compared using a between-participants t-test, $t(235) = 12.06$, $p < .001$, $d = 1.57$. These results indicate that participants were less willing to fly when there was no human in the cockpit to monitor the autopilot. Importantly, this provides strong evidence of the willingness to fly scale's sensitivity to differences in the data.

Methodology – Willingness to Pilot Scale (WFP) For Use with Pilots

Participants. 482 (63 females) pilots were recruited from a large southeastern flight school. The mean age was 21.54 ($SD = 4.17$) years.

Procedure, Materials, and Stimuli. Participants first gave written consent and then were presented with the two following hypothetical scenarios:

Scenario One

> *Imagine that you are supposed to pilot a Cessna 172 on a 2-hour flight from Daytona Beach, FL to Naples, FL, crossing over the central part of the state. There are 2 passengers on board. You are told that weather for the route of this flight is marginal VFR with 3 miles visibility and 2,000 foot cloud ceilings. The turn coordinator is placarded inoperative so conducting the flight under instrument flight rules will not be possible.*

Scenario Two

> *Imagine a scenario where you have to conduct a ferry flight for a new aircraft. While you have experience in the type of aircraft, you have never flown this particular version of a "glass cockpit" or avionics.*

Following this, they were presented with the Willingness to Pilot (WTP) scale (see Appendix B). The only difference between the WTP scale and the previously validated WTF scale is the use of "piloting" instead of "flying." Lastly, participants were asked for provide basic demographic information, along with their total flight hours, certificates and ratings. They were then debriefed and dismissed.

Results – WTP Scale

For the first scenario, a factor analysis using the principal components and varimax rotation was conducted on the data. This analysis revealed that all items loaded strongly on a single factor, with 70.02% of the variance being explained by this model (see Figure 3 for the Scree Plot analysis). To test internal consistency of the scale, a Cronbach's Alpha test was conducted, which resulted in a value of .927, indicating extremely high consistency between items. To rest reliability, a Guttmann's Split Half test was conducted, resulting in a value of .886, indicating extremely high reliability.

For the second scenario, a factor analysis using the principal components and varimax rotation was conducted on the data. This analysis revealed that all items loaded strongly on a single factor, with 74.42% of the variance being explained by this model (see Figure 4 for the Scree Plot analysis). To test internal consistency of the scale, a Cronbach's Alpha test was conducted, which resulted in a value of .942, indicating extremely high consistency between items. To test reliability, a Guttmann's Split Half test was conducted, which resulted in a value of .918, indicating extremely high reliability.

Discussion

The current study sought to develop and validate two unique instruments for accurately measuring passengers' and pilots' willingness, respectively, within the aviation field. The first instrument captures passengers' willingness to a fly in different scenarios. This scale allows researchers to better understand the factors influencing passengers' decision-making process, which could have a tremendous impact on the commercial aviation sector. Organizations such as Uber, Bell, Boeing, and Airbus are heavily investing in new aircraft and technology prototypes for the future of urban air mobility (Uber, 2016). Unless they understand the passenger perceptions and concern, successful implementation and adoption of this technology will be difficult.

Additionally, the second scale gauges pilots' willingness to pilot an aircraft, which provides a marked departure from previous scales focused on pilot perceptions and resultant behaviors. Pilot training focuses on skills, knowledge, and attributes, which includes competency based and evidence-based training techniques and models. Non-technical skills training focus on attributes and includes behavior management indicators, which includes attitudes and emotional responses to tasks and challenges. While there are abundant tools and frameworks on these factors, for example; Human Factors Accident Classification System (HFACs) (Wiegmann & Shappell 2003), no such tool or scale exists for pilot perceptions towards willingness.

The willingness to fly scale provides the aviation research community with a valid instrument, which can be used by researchers to help better understand human performance and decision-making processes. When conducting research, the use of an existing valid scale provides an advantage over creating a new instrument. The development and validation of new instruments require large sample sizes to conduct psychometrics, and without ensuring the validity of the instrument beforehand, it is possible that the findings could be flawed or corrupt (Edmonds & Kennedy, 2017).

Through the creation of the willingness to fly scale, a five-stage validation process resulted in a seven-item, valid measure of this construct (Hinkin, 1998; Rice, Mehta, Steelman, & Winter, 2014; Rice, Mehta, Winter & Oyman, 2015). This scale can help researchers examine consumer perceptions of potentially new situations, such as passenger willingness to fly in autonomous aircraft (Ragbir, Baugh, Rice, & Winter, 2018; Rice, Winter, Metha, & Ragbir, 2019), new pilot policies, such as controlled rest in position (Rice, Winter, Tamilselvan, & Milner 2017), use of pilot medications (Winter, Rice, Rains, Milner, & Mehta, 2017), or programs like the Federal Flight Deck Officer program (Winter, Rice, Friedenreich, Metha, & Kaiser, 2017). Additionally, since the scale consists of seven-items rated on a strongly disagree to strongly agree bipolar rating system, its brevity lends itself to exploring mediating, moderating or dependent variables.

Research considering pilots' thoughts, behaviors, and perceptions involves a wide range of domains, and the willingness to pilot scale provides opportunities for further exploration of pilots' decision-making process. Studying pilot's flight performance often utilizes root mean square error (RMSE), mean error of scores (ME), and standard deviation (Hubbard et al., 1989; Meister & Gawron, 2010). When collecting non-flight performance measures, there are several different factors to consider, such as the Task Load Index (TLI) (Vidulich & Tsang, 1985), situational awareness (Endsley, 1988), hazardous attitudes (Hunter, 2005), risk perception (Hunter, 2006), and locus of control (Hunter, 2002).

However, no previous measures provided a valid scale of willingness to pilot, which can help researchers investigate pilots' willingness to pilot into

deteriorating weather conditions, under various levels of stress/fatigue, or when dealing with new levels of technology in the aircraft. Similar to the willingness to fly scale, the brevity of the seven-item willingness to pilot scale makes it useful to administer to pilots as a mediating, moderating or dependent variable, especially in longitudinal studies where a researcher wishes to examine the measure over a period of time. Therefore, this willingness to pilot scale will provide researchers with a valid measure to collect data evaluating the construct of willingness in a number of experimental and non-experimental studies.

Practical Applications

The importance of developing valid scales relevant to passenger and pilot willingness to fly/pilot can be attributed to innovation, modernization, and expected growth within the aviation industry. Aircraft have become operationally more autonomous, while passenger expectations for safe and efficient carriage have become necessary elements of modern travel. Studying these variables is a necessary step in developing insights and perceptions related to issues, such as training, reliability, and safety. The use of data derived by the willingness to fly and willingness to pilot scales provide researchers with tools that can be tailored and applied to a myriad of aviation related issues and scenarios, while providing organizations with applied research in the areas of human factors, airline operations, business management, and marketing.

This study offers unique insight of both passenger and pilot willingness to fly/pilot under certain circumstances, along with valid instruments which can be used by the aviation research community. Continuation of this research is critical to understanding both passenger and pilot thresholds used in the determination of the future needs and trends within the aviation industry. An understanding of related variables is necessary to define the boundaries, decision making process, and emotional factors involved in the current and future states of air travel.

Limitations

This current study is subject to the following discussed limitations. First, self-reported credentials and experience levels of participants could not be validated by the research team. While the study provides insight into passenger and pilot motivations for willingness to fly/pilot, the conclusions are limited by convenience sampling techniques, which may not be representative of the overall population. Future research studies can enhance the generalizability of the scales to wider audiences and demonstrate concurrent validity, resulting in a more robust series of instruments. Additionally, participants reported being located in the United States, which may limit international perspective. Lastly, the mean age for consumers was around the mid-30's for passengers and early 20's for pilots, which

suggests that the sampling may have limited to the younger segments of the flying population.

Conclusions

The purpose of the current study was to create and validate two scales for use within aviation-related research. The first scale was a valid measure of a passenger's willingness to fly. Through the use of a five-stage process, a seven-item scale was produced and shown to measure willingness to fly. This scale can be used by researchers in a number of inquiries to measure the willingness to fly of potential passengers. The second instrument produced was a valid measure of willingness to pilot. Prior scales have been developed to measure a number of constructs in pilots, but no prior scale has offered a valid measure of willingness to pilot. The willingness to pilot scale is a seven-item scale which can be used to provide a valid measure in a number of various scenarios for experimental and non-experimental studies. The brevity of both scales makes them ideal for use as mediating, moderating or dependent variables, especially in longitudinal studies where a researcher desires to evaluate these values over a period of time.

References

Anania, E. C., Rice, S., Winter, S. R., Milner, M. N., Walters, N. W., & Pierce, M. (2018). Why people are not willing to let their children ride on driverless school buses: A gender and nationality comparison. *Social Sciences, 7*(3), 1-17. doi: 10.3390/socsci7030034

Billings, C. E. (1996). *Aviation automation: The search for a human-centered approach* (1st ed.). Mahwah, N.J: Lawrence Erlbaum Associates Publishers. doi:10.1201/9781315137995

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.

Cook, M., Noyes, J. M., & Masakowski, Y. (2017). *Decision making in complex environments* (1st ed.). Aldershot: Ashgate. doi:10.1201/9781315576138

Edmonds, W. A., & Kennedy, T. D. (2017). *An applied guide to research designs: Quantitative, qualitative, and mixed methods.* Thousand Oaks, CA: Sage.

Ekman, P. & Friesen, W.V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–29. doi:10.1037/h0030377. PMID 5542557.

Endsley, M. R. (2017). From here to autonomy: Lessons learned from human–automation research. *Human Factors: The Journal of Human Factors and Ergonomics Society, 59*(1), 5-27. doi:10.1177/0018720816681350

European Aviation Safety Agency (Ed.). (2013). *EASA: Automation policy bridging design and training principles*. Brussels: European Aviation Safety Agency. Retrieved from https://www.easa.europa.eu/sites/default/files/dfu/sms-docs-EASp-SYS5.6---Automation-Policy---28-May-2013.pdf

Flin, R. H., O'Connor, P., & Crichton, M. (2008). *Safety at the sharp end*. Surrey UK: Ashgate. Retrieved from http://digitool.hbz-nrw.de:1801/webclient/DeliveryManager?pid=2336761&custom_att_2=simple_viewer

Gawron, V. J. (2008). *Human performance, workload, and situational awareness measures handbook.* Boca Raton, FL: CRC Press.

Germine, L., Nakayama, K., Duchaine, B. C., Chabris, C. F., Chatterjee, G., & Wilmer, J. B. (2012). Is the Web as good as the lab? Comparable performance from Web and lab in cognitive/perceptual experiments. *Psychonomic Bulletin & Review*, *19*(5), 847-857.

Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods, 1*(1), 104–121. http://dx.doi.org/10.1177/109442819800100106

Hubbard, D. C., Rockway, M. R., & Waag, W. L. (1989). Aircrew performance assessment. In R. S. Jensen (Ed.), *Aviation psychology* (pp. 342-377). Brookfield, IL: Gower Technical.

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

Hunter, D. R. (2003). Measuring general aviation pilot judgment using a situational judgment technique. *The International Journal of Aviation Psychology, 13*(4), 373-386.

Hunter, D. R. (2005). Measurement of hazardous attitudes among pilots. *The International Journal of Aviation Psychology, 15*(1), 23-43.

Hunter, D. R. (2006). Risk perception among general aviation pilots. *The International Journal of Aviation Psychology, 16*(2), 135-144.

International Air Transport Association. (2019). *IATA safety report 2018.* Montreal, Canada: Retrieved from https://libraryonline.erau.edu/online-full-text/iata-safety-reports/IATA-Safety-Report-2018.pdf

International Civil Aviation Organization. (2013). *Manual of evidence-based training*. Montreal, Canada: International Civil Aviation Organization.

Lerner, J. S., Lerner, J. S., & Tiedens, L. Z. (2001). Portrait of the angry decision maker: How appraisal tendencies shape anger's influence on cognition. *Journal of Behavioral Decision Making, 19*(2), 115-137.

Lerner, J. S., Li, Y., Valdesolo, P., & Kassam, K. S. (2015). Emotion and decision making. *Annual Review of Psychology, 66*(1), 799-823. doi:10.1146/annurev-psych-010213-115043

Mehta, R., Rice, S, Winter, S. R., Moore, J. C., & Oyman, K. (2015, April). *Public perceptions of privacy toward the usage of unmanned aerial systems: A valid and reliable instrument.* Poster presented at the 8th annual Human Factors and Applied Psychology Student Conference, Daytona Beach, FL.

Mehta, R., Rice, S., Winter, S. R. & Oyman, K. (2014). Consumers' perceptions about autopilots and remote-controlled commercial aircraft. *Proceedings of the 57thAnnual Meeting of the Human Factors and Ergonomics Society, Chicago, IL.*

Meister, D., & Gawron, V. (2010). Measurement in aviation systems. In J. A. Wise, V. David Hopkins, & Daniel J. Garland (Eds.), *Handbook of aviation human factors* (pp. 3/1-3/15). Boca Raton, FL: CRC Press.

The National Transportation Safety Board. (2019). *Assumptions used in the safety assessment process and the effects of multiple alerts and indications on pilot performance*. Washington, DC: National Transportation Safety Board (NTSB). Retrieved from https://www.ntsb.gov/investigations/AccidentReports/Reports/ASR1901.pdf

Ragbir, N. K., Baugh, B. S., Rice, S., & Winter, S. R. (2018). How nationality, weather, wind, and distance affect consumer willingness to fly in autonomous airplanes. *Journal of Aviation Technology and Engineering*, *8*(1), 2-10.

Rains, T., Winter, S. R., Rice, S., Milner, M. N., Bledsaw, Z., & Anania, E. (2017). Biofuel and commercial aviation: Will consumers pay more for it? *International Journal of Sustainable Aviation, 3*(3), 217-232. doi:10.1504/IJSA.2017.086846

Rice, S., Mehta, R., Steelman, L., & Winter, S.R. (2014). A trustworthiness of commercial airline pilots (T-CAP) scale for Indian consumers. *International Journal of Aviation, Aeronautics, and Aerospace, 1(*3), 1-15.

Rice, S., Mehta, R., Winter, S. R., & Oyman, K. (2015). A trustworthiness of commercial airline pilots (T-CAP) scale for American consumers. *Journal of Aviation Technology and Engineering, 4*(2), 55-63.

Rice, S., & Winter. S. R. (2015, June). *A quick affect scale: Providing evidence for validity and reliability.* Proceedings from the 10th International Conference on Interdisciplinary Social Sciences, Split, Croatia.

Rice, S., & Winter, S. R. (2019). Do gender and age affect willingness to ride in driverless vehicles: If so, then why? *Technology in Society, 58,* 1-7.

Rice, S., Winter, S. R., Doherty, S., & Milner, M. N. (2017). Advantages and disadvantages of using internet-based survey methods in aviation-related research. *Journal of Aviation Technology and Engineering, 7*(1), 58-65.

Rice, S., Winter, S. R., Mehta, R., & Ragbir, N. K. (2019). What factors predict the type of person who is willing to fly in an autonomous commercial airplane? *Journal of Air Transport Management, 75*, 131-138.

Rice, S., Winter, S.R., Tamilselvan, G., & Milner, M.N. (2017). Attitudes toward controlled rest in position (CRIP): A gender comparison between pilots and non-pilots. *International Journal of Aviation, Aeronautics, and Aerospace, 4*(3), 1-33.

Uber Elevate. (2016). *Fast-forwarding to a future of on-demand urban air transportation*. Retrieved from https://www.uber.com/elevate.pdf/

Vidulich, M. A., & Tsang, P. S. (1985, September). Assessing subjective workload assessment: A comparison of SWAT and the NASA-bipolar methods. *Proceedings of the Human Factors Society* (pp. 71-75). Baltimore, MD.

Wickens, C. D., Mavor, A. S., & McGee, J. (1997). *Flight to the future: Human factors in air traffic control*. Washington, D.C: National Academy Press. doi:10.17226/5493

Wiegmann, D. A., & Shappell, S. A. (2003). *A human error approach to aviation accident analysis: The human factors analysis and classification system* (1st ed.). Burlington, VT: Ashgate. doi:10.4324/9781315263878

Winter, S. R., Rice, S., Friedenreich, K., Mehta, R., & Kaiser, B. (2017). Passenger perceptions toward the federal flight deck officer program and their willingness to fly. *Aviation Psychology and Applied Human Factors, 7*(1), 7-17.

Winter, S. R., Rice, S., Rains, T., Milner, M., & Mehta, R. (2017). A longitudinal study on the alteration of consumer perceptions and the use of pilot medication. *Journal of Air Transport Management, 59,* 100-106.

Figure 1. Screen plot of the factor analysis from the first hypothetical scenario for WTF scale.
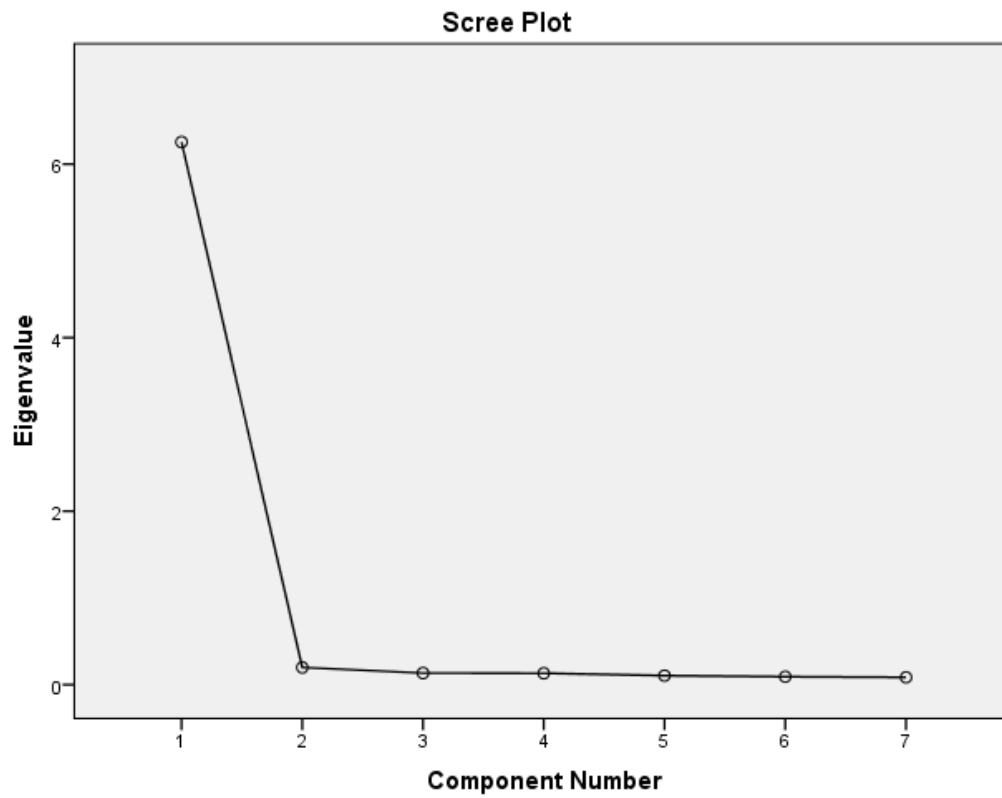
Figure 2. Screen plot of the factor analysis from the second hypothetical scenario for WTF scale.
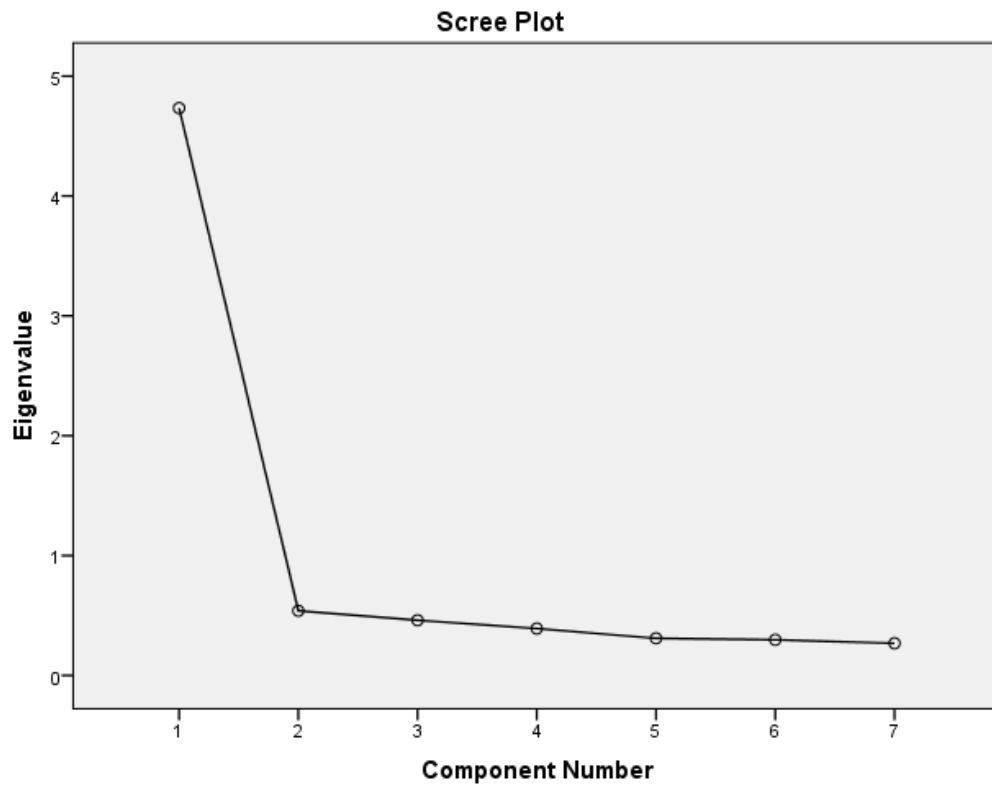
Figure 3. Screen plot of the factor analysis from the first hypothetical scenario for WTP scale.

Figure 4. Screen plot of the factor analysis from the second hypothetical scenario for WTP scale.

Appendix A – Willingness to Fly Scale

Please respond how strongly you agree or disagree with the following statements.

1. I would be willing to fly in this situation.

       Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

2. I would be comfortable flying in this situation.

       Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

3. I would have no problem flying in this situation.

       Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

4. I would be happy to fly in this situation.

       Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

5. I would feel safe flying in this situation.

       Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

6. I have no fear of flying in this situation.

       Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

7. I feel confident flying in this situation.

       Strongly Disagree   Disagree   Neutral   Agree   Strongly Agree

Appendix B – Willingness to Pilot Scale
Please respond how strongly you agree or disagree with the following statements.

1. I would be willing to pilot in this situation.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

2. I would be comfortable piloting in this situation.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

3. I would have no problem piloting in this situation.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

4. I would be happy to pilot in this situation.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

5. I would feel safe piloting in this situation.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

6. I have no fear of piloting in this situation.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree

7. I feel confident piloting in this situation.

Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree