PANTEION UNIVERSITY OF SOCIAL AND POLITICAL SCIENCES

SCHOOL OF SOCIAL SCIENCE

DEPARTMENT OF PSYCHOLOGY

POSTGRADUATE PROGRAM

“PSYCHOLOGY”

DIRECTION: “ APPLIED COGNITIVE AND DEVELOPMENTAL PSYCHOLOGY ”

The Cognitive Processes Involved in Analogue to Digital Cockpit Transition in Pilots

A postgraduate thesis submitted by

Vasiliki Betsiou

Athens, 2023

Three-member advisory committee

Argiro Vatakis, Associate Professor, Panteion University (Supervisor)

Stavroula Samartzi, Professor, Panteion University

Fotis Fotiadis, Laboratory Teaching Staff, Panteion University

The inspiration of this thesis topic was an observation in army aviation field, according to which the performance of helicopter pilots during the transition from analogue to digital cockpit presented significant differences. Specifically, an army aviation flight instructor (the author's husband) noticed that pilots with an appeal to technology, seemed to have better adaptation to new digital cockpits and greater performance especially in high-risk special army aviation operations. Therefore, it was hypothesized that the profile of a pilot with better transition to digital cockpits may be predicted, to lead to a better use of manpower in aviation.

Aut viam inveniam aut faciam.

Annivas, 212 bc

Acknowledgments

I am most grateful that my study was supported by Global Aviation Academy. Global Aviation academy(GR-ATO-102) is certified to provide pilot training by the Civil Aviation Authority (CAA) and is a member of the European Aviation Safety Agency (EASA). It purveys high-level training for civil aviation pilots and since 2018 has been the training provider for the scholarship program of AEGEAN Airlines, as well as EPST (European Pilot Selection & Training). The academy provided both analogue and digital simulators, the facilities required for the briefing before the flights at its modern simulator center at Megara airbase and its experienced instructors in order to sponsor my master's studies. Therefore, I would like to thank the staff at Global Aviation Training Centre for their support which contributed to the completion of this research project.

There are also several individuals who contributed to this thesis provided guidance and support throughout my academic endeavors.

First, I would like to acknowledge my supervisor, Dr Argiro Vatakis, for her patience, guidance, support, constructive feedback, insights, and suggestions. For their input and guidance, I also extend my appreciation to my committee members, Dr Stavroula Samartzi, and Dr Fotis Fotiadis.

For his inspiration, useful observations, motivational talent, unyielding support, and affection, I thank my husband, George. He was the first to observe in the field what the study investigated. I cannot adequately express how much I appreciate his continual encouragement, and belief in me. I am most thankful to my daughters Evaggelia and Christiana, for reminding me of life outside of academia with love, laughter, joy, and patience.

Finally, I would like to express my appreciation to all my academic cohorts, for their friendship, support, and study group.

Table of Contents

Abstract

Background. Since the explosive growth in aircraft complexity, when the Wright brothers' first powered flight in 1903 the cockpit's visual complexity had reached a point where there was no more space for new instruments. In traditional aircraft instrumentation the flight instruments (Airspeed Indicator, Attitude Indicator, Altitude Indicator, Heading Indicator, and Vertical Speed Indicator,) are displayed in front of a pilot, making it easy for him or her to scan and obtain information from them, to maintain a safe flight (Mumaw, Sarter, Wickens, 2001). Nevertheless, technology allowed for the introduction of digital displays, which show the same information in a smaller area (Curtis et al. 2010). These current flexible multifunction displays require pilots to modify their cognitive processes to safely accomplish flight in instrument conditions.

Aims. The transition from analogue to digital technology affects the cognitive demands and processes associated with extracting meaning from large fields of data (Hamblin, Gilmore, & Chaparro, 2006). The pilot's scanning technique is maybe where we should start to encounter uncharted territory. A pilot uses a specific scan path when flying in instrument conditions (Jones, 1985). He / she begins his or her scan at the attitude indicator, then scans another instrument and returns to the attitude indicator (Pennington, 1979). This type of scanning pattern is called the 'T' scan path and is commonly used in instrument flying conditions, in analogue cockpit (Mumaw, Sarter, Wickens, 2001). According to Mumaw, et al. (2001), "*There are no documented strategies for effectively monitoring this diverse set of indications, and, as a result, pilots often develop their own not necessarily effective approaches to the task.*" (p. 2). As a result, any new interface which increases cognitive demand compared to traditional systems may not be used to its full potential (disuse or misuse). Considering Woods's (1996) reports that pilots cope with technology by using only a few of the available functions, especially during high workload periods, it is worthwhile to

determine potential factors that allow pilots to shift to digital instrumentation faster and more efficiently, to reduce training times and potential risks during challenging operations.

Hypothesis. The purpose of the current research was to test the idea that Affinity for Technology Interaction scale (ATI) can be applied to an academic aviation setting to predict pilots' performance in the transition from analogue to digital displays. In addition, our study investigated what causal direction might exist between the ATI and workload variables and portray the profile of the pilot who is going to transit better from analogue to digital displays. Given the disposition of higher-ATI pilots to figure out systems on their own-in the transition from analogue to digital displays - whereas lower-ATI pilots need more assistance, measures supporting adaptation processes in familiarizing with new technology (e.g., trainings, tutoring systems, adaptive user interfaces) could become more efficient and effective by taking individual differences into account (e.g., adapting speed of trainings or learning demands to user diversity).

Definition of Key Terms

Traditional (analogue) instrumentation means the flight condition information available to the pilot is in the form of non-electronic flight instrumentation and positional information is obtained from ground based navigational information sources (FAA, 2012).

Glass (digital) cockpit instrumentation means displays driven by computer graphic systems (Mitchell et al., 2010).

Περίληψη

Ιστορικό. Με την εκρηκτική ανάπτυξη της τεχνολογίας στα αεροσκάφη, από την πρώτη πτήση των αδελφών Wright το 1903, τα πιλοτήρια είχαν φτάσει στο σημείο όπου δεν υπήρχε πλέον χώρος για νέα όργανα. Σε αεροσκάφη με αναλογικά όργανα, τα όργανα πτήσης (Ενδείκτης Ταχύτητας, Τεχνητός Ορίζοντας, Ενδείκτης Ύψους, Ενδείκτης Πορείας και Ενδείκτης Ανόδου / Καθόδου) παρουσιάζονται μπροστά στον πιλότο, διευκολύνοντάς τον να σαρώσει και να λάβει πληροφορίες από αυτά, με άλλα λόγια να διατηρήσει μια ασφαλή πτήση (Mumaw, Sarter, Wickens, 2001). Παρ'όλα αυτά, η τεχνολογία βοήθησε στην εισαγωγή ψηφιακών οθονών, οι οποίες εμφανίζουν τις ίδιες πληροφορίες σε μικρότερο χώρο (Curtis et al. 2010). Αυτές οι οθόνες πολλαπλών λειτουργιών απαιτούν από τους πιλότους να τροποποιήσουν τις γνωστικές νοητικές τους διεργασίες για να πραγματοποιήσουν με ασφάλεια την πτήση σε συνθήκες χρήσης οργάνων.

Στόχοι. Η μετάβαση από την αναλογική στην ψηφιακή τεχνολογία επηρεάζει τις γνωστικές απαιτήσεις και διαδικασίες που σχετίζονται με την αντίληψη δεδομένων (Hamblin, Gilmore, & Chaparro, 2006). Η τεχνική σάρωσης του πιλότου είναι ίσως η αχαρτογράφητη περιοχή που θα πρέπει να αρχίσουμε να μελετάμε. Ένας πιλότος χρησιμοποιεί μια συγκεκριμένη διαδρομή σάρωσης όταν πετάει σε συνθήκες οργάνων (Jones, 1985). Αυτός / αυτή ξεκινά τη σάρωση από τον Τεχνητό Ορίζοντα, στη συνέχεια σαρώνει ένα άλλο όργανο και επιστρέφει στον Τεχνητό Ορίζοντα (Pennington, 1979). Αυτός ο τύπος σάρωσης ονομάζεται διαδρομή σάρωσης «Τ» και χρησιμοποιείται συνήθως σε συνθήκες πτήσης με όργανα, σε αναλογικό πιλοτήριο. Σύμφωνα με τους Mumaw, et al. (2001), «*Δεν υπάρχουν τεκμηριωμένες στρατηγικές για την αποτελεσματική παρακολούθηση αυτού του διαφορετικού συνόλου ενδείξεων και, ως αποτέλεσμα, οι πιλότοι συχνά αναπτύσσουν τις δικές τους, όχι απαραίτητα αποτελεσματικές στρατηγικές*» (σελ. 2). Ως αποτέλεσμα, οποιαδήποτε νέα διεπαφή που αυξάνει τη γνωστική απαίτηση σε σύγκριση με τα παραδοσιακά συστήματα μπορεί να μην

χρησιμοποιείται στο μέγιστο των δυνατοτήτων της (μη ή κακή χρήση της). Λαμβάνοντας υπόψη τις αναφορές του Woods (1996) ότι οι πιλότοι αντιμετωπίζουν την τεχνολογία χρησιμοποιώντας μόνο μερικές από τις διαθέσιμες λειτουργίες, ειδικά σε περιόδους υψηλού φόρτου εργασίας, αξίζει τον κόπο να προσδιοριστούν πιθανοί παράγοντες που επιτρέπουν στους πιλότους να μεταβούν στα ψηφιακά όργανα γρηγορότερα και πιο αποτελεσματικά, ώστε να περιοριστεί ο χρόνος εκπαίδευσης και οι πιθανοί κίνδυνοι κατά τη διάρκεια απαιτητικών επιχειρήσεων.

Υπόθεση. Ο σκοπός της παρούσας έρευνας ήταν να διερευνηθεί αν η εφαρμογή του ερωτηματολογίου ATI (Affinity for Technology Interaction) μπορεί να εφαρμοστεί σε ένα ακαδημαϊκό αεροπορικό περιβάλλον για την πρόβλεψη της καλύτερης απόδοσης των πιλότων, στη μετάβαση από τις αναλογικές σε ψηφιακές οθόνες. Επιπλέον, η μελέτη μας διερεύνησε ποια αιτιολογική κατεύθυνση μπορεί να υπάρχει μεταξύ των μεταβλητών ATI και φόρτου εργασίας και απεικόνισε το προφίλ του πιλότου που πρόκειται να μεταβεί βέλτιστα από τις αναλογικές σε ψηφιακές οθόνες. Δεδομένης της ένδειξης ότι οι πιλότοι με υψηλότερα σκορ ATI μπορούν να αντιληφθούν γρηγορότερα και καλύτερα τα ψηφιακά συστήματα - κατά τη μετάβαση από τις αναλογικές σε ψηφιακές οθόνες - ενώ οι πιλότοι με χαμηλότερα σκορ ATI χρειάζονται περισσότερη βοήθεια, θα μπορούσαν να υπάρξουν πιο αποτελεσματικά μέτρα που να υποστηρίζουν διαδικασίες προσαρμογής στην εξοικείωση με τη νέα τεχνολογία (π.χ. εκπαίδευση, συστήματα διδασκαλίας, προσαρμοστικές διεπαφές χρήστη) λαμβάνοντας υπόψη την αξιολόγηση με το ερωτηματολόγιο ATI (π.χ. προσαρμογή της ταχύτητας της εκπαίδευσης ή των απαιτήσεων μάθησης στην ποικιλομορφία των χρηστών).

Ορισμός Βασικών Όρων

Αναλογικά όργανα πιλοτηρίου: οι πληροφορίες για την κατάσταση πτήσης που είναι διαθέσιμες στον πιλότο είναι με τη μορφή μη ηλεκτρονικών οργάνων πτήσης και οι πληροφορίες θέσης λαμβάνονται από πηγές πληροφοριών πλοήγησης στο έδαφος (FAA, 2012).

Ψηφιακά όργανα πιλοτηρίου: οι πληροφορίες για την κατάσταση πτήσης που είναι διαθέσιμες στον πιλότο παρουσιάζονται σε οθόνες που οδηγούνται από γραφικά υπολογιστικά συστήματα (Mitchell et al., 2010).

Introduction

As technological change produces complexity in the aviation field, it is tempting to think of the powers of technology as the source of solutions to the problems that accompany these complexities. Ironically, technological capabilities, which on the surface seem to offer the potential for expanding human interpretative capabilities, have in practice contributed new complexities to the world of the aviators. Specifically with the transition from analogue to digital cockpit, variety of information is available to provide pilots with comprehensive and accurate data about cockpit conditions and the external environment (Cheng et al., 2019), but these advancements led to a new problem; A syndrome, which Wiener in 1989 termed clumsy automation, which is a form of poor coordination between the human and machine in the control of dynamic processes where the benefits and the costs or burdens imposed by the technology occur during periods of peak workload, high criticality or high tempo operations (Cook et al, 1990; Sarter & Woods, 1992). The problem occurs because of a fundamental relationship: the higher the tempo of operations, the greater the information processing activities required to cope with the trouble or pace of activities (Woods et al., 1994). Studies revealed a variety of ways in which the clumsy use of technology creates new complexities that increase the potential for erroneous assessments and actions under certain circumstances and given the presence of other factors, creates new paths to system breakdown, (Woods, Johannesen; Cook & Sarter, 1994). It seems like pilots tailor both the system and their own cognitive strategies to cope with this bottleneck. Some of them are observed to constrain the display of data into a fixed spatially dedicated default organization rather than exploit device flexibility. They force scheduling of device interaction to low criticality self-paced periods to try to minimize any need for interaction at high workload periods. They develop stereotypical routines to avoid getting lost in the network of display possibilities and complex menu structures (Woods et al., 1994). They cope with new burdens associated with clumsy technology by learning only a subset of stereotypical methods, underutilizing system functionality. They also convert interface flexibility into fixed, spatially dedicated displays to avoid interacting with the interface system during busy periods (Woods et al., 1994). It seems like they escape from flexible but complex modes of automation and switch to less automated, more direct means of accomplishing their tasks when the pace of operations increases. As a result, they tailor their activities to insulate the larger system from device deficiencies (Cook and Woods, 1994). While wide sampling of cases within the aviation field is a necessary condition it is probably not a sufficient condition for acquisition of flexibility. Practicing with lots of variations of a task is one way to inculcate a kind of flexibility, the structural kind: 1) through practice, 2) on a wide sampling of cases, and 3) in a relatively constricted (and stable) domain of activity. Although some pilots have the same opportunity in practicing, they end up thinking in an ossified manner. Or alternatively, by becoming progressively ossified, their performance may progressively deteriorate, leading to their loss of constituency leading to loss of opportunity to engage in rich experiences affording the opportunity to progress. In short, they may just not be able to succeed in the operation, and therefore they are not selected for very difficult missions (i.e., high risk in military operations). On the contrary pilots who remain successful in the operation field, are observed to overcome their considerable pressure to oversimplify. At the heart of such an effort is the need to develop an underlying epistemic stance. That is the individual comes to expect variability, novelty and interdependence in knowledge and its uses. He or she looks for connection as well as for legitimate ways to compartmentalize for change and patterns of change as well as for what remains the same, for exceptions as well as rules, for context sensitivity (and its basic determinants), as well as more universal application of concepts and principles. There are no additional studies that investigated which pilots use technological flexibilities skillfully.

To examine the ways in which pilots deal with the switch from analogue to digital, we considered Lewin's (1939) reports about Behavior as a function of the Person and Environment $B = f\ (P \times E)$, meaning that coping with technology is a function of personal resources and system resources. From an analytical standpoint, the influence of personal resources on successful coping with technology is twofold. First, the higher the skills and knowledge regarding interaction with specific systems, the easier it is to cope with similar new systems. Second, users' personality characteristics also play an important role to the extent that they manifest in general interaction styles. A key dimension of pilots' resources is the way he/she approaches technical systems, namely his/her Affinity for Technology Interaction (ATI). ATI can explain differing human behavior and usability ratings, as in general, higher ATI users like to explore new technology, while lower ATI users are likely stumped by it. In the aviation field it can be used to describe whether a pilot tends to actively approach interaction with technical systems or, rather, tends to avoid intensive interaction with new systems and prefer to continue with their habitual use, avoiding the need for a detailed preoccupation with technical systems (Franke, Attig, & Wessel, 2018). This individual-difference dimension is what is conceptualized as ATI.

The present study is designed to measure pilots' ATI as it is hypothesized that it is correlated with successful transition to digital displays, which is relevant for mastering daily aviation life. Considering that proper information acquisition is a vital skill and lays the foundation for safe flying skills, scanning pattern, the time to accomplish information, and workload measures were collected, and compared to ATI scores, between an analogue and a digital cockpit.

1.1 Performance

In 2011 McCracken's survey nearly half (43%) of the participants found the digital cockpit aircraft more difficult to fly than the conventional aircraft and over half of the students obtained a lower check ride score in the glass – digital cockpit aircraft than in their previous check ride. The results showed that pilots using the glass – digital cockpit took longer to recover from unusual attitudes than pilots using the traditional cockpit. There are also results of a study examining performance in paired simultaneous approaches that illustrate the trade-offs involved in mixing old and new procedures and automation (Verma et al., 2011). Perhaps the most interesting finding was the interaction between display and automation with respect to workload. These results suggest that pilots need to be very careful when applying old mental models or 'how to' knowledge to new displays and automation. An expert pilot is one who brings more and more of his or her world into the realm of the familiar. Even though much in this realm is context dependent, the expert has to be able to see and codify (in schema – like structures) much of these relevant contexts and their effects (Chase and Simon 1973; Feltovich 1983). This enables a kind of rich knowledge-based flexibility, so long as the expert is functioning broadly within his or her usual domain and that domain is relatively stable. However, there a more fundamental expertise has been noted by Woods and colleagues in their studies of cognition in the workplace: '*Interestingly practitioners are acutely aware of how deficient their rules of thumb may be and how certain situations may require abandoning the cognitively easy method in favor of more cognitively demanding deep thinking*' (Woods, Johannesen, Cook, Sarter 1994, p.66). However, for this more basic kind of reasoning to be engaged it appears that there must be some tip-off to the expert that the current situation is outside the normal realm of inquiry. Otherwise, it is schema-driven processing as usual, even when this leads to bad outcomes. In this regard, Woods and Colleagues go on to point out that failure to recognize when simplification strategies are not adequate and when a situation is not within routinized normal practice and, hence, requires "*deeper thinking*" are upon the major contributors to workplace tragedies and mishaps (even when these involve highly experienced people). Experience can induce a false sense of coherence, or the tendency to see what one expects to see rather than what is there, as illustrated by the "phantom memory" phenomenon found in part-task simulation data (Mosier et al., 1998; Mosier, Skitka, Dunbar, & McDonnell, 2001). New automation (i.e., digital instrumentation) requires coherent data-processing displays to be examined thoroughly to ensure accurate comprehension, particularly when displays are coupled with new systems and procedures. This represents the need to shift to knowledge-based mode of problem solving to guarantee better performance in the transition to digital displays that concludes accordingly in safer flights. In that way we may see more experienced pilots to make better decisions in terms of speed and accuracy, allocate more attention to relevant cues when failures are present and show better performance in motion anticipation (Schriver, Morrow, Wickens, & Talleur, 2008). Consequently, as Damos (1996) noted, performance is not an easy criterion to measure or use. It necessitates a comprehensive job analysis, and clearly defined and articulated measures of performance. What is known is that the role of a pilot continues to evolve with the advancement of technology. Pilot selection tests need to evolve to maintain pace with these changes. As highlighted in a recent report from the Federal Aviation Administration, adhering to best practices for the development of selection methods is the optimum pathway to success (Broach et al., 2019).

1.2 Individual Differences

The speed of technological innovation in the aviation field is steadily increasing. Thus, pilots need to learn to cope with new technology at a faster pace and understanding how to optimally utilize current technology becomes more and more relevant. Consequently, pilots do not only differ in their technology usage but also in their success with utilizing new technology. Hence, it becomes increasingly important to take the individual fit between persons and technical systems in the focus of psychology research. Adolphe, et al' findings (2022) indicates that complex interaction lies on the cognitive mechanisms depending on stimulus parameters even in the same task. Of relevance to displays are the resources associated with attention, spatial orientation, and working memory. Attention involves focusing on some feature of a message, the environment, or even an internal thought. This focus can be driven by external factors that draw attention automatically (e.g., color, a blinking light) or by the individual who makes deliberate decisions about where to direct attentional resources. Attention directed toward a display, whether driven in a top-down (by the learner) or bottom-up (by features of the display) manner, is necessary to initiate and maintain focus on stimuli (Hegarty et al. 2010). This focus can be a challenge though as attentional parameters are limited and can only be directed toward a narrow range of information contained within an instructional message. Information that is given sufficient attention and recognition resides in working memory (WM), a term which some researchers use to refer to the *mental desktop* where thought occurs (Baddeley 2007; Mayer 2009). Working memory is a fixed resource that can be used to process an instructional message. For example, interpreting a display's contents and determining the usefulness of those contents given a person's goals, involves active processing in working memory (e.g., encoding information into memory, retrieving information from memory, etc.). Thus, actively maintaining and using information in working memory consumes precious attentional resources, making it crucial that this limited pool be leveraged in a way that contributes to comprehension and understanding of important information in an instructional message. A display may improve processing efficiency when it helps a learner organize important information more quickly (e.g., related ideas are near one another). For instance, spatially integrating important information (digital cockpit) can make it easier to see relations among important content than when that information is spatially separated (Sweller et al., 1998). The spatial design of a display can thus potentially facilitates or impede organizational inferences of presented content. For instance, a display can minimize the need to hold facts in working memory during a search for related information as would occur when searching a text for disparate pieces of information that need to be related to one another such as in the analogue instrumentation (McCrudden & Rapp, 2015). Hence, a display may support processing, by minimizing the resources necessary to engage with information. Nevertheless, people might have different amounts of spatial ability (Hoffler, 2010) that influence the ease with which they process a visual display. Or individuals might differ in how they utilize those resources (Just & Carpenter, 1992); that is, people might exhibit different strategies and tendencies when they process displays (Ponce and Mayer 2014). Further, individuals can differ in both the quantity and quality of prior knowledge they possess, which could influence the ways in which integration operates during comprehension (Hegarty et al., 2010; Mason et al., 2013). Visual perceptual skills are crucial abilities accounting for the advantage of highly trained experts in many domains (Li et al., 2012). Indeed, expertise exerts a top-down modulation on gaze behavior and strategies. In this sense, experts with extensive training, domain knowledge, and experience can perceive important relationships among multiple information, enabling them to orient their attention toward relevant information and identify abnormalities with a high efficiency (Hoffman and Fiore 2007; Palmeri et al. 2004). Multiple studies investigated differences in scan paths and scan patterns between novices and experts from different domains (Law et al, 2004; Ooms et al., 2014). In aviation, the literature also emphasizes different visual scanning strategies in novices vs experts pilots (Kasarskis et al., 2001; Yang et al., 2013). As experts show more flexible scanning strategies and they are more focused on relevant information and allocate their attention more efficiently, they adjust their scanning behaviors more effectively to the situational demands.

Besides these characteristics, other kinds of individual differences matter. Consider that the kinds of expectations or goals that individuals have when they approach a visual display could guide particular kinds of interactions with the material. For instance, learners who seek to understand content might work harder to organize and integrate what they are seeing, in contrast to learners who seek to peruse a display for fun in a more cursory way. Different learners might have different motivations to engage in the processing of a display, which can influence the extent to which they attempt to make connections or derive understandings from what they are viewing. Other individual differences could also play important roles such as cultural considerations (Guiterrez & Rogoff, 2003), learner preferences (Kozhevnikov et al., 2014), and the need or desire for competency (Stroet et al., 2015).

Across all these characteristics, the ways in which an individual engages with, processes, and derives an understanding of the display could be related to features of the learner. Thus, it is important to identify learner characteristics and to carefully consider how they might interact with display experiences. Individual differences in pilots may be unknown factors that would influence skill development, making the interpretation of our results difficult, presumably due to small numbers of participant sampling. Using a cognitive test battery (COMPASS) that includes six tasks (Control, Slalom, Memory, Math, Spatial Orientation and Task management) we were able to measure diverse individual differences. In general, students pilots with sufficient cognitive abilities can successfully complete the theoretical part of the flight training within the given time. Furthermore, for pilots who have their license, the cognitive abilities are important, especially for captains, to get a quick overview and take good decisions (in non-routine situations) within a short period of time.

As well as cognitive abilities, good sensory-motor skills are important to learn to fly. Individuals with good flying aptitude will show as students that they learn more easily and faster than those who lack potential flying aptitude. For flying training organizations and airlines newly-trained and licensed pilots will have to adjust to new (technical) situations and aircraft types. Pilots with a good flying aptitude learn new tasks faster, thus saving valuable training time and expenses. The operational flying capabilities of candidates is assessed by flying aptitude tests, such as multi-tasking (Task Manager), spatial orientation, and eye-hand-(foot) co-ordination (Control and Slalom). Selection tests such as COMPASS test, are widely used to minimize the time and costs associated with equipping personnel with the knowledge and skills required to perform a specific job (Stabile, 2002). They are generally founded on the perceived intellectual ability required to succeed in the position, as well as the desired personality, attitude, and aptitude of the candidate (Hunter & Burke, 1994; Stabile, 2002). For high-hazard, high-consequence industries such as aviation, the impetus for effective tests is greater than for less safety critical industries (Broach et al., 2019). Further adding to this need, is the shortage of pilots worldwide (Boeing Commercial Airplanes, 2019).

Accordingly, the present study aimed to use the cognitive test battery (COMPASS) to capture diverse individual differences, to avoid only assuming that participants with better performance (better response times) are superior in a specific cognitive ability. It was designed to prove that highly motivated in technology pilots (High and Very High ATI participants) score better in any subscale of the test battery (i.e., spatial orientation, memory).

1.3 Relation among Affinity in Technology Interaction and Need for Cognition scales

Early studies examining ATI have shown that the scale is applicable in highly heterogeneous populations and have found higher ATI scores to be related to higher intrinsic motivation for technical device usage, and lower subjective workload while interacting with new technical devices (Wessel et al., 2019). Theoretically, ATI scale is rooted in the construct Need for Cognition (NFC), as the need to actively explore new technical systems and the tendency to cognitively engage with the systems. Thus, while NFC can be seen as the relatively stable tendency to enjoy intensive thinking and effortful cognitive activity, ATI can be seen as the relatively stable tendency to enjoy intensive technology interaction. Viewing technology interaction as a type of problem-solving task (Beier, 1999) the construct NFC appears particularly well suited to ground ATI theoretically (Schmettow, Noordzij, & Mundt, 2013). NFC denotes that individuals differ regarding their tendency to engage in cognitive activities (Cacioppo & Petty, 1982; Cacioppo, Petty, Feinstein, & Jarvis, 1996). Actively exploring new systems also needs a tendency to cognitively engage with the systems. Every new technical system requires adaptation and learning by its users (e.g., because of new functions, interfaces, interaction paradigms; Hawk, 1989; Tyre & Orlikowski, 1996). That is, for successful adaptation to new systems, users need to have certain personal coping resources (Beaudry & Pinsonneault, 2005; Chen, Westman, & Eden, 2009). Existing skills for interacting with similar systems (e.g., computer literacy, Poynton, 2005; e-Health literacy, Norman & Skinner, 2006) can directly facilitate coping by reducing adaptation demands. However, general interaction styles (i.e., facets of user personality) can also drive users' adaptation to technical systems and therefore act as coping resources for successful technology interaction. Studies concluded that individuals high in the need for cognition tend to seek out and reflect on information to make sense of stimuli and events, whereas individuals low in the need of cognition tend to use other sources such as heuristics to make sense of the world (Cacioppo & Petty, 1982). Thus, given this tendency to seek out and enjoy effortful cognitive activity, those higher in need for cognition are generally expected to have more positive attitudes toward situations that require reasoning and problem solving, and to respond more substantively to such situations. Studies also confirmed that individuals with high scores on the NFC scale (cognisers) tend to be flexible in their choice of learning strategies (Cacioppo, Petty, Feinstein, 1996). In addition, they are usually highly motivated for challenging tasks, not strongly influenced by surface features, and they have excellent control over their attentional resources. In contrast, individuals low in need for cognition (cognitive misers) show little affection for complex thought and are considered to rely more on others to find meaning in outside events (Evans, Kirby, and Fabrigar, 2003). Cacioppo et al., in 1996, pinpointed that although everyone must make sense of their world, those who are high in need of cognition (cognisers) tend to seek, acquire, think about, and reflect on information to make sense of stimuli and events. Winne's (1995) discussion of mental resources required for cognitive monitoring, suggests that cognisers, with their habit of engaging in mental reflection, would have the advantage over cognitive misers, and suggested that cognitive monitoring processes would be more likely to have become automatic for those who are high in need for cognition. The need for cognition is furthermore positively linked with openness to experience and intelligence (Furnham & Thorne, 2013), intrinsic motivation (Cacioppo et al., 1996), and information processing (Sicilia, Ruiz, & Munuera, 2005). Importantly, it also predicts a range of attitudinal and behavioral outcomes, including preferences for a complex number-circling task over a simple one (Cacioppo & Petty, 1982) and achievements of higher-grade point averages (Aquino et al., 2016; Wolf et al., 2017). NFC has been related to more intensive flow states regarding website interaction (Sicilia, Ruiz, & Munuera, 2005), lower computer anxiety (Maurer & Simonson, 1993), higher technological innovativeness (Hoffmann & Soyez, 2010), and a stronger tendency to search for more efficient problem-solving procedures when interacting with computers (Ebelhäuser, 2015, Keil, 2015). ATI is conceptualized in close relationship to NFC (Schmettow & Drees, 2014). Hence, the ATI scale provides a tool to discriminate between pilots based on their differing tendency to actively engage in intensive (i.e., cognitively demanding) technology interaction.

1.4 Workload

A flight environment requiring pilot to hold a large amount of information in his/her working memory while seeking more information or while attending to a secondary task (i.e., answering a radio call) clearly describes a potential workload problem. This could be especially true during the unexpected occurrence of new operative demands (i.e., transition from analogue to digital). It is indicative that pilots gather information and might respond differently to a particular interface (analogue or digital) depending on workload levels. The perception of the level of workload can also be affected by the experience, the skills or simply the individual differences between pilots. For example, novice and expert aircraft pilots will clearly experience different levels of workload when performing the same task (Borghini et al., 2011; Parasuraman and Jiang, 2012; Doppelmayr et al., 2008). In fact, skill development and expertise produce both an economy of action and automated "motor programs" that do not require conscious effort. In aviation these motor programs are called Boldface and they are the steps (emergency procedure memory items) necessary to deal with in-flight requirements promptly and completely. Measurement of pilot workload during flight, under different flying conditions, is necessary to evaluate pilot performance as it is a fact that sometimes a pilot's performance decrements may result more from a different interface than from a depletion of mental resources (A. Law & S. Jennings, 2019). Workload can be defined as "the relative capacity to respond" (Lysaght et al., 1989). "Workload is also a construct that is used to describe the extent to which an operator has engaged the cognitive and physical resources required for a task performance" (Backs, Ryan, & Wilson, 1994). These definitions show that workload which is a difficult to define concept consists of several components: (1) there is an operator, using his or her resources to respond to (2) external physical or cognitive demands to (3) perform a certain task. Several subjective rating techniques are available to measure operator-based workload. These rating techniques are called subjective to set them apart from "objective techniques" such as physiological measures. As subjective techniques can be quick and inexpensive to administer and analyze (Hill *et al.*, 1992), we will use the NASA-TLX which is one of the most widely used instruments to assess overall subjective workload. (Hart, 2006). In TLX, workload is defined as the *cost incurred by human operators to achieve a specific level of performance.* The subjective experience of workload is defined as an integration of subjective responses (emotional, cognitive) and evaluation of behaviors. Hart and Staveland (1987) concluded that the TLX provides a sensitive indicator of overall workload as it differed among tasks of various cognitive and physical demands. Battiste and Bortolussi (1988) reported significant workload effects as well as a test – retest correlation of +.769.

1.5 Situation Awareness

As digital flight is qualitatively different than flight in traditional cockpits, new skills must be developed and practiced. These are described as "cognitive skills," including an emphasis on planning, alternative selection, and predicting and monitoring the performance of the automation. This is what is called situational awareness. Parasuraman et al. (2008, p.144) described situation awareness as the "*continuous diagnosis of the state of a dynamic world*". Bolstad et al. (2010) evaluated a computer-based situation awareness training system for general aviation pilots and found that lower situation awareness scores contributed to poorer simulated flight outcomes. Endsley and Bolstad (1994) investigated individual differences in pilot situation awareness and found that differences pertaining to perception and spatial skills were most associated with situation awareness scores. In today's flight deck, situation awareness is not an easy task as pilots report spending a lot of time working on it, even after extensive use of the systems (Endsley, Jones 2011). Regular attention or scanning of available information is a vital skill that pilots must learn and maintain to safely fly an aircraft. Research (Hiremath et al., 2009) proposed that in the traditional cockpit the position of the airspeed and altitude indicator needles could be picked out at a glance. In contrast, digital cockpit displays do not present the whole data range, so to get an idea of the airspeed or altitude, the pilot must focus longer on the numerical readout thus, subsequently increasing workload. Although digital cockpits are designed to enhance situational awareness (SA) and make flying simpler and easier, it requires a greater effort to maintain "situational awareness," which can easily be sacrificed in highly automatic operations (high risk). The cause of several aircraft accidents has been attributed to lack of SA due to cockpit automation. A survey conducted in 1996 showed that approximately three-quarters of the situational awareness errors made by a pilot were due to a failure to monitor and obtain data from the instruments and the outside world (Jones & Endsley, 1996). In other words, this is a failure at the first level of situational awareness. This occurred because pilots were *out-of-the-loop*, that is, they did not know what the system was doing or why (Endsley, Jones 2011). In cases like this, pilots tend to dismiss conflicting information (the *confirmation bias*) and may never realize the error they are making. In general, the less direct access the operator has to the system, the more important feedback is to maintain SA. Two related factors that influence whether automation is used, and how, are trust and reliability. Too much trust leads to complacency or over-reliance (Parasuraman et al, 2008). Errors resulting from this bias are generally split into omission and commission errors (Mosier et al., 1998). Among some of the decisions a pilot needs to make is when to attend to information (which is always available), where to look for it among all the different menus options, and how to interpret that information (Hollnagel 2012). Bainbridge in 1983, called this outcome the *ironies of automation*, implying that automation may sometimes be more time consuming and/or incomprehensible than the manual operation of a system. The term *automation surprise* was introduced to designate all those occasions when humans were left astonished and confused by the machine's behavior. In a pair of experiments, Mosier and colleagues (Mosier et al., 1998; 2001) identified automation bias as a threat to SA.

As lack of situational awareness have resulted in several aviation incidents and accidents (e.g., Sarter & Woods, 1994a,b; Woods & Sarter, 2000), it is worthwhile to investigate pilot's situational awareness (response times-performance in different workload conditions (Climbing Leg, Level Flight Leg, 360 Turn, Descending Leg are the correspondence of situational awareness in the study). As pilot's scores in Affinity for Technology Interaction Scale affected the strength of the relation between performance (response times) and workload, we were able to predict a pilot's profile of a better transition from analogue to digital displays that abridges the automation surprises.

Figure 1. Flight instruments : Airspeed Indicator (a), Attitude Indicator (b), Altitude Indicator (c), Heading Indicator (d), and Vertical Speed Indicator (e) in Analogue and Digital Instrumentation.

Method

2.1 Participants

The purpose of the study was to investigate the cognitive processes engaged in a successful transition to digital cockpit technology in the aviation field and more specifically to find the relationship between cognitive criterion predictors and successful transition performance in an academy flight program. All subjects recruited (*N*=14), were students from an aviation academy who were attended an aviation training course respectively novices (*n*=7, no additional experience in any kind of cockpit, analogue or digital) and trainees (*n*=7, experience in analogue cockpit or both analogue and digital cockpit).

a. The Initial Training Group with no flying experience in a flying simulator or aircraft, who were in the stage of attending theoretical courses.

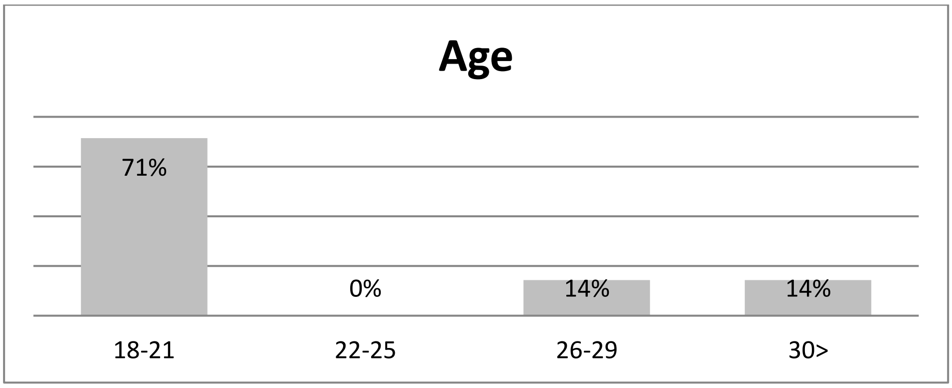


Figure 2. The ages of the Initial Training Group.

b. The Continuous Training Group who had finished theoretical courses and analogue or both analogue and digital flight simulator's courses.

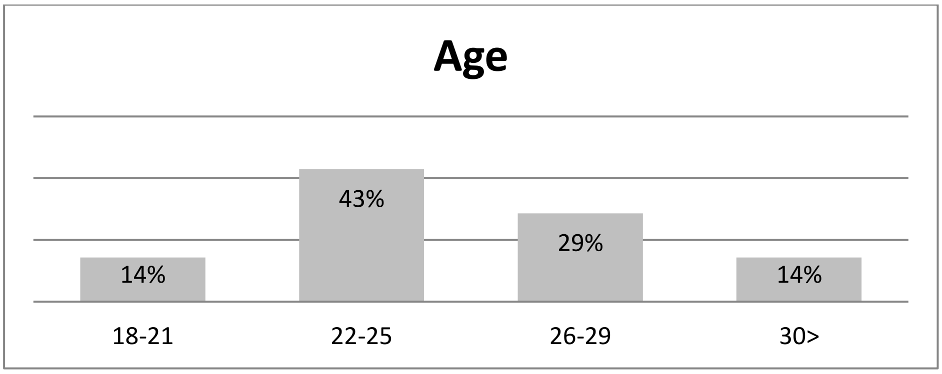


Figure 3. The ages of the Continuous Training Group.

Flight novice undergraduates with no previous experience were examined on flight displays to maximize internal validity. Their exposure and familiarity with both kinds of display was, thus, completely controlled. Using this group before their initial training in analogue displays clearly involves a potential trade off with external validity or the ability to generalize the results to the population of pilots (Rogers et al., 2012). Nevertheless, predictive validity played the most important role, because the overall objective of the tests used in this study were to predict the future performance of pilots, as it measured the ability of participants to manage workload and have better response times in the tasks. Thus, the first test results were collected before the beginning of training of the Initial Training Group, and they were related to the measurements of Continuous Training Group who completed training courses, analogue or both analogue and digital simulator courses, respectively. The convenience sample of flight students was representative of this limited population in the academy and was conveniently obtainable. To some extent, convenience sample is a random sample as the participants are representative of a specific population (Heiman, 2002), for example, the flight students who are in primary flight training courses. Trainees participate in this study primarily in analogue and then in digital displays as non–flying pilots and flying pilots respectively. One of the primary roles of a non-flying pilot is to provide back up for the flying pilot, so it is critical that both pilots (flying and non-flying) maintain a high level of situational awareness at all times. Having both pilots in the loop and cognizant of the current state of the aircraft is a critical aspect of the safe operation of a multi-crew platform.

The implementation of our study took place at a large aviation academy, Global Aviation S.A. at Megara Airport area. All pilots enrolled in the aviation training program were informed of the experiment via the academy's site and their participation in the study was voluntary (no compensation was provided). Confidentiality was maintained through the secured electronic record keeping system (computerized software program) in which student

ID numbers were used for flight evaluation purposes. Pilot candidates participating in the research were from different countries (Greeks including) and they were informed by email which provided a description of the study and the requirements of the participation (Information Statement in Appendix A). They were asked to agree in the participation by signing their consent in Demographic Questionnaire (Appendix B).

2.2 Research questions and hypotheses

The following research questions and null and alternate hypotheses were used to analyze the predictors of successful pilot performance and therefore predict a pilot's profile of a better transition from analogue to digital displays:

a. To what extent is there a relationship between performance (response times) in a flight program and trainee's cognitive abilities score of Compass battery test?

H1o. There is no significant correlation between any of the cognitive and sensomotor measure scores of Compass test and performance scores (better response times).

H1. There are significant correlations between cognitive and sensomotor measure scores of Compass test and performance scores (better response times).

b. To what extent is there a relationship between the highly motivated in technology pilots (High and Very High ATI participants) and their better scores in any subscale of the Compass test battery (i.e., spatial orientation, memory)?

H2o. There is no significant correlation between high and very high scores in Affinity for Interaction Scale and better scores in Compass measurements.

H2. There is significant correlation between high and very high scores in Affinity for Interaction Scale and better scores in Compass measurements.

c. To what extent is there a relationship between pilot's scores in Affinity for Technology Interaction Scale and overall workload?

H3o. There is no significant correlation between high and very high scores in Affinity for Interaction Scale and overall workload.

H3. There is significant correlation between high and very high scores in Affinity for Interaction Scale and overall workload.

d. To what extent is there a relationship between higher ATI Scale scores and better response times in performance?

H4o. There is no significant effect between higher scores in Affinity for Interaction Scale and response times.

H4. There is significant effect between higher scores in Affinity for Interaction Scale and response times.

2.3 Equipment

2.3.1 Demographic Questionnaire

A demographic questionnaire was used in this study that included six questions (i.e., age, country, gender, language background, pilot license, flight experience), as shown in Appendix B.

2.3.2 COMPASS cognitive battery test (Version 3.0, January 2013)

COMPASS is a battery test, hosted by European Pilot Selection and Training (EPST) that comprises seven tasks, with the first six only counting toward the total score. The norm EPST uses for cutting off is based on the total score of COMPASS, computed by the sum of the 6 basic tests: hand-eye coordination (Control and Slalom), Memory, Mathematics, Spatial Orientation and Task manager. Based on 15 years of experience, EPST has set the cut off at the total sum of 24 in general, (for Europe the total sum is 25). The total of the grades indicates the overall strength of the candidate's performance. A total of 24 indicate an average overall performance. A grade total of 24 or above combined with grades of 3 or better in all the tests indicates a general aptitude for typical piloting tasks. A student with a sum less than 24 is highly likely to fail as it is shown in the figure 2. When COMPASS is used as a solely selection tool, EPST sets the cut off at a total score of 32, which is a good indicator for a successful training of a student pilot. The maximum score obtainable on COMPASS is 42 while in each subtest the maximum score is 7. The six tasks are described below :

a. Control. This task involves tracking a needle on a dial with hand and foot control inputs. The task examines candidates' hand-eye-foot coordination and scan rate. An individual who has bad grades on this test, has difficulties in recognizing quickly enough to perform this sensomotor task.

b. Slalom. This task involves following a slalom path with stick input. The task assesses candidates' hand/eye coordination by tracking. An individual who has bad grades on this test, has difficulties in processing and reacting quickly enough to perform this sensomotor task.

c. Memory. This task involves memorizing numbers and categories. The task assesses candidates' short-term memory and ability to "chunk" information. If candidates' memory is not sufficient, pilots may experience difficulties in putting the information together to make decisions. They will have to put more effort and need more time to get the mental picture during the flight operation.

d. Mathematics (Math). This task involves solving arithmetic problems. The task assesses candidates' basic arithmetic ability and mental agility. Consequences of a lack of numerical ability can be that a candidate needs more time to set the required setting during the flight and has difficulties in getting an overview quickly.

e. Orientation (Spatial Orientation). In this task, the participant matches instrument readings with the corresponding relative position of an airplane. The task assesses candidates' ability to read directional instruments, the speed of comprehension and spatial orientation. Pilots who lack good orientation are slower in determining their position and therefore are slower in making decisions.

f. Task Management. This task involves two sub-tasks, namely, to update autopilot settings and to react to a periodical signal. The task assesses candidates' ability to manage and prioritize demands from an input task and monitoring task. It also assesses multi-tasking through diverting attention to two concurrent tasks. A candidate who scores well on the Task Manager test, is able to follow procedures, and also puts a demand on the cognitive abilities to process information quickly and acts upon it.

The tests are thought to have high face validity, primarily because several exercises employed in COMPASS are based on tasks pilot typically perform (e.g., navigation, memory for material to be entered in the flight management system). A recent search of the European Pilot Selection and Training (EPST) website reveals the 'total validity of the COMPASS test to be .761′ (European Pilot Selection & Training, 2021).

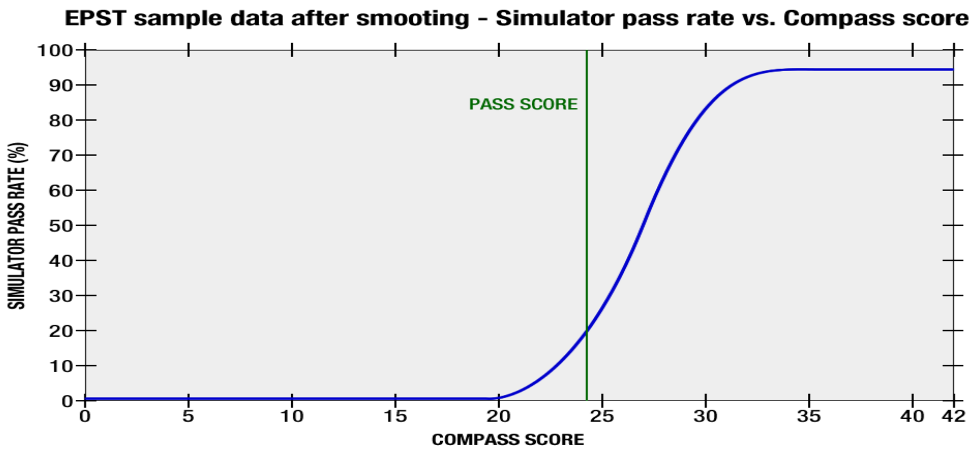


Figure 4. Pass rate compass score. (*EPST, Compass Version 3 User's manual*)

2.3.3 Affinity for technology interaction Scale (ATI) (Franke, Attig, & Wessel, 2018)

Given the importance of affinity for technology interaction during the transition from analogue to digital displays the 9-item ATI-scale (Appendix F) was administered to all participants, to investigate diverse facets of an active cognitive engagement in technology interaction (i.e., exploring and testing functions, devoting time, occupying oneself in greater detail, trying to understand systems, utilizing system capabilities). Responses were given on a 6-point scale (completely disagree, largely disagree, slightly disagree, slightly agree, largely agree, completely agree). We used the Questionnaire Scale at the address https://ati-scale.org/. The 9-item affinity for technology interaction (ATI) scale is designed to assess a person's tendency to actively engage in intensive technology interaction or to avoid it. ATI can be seen as a core personal resource for users' successful coping with technology. Studies examining ATI have shown that the scale is applicable in highly heterogeneous populations and have found higher ATI to be related to higher intrinsic motivation for technical device usage, and lower subjective workload while interacting with new technical devices (Wessel., Attig, Franke, 2019). ATI scale can differentiate between higher- and lower-ATI participants and there are no marked floor or ceiling effects. «Average» ATI varies between populations. Groups which are self-selected for their interest in technology (e.g., computer scientists) will have higher ATI values, so a person might be below average in the sample but above average in the population. Below there are the results of ATI's Cronbach's Alpha in other studies.

Cronbach's Alpha studies results

Research	Cronbach's Alpha	N of Items	Mean
Usage motives in interaction with activity trackers	,94	58	4,28
ATI construct validity study	.88	300	4,14

2.3.4 Pilot Performance Questionnaire (PPQ)

It is a type of rating which involves using independent, knowledgeable observer to rate the quality of a participant's situation awareness (i.e, instructor). Efforts were made to limit potential bias due to subjective measurement by employing only one instructor for the assessment. It is indicated that for some data collection tasks human observations may be robust and free from bias. This external observer's report (PPQ) can be interpreted as objective in contrast to the acknowledged subjectivity of self-report. It is common practice to use human expert ratings as objective data as Waag, Eddowes, Fuller, and Fuller (1975) reported a high degree of correlation between observer ratings and objective performance measures in standard flight maneuvers. It is recommended that observer ratings may be used as performance criteria in the development and validation of automated performance measures (Kelly et al, 1979 ; Stiffler, 1987). Observer ratings have also been used extensively in assessing crew awareness (Stout, Carson, and Salas, 1991; Brannick, Prince, Prince and Salas, 1992).

The Pilot Performance Questionnaire (PPQ, Appendix C) was structured and used to measure the participant's Situation Awareness (SA) score. Measures were recorded in this form, addressing the evaluations of the pilot's performances in a flight simulator. This Real-Time probe technique which was applied 'in-the-field' was preferred as it reduced the level of intrusion imposed by task freezes in the freeze-probe techniques of SA measurement and assessment methods that have been used in previous studies (Nguyen et al., 2019). Its main advantage is that it has no impact on the task being executed.

2.3.5 Flight simulators

The flight simulators that were used are located at the academy's facilities in Pachi Megaron. The academy has a total of 3 flight simulators on its premises. To carry out the research, the following two simulator models have been selected, in collaboration with the academy:

Elite Evolution S923 FNPT II (with the analogue cockpit). The Elite Evolution S923 FNPT II MCC simulator (119-60118-C-1EX) can simulate two types of aircraft, a twin-engine aircraft based on the Piper-Seneca III PA-34-220T and a twin-engine aircraft based on the Beech King Air B200.

Figure 5. Analogue display cockpit in simulator

Simnest A320 FNPT II MCC (with the digital cockpit). The MCN FNPT II MCS simulator is based on the Airbus A320, the aircraft used by many airlines worldwide including the domestic Aegean Airlines, which is a partner of the academy. The simulator introduces Fly-by-wire logic and ECAM systems while training prospective pilots to operate in a multi-Crew environment. In the A320 simulator, Advanced MCC (APS-Airline Pilot Standards), PBN (Procedure Based Navigation) training and part of the Instrument Rating (IR) in a multi-pilot environment are carried out, which are of particular interest to airlines.

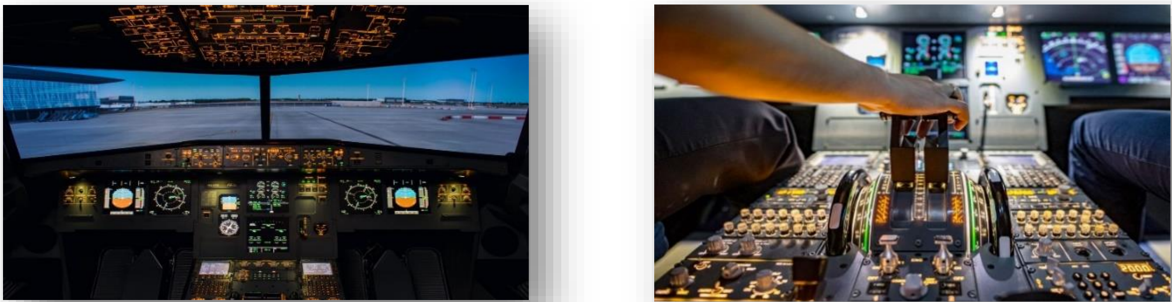


Figure 6. Digital display cockpit in simulator

2.3.6 Stopwatch

A stopwatch capable of measuring centiseconds was used to ensure accuracy in recorded timings.

2.3.7 Workload measures

Several arguments can be made for the usefulness of subjective rating techniques. According to some researchers, operator ratings are the most direct indicators of operator workload (Sheridan, 1980). That gives the approach more validity. Operator ratings are among the least intrusive of all techniques because they can be administered after the task is completed without disturbing the operator during task performance. The subjective techniques are flexible and portable; no equipment or special data collection devices are needed. In our study perceptions of workload were measured with the widely used NASA-TLX scale (Hart and Staveland, 1988).

2.3.8 NASA-TLX (Task Load Index): workload questionnaire (Hart and Staveland, 1988)

In 1988, Sandra G. Hart of NASA's Human Performance Group and Lowell E. Staveland of San Jose State University introduced the Task Load Index. With more than 8,000 citations since 1988, the NASA-TLX is applicable to several domains (air traffic control, civilian and military cockpits, robotics, and unmanned vehicles). In later years, studies in the automotive, healthcare, and technology domains used the TLX (Hart, 2006). The NASA-TLX is a multidimensional rating procedure that assesses a participant's subjective workload on six 100-point scales related to a different aspect of workload. It allows the determination of the subjective mental workload of a participant while he/she is performing a task. It rates performance across six dimensions to determine an overall workload rating. The six dimensions are as follows :

Mental demand (MD). How much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?

Physical demand (PD). How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?

Temporal demand (TD). How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?

Effort (EF). How hard did you have to work (mentally and physically) to accomplish your level of performance?

Performance (PE). How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?

Frustration (FR). How insecure, discouraged, irritated, stressed, and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task? (NASA Task Load Index, p.13).

Hart, in 1988, showed that the six dimensions correlate with each other. Other researchers have also found the subscales are significantly correlated with each other and Hart generally thinks the items "*are all measuring some aspect of the same underlying entity*." The most common modification made to NASA-TLX has been to eliminate the weighting process all together, which reduces the amount of time needed to administer the TLX and analyze the raw TLX responses. The former has been referred to as Raw TLX (RTLX) and has gained some popularity because it is simpler to apply; the ratings are simply averaged or added to create an estimate of overall workload. In the 29 studies in which

RTLX was compared to the original version, it was found to be either more sensitive (Hendy, Hamilton, & Landry, 1993), less sensitive (Liu & Wickens, 1994), or equally sensitive (Byers, Bittner, Hill, 1989). The NASA-TLX and its sub-scales sufficiently represent sources of cognitive workload among different tasks. Workload, like usability, is a complex construct but essentially means the amount of effort people must exert both mentally and physically to use the interface. Hart and her colleagues make a compelling case that the perception of workload may be a better measure than trying to find an objective measure of workload (such as heart rate) that may vary too much based on the nature of the task. Not only did Hart and Staveland validate their measure in their 1988 paper, but independent studies also found the TLX to be a valid measure of subjective workload (Hart & Staveland, 1988; Rubio, et al., 2004; Xiao, et al., 2005). The underlying assumption is that the combination of these 6 dimensions is likely to represent "workload" (Overall workload – OW) experienced by operators (Hart, 2006). Each response scale is essentially a line with 21 marks (Appendix D). To score TLX scale, someone can count the number of lines a participant marked, subtract 1, and multiply by 5. The overall workload estimated by RTLX is attractive for obvious reasons that it is as simple as combining the scores of each sub-scale. In other words, no calculations besides a plain sum are necessary (Hart, 2006). In our study at the conclusion of every test flight a RTLX scale was given to all participants to measure perceived workload. Prior to completing the RTLX scale, participants were instructed to ignore any of the effects the secondary tasks may have on their flight experience. Ratings for each subscale were summed and averaged to provide an overall workload score.

2.3.9 Voice Recorder

It is necessary to use a special application to record voice, in correlation with the time, thus giving the possibility of verifying the time, during the answers, with great accuracy.

Samsung Voice Recorder is designed to provide an easy recording experience with high quality sound, while also offering playback and editing capabilities.

2.4 Procedure

Each subject was given a detailed explanation of the study (Information Statement in Appendix A) before he or she participated in the experiment. She or he was provided with a Demographic Questionnaire (Appendix B) to complete and confirm his or her agreement to participate in the study. All participants were given the COMPASS Battery Test which was administered in flying operations unit of Global Aviation Academy. The COMPASS test took approximately 90 minutes. After their evaluation with the COMPASS test, they were all provided with the flight plan, maps, airport diagrams, frequencies, checklists, and other information required to complete the flight project. During the simulator sessions recording of times were performed by an experienced flight instructor (more than 2967hr of flight experience) who took the measures in each stage and phase of the experiment and recorded them at the Pilot Performance Questionnaire (PPQ in Appendix C). At the end of the simulation flights participants were asked to fill out the TLX scale Questionnaire (Appendix D).

The simulation flight was conducted in day visual flight rules (VFR) condition (i.e., visibility had to be greater than five nautical miles), in a flight simulator. Due to safety and resource (cost, aircraft, pilots, etc.) concerns, having aircraft deployed primarily for research purposes is uncommon and data collections normally to be performed during flight operations planned for other purposes (Wilson, 2002a). Gaining access to simulators for research is often as difficult. According to Salas, Bower, and Rhodenizer (1998), "*Simulators are typically booked for training and practice continuously, not leaving time for research or other experimental purposes*" (p. 201). For use of any recording equipment the primary criteria are that it is nonintrusive and in no way interferes with safety or pilot performance (Wilson, 2001). For basic training even a remote risk of recording equipment interference can render student pilots and flight instructors reluctant to participate. This difficulty was encountered in our study resulted in its long time to be accomplished.

The flight route was in Attika's area. It was selected since all subjects were familiar with this airspace and, therefore, avoided any unwanted navigational challenges (e.g., navigation in an unfamiliar area). Participants in this study experienced the same environmental flight operation conditions while in training. Confounding variables in this instance were within the normal operating range of flight operations for this training environment and considered to have minimal impact on the results. The average time taken to complete a flight was approximately 30 minutes. The procedure of the experiment was completed in two stages, each of them involving two phases. The 2nd Stage of the experiment gave predictive validity (criterion validity) in our study as predictive validity refers to a relationship between test scores and a measure of performance at some later time.

2.4.1 Stage 1

At the first stage the Initial Training Group (group IT) with no flying experience in a flying simulator or aircraft, who were in the stage of attending theoretical courses answered the Affinity for Technology Interaction (ATI) questionnaire in a computer. The process took place in one of the academy's rooms. After evaluating all the answers of the ATI test, two teams we assumed would emerge:

a. Technology-oriented group (ATI – High), which will be called for standardization purposes "IT-High TECH" (Initial Training high Tech).

b. Technology-oriented group (ATI – Very High), which will be called for standardization purposes "IT-very high TECH" (Initial Training very high Tech).

Both teams were invited to the 1st stage of the research, which was conducted in two phases. During the 1st phase the participants participated in a flight at the simulator with analogue instruments, while in the 2nd phase, they participated in a flight at the simulator with the digital instruments.

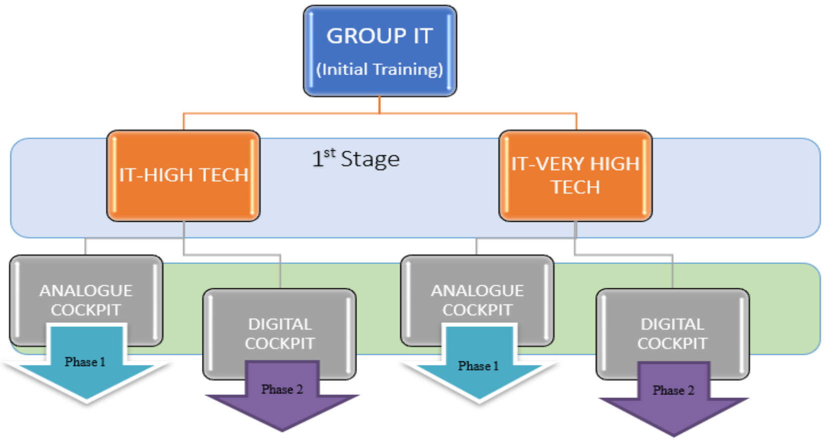


Figure 7. Diagram of 1st stage of the experiment.

Phase 1. At this stage both groups (IT-HIGH TECH and IT-VERY HIGH TECH) had the opportunity to participate in a flight as pilots not flying (PNF – Pilot Non-Flying) in the flight simulator with analogue instruments. The flight was conducted by the flight instructor (PF – Pilot Flying) and participants performed PNF duties. During the flight, the observer was present in a special area on the cockpit and recorded the responses of the participants. The process was performed as below:

a. Training was carried out by a flight instructor, who had to explain in detail the operation of the analogue cockpit of the simulator (position of all flight instruments) and the procedure of the study. Participants then waited in a waiting room until they entered the simulator.

b. Both teams participated in a flight on a fixed route (flying in the traffic pattern - Appendix E). At certain points - four distinct flight segments- (Climbing Leg/Descending Leg/Level Flight Leg/Level Flight Leg 360Turn), the flight instructor asked the same questions to all participants and the observer recorded the answers (correct answer – accurracy) and the time required to answer in the evaluation form PPQ (Pilot Performance Questionnaire-Appendix C). Each participant was asked three questions at a different distinct point in the route (S1, S2, S3 and S4) : about the Airspeed, the Climb rate, the Heading at Climbing Leg – S1, about the Airspeed, the Descent rate, the Heading at Descenting Leg - S2, about the Airspeed, the Altitude, the Heading at Level Flight Leg - S3, about the Airspeed, the Altitude, the Angle of Bank at Level Flight Leg 360 Turn – S4. Participants' performance was defined as the correct answer at the time given (response time began at the end of the question given from the flight instructor until the beginning of trainee's answer). Response times were classified as speed (reaction time), and accuracy - correct items (Kay, 1995). A common model to use was the Simple Reaction Time (SRT) model, which predicts response time in simple, single-stimulus reaction tasks. The Simple Reaction Time (SRT) model (Wundt, 1873; Johnson et al., 1985) is a basic model used to predict the time it takes for an individual to respond (correctly) to a simple, single stimulus. The model assumes that the time between the presentation of the stimulus and the initiation of the response is primarily determined by the time required to process the stimulus and make a response, and that the time taken to make the response is relatively constant. This model is commonly used in experimental psychology to study the basic processes involved in perception and reaction. The SRT model provides a baseline measure of an individual's basic processing speed and can be used to investigate the effects of various factors (such as the independent variables of this study) on reaction time. From a methodological perspective, it has to be mentioned that the test scenarios were presented in an ascending order. This was done to increase the task difficulty step by step.

At the conclusion of test flight (1st phase), a TLX scale questionnaire was given to measure participants' perceived workload. Participants rated each task performed on each of the six subscales. After that, each participant moved away from the simulator area to avoid communicating with other participants.

Phase 2. The process was repeated next day in the same way, with the only difference being the examination of flight participants in the simulator with a digital cockpit.

2.4.2 Stage 2

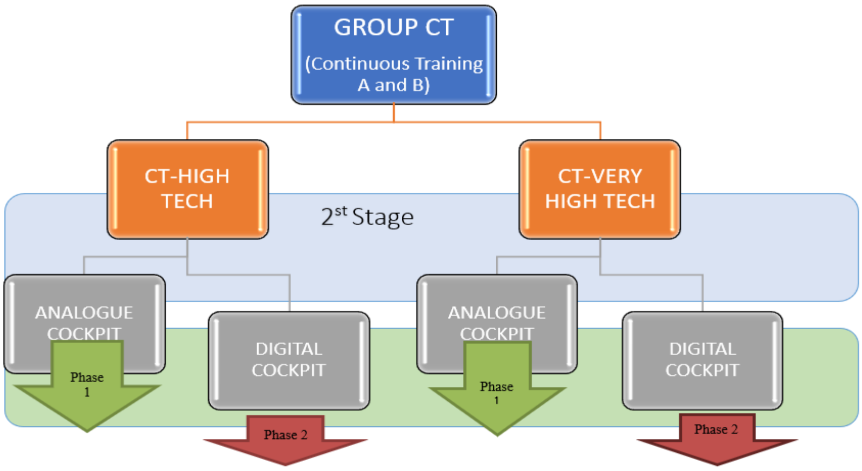


Figure 8. Diagram of 2nd stage of the experiment.

At the second stage of the experiment the Continuous Training Group (group CT), who have completed analogue flight simulator training or both analogue and digital cockpit simulator training, answered the Affinity for Technology Interaction (ATI) questionnaire in a computer. The evaluation of their answers to the ATI test was in accordance with the initial training group. There were only High and Very High ATI participants, consequently two teams emerged:

a. Technology-oriented group (ATI – High), which will be called for standardization purposes "CT- HIGH TECH" (Continuous Training high Tech).

b. Technology-oriented group (ATI – Very High), which will be called for standardization purposes "CT-VERY HIGH TECH" (Continuous Training very high Tech).

The 2nd stage of the research was conducted again in two phases. During the 1st phase the participants participated in a flight in the flight simulator with analogue instruments, while in the 2nd phase, they participated in a flight in the simulator with the digital flight instruments. At this stage both groups (CT- HIGH TECH and CT-VERY HIGH TECH) had the opportunity to participate in a flight as pilots flying (PF – Pilot Flying) in the flight simulator, while the flight instructor was the pilot non-flying (PNF – Pilot Non-Flying). During the flight, the observer was present in a special area at the cockpit and recorded the responses of the participants. The process was performed as described in the 1st stage. Accordingly each participant was asked three questions at a different distinct point in the route (S1’, S2’, S3’ and S4’) : about the Airspeed, the Climb rate, the Heading at Climbing Leg – S1’, about the Airspeed, the Descent rate, the Heading at Descenting Leg - S2’, about the Airspeed, the Altitude, the Heading at Level Flight Leg - S3’, about the Airspeed, the Altitude, the Angle of Bank at Level Flight Leg 360 Turn – S4’. Participants' performance was defined as the correct answer at the time given (response time began at the end of the question given from the flight instructor till the beginning of trainee’s answer). At the conclusion of test flight (2st phase), a TLX scale questionnaire was given to measure participants’ perceived workload.

Data was tabulated into a PC-based spreadsheet program, Microsoft Excel. For this study, only flight instruments scanned inside the plane were examined. The primary flight instruments in the cockpit: the airspeed indicator, attitude indicator, altitude indicator, heading indicator, vertical speed indicator, and turn and bank indicator are different between an analogue cockpit and a digital cockpit. The full flight route was divided into four different levels: climb, descent, cruise and cruise with 360 turns. The climb level started immediately after the take-off phase and included the plane climbing to the assigned cruising altitude. The cruise phase started once the plane was at the assigned cruising altitude and navigating towards the destination era. The descent phase started as soon as the plane reduced power and began its approach into the destination era.

Workload data were collected using the subjective questionnaire NASA TLX. The workload data were compared between Initial and Continuous Training groups to an analogue and a digital cockpit and were correlated with response times in performance.

Results

This chapter includes the descriptive results of sample demographic data, a report of the testing of statistical assumptions, and the measures used to test the four hypotheses that were investigated, to describe the results of the study.

Independent-sample /-tests were used to determine relationships between the group means and specifically the Mann-Whitney test (Mann &Whitney, 1947), which is the equivalent of the independent t-test. A Spearman's rho correlation was also used to examine the relationships of the hypotheses of successful flight performance (response times). Nonparametric testing was used to minimize the effects of extreme scores that may occur because of the small sample in the study. By using the ranks of non-parametric tests, we eliminated the effect of outliers. It must be mentioned that in our data the sampling distribution was normal, meaning that Type I error rate of tests based on this distribution is indeed 5% (Field, 2013). Thus, it was possible to define the power of the test. Using Mann-Whitney test, correlation, and regression statistical designs is consistent with previous aviation research on the relationship between criterion predictors, cognitive factors, and flight performance (Callister, 1996; Kole, 2006; Lehenbauer, 2003; Olson, 2002; Taylor, et al., 2000). Correlations were used to provide prediction of two or more variables; however, causality in a correlation design cannot be inferred. Pedhaszur and Schmelkin (1991) indicated that a correlation design is useful to analyze relationships of potential predictor variables with the absence of manipulation and randomization. Multiple correlations were used to observe and examine the relationships of several variables, or scores between two or more variables. The correlational procedures were used to provide the linear relationship of the direction, magnitudes, and strengths among the variables (Heiman, 2002; Moore, 2003). Furthermore, regression analyses were used (there isn’t any suitable non-parametric analysis) to investigate the predictor significance and variances of the predictors (i.e., independent variables) and the dependent criterion variable, accordingly to the normal distribution. Simple linear regression analyses were used to determine the level of effect between the independent variables and the dependent variable within the measures and to strengthen the null hypotheses testing. Researchers frequently use regression analyses if there is probability of multiple variable predictors of a behavior (Heiman, 2002). The assumption of homogeneity of variance between the groups was examined with a Levene's test (Levene, 1960). An assumption of homogeneity of variance indicates that variability of scores for both groups was analogous (Pallant, 2001). Nevertheless, it won't be presented here as it was not necessary in statistical analysis with non-parametric statistics which were used to exclude the outliers (small sample). Assumptions related to regressions consist of collinearity, singularity, normal distribution, linearity, and homoscedasticity (Tabachnick & Fidell, 2001).

3.1 Descriptive Statistics

The statistical processing was carried out with the help of the IBM SPSS (Statistical Package for the Social Sciences) which is a statistical software package developed by IBM Corporation. Due to the large number of variables of interest in this study, the statistical analysis was organized by hypotheses, including the descriptive statistics as well as correlations, regression analyses and moderation analyses.

3.1.1 Demographic Analyses

The demographic characteristics examined in the study were age, gender, country of origin, main language, secondary language, and flight experience.

Table 1

Demographics

		Group			
		Initial training		Continuous training	
		Count	Column N %	Count	Column N %
Country of Origin	Greece	3	42.9%	6	85.7%
	Other countries	4	57.1%	1	14.3%
Gender	Male	6	85.7%	7	100.0%
	Female	1	14.3%	0	0.0%
Main Language	Greek	3	42.9%	6	85.7%
	Dutch	4	57.1%	0	0.0%
	Spanish	0	0.0%	1	14.3%
Secondary Language	None	1	14.3%	0	0.0%
	English	6	85.7%	7	100.0%
Flight Experience	No	7	100.0%	0	0.0%
	Yes	0	0.0%	7	100.0%

3.1.2 Dependent Variables descriptive statistics

The variables response times which were used as dependent variables in the study are reported in groups of Initial and Continuous Training in an Analogue and Digital Instrumentation respectively. Each variable represents participant’s response time to answer PPQ ‘s questions, during flight simulation (12 variables for analogue instrumentation and 12 variables for digital instrumentation). The total response time variables in the 4 conditions of the flight experiment (Climbing leg, Level leg, 360 Turn, Descending Leg) in Analogue and Digital instrumentation were also measured as dependent variables. The data is presented in centi seconds.

Table 2

Analogue Both Groups

	Group							
	Initial training				Continuous training			
	Mean	Std Deviation	Min	Max	Mean	Std Deviation	Min	Max
Airspeed- Climbing leg – Analogue	352.14	165.37	52.00	556.00	149.71	33.55	96.00	191.00
Climbing Rate- Climbing leg - Analogue	405.71	145.05	232.00	654.00	124.57	41.85	72.00	175.00
Heading-Climbing leg - Analogue	301.00	67.45	185.00	392.00	142.00	24.09	114.00	186.00
Airspeed- Level flight leg - Analogue	310.71	70.25	199.00	380.00	123.29	58.79	66.00	227.00
Altitude- Level flight leg - Analogue	279.86	84.98	159.00	380.00	130.14	36.09	78.00	183.00
Heading- Level flight leg - Analogue	552.29	263.32	309.00	935.00	167.29	60.84	97.00	252.00
Airspeed- 360 turn level flight - Analogue	259.57	70.58	167.00	335.00	130.29	49.25	63.00	219.00
Altitude- 360 turn level flight - Analogue	333.14	140.76	192.00	552.00	119.14	32.34	69.00	146.00
Bank Angle- 360 turn level flight - Analogue	338.71	221.46	119.00	651.00	269.86	85.24	151.00	381.00
Airspeed- Descending leg - Analogue	274.43	134.50	164.00	551.00	124.29	56.26	63.00	197.00
Rate of Descent- Descending leg - Analogue	344.86	50.14	260.00	406.00	152.57	97.35	69.00	354.00
Heading- Descending leg - Analogue	438.14	142.26	299.00	665.00	163.86	56.16	93.00	252.00

Table 3

Digital Both Groups

	Group							
	Initial training				Continuous training			
	Mean	Std Deviation	Min	Max	Mean	Std Deviation	Min	Max
Airspeed- Climbing leg - Digital	179.86	78.41	120.00	351.00	125.57	96.17	43.00	332.00
Climbing Rate- Climbing leg - Digital	160.29	117.59	80.00	409.00	99.71	33.66	50.00	157.00
Heading-Climbing leg - Digital	157.86	67.78	71.00	279.00	252.86	208.82	83.00	598.00
Airspeed- Level flight leg - Digital	195.14	68.22	126.00	324.00	125.00	37.66	81.00	189.00

Altitude- Level flight leg - Digital	191.71	180.55	63.00	576.00	99.86	31.60	46.00	132.00
Heading- Level flight leg - Digital	239.43	87.11	111.00	383.00	155.00	39.83	96.00	229.00
Airspeed- 360 turn level flight - Digital	200.57	77.27	123.00	337.00	75.00	28.08	48.00	109.00
Altitude- 360 turn level flight - Digital	202.57	103.72	87.00	346.00	94.29	28.08	61.00	133.00
Bank Angle- 360 turn level flight - Digital	411.71	236.96	206.00	771.00	285.14	125.50	100.00	432.00
Airspeed- Descending leg - Digital	150.71	59.10	99.00	255.00	97.57	43.20	60.00	185.00
Rate of Descent- Descending leg - Digital	188.57	62.88	118.00	290.00	80.71	22.98	50.00	113.00
Heading- Descending leg - Digital	172.86	54.99	107.00	257.00	124.57	45.15	78.00	201.00

Table 4

Initial training - Analogue

	N		Mean	Median	Std. Deviation	Variance	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis	Min	Max
	Valid	Missing										
Total time Climbing leg - Analogue	7	0	1058.8571	1031.0000	306.17937	93745.810	-.051	.794	.065	1.587	575.00	1509.00
Total time Level leg - Analogue	7	0	1142.8571	1017.0000	272.21403	74100.476	1.087	.794	-.020	1.587	869.00	1617.00
Total time 360 turn - Analogue	7	0	931.4286	1095.0000	339.95777	115571.286	-.130	.794	-2.093	1.587	533.00	1376.00
Total time descending - Analogue	7	0	1057.4286	1036.0000	200.02905	40011.619	.329	.794	-.620	1.587	792.00	1363.00
Total Time All Tests - Analogue	7	0	4190.5714	4429.0000	860.43726	740352.286	.343	.794	-.984	1.587	3209.00	5551.00

Table 5

Continuous training - Analogue

	N		Mean	Median	Std. Deviation	Variance	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis	Min	Max
	Valid	Missing										

Total time Climbing leg - Analogue	7	0	416.2857	412.0000	50.11225	2511.238	.619	.794	.560	1.587	347.00	502.00
Total time Level leg - Analogue	7	0	420.7143	383.0000	111.32792	12393.905	1.363	.794	1.596	1.587	320.00	634.00
Total time 360 turn - Analogue	7	0	519.2857	577.0000	120.44738	14507.571	-.210	.794	-2.473	1.587	377.00	646.00
Total time descending - Analogue	7	0	440.7143	466.0000	182.95511	33472.571	.749	.794	.512	1.587	243.00	766.00
Total Time All Tests - Analogue	7	0	1797.0000	1745.0000	380.66608	144906.667	-.226	.794	-1.917	1.587	1318.00	2204.00

Table 6

Initial training - Digital

	N		Mean	Median	Std. Deviation	Variance	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis	Min	Max
	Valid	Missing										
Total time Climbing leg - Digital	7	0	498.0000	426.0000	196.27107	38522.333	2.086	.794	4.670	1.587	338.00	918.00
Total time Level leg - Digital	7	0	626.2857	525.0000	275.86211	76099.905	.958	.794	-.610	1.587	322.00	1065.00
Total time 360 turn - Digital	7	0	814.8571	651.0000	368.93786	136115.143	1.287	.794	-.184	1.587	542.00	1449.00
Total time descending - Digital	7	0	512.1429	506.0000	87.82260	7712.810	-.169	.794	-2.077	1.587	403.00	606.00
Total Time All Tests - Digital	7	0	2451.2857	2157.0000	829.01964	687273.571	1.252	.794	.338	1.587	1693.00	3935.00

Table 7

Continuous training - Digital

	N		Mean	Median	Std. Deviation	Variance	Skewness	Std. Error of Skewness	Kurtosis	Std. Error of Kurtosis	Min	Max
	Valid	Missing										
Total time Climbing leg - Digital	7	0	478.1429	376.0000	235.52596	55472.476	.262	.794	-2.597	1.587	244.00	742.00

Total time Level leg- Digital	7	0	379.8571	356.0000	81.72602	6679.143	.354	.794	-.060	1.587	263.00	510.00
Total time 360 turn - Digital	7	0	454.4286	435.0000	120.63424	14552.619	-.096	.794	-.427	1.587	267.00	625.00
Total time descending - Digital	7	0	302.8571	265.0000	89.42302	7996.476	.631	.794	-1.418	1.587	217.00	440.00
Total Time All Tests - Digital	7	0	1615.2857	1589.0000	247.47034	61241.571	-1.088	.794	1.877	1.587	1144.00	1912.00

3.1.3 Independent Variables descriptive statistics

Following presented are all the independent variables descriptive statistics in Initial and Continuous groups in analogue and digital instrumentation.

Table 8

Workload Analogue

	Group							
	Initial training				Continuous training			
	Mean	Standard Deviation	Min	Max	Mean	Standard Deviation	Min	Max
Mental Demand - Analogue	22.14	19.71	2.5	55.00	35.00	36.61	5.00	100.00
Physical Demand- Analogue	6.07	6.90	00	20.00	24.29	35.29	00	100.00
Temporal Demand- Analogue	15.71	13.67	5.00	35.00	25.71	18.13	00	50.00
Performance- Analogue	21.07	23.71	00	65.00	22.86	27.67	00	80.00
Effort- Analogue	25.00	16.58	00	45.00	39.29	30.20	00	80.00
Frustration- Analogue	3.21	4,72	00	10.00	7.14	6.99	00	20.00
Overall Workload - Analogue	15.54	11.58	3.30	33,80	25.71	10.83	14.20	44.20

Table 9

Workload Digital

	Group							
	Initial training				Continuous training			
	Mean	Standard Deviation	Min	Max	Mean	Standard Deviation	Min	Max
Mental Demand - Digital	11.79	13.75	.00	40.00	37.14	30.12	5.00	75.00
Physical Demand- Digital	2.14	3.93	.00	10.00	23.57	28.54	.00	70.00
Temporal Demand- Digital	3.93	5.18	.00	15.00	24.29	20.30	5.00	55.00
Performance- Digital	14.64	16.80	2.50	45.00	22.86	23.07	.00	60.00
Effort- Digital	9.29	10.38	2.50	30.00	49.29	23.17	15.00	80.00
Frustration- Digital	8.21	9.21	.00	25.00	13.57	10.29	5.00	30.00
Overall Workload - Digital	7.74	7.02	1.30	18.30	28.44	10.27	10.00	40.00

Table 10

Compass Test

	Group							
	Initial training				Continuous training			
	Mean	Standard Deviation	Min	Max	Mean	Standard Deviation	Min	Max
Control	4.57	2.30	2.00	7.00	3.29	1.38	1.00	5.00
Slalom	6.57	.53	6.00	7.00	6.29	.49	6.00	7.00
Memory	5.14	1.07	3.00	6.00	5.29	.95	4.00	7.00
Mathematic	3.14	1.77	1.00	5.00	2.71	.76	2.00	4.00
Orientation	3.14	1.57	1.00	5.00	2.29	1.11	1.00	4.00
Task Management	4.86	1.95	1.00	7.00	4.57	.53	4.00	5.00
Total Compass Score	27.43	6.73	18.00	36.00	24.43	2.57	20.00	27.00

Table 11

ATI Scale Scores

	Group							
	Initial training				Continuous training			
	Mean	Standard Deviation	Minimum	Maximum	Mean	Standard Deviation	Minimum	Maximum
ATI Scale Scores	4.78	.47	4.22	5.67	4.92	.39	4.33	5.56

3.2 Inductive Statistics

3.2.1 Assumptions made in the statistical analyses.

There were several assumptions that underlay the statistical analyses. The crucial assumptions of tests are primarily that the population data from the sample data are normally distributed (Choudhury, 2009).

3.2.2 Normality analyses

In order to choose the inductive statistical analysis, it was first investigated whether the variables under investigation follow the pattern of normal distribution or not. Nevertheless, we have had a presentation of the distribution of the variables through the histogram graphs in the section of descriptive statistics. In case the distribution of the variables can be considered normal, parametric tests such as Anova and Pearson correlations are indicated, while if their understanding is not normal, non-parametric tests such as Kruscal-Wallis, Mann-Whitney and Spearman Correlations should be used. To determine the presence or absence of non-normality, the Kolmogorov-Smirnov normality test was chosen. To consider the distribution of a variable as normal, the statistical significance test of the Kolmogorov-Smirnov test must give a p greater than 0.05 ($p>0.05$). The results are presented in the following tables in Initial training and Continuous Training groups respectively.

Table 12

Group = Initial training

One-Sample Kolmogorov-Smirnov Test[a]

	N	Normal Parameters[b,c]		Test Statistic	Exact Sig. (2-tailed)
		Mean	Std. Deviation		
ATI Scale Scores	7	4.7786	0.46870	0.213	0.849
Airspeed- Climbing leg - Analogue	7	352.1429	165.36972	0.175	0.957
Climbing Rate- Climbing leg - Analogue	7	405.7143	145.05254	0.129	0.999
Heading-Climbing leg - Analogue	7	301.0000	67.45122	0.155	0.985
Airspeed- Level flight leg - Analogue	7	310.7143	70.25362	0.249	0.695
Altitude- Level flight leg - Analogue	7	279.8571	84.97535	0.214	0.846
Heading- Level flight leg - Analogue	7	552.2857	263.31712	0.374	0.219
Airspeed- 360 turn level flight - Analogue	7	259.5714	70.57822	0.259	0.647
Altitude- 360 turn level flight - Analogue	7	333.1429	140.75561	0.222	0.814
Bank Angle- 360 turn level flight - Analogue	7	338.7143	221.46385	0.232	0.770
Airspeed- Descending leg - Analogue	7	274.4286	134.50385	0.330	0.354
Rate of Descent- Descending leg - Analogue	7	344.8571	50.14455	0.150	0.990
Heading- Descending leg - Analogue	7	438.1429	142.25848	0.219	0.824
Total time Climbing leg - Analogue	7	1058.8571	306.17937	0.155	0.986
Total time Level leg - Analogue	7	1142.8571	272.21403	0.250	0.691
Total time 360 turn - Analogue	7	931.4286	339.95777	0.256	0.659
Total time descending - Analogue	7	1057.4286	200.02905	0.187	0.931
Total Time All Tests - Analogue	7	4190.5714	860.43726	0.200	0.894
Airspeed- Climbing leg - Digital	7	179.8571	78.40797	0.321	0.385
Climbing Rate- Climbing leg - Digital	7	160.2857	117.58928	0.377	0.211
Heading-Climbing leg - Digital	7	157.8571	67.78011	0.166	0.972

Airspeed- Level flight leg - Digital	7	195.1429	68.22372	0.193	0.915
Altitude- Level flight leg - Digital	7	191.7143	180.54521	0.344	0.305
Heading- Level flight leg - Digital	7	239.4286	87.10885	0.158	0.983
Airspeed- 360 turn level flight - Digital	7	200.5714	77.27194	0.253	0.673
Altitude- 360 turn level flight - Digital	7	202.5714	103.72216	0.245	0.713
Bank Angle- 360 turn level flight- Digital	7	411.7143	236.96252	0.346	0.298
Airspeed- Descending leg - Digital	7	150.7143	59.09516	0.260	0.644
Rate of Descent- Descending leg - Digital	7	188.5714	62.87516	0.175	0.957
Heading- Descending leg - Digital	7	172.8571	54.99221	0.155	0.985
Total time Climbing leg - Digital	7	498.0000	196.27107	0.285	0.528
Total time Level leg - Digital	7	626.2857	275.86211	0.353	0.275
Total time 360 turn - Digital	7	814.8571	368.93786	0.385	0.192
Total time descending - Digital	7	512.1429	87.82260	0.235	0.759
Total Time All Tests - Digital	7	2451.2857	829.01964	0.350	0.287
Mental Demand - Analogue	7	22.1429	19.70769	0.160	0.980
Physical Demand - Analogue	7	6.0714	6.90066	0.276	0.568
Temporal Demand - Analogue	7	15.7143	13.67131	0.235	0.757
Performance- Analogue	7	21.0714	23.71081	0.251	0.683
Effort- Analogue	7	25.0000	16.58312	0.214	0.844
Frustration- Analogue	7	3.2143	4.72456	0.323	0.376
Overall Workload - Analogue	7	15.5429	11.58258	0.185	0.937
Mental Demand - Digital	7	11.7857	13.74729	0.265	0.620
Physical Demand- Digital	7	2.1429	3.93398	0.421	0.122
Temporal Demand- Digital	7	3.9286	5.17549	0.323	0.377
Performance- Digital	7	14.6429	16.79711	0.288	0.514
Effort- Digital	7	9.2857	10.37970	0.315	0.408
Frustration- Digital	7	8.2143	9.20985	0.208	0.868
Overall Workload - Digital	7	7.7429	7.01946	0.265	0.620
Age	7	22.43	4.860	0.330	0.353
Flight Experience Duration (Hours)	7	0.0000	.00000[f]		
Control	7	4.5714	2.29907	0.324	0.373
Slalom	7	6.5714	0.53452	0.360	0.256
Memory	7	5.1429	1.06904	0.304	0.450
Mathematic	7	3.1429	1.77281	0.257	0.655
Orientation	7	3.1429	1.57359	0.278	0.557
Task Management	7	4.8571	1.95180	0.243	0.719
Total Compass Score	7	27.4286	6.72947	0.131	0.998

a. Group = Initial training

b. Test distribution is Normal.

Table 13

Group = Continuous training

One-Sample Kolmogorov-Smirnov Test[a]

	N	Normal Parameters[b,c]		Test Statistic	Exact Sig. (2-tailed)
		Mean	Std. Deviation		
ATI Scale Scores	7	4.9186	0.39134	0.210	0.861
Airspeed- Climbing leg - Analogue	7	149.7143	33.55450	0.158	0.982
Climbing Rate- Climbing leg - Analogue	7	124.5714	41.84837	0.230	0.780
Heading-Climbing leg - Analogue	7	142.0000	24.09011	0.280	0.549
Airspeed- Level flight leg - Analogue	7	123.2857	58.78694	0.238	0.744
Altitude- Level flight leg - Analogue	7	130.1429	36.08984	0.147	0.992
Heading- Level flight leg - Analogue	7	167.2857	60.83780	0.236	0.754
Airspeed- 360 turn level flight - Analogue	7	130.2857	49.25347	0.172	0.963
Altitude- 360 turn level flight - Analogue	7	119.1429	32.34413	0.298	0.473
Bank Angle- 360 turn level flight - Analogue	7	269.8571	85.23581	0.153	0.987
Airspeed- Descending leg - Analogue	7	124.2857	56.25749	0.257	0.657
Rate of Descent- Descending leg - Analogue	7	152.5714	97.34793	0.241	0.729
Heading- Descending leg - Analogue	7	163.8571	56.16472	0.161	0.979
Total time Climbing leg - Analogue	7	416.2857	50.11225	0.240	0.733
Total time Level leg - Analogue	7	420.7143	111.32792	0.253	0.676
Total time 360 turn - Analogue	7	519.2857	120.44738	0.257	0.655
Total time descending - Analogue	7	440.7143	182.95511	0.174	0.960
Total Time All Tests - Analogue	7	1797.0000	380.66608	0.224	0.806
Airspeed- Climbing leg - Digital	7	125.5714	96.17494	0.298	0.476
Climbing Rate- Climbing leg - Digital	7	99.7143	33.65865	0.151	0.989
Heading-Climbing leg - Digital	7	252.8571	208.81924	0.303	0.454
Airspeed- Level flight leg - Digital	7	125.0000	37.65634	0.175	0.957
Altitude- Level flight leg - Digital	7	99.8571	31.59867	0.254	0.670
Heading- Level flight leg - Digital	7	155.0000	39.83298	0.229	0.783
Airspeed- 360 turn level flight - Digital	7	75.0000	28.07727	0.299	0.470
Altitude- 360 turn level flight - Digital	7	94.2857	28.08151	0.204	0.880
Bank Angle- 360 turn level flight - Digital	7	285.1429	125.50489	0.204	0.882
Airspeed- Descending leg - Digital	7	97.5714	43.20053	0.221	0.818
Rate of Descent- Descending leg - Digital	7	80.7143	22.97618	0.227	0.794
Heading- Descending leg - Digital	7	124.5714	45.15476	0.213	0.848
Total time Climbing leg - Digital	7	478.1429	235.52596	0.250	0.689

Total time Level leg - Digital	7	379.8571	81.72602	0.186	0.933
Total time 360 turn - Digital	7	454.4286	120.63424	0.151	0.989
Total time descending - Digital	7	302.8571	89.42302	0.235	0.756
Total Time All Tests - Digital	7	1615.2857	247.47034	0.266	0.615
Mental Demand - Analogue	7	35.0000	36.51484	0.279	0.552
Physical Demand - Analogue	7	24.2857	35.28793	0.318	0.396
Temporal Demand - Analogue	7	25.7143	18.12654	0.195	0.909
Performance- Analogue	7	22.8571	27.66724	0.255	0.664
Effort- Analogue	7	39.2857	30.19776	0.218	0.830
Frustration- Analogue	7	7.1429	6.98638	0.198	0.899
Overall Workload - Analogue	7	25.7143	10.83027	0.144	0.993
Mental Demand - Digital	7	37.1429	30.11881	0.228	0.788
Physical Demand- Digital	7	23.5714	28.53569	0.224	0.804
Temporal Demand- Digital	7	24.2857	20.29544	0.298	0.475
Performance- Digital	7	22.8571	23.06822	0.205	0.879
Effort- Digital	7	49.2857	23.17121	0.160	0.980
Frustration- Digital	7	13.5714	10.29332	0.226	0.796
Overall Workload - Digital	7	28.4429	10.26773	0.178	0.952
Age	7	25.14	3.532	0.198	0.900
Flight Experience Duration (Hours)	7	150.2857	35.16492	0.147	0.992
Control	7	3.2857	1.38013	0.269	0.600
Slalom	7	6.2857	0.48795	0.435	0.101
Memory	7	5.2857	0.95119	0.332	0.344
Mathematic	7	2.7143	0.75593	0.256	0.659
Orientation	7	2.2857	1.11270	0.173	0.962
Task Management	7	4.5714	0.53452	0.360	0.256
Total Compass Score	7	24.4286	2.57275	0.302	0.457

a. Group = Continuous training

b. Test distribution is Normal.

Consequently, all variables met the assumption of normality and were used with confidence in further grouped analyses where parametric inductive controls used for them.

3.3 Testing of Hypotheses

The following section contains the results of the study research questions and hypotheses.

3.3.1 Question and Hypothesis 1

Response times and Scores in Compass Battery Test. To what extent is there a relationship between performance (response times) in a flight program and trainee's cognitive abilities score of Compass battery test?

H1o. There is no significant correlation between any of the cognitive and sensomotor measure scores of Compass test and performance scores (better response times).

H1. There are significant correlations between cognitive and sensomotor measure scores of Compass test and performance scores (better response times).

During the study, the relationship between Compass test scores and response times (flight performance) among aviation university students in a flight program was examined. It was hypothesized that there would be a relationship between the variables. In this first question of the research the independent variables derived of Compass test scores were Control, Slalom, Memory, Mathematic, Orientation, Task Management, and Total Compass Score.

Correlation Analysis. Addressing the issue of small size sample of the study and although the normality assumption was met for all variables, it was recommended Spearman's rank correlation coefficient to be used. This is because it is a non-parametric test, and it is less sensitive to deviations from normality compared to Pearson's correlation. Given the non-normality distribution of some variables, the non-parametric Spearman's rho test was chosen for correlation analysis in the 2 groups of participants, Initial Training and Continuous

Training, respectively and was computed to assess the relationship between response time variables and Total Compass Score variables.

Of the variables analyzed, the following showed significant ($p<0.05$) and ($p<0,01$) correlations, in Initial Training Group.

There was a negative correlation between Climbing Rate- Climbing leg - Analogue, and Slalom variables, $r(5) = -.87$, $p = .012$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Initial Training participants with a higher score in Slalom task of Compass test had lower response time score in Climbing Rate- Climbing leg – Analogue.

There was a positive correlation between Altitude- Level flight leg - Analogue, and Memory variables, $r (5) = .77$, $p = .042$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that in Initial Training participants group higher score in Memory task of Compass test, relates with high response time score in Altitude- Level flight leg – Analogue and vice versa.

There was a negative correlation between, Heading- Level flight leg - Analogue and Task Management variables, $r (5) = -.87$, $p = .010$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Initial Training participants with a higher score in Task Management task of Compass test had lower response time score in Heading- Level flight leg – Analogue.

There was a positive correlation between, Airspeed- 360 turn level flight – Analogue and Mathematics variables, $r (5) = .90$, $p = .006$. This association was statistically significant with a margin of error of less than 1%. Based on the coding of the variables, this correlation demonstrates that in Initial Training participants group higher score in Mathematics task of

Compass test, relates with high response time score in Airspeed- 360 turn level flight – Analogue and vice versa.

There was a negative correlation between, Altitude- 360 turn level flight – Analogue and Slalom variables, $r(5) = -.87$, $p = .012$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Initial Training participants with a higher score in Slalom task of Compass test had lower response time score in Altitude- 360 turn level flight – Analogue.

There was a positive correlation between, Airspeed- Descending leg – Analogue and Slalom variables, $r(5) = .87$, $p = .010$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that in Initial Training participants group higher score in Slalom task of Compass test, relates with high response time score in Airspeed- Descending leg – Analogue and vice versa.

There was a positive correlation between, Heading- Descending leg – Analogue and Total Compass score variables, $r(5) = .86$, $p = .014$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that in Initial Training participants group higher score in Total Compass score, relates with high response time score in Heading- Descending leg – Analogue and vice versa.

There was a positive correlation between, Heading- Descending leg – Analogue and Mathematics variables, $r(5) = .97$, $p = .000$. This association was statistically significant with a margin of error of less than 1%. Based on the coding of the variables, this correlation demonstrates that in Initial Training participants group higher score in Mathematics relates with high response time score in Heading- Descending leg – Analogue and vice versa.

There was a positive correlation between, Heading- Descending leg – Analogue and Orientation variables, $r(5) = .81$, $p = .029$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that in Initial Training participants group higher score in Orientation, relates with high response time score in Heading- Descending leg – Analogue and vice versa.

There was a positive correlation between, Heading- Descending leg – Analogue and Task management variables, $r(5) = .76$, $p = .046$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that in Initial Training participants group higher score in Task Management, relates with high response time score in Heading- Descending leg – Analogue and vice versa.

There was a negative correlation between, Total time 360 turn – Analogue and Slalom variables, $r(5) = -.87$, $p = .012$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Initial Training participants with a higher score in Slalom task of Compass test had lower response time score in Total time 360 turn – Analogue.

There was a positive correlation between, Total time descending – Analogue and Memory variables, $r(5) = .77$, $p = .042$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that in Initial Training participants group higher score in Memory, relates with high response time score in Total time descending – Analogue and vice versa.

There was a negative correlation between, Total Time All Tests – Analogue and Slalom variables, $r(5) = -.87$, $p = .012$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Initial Training participants with a higher score in Slalom task of Compass test had lower response time score in Total Time All Tests – Analogue.

There was a negative correlation between, Climbing Rate- Climbing leg – Digital and Total Compass score variables, r(5) = -.78, p = .041. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Initial Training participants with a higher score in Total Compass score had lower response time score in Climbing Rate- Climbing leg – Digital.

There was a positive correlation between, Airspeed- Descending leg – Digital and Memory variables, r(5) = .81, p = .027. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that in Initial Training participants group higher score in Memory, relates with high response time score Airspeed- Descending leg – Digital and vice versa.

There was a negative correlation between, Total time Climbing leg – Digital and Total Compass score variables, r(5) = -.86, p = .014. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Initial Training participants with a higher score in Total Compass score had lower response time score in Total time Climbing leg – Digital.

There was a negative correlation between, Total time Climbing leg – Digital and Orientation variables, r(5) = -.81, p = .029. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Initial Training participants with a higher score in Orientation Task of Compass test had lower response time score in Total time Climbing leg – Digital.

There was a negative correlation between, Total time Climbing leg – Digital and Task management variables, r(5) = -.86, p = .014. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Initial Training participants with a higher score in Task management of Compass test had lower response time score in Total time Climbing leg – Digital.

There was a positive correlation between, Total time descending – Digital and Memory variables, r(5) = .93, p = .003. This association was statistically significant with a margin of error of less than 1%. Based on the coding of the variables, this correlation demonstrates that in Initial Training participants group higher score in Memory, relates with high response time score Total time descending – Digital and vice versa.

Spearman's rank correlation was computed to assess the relationship between response time variables and Total Compass Score variables in Continuous Training Group. Of the variables analyzed, the following showed significant ($p<0.05$) and ($p<0,01$) correlations.

There was a positive correlation between Climbing Rate- Climbing leg - Analogue, and Control variables, r(5) = .82, p = .025. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that in Continuous participants group higher score in Control, relates with high response time score Climbing Rate- Climbing leg – Analogue and vice versa.

There was a negative correlation between Heading-Climbing leg - Analogue, and Slalom variables, r(5) = -.79, p = .034. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Continuous Training participants with a higher score in Slalom task of Compass test had lower response time score in Heading-Climbing leg - Analogue.

There was a negative correlation between Heading- Level flight leg - Digital, and Mathematics variables, r(5) = -.85, p = .016. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Continuous Training participants with a higher score in Mathematics task of Compass test had lower response time score in Heading- Level flight leg – Digital.

There was a positive correlation between Airspeed- Descending leg - Digital, and Slalom variables, $r(5) = -79$, $p = .034$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that in Continuous Training participants group higher score in Slalom, relates with high response time score Airspeed- Descending leg – Digital and vice versa.

Consequently, answering to the first research question there were significant effects between scores of Compass test and response times. Most of them were detected in Initial Training group.

3.3.2 *Question and Hypothesis 2*

ATI Scale scores and Scores in Compass Battery Test. To what extent is there a relationship between the highly motivated in technology pilots (High and Very High ATI participants) and their better scores in any subscale of the Compass test battery (i.e., spatial orientation, memory)?

H2o. There is no significant correlation between high and very high scores in Affinity for Interaction Scale and better scores in Compass measurements.

H2. There is significant correlation between high and very high scores in Affinity for Interaction Scale and better scores in Compass measurements.

To test the second question of the research and the associated hypotheses, the Mann and Whitney (Mann & Whitney, 1947) was used. It is a non-parametric statistical test that is used to compare the distributions of two independent samples (with different entities) when the sample sizes are small or when the data is not normally distributed. This is a class difference test applied in cases where the sample includes two different groups, each member of which corresponds to an observation (Gnardellis, 2003; Dafermos, 2011). The group with a higher Mean Rank has higher scores on that variable than one with a lower one, this is often reflected in the mean and median except in cases of extreme non-normality. This test is the non-parametric equivalent of independent t-test and works by looking at differences in the ranked positions of scores in different groups. Mann–Whitney test relies on scores being ranked from lowest to highest; therefore, the group with the lowest mean rank is the group with the greatest number of lower scores in it. Similarly, the group that has the highest mean rank should have a greater number of high scores within it (Field, 2013). Mann-Whitney U test is used to compare differences between two independent groups when the dependent variable is either ordinal or continuous, but not normally distributed. To use Mann-Whitney U test data must pass four assumptions that are required to have valid results. Specifically, the four assumptions are:

Assumption 1. Dependent variable should be measured at the ordinal or continuous level (in this study Compass tests scores).

Assumption 2. Independent variable should consist of two categorical, independent groups (in this study ATI groups High and Very high).

Assumption 3. There should be independence of observations, which means that no relationship between the observations in each group or between the groups themselves is allowed (this study meets this assumption).

Assumption 4. A Mann-Whitney U test can be used when the variables are not normally distributed. However, to know how to interpret the results from a Mann-Whitney U test, it must be determined whether the distributions for both groups of the independent variable have the same shape. If they do have the same shape, the medians of the dependent variable may be compared. However, if the distributions have a different shape, only mean ranks of the test can be compared (in this study mean ranks were compared).

Following there are the results of Mann-Whitney first for all participants (sample *N*=14) and then presented in groups Initial Training (*n*=7) and Continuous Training (*n*=7) respectively.

Table 14

Mann-Whitney Test

Ranks Group ATI		N	Mean Rank	Sum of Ranks
Control	High	10	7.85	78.5
	Very High	4	6.63	26.5
	Total	14		
Slalom	High	10	7.3	73
	Very High	4	8	32
	Total	14		
Memory	High	10	7	70
	Very High	4	8.75	35
	Total	14		
Mathematic	High	10	7.45	74.5
	Very High	4	7.63	30.5
	Total	14		
Orientation	High	10	7.5	75
	Very High	4	7.5	30
	Total	14		
Task Management	High	10	8.1	81
	Very High	4	6	24
	Total	14		
Total Compass Score	High	10	7.55	75.5
	Very High	4	7.38	29.5
	Total	14		

Table 15

Test Statistics for Group ATI

	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (2-tailed)
Control	16.5	26.5	-0.506	0.662
Slalom	18	73	-0.329	1
Memory	15	70	-0.765	0.549
Mathematic	19.5	74.5	-0.072	0.999
Orientation	20	30	0	1
Task Management	14	24	-0.895	0.441
Total Compass Score	19.5	29.5	-0.071	0.972

From the table above, it seems like Very high ATI group participants scored higher in Slalom, Memory, and Mathematics compared to High ATI group participants. Nevertheless, based on the results, it can be concluded that there was no significant difference between the mean ranks of the "High" and "Very High" ATI groups for the six variables of Compass test, as all p-values were greater than .05.

Table 16

Mann-Whitney Test for Group Initial training

Group ATI		N	Mean Rank	Sum of Ranks
Control	High	5	3.8	19
	Very High	2	4.5	9
	Total	7		
Slalom	High	5	3.4	17
	Very High	2	5.5	11
	Total	7		
Memory	High	5	3.8	19
	Very High	2	4.5	9
	Total	7		
Mathematic	High	5	4	20
	Very High	2	4	8
	Total	7		
Orientation	High	5	4.3	21.5
	Very High	2	3.25	6.5
	Total	7		
Task Management	High	5	4.5	22.5
	Very High	2	2.75	5.5
	Total	7		
Total Compass Score	High	5	4	20
	Very High	2	4	8
	Total	7		

Table 17

Test Statistics for Group Initial training

	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (2-tailed)
Control	4	19	-0.418	1
Slalom	2	17	-1.342	0.429
Memory	4	19	-0.418	1
Mathematic	5	8	0	1
Orientation	3.5	6.5	-0.609	0.762
Task Management	2.5	5.5	-0.986	0.476
Total Compass Score	5	8	0	1

From the table above, it seems like Very high ATI Initial training group scored higher in Control, Slalom, and Memory compared to High ATI group participants and lower in Orientation and Task management. Nevertheless, based on the results, it can be concluded that there was no significant difference between the mean ranks of the "High" and "Very High" ATI groups for the all six variables of Compass test, as all p-values were greater than .05.

Table 18

Mann-Whitney Test for Group Continuous Training

Group ATI		N	Mean Rank	Sum of Ranks
Control	High	5	4.6	23
	Very High	2	2.5	5
	Total	7		
Slalom	High	5	4.4	22
	Very High	2	3	6
	Total	7		
Memory	High	5	3.7	18.5
	Very High	2	4.75	9.5
	Total	7		
Mathematic	High	5	3.8	19
	Very High	2	4.5	9
	Total	7		
Orientation	High	5	3.4	17
	Very High	2	5.5	11
	Total	7		
Task Management	High	5	4.1	20.5
	Very High	2	3.75	7.5
	Total	7		
Total Compass Score	High	5	3.9	19.5
	Very High	2	4.25	8.5
	Total	7		

Table 19

Test Statistics for Group Continuous training

Test statistics	Mann-Whitney U	Wilcoxon W	Z	Exact Sig. (2-tailed)
Control	2	5	-1.206	0.286
Slalom	3	6	-0.98	0.524
Memory	3.5	18.5	-0.641	0.619
Mathematic	4	19	-0.418	1

Orientation	2	17	-1.194	0.381
Task Management	4.5	7.5	-0.224	1
Total Compass Score	4.5	19.5	-0.203	0.857

From the table above, it seems like Very high ATI Initial training group scored higher in Memory, Mathematics, Orientation and Total Compass Score, compared to High ATI group participants and lower in Control, Slalom and Task management. Nevertheless, based on the results, it can be concluded that there was no significant difference between the mean ranks of the "High" and "Very High" ATI groups for the all six variables of Compass test, as all p-values were greater than 0.05.

Consequently, the alternative - experimental hypothesis was not confirmed, that is there is no significant effect between high and very high scores in Affinity for Interaction Scale and better scores in Compass scores.

3.3.3 *Question and Hypothesis 3*

ATI Scale scores and Workload scores. To what extent is there a relationship between pilot's scores in Affinity for Technology Interaction Scale and overall workload?

H3o. There is no significant correlation between high and very high scores in Affinity for Interaction Scale and overall workload.

H3. There is significant correlation between high and very high scores in Affinity for Interaction Scale and overall workload.

Correlation Analysis. Addressing the issue of small size sample of the study and although the normality assumption was met, it was recommended Spearman's rank correlation coefficient to be used for this correlation analysis.

Following there are the tables with the results as they appear first for all participants (sample N=14) and then presented in groups Initial Training (*n*=7) and Continuous Training (*n*=7), respectively.

Table 20

Correlations All Participants

Variable	N	M	SD	1	2	3	4
1.Group ATI	14	1,29	.469				
2.ATI Scale Scores	14	4.85	.421	.715**			
3.Overall Workload - Analogue	14	20.63	12.00	0.118	0.281		
4.Overall Workload - Digital	14	18.09	13.67	-0.157	0.043	0.291	

****. p< 0.01 level (2-tailed); N=14**

There was no significant correlation between Overall Workload Analogue and Overall Workload Digital, and between Group ATI and either Overall Workload Analogue or Overall Workload Digital.

Table 21

Correlations Group Initial Training

Variable	N	M	SD	1	2	3	4
1.Group ATI	7	1,29	.488				
2.ATI Scale Scores	7	4.78	.469	.798*			
3.Overall Workload - Analogue	7	15.54	11.58	0	0.342		
4. Overall Workload - Digital	7	7.74	7.02	-0.479	-0.591	-0.252	

***. P<0.05 level (2-tailed); n=7**

There was no significant correlation between Overall Workload Analogue and Overall Workload Digital, and between Group ATI and either Overall Workload Analogue or Overall Workload Digital.

Table 22

Correlations. Group Continuous training

Variable	N	M	SD	1	2	3	4
1. Group ATI	7	1.29	.488				
2. ATI Scale Scores	7	4.92	.391	0.638			
3. Overall Workload - Analogue	7	25.71	10.83	0.479	0.236		
4. Overall Workload - Digital	7	28.44	10.27	-0.158	-0.09	0.342	

N=7

There was no significant correlation between any variables tested in this correlation's analysis.

Furthermore, the alternative – experimental hypothesis was not confirmed, that is there is no significant effect between high and very high scores in Affinity for Interaction Scale and Workload scores.

3.3.4 *Question and Hypothesis 4*

ATI Scale Scores and Performance (response times). To what extent is there a relationship between higher ATI Scale scores and better response times in performance?

H4o. There is no significant effect between higher scores in Affinity for Interaction Scale and response times.

H4. There is significant effect between higher scores in Affinity for Interaction Scale and response times.

Correlation Analysis. Addressing the issue of small size sample of the study and although the normality assumption was met for all variables, it was recommended Spearman's rank correlation coefficient to be used. This is because it is a non-parametric test and is less sensitive to deviations from normality compared to Pearson's correlation. This is because it is a non-parametric test, and it is less sensitive to deviations from normality compared to Pearson's correlation.

Spearman's rank correlation was computed to assess the relationship between response time variables and ATI Scale Score variables in all groups of interest (all participants, initial training group, continuous training group). Of the variables analyzed, the following showed significant ($p<0.05$) and ($p<0,01$) correlations.

There was a negative correlation in Continuous Training group between, Climbing Rate-Climbing leg – Analogue and ATI scale scores variables, $r(5) = -.81$, $p = .027$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Continuous Training participants with a higher score in ATI scale had lower response time score in Climbing Rate- Climbing leg – Analogue.

There was a negative correlation in Initial Training group between, Airspeed- Level flight leg – Analogue and ATI scale scores variables, $r(5) = -.78$, $p = .041$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Initial Training participants with a higher score in ATI scale had lower response time score in Airspeed- Level flight leg – Analogue.

There was a negative correlation in Continuous Training group between, Altitude- Level flight leg – Analogue and ATI scale scores variables, $r(5) = -.85$, $p = .016$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Continuous Training participants with a higher score in ATI scale had lower response time score in Altitude- Level flight leg – Analogue.

There was a positive correlation in Initial Training group between, Heading- Level flight leg - Analogue and ATI scale scores variables, $r(5) = .79$, $p = .033$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that in Initial Training participants group higher score in ATI scale, relates with high response time score in Heading- Level flight leg - Analogue and vice versa.

There was a negative correlation in Continuous Training group between, Airspeed- 360 turn level flight – Analogue and ATI scale scores variables, $r(5) = -.76$, $p = .049$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that Continuous Training participants with a higher score in ATI scale had lower response time score in Airspeed- 360 turn level flight – Analogue.

There was a negative correlation in all participants group between, Rate of Descent- Descending leg – Analogue and ATI scale scores variables, $r(12) = -.57$, $p = .034$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that all participants groups with a higher score in ATI scale had lower response time score in Rate of Descent- Descending leg – Analogue.

There was a negative correlation in Continuous Training group between, Rate of Descent- Descending leg – Analogue and ATI scale scores variables, $r(5) = -.80$, $p = .034$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that all participants groups with a higher score in ATI scale had lower response time score in Rate of Descent- Descending leg – Analogue.

There was a positive correlation in Initial Training group between, Total time Level leg– Analogue and ATI scale scores variables, $r(5) = .81$, $p = .027$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that in Initial Training participants group higher score in ATI scale, relates with high response time score in Total time Level leg – Analogue and vice versa.

There was a negative correlation in Continuous Training group between, Total time descending – Analogue and ATI scale scores variables, $r(5) = -.78$, $p = .041$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that all participants groups with a higher score in ATI scale had lower response time score in Total time descending – Analogue.

There was a positive correlation in Initial Training group between, Airspeed- 360 turn level flight – Digital and ATI scale scores variables, $r(5) = .81$, $p = .027$. This association was statistically significant with a margin of error of less than 5%. Based on the coding of the variables, this correlation demonstrates that in Initial Training participants group higher score in ATI scale, relates with high response time score in Airspeed- 360 turn level flight – Digital and vice versa.

The findings of Spearman's correlation analysis indicated that there were statistically significant results and that the investigation of fourth question hypotheses should continue with regression analyses. Thus, regression analyses were performed to further investigate the alternative hypotheses that there are significant effects between higher scores in Affinity for Interaction Scale and response times.

3.4 Regression Analyses

The presentation of regression analyses involving only the significant findings of Spearman 's Correlation analyses, is addressing below that proved ATI's statistical significance as a predictor for various variables (response times).

3.4.1 Regression 1

Regression analysis was used after Spearman's correlation demonstrated that Initial Training group (Airspeed- Level flight leg – Analogue and ATI scale scores variables $r(5) = -.78$, $p = .041$), with a higher score in ATI scale had lower response time score in Airspeed- Level flight leg – Analogue. The analysis examined the relationship between the dependent variable Airspeed- Level flight leg - Analogue and the predictor variable ATI Scale Scores and showed that the Airspeed- Level flight leg - Analogue is significantly predicted by the ATI Scale Scores. More specifically the multiple regression model had an R value of .779

and an R square of .607, indicating that 60,7% of the variance in the dependent variable (Airspeed- Level flight leg - Analogue) was explained by the predictor variable (ATI Scale Scores). The adjusted R square of .529 indicated that after adjusting for the number of predictors in the model, 52,9% of the variance in the dependent variable was explained by the predictor (ATI Scale Scores). The standard error of the estimate was 48.21758, which is the average difference between the observed and predicted values. The Durbin-Watson statistic was 2.672, which indicated that there was no significant autocorrelation present in the residuals.

From the ANOVA table the F-statistic for the regression model was 7.737 and the p-value was .039, indicating that the regression model was statistically significant at a significance level of .05. $F(1, 5) = 7.737$, $p = 0.039$, $\eta^2 = 0{,}61$, indicating that 61% of the variance in the dependent variable Airspeed- Level flight leg – Analogue was explained by the predictor ATI Scale Scores in the regression model. Further analysis, such as residual plots and diagnostic tests, were performed to ensure that the assumptions of the regression model are met.

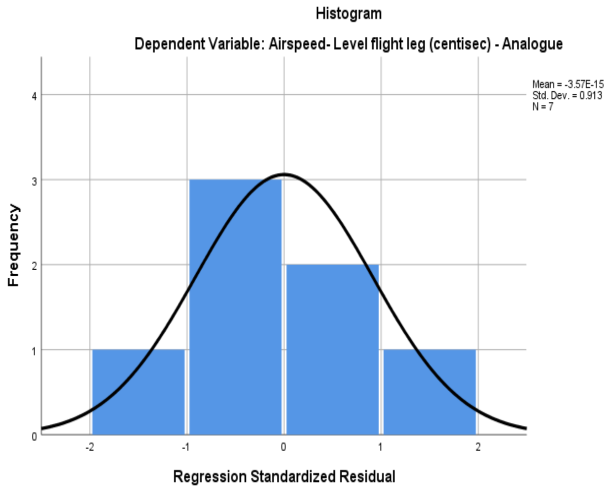
The coefficients table showed that the intercept of the regression line was 868.962 and the coefficient for the predictor variable (ATI Scale Scores) was -116.823, indicating that a unit increase in ATI Scale Scores was associated with a decrease of 116.823 units in the rate of descent. The t-statistic for the predictor variable was -2.782 and the p-value was .039, indicating that the predictor was statistically significant at a significance level of .05.

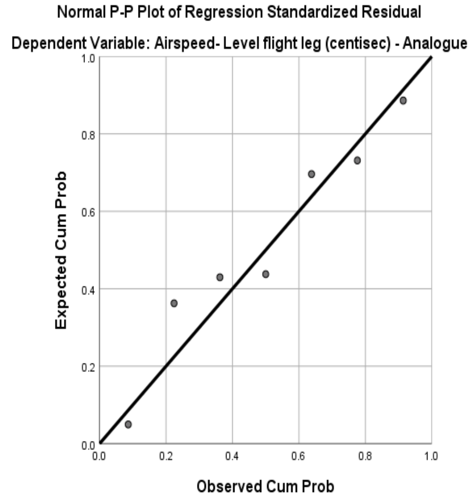
The collinearity diagnostics indicated that there was no multicollinearity present in the model, as the condition index was 1.000 for both dimensions of the model.

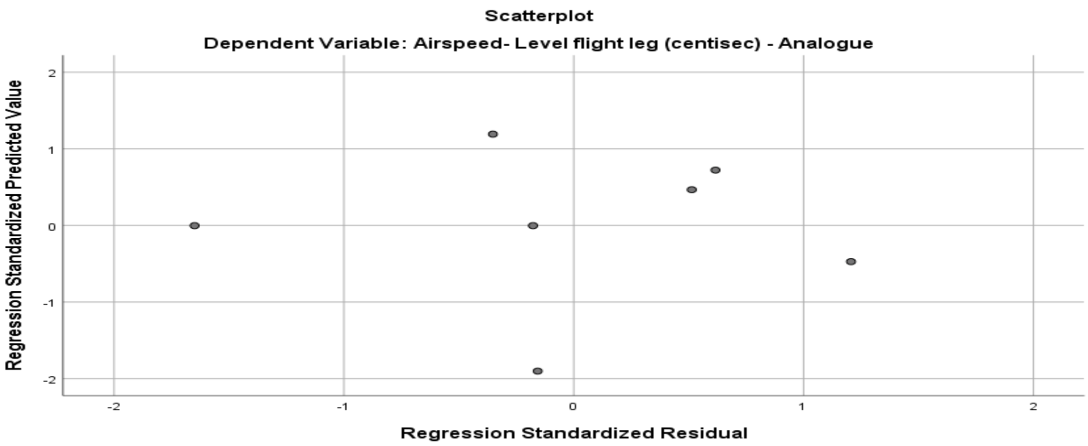
The residuals statistics showed that the minimum and maximum residuals were -79.54739 and 58.15369, respectively, and the mean of the residuals was 0. These residual statistics suggest that the prediction model for the Airspeed-Level flight leg had a mean residual of

0.00000 and a standard deviation of 44.01643. The standardized residuals indicated a good fit between the observed and predicted values.

Charts







3.4.2 Regression 2

Regression analysis was used after Spearman's correlation demonstrated that Continuous Training group (Climbing Rate- Climbing leg - Analogue and ATI scale scores variables, r(5) = -.81, p = .027), with a higher score in ATI scale had lower response time score in Climbing Rate- Climbing leg – Analogue. The analysis examined the relationship between the dependent variable Climbing Rate- Climbing leg - Analogue and the predictor variable ATI Scale Scores and showed that the Climbing Rate- Climbing leg - Analogue was significantly predicted by the ATI Scale Scores. More specifically the regression model had an R value of

.794 and an R square of .630, indicating that 63% of the variance in the dependent variable (Climbing Rate- Climbing leg – Analogue) was explained by the predictor variable (ATI Scale Scores). The Adjusted R Square value of .556 adjusted for the number of predictors 55.6% in the model. The Std. Error of the Estimate was 27.88412, which represented the standard deviation of the residuals and provided an estimate of the accuracy of the predictions made by the model. The Durbin-Watson value of .793 was a statistical test that measured the autocorrelation of the residuals in the regression model. The predictions made by the model had an estimated standard deviation of 27.88412 and there was no significant autocorrelation in the residuals.

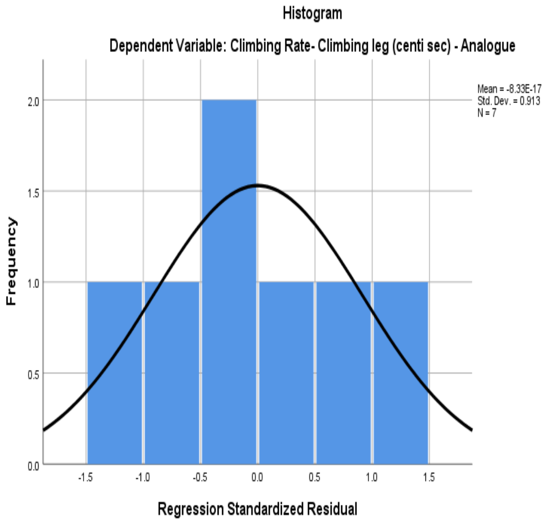
From the ANOVA table the F-statistic for the regression model was 8.514 and the p-value was .033, indicating that the regression model was statistically significant at a significance level of .05. $F(1, 5) = 8.514$, $p = 0.033$, $\eta^2 = 0,63$, indicating that 63% of the variance in the dependent variable Climbing Rate- Climbing leg - Analogue was explained by the predictor ATI Scale Scores in the regression model.

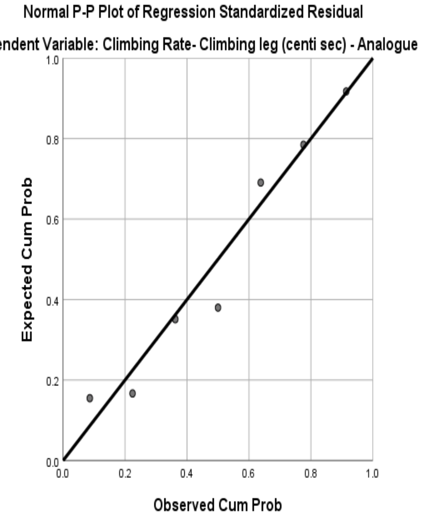
From the coefficients table the constant coefficient was 542.056 with a standard error of 143.463 and a t-statistic of 3.778, with a p-value of .013, which was significant at a level of .05. The ATI Scale Scores coefficient was -84.879 with a standard error of 29.089 and a t-statistic of -2.918, with a p-value of .033, which was significant at a level of .05. The standardized coefficient (beta) was -.794, indicating the strength of the relationship between the ATI Scale Scores and the dependent variable (Climbing Rate- Climbing leg - Analogue). The collinearity diagnostics indicated that there was no multicollinearity present in the model, as the condition index was 1.000 for both dimensions of the model.

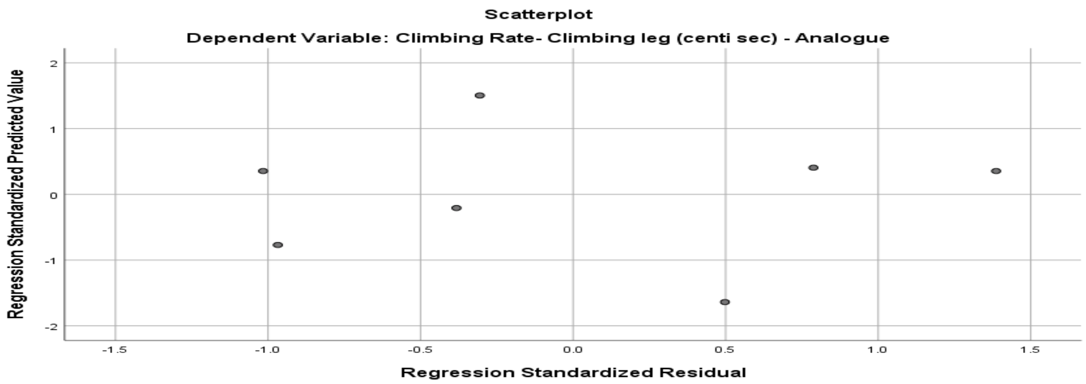
The residuals statistics showed that the minimum value was -28.33326 and the maximum value was 38.66674. The mean of the residuals was 0 and the standard deviation was

25.45461. The standardized residuals indicated a good fit between the observed and predicted values.

Charts







3.4.3 Regression 3

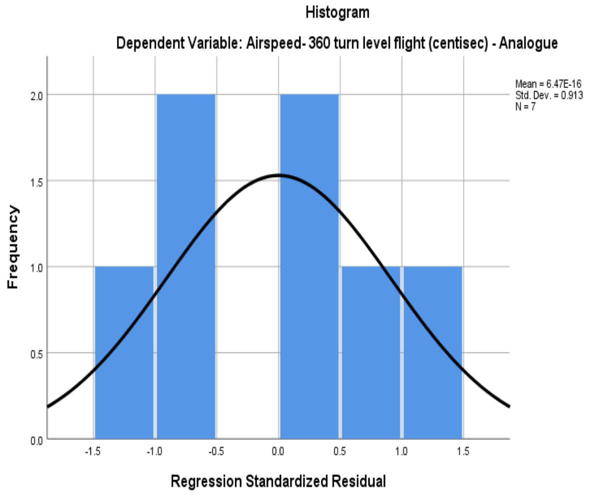
Regression analysis was used after Spearman's correlation demonstrated that Continuous Training group (Airspeed- 360 turn level flight - Analogue and ATI scale scores variables r(5) = -.76, p = .049), with a higher score in ATI scale had lower response time score in Airspeed- 360 turn level flight – Analogue.

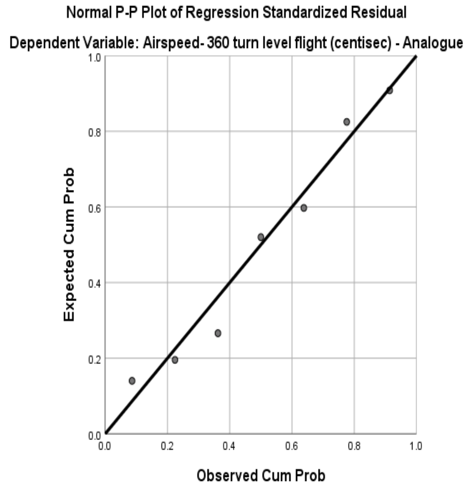
The analysis examined the relationship between the dependent variable Airspeed- 360 turn level flight - Analogue and the predictor variable ATI Scale Scores and showed that the Airspeed - 360 turn level flight - Analogue was significantly predicted by the ATI Scale Scores. More specifically the regression model had an R value of .739, which indicated a strong positive linear relationship between the predictor variable (ATI Scale Scores) and the dependent variable. The R-squared value of .546 suggested that 54.6% of the variance in airspeed - 360 turn level flight - Analogue was explained by the predictor variable. The adjusted R-squared value of .455 indicated that this proportion was slightly lower 45.5% after adjusting for the sample size. The standard error of the estimate was 36.34827, which gave an idea of the average difference between the predicted values and the actual values. The Durbin-Watson statistic of 2.259 there was no significant autocorrelation in the residuals. From the ANOVA table the F-statistic for the regression model was 6.017 and the significance level of .058 suggested that there was some evidence to support the presence of a relationship between the predictor variable and the dependent variable. $F(1, 5) = 6{,}017$, $p = .058$, $\eta^2 = .55$, indicated that 55% of the variance in the dependent variable Airspeed- 360 turn level flight - Analogue was explained by the predictor ATI Scale Scores in the regression model.

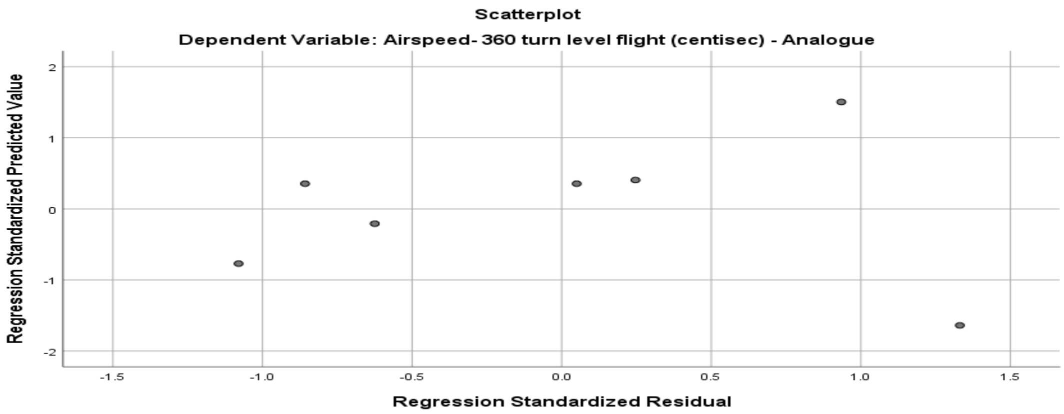
The unstandardized coefficient of -93.012 for ATI Scale Scores suggested that for each unit increase in the ATI Scale Scores, the Airspeed - 360 turn level flight decreases by 93.012 centiseconds, on average. The standardized coefficient of -0.739 indicated that the effect of the predictor variable on the dependent variable was strong and negative. The t-value of -2.453 and the significance level of .058 suggested that the effect was statistically significant. From the Collinearity Diagnostics table the condition index was 1.000 for both dimensions, indicating that there was no multicollinearity present in the model. The variance proportions showed that the predictor variable explained 100% of the variation in the model.

From the Residuals Statistics table the residuals had a minimum value of -39.24932, a maximum value of 48.37466, a mean of 0, and a standard deviation of 33.18128. The standardized residuals indicated a good fit between the observed and predicted values.

Charts







3.4.4 Regression 4

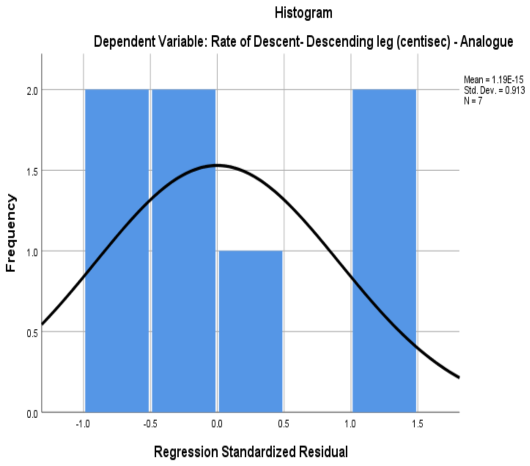
Regression analysis was used after Spearman's correlation demonstrated that Continuous Training group (Rate of Descent- Descending leg - Analogue and ATI scale scores variables r(5) = -.80, p = .034), with a higher score in ATI scale had lower response time score in group Rate of Descent- Descending leg – Analogue.

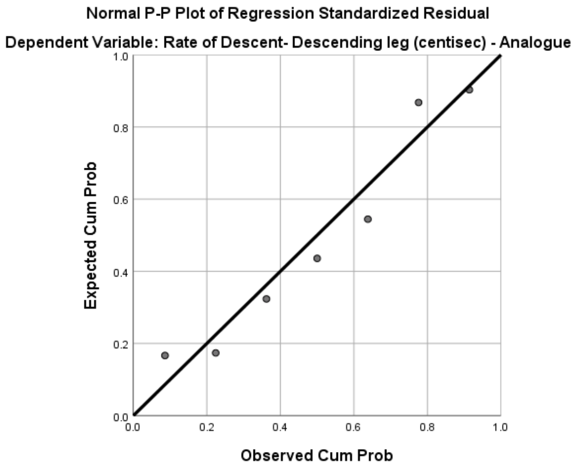
The analysis examined the relationship between the dependent variable Rate of Descent- Descending leg - Analogue and the predictor variable ATI Scale Scores and showed that the Rate of Descent- Descending leg - Analogue was significantly predicted by the ATI Scale Scores. More specifically the regression model had an R value of .749 and the R-squared (coefficient of determination) was .561, indicating that 56.1% of the variability in the dependent variable was explained by the ATI Scale Scores. The adjusted R-squared (corrected for degrees of freedom) was .473 adjusted for the number of predictors 47.3% in the model. The standard error of the estimate was 70.68243. The Durbin-Watson statistic was 2.843 indicating there was not a significant presence of autocorrelation in the residual. The F-value for the regression was 6.381 and the significance level (p-value) was .053, indicating that the predictor was significantly associated with the dependent variable. $F(1, 5) = 6{,}381$, $p = .053$, $\eta^2 = .56$, indicating that 56% of the variance in the dependent variable was explained by the predictor ATI Scale Scores in the regression model.

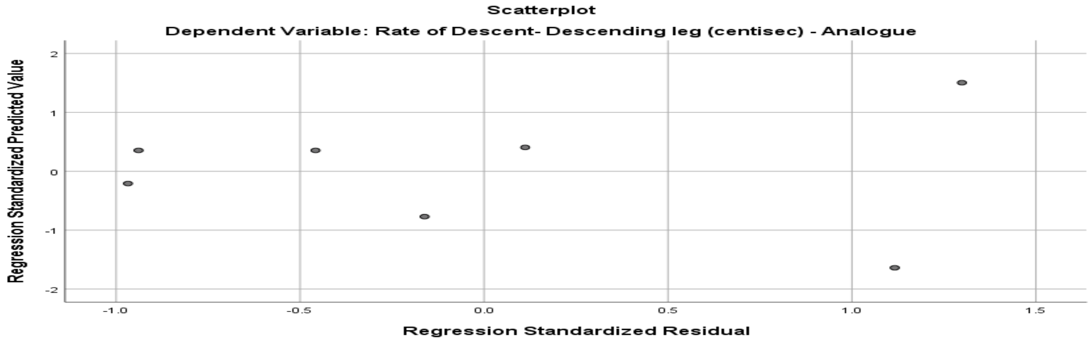
From the coefficients table the unstandardized coefficient (B) for the predictor (ATI Scale Scores) was -186.263 and the standardized coefficient (beta) was -0.749. The t-value was -2.526 and the significance level (p-value) was .053, indicating that the predictor was significantly associated with the dependent variable. The collinearity statistics indicated that the tolerance (1 / VIF) for the predictor was 1.000, indicating that the predictor was not collinear with any other predictors. The variance proportions showed that 100% of the variability in the dependent variable was explained by the predictor (ATI Scale Scores).

The residuals statistics showed that the minimum predicted value was 33.0971, the maximum predicted value was 262.2004, and the mean predicted value was 152.5714. The minimum residual was -68.40431, the maximum residual was 91.79957, and the mean residual was 0. The standard deviation of the predicted values was 72.89226 and the standard deviation of the residuals was 64.52393. The standardized residuals indicated a good fit between the observed and predicted values.

Charts







3.4.5 Regression 5

Regression analysis was used after Spearman's correlation demonstrated that Continuous Training group (Total time descending - Analogue and ATI scale scores variables r(5) = -.78, p = .041), with a higher score in ATI scale had lower response time score in group Total time descending – Analogue. The analysis examined the relationship between the dependent variable Total time descending - Analogue and the predictor variable ATI Scale Scores and showed that the Total time descending - Analogue was significantly predicted by the ATI Scale Scores. More specifically the regression model had an R value of .816 that suggested a strong positive correlation between the predictor and the dependent variable "Total time descending - Analogue". The R squared value of .666 meant that 66.6% of the variation in the dependent variable was explained by the predictor. The adjusted R squared of .599

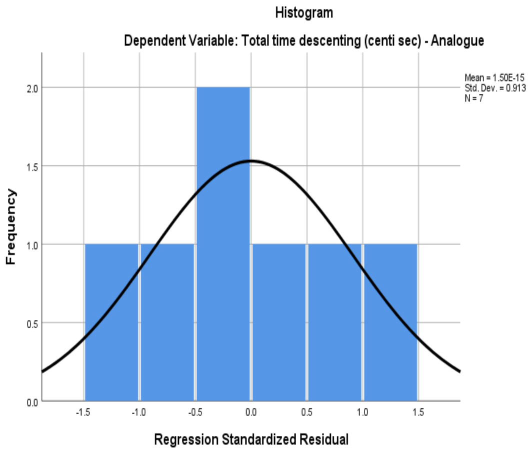
adjusted for the number of predictors in the model and indicated that 59.9% of the variation in the dependent variable was explained by the predictor. The standard error of the estimate (115.88) was the average distance that the observed values deviate from the regression line. A low standard error of the estimate indicated a good fit of the model to the data. The Durbin-Watson statistic of 2.22 indicated that there was no significant autocorrelation in the residuals, which was a desirable property in a regression model.

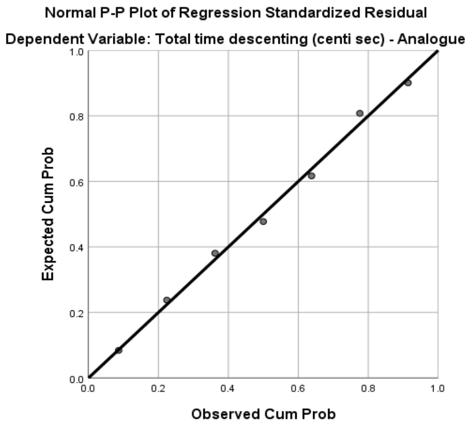
From the ANOVA table which summarized the analysis of variance for the regression model the p-value of .025 indicated that the predictor was significantly related to the dependent variable at a significance level of .05. $F(1, 5) = 9.955$, $p = .025$, $\eta^2 = .67$, indicating that 67% of the variance in the dependent variable Total time descending - Analogue was explained by the predictor ATI Scale Scores in the regression model.

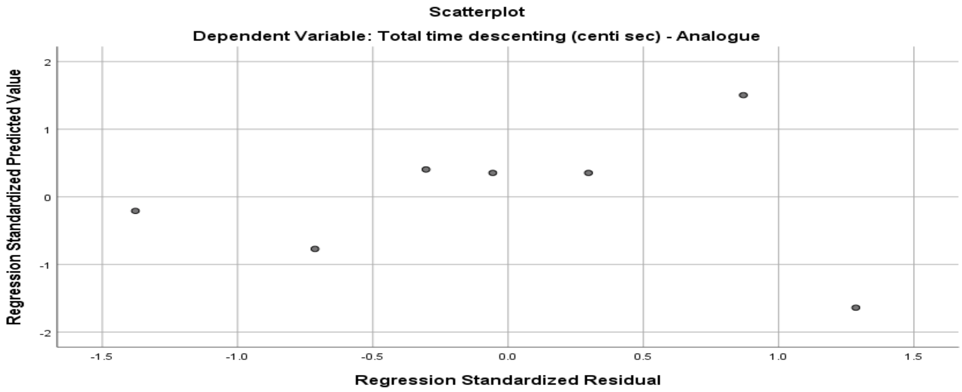
From the coefficients table the predictor ATI Scale Scores had a regression coefficient of -381.432, indicating that for a one-unit increase in ATI Scale Scores, the Total time descending - Analogue decreases by 381.432 units. The p-value of .025 indicated that the predictor was significantly related to the dependent variable.

The "Tolerance" and "VIF" columns provided information about collinearity, with a high tolerance and a low VIF indicating that there was no multicollinearity in the model. In this case, the tolerance was 1.000 and the VIF was also 1.000, indicating that there was no multicollinearity in the model. The mean of the residuals was close to zero, and the standard deviation of the residuals was 105.78, indicating that the residuals had a moderate spread around the mean.

Charts







3.4.6 Regression 6

Regression analysis was used after Spearman's correlation demonstrated that Continuous Training group (Altitude- Level flight leg - Analogue and ATI scale scores variables r(5) = -.85, p = .016 , with a higher score in ATI scale had lower response time score in group Altitude- Level flight leg – Analogue. The analysis examined the relationship between the dependent variable Altitude- Level flight leg - Analogue and the predictor variable ATI Scale Scores and showed that the Altitude- Level flight leg - Analogue was significantly predicted by the ATI Scale Scores. More specifically the regression model had an R value of .890, indicating a strong relationship between the predictor and the dependent variable. The "R Square".792 column gave the coefficient of determination, which represented the proportion of variance in the dependent variable that was explained by the predictor. The "Adjusted R

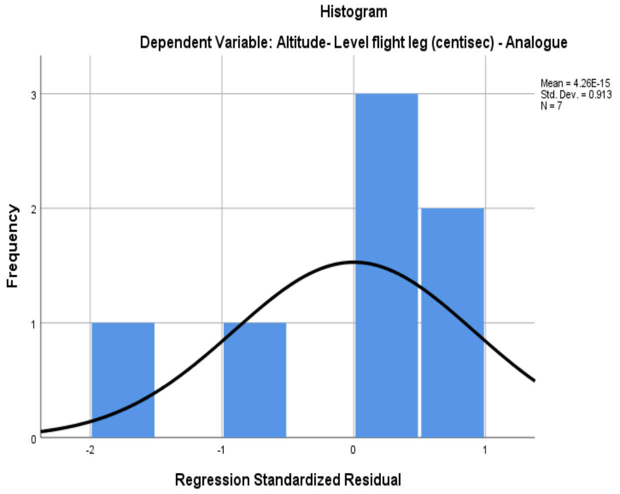
Square" column gave a modified version of R Square that adjusted for the number of predictors in the model. In this case, the adjusted R-square was .750, indicating that the model fitted the data well and that the predictor ATI Scale Scores explained 75% of the variance in the dependent variable Altitude- Level flight leg - Analogue. The "Std. Error of the Estimate" (18.03571) gave the standard error of the estimate, which was a measure of the average difference between the observed values and the values predicted by the regression model. The "Durbin-Watson" (1.450) with a value close to 2 indicating the absence of autocorrelation in the residuals.

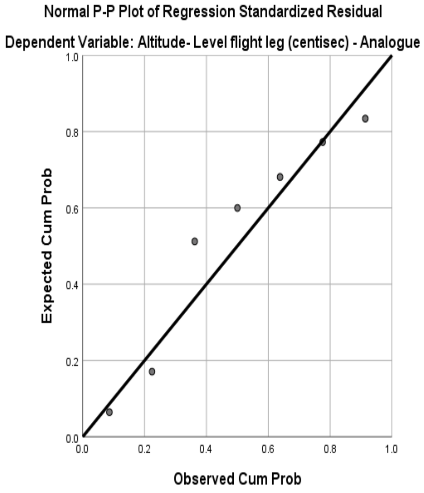
From the ANOVA table the F-statistic was 19.024 with a significance level of .007, indicating that the regression model was a good fit to the data and that the predictor ATI Scale Scores was a significant predictor of the dependent variable Altitude- Level flight leg - Analogue. $F(1, 5) = 19024$, $p = .007$, $\eta^2 = .79$, indicating that 79% of the variance in the dependent variable Altitude- Level flight leg - Analogue was explained by the predictor ATI Scale Scores in the regression model.

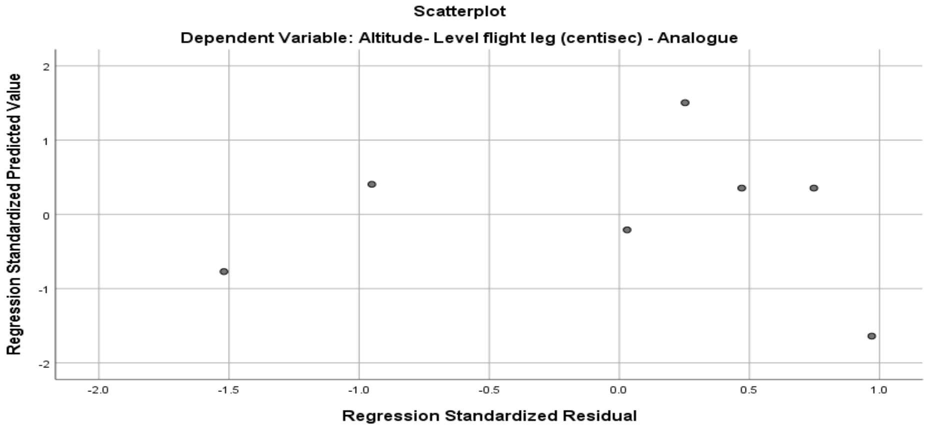
The predictor "ATI Scale Scores" had a coefficient of -82.065 with a t-statistic of -4.362 and a significance level of .007, indicating that it was a significant predictor of the dependent variable Altitude- Level flight leg - Analogue.

The collinearity diagnostics table showed that the first dimension had an eigenvalue of 1.997, with a condition index of 1.000. This suggested that there was no multicollinearity present in the model. The standardized residuals indicated a good fit between the observed and predicted values.

Charts







Summarizing the investigation of the fourth question ATI may predict six dependent variables which are listed below :

1.R^2_adj=. 529, F(1, 5) = 7.737, p = .039, η^2 =.61, indicating that 61% of Airspeed- Level flight leg – Analogue was explained by the predictor ATI Scale Scores (Initial training pilots with higher score in ATI scale had lower response time score in Airspeed- Level flight leg – Analogue**,** r(5) = -.78, p = .041)

2. R^2_adj =.556, F(1, 5) = 8.514, p =.033, η^2 =.63, indicating that 63% of Climbing Rate-Climbing leg - Analogue was explained by the predictor ATI Scale Scores (Continuous

training pilots with a higher score in ATI scale had lower response time score in Climbing Rate- Climbing leg – Analogue, r(5) = -.81, p = .027)

3. R^2_adj = .455, F(1, 5) = 6,017, p =.058, η^2 =.55, indicating that 55% of Airspeed- 360 turn level flight - Analogue was explained by the predictor ATI Scale Scores (Continuous Training pilots with a higher score in ATI scale had lower response time score in Airspeed- 360 turn level flight – Analogue, r(5) = -.76, p = .049)

4. R^2_adj = .473, F(1, 5) = 6,381, p =.053, η^2 =.56, indicating that 56% of Rate of Descent- Descending leg – Analogue was explained by the predictor ATI Scale Scores (Continuous training pilots with a higher score in ATI scale had lower response time score in Rate of Descent- Descending leg – Analogue**,** r(12) = -.57, p = .034)

5. R^2_adj = .599, F(1, 5) = 9.955, p =.025, η^2 =.67, indicating that 67% of Total time descending - Analogue was explained by the predictor ATI Scale Scores (Continuous training pilots with a higher score in ATI scale had lower response time score in Total time descending – Analogue**,** r(5) = -.78, p = .041)

6. R^2_adj = .750, F(1, 5) = 19024, p =.007, η^2 =.79, indicating that 79% of Altitude- Level flight leg - Analogue was explained by the predictor ATI Scale Scores (Continuous training pilots with a higher score in ATI scale had lower response time score in Altitude- Level flight leg – Analogue, r(5) = -.85, p = .016)

3.5 Moderation Analyses

Though there were some interesting results about ATI 's effects on performance (response times) it was worthwhile to extend our investigation in examining whether ATI may contribute as a moderator including workload variable to performance.

Furthermore, to proceed, Spearman's rank correlation was computed to assess the relationship between performance (response time) variables and Workload variable in All participants, Initial and Continuous Training Group, respectively. Of the variables analyzed, only the variables that showed significant correlations are presented. Their code names are also presented, as they were computed for moderation analyses.

DBank360
Bank Angle- 360 turn level flight – Digital, Initial Training Group, r(5) = .85, p = .016

AAirDe
Airspeed- Descending leg – Analogue, Continuous Training Group, r(5) = -.78, p = .041

AHeDe
Heading- Descending leg – Analogue, Continuous Training Group, r(5) = -.83, p = .021

DHeCli
Heading-Climbing leg – Digital, Continuous Training Group, r(5) = .79, p = .036

DAir360
Airspeed- 360 turn level flight – Digital, Continuous Training Group, r(5) = .70, p = .086

DTCliL,
Total time Climbing leg – Digital, Continuous Training Group, r(5) = .79, p = .036

For the moderation analyses PROCESS command tool (Hayes & Matthes, 2009; Preacher & Hayes, 2004, 2008a ; Hayes, 2012), which is a custom dialog box in SPSS Version 4.1, was used (written by Andrew F. Hayes, Ph.D - Documentation available in Hayes, 2022).

It was used to model the relationship between an outcome variable (performance) and a predictor (workload) while considering the effect of a moderator variable (ATI scale scores) on that relationship. The combined effect of two variables on another is known conceptually as moderation, and in statistical terms as an interaction effect.

Following, only the significant results of moderation analyses are presented :

Moderation Analysis 1

Initial Training Group (n=7)

According to the model 85.79% of the variability in the dependent variable (DBank360) was explained by the independent variables (D_Ov_Wo and ATIRAW).

Model

	coeff	se	t	p	LLCI	ULCI
constant	578.4988	71.8664	8.0496	.0040	348.9458	808.0518
D_Ov_Wo	46.4240	11.5431	4.0218	.0276	9.5533	83.2947
ATIRAW	629.8424	235.2947	2.6768	.0753	-121.7271	1381.4120
Int_1	**123.5759**	**39.7950**	**3.1053**	**.0531**	-3.5358	250.6876

The interaction term was statistically significant (b=123.5759, s.e.=39.7950, p=.0531), consistent with the hypothesis that the ATI Raw (scale scores) variable moderated the effect of workload on performance. The p-values for D_Ov_Wo were less than 0.05, indicating that there was evidence that this independent variable was significant predictor of DBank360. The p-value for the constant was lower than 0.05, indicating that the constant term was significant.

At -1 sd (i.e., at -.4687) on the centered ATIRAW variable (representing low ATIRAW), the relationship between D_Ov_Wo and performance was negative and not significant (b=-11.4962, s.e.= -.8463, p=.4596). Similarly, at the mean (i.e., at 0) on the centered moderator variable (representing medium ATIRAW), the relationship was positive and significant (b=46.4240 , s.e.= 11.5431, p=.0276). Finally, at +1sd (i.e., +.4687) on the centered D_Ov_Wo variable (represent high ATIRAW), the relationship was positive and significant (b=104.3442, s.e.= 27.8880, p=.0333).

The results of a regression analysis examining the interaction between two variables: "D_Ov_Wo" and "ATIRAW", indicated that the highest-order interaction term "X * W" was significant (p = .0531) with an R-squared change of .4568. A Johnson-Neyman significance region was identified where the moderator "ATIRAW" was below -.1544.

The effect size was positive and increased as the value of the moderator increased, meaning that when ATI values were high Workload was increased, predicting lower response times (higher scores in ATI can contribute to higher scores in workload and therefore predict lower response time – better performance).

Moderation Analysis 2

Continuous Training Group (*n*=7)

According to the model 92.48% of the variability in the dependent variable (DTCliL) was explained by the independent variables (D_Ov_Wo and ATIRAW).

Model

	coeff	se	t	p	LLCI	ULCI
constant	517.9696	36.4315	14.2176	.0008	401.6015	634.3377
D_Ov_Wo	1.2215	5.8871	.2075	.8489	-17.5827	20.0258
ATIRAW	-840.4466	192.1676	-4.3735	.0221	-1454.2611	-226.6321
Int_1	**-65.8382**	**19.2483**	**-3.4205**	**.0418**	-127.3205	-4.3559

The interaction term was statistically significant (b= -65.8382, s.e.=19.2483, p=.0418), consistent with the hypothesis that the ATI Raw (scale scores) variable moderated the effect of workload on performance. The p-values for ATIRAW were less than 0.05, indicating that there was evidence that this independent variable was significant predictor of DTCliL. The p-value for the constant was lower than 0.05, indicating that there was evidence that the constant term was significant.

At -1 sd (i.e., at -.3913) on the centered ATIRAW variable (representing low ATIRAW), the relationship between D_Ov_Wo and performance was negative and significant (b=26.9867, s.e.= 4.7199 , p=.0106). At the mean (i.e., at 0) on the centered moderator variable (representing medium ATIRAW), the relationship was positive and not significant (b= 1.2215, s.e.= 5.8871, p=.8489). Finally, at +1sd (i.e., +.3913) on the centered D_Ov_Wo variable (represent high ATIRAW), the relationship was negative and not significant (b=-24.5436, s.e.= 12.6696, p=.1481).

The statistical analysis examining the interaction between two variables, "D_Ov_Wo" and "ATIRAW", indicated that he highest order unconditional interaction was found to be significant (p = .0418), with D_Ov_Wo as the focal predictor and ATIRAW as the moderator. The Johnson-Neyman significance region (JNSR) for the moderator was found to be between -.1711 and .1711, meaning that below -.1711, the effect of D_Ov_Wo becomes negative, while above .1711, the effect becomes positive. The JNSR is defined as the range of values of the moderator (ATIRAW) for which the effect of the focal predictor was statistically significant. The effect size of ATIRAW on the outcome variable was only significant when the value of the moderator was above -.1711, meaning that lower scores in ATI scale gave lower workload in Continuous training group in digital cockpit and therefore predict higher response times (worse performance).

Summarizing the results from moderation analyses :

-ATI scale moderated overall workload in Digital cockpit in Initial training group to performance (response time in Bank Angle – 360 turn level flight / DBank360).

R^2 *= .8579, b=123.5759, s.e.=39.7950, p=.0531*

-ATI scale moderated overall workload in Digital cockpit in Continuous training group to performance (response time in Total time Climbing leg- DTCliL)

R^2 *= .9248, b= -65.8382, s.e.=19.2483, p=.0418*

3.6 Conclusions

The purpose of this research was to examine whether three independent variables are predictive of successful performance in analogue and digital instrumentation between two groups, Initial and Continuous Training, respectively. Accordingly, to descriptive statistics, participants had better recorded times (lower means) in digital displays. It seems that digital instrumentation might improve processing efficiency as it helps participants organize important information more quickly, in accordance with Sweller' s (1998) findings that spatially integrating important information (as it is in digital cockpit) can make it easier for someone to see relations among important content than when that information is spatially separated (i.e., analogue cockpit).

Overall workload (mean) was lower in Initial Training Group than in Continuous Training Group. This finding was probably due to the lowest requirement of flight operations and stable environment during the performance of Initial Training group, as participants were asked to perform as Pilot Not Flying. In contrast, Continuous Training Group performed as Pilot Flying and had higher means in overall workload.

Multiple studies investigated differences in scan paths and scan patterns between novices and experts from different domains (Law et al, 2004; Ooms et al., 2014). In aviation, the literature also emphasizes different visual scanning strategies in novices vs experts pilots (Kasarskis et al., 2001; Yang et al., 2013). As experts show more flexible scanning strategies and they are more focused on relevant information and allocate their attention more efficiently, they adjust their scanning behaviors more effectively to the situational demands. In line with Kang and Landrey research in 2014, experts (Continuous Training Group) showed more flexible scanning strategies, were more focused on relevant information and allocated their attention more efficiently. Differences in monitoring patterns between continuous (CT) and initial training group (IT) could have been caused by the continuous training group having a more salient cognitive picture. During the orientation phase at the beginning of the new situation, the Continuous Training group monitored the relevant information intensively. As a result, they might have used this phase more efficiently to build up a salient picture. Consequently, they were able to focus their attention more and intensively on the relevant patterns at the right time, and therefore had lower response times. Regarding the first research question, it was determined that there were significant effects between scores of Compass test and response times variables.

Initial Training group (n=7)

Slalom ↑	Memory ↑	Task Management ↑	Mathematics ↑	Total Compass score ↑	Orientation ↑
Climbing Rate- Climbing leg – Analogue↓	Altitude- Level flight leg – Analogue ↑	Heading- Level flight leg – Analogue ↓	Airspeed- 360 turn level flight – Analogue ↑	Heading- Descending leg – Analogue ↑	Total time Climbing leg – Digital ↓
Altitude- 360 turn level flight – Analogue ↓	Total time descenting – Analogue ↑	Heading- Descending leg – Analogue ↑	Heading- Descending leg – Analogue ↑	Climbing Rate- Climbing leg – Digital ↓	Heading- Descending leg – Analogue ↑
Total Time All Tests – Analogue ↓	Airspeed- Descending leg – Digital ↑	Total time Climbing leg – Digital ↓		Total time Climbing leg – Digital ↓	
Airspeed- Descending leg – Analogue ↑	Total time descenting – Digital ↑				

Continuous Training group (n=7)

Slalom ↑	Control ↑	Mathematics ↑
Heading-Climbing leg – Analogue ↓	Climbing Rate- Climbing leg – Analogue ↑	Heading- Level flight leg – Digital ↓
Airspeed- Descending leg – Digital ↑		

Summarizing the results, it was interesting to find that most correlations were found regarding the analogue instrumentation in Initial Training Group. Nevertheless, Slalom task which is associated in processing and reacting quickly enough to perform sensomotor tasks, was correlated with all performance response times in Analogue instrumentation. This indicates the need to further investigate if Slalom may potentially be a predictor for better (lower) response time in Situation Awareness Tasks.

As far as the second question of the research is concerned there was no significant effect between high and very high scores in Affinity for Interaction Scale and better scores in Compass scores.

Accordingly, answering the third question there was no significant effect between high and very high scores in Affinity for Interaction Scale and Workload scores.

Summarizing the results of the fourth research question, there were evidence that ATI may predict 6 variables which are listed below :

1.Initial training pilots with higher score in ATI scale had lower response time score in Airspeed- Level flight leg – Analogue.

2.Continuous training pilots with a higher score in ATI scale had lower response time score in Climbing Rate- Climbing leg – Analogue.

3.Continuous Training pilots with a higher score in ATI scale had lower response time score in Airspeed- 360 turn level flight – Analogue.

4.Continuous training pilots with a higher score in ATI scale had lower response time score in Rate of Descent- Descending leg – Analogue.

5.Continuous training pilots with a higher score in ATI scale had lower response time score in Total time descending – Analogue.

6.Continuous training pilots with a higher score in ATI scale had lower response time score in Altitude- Level flight leg – Analogue.

However, the most interesting finding was resulted from moderation analyses which gave evidence that ATI scale score served as moderator to Digital Overall Workload.

In Initial training group in Digital cockpit when ATI values were high, Digital Overall Workload was increased, predicting lower response times, therefore better performance, better situation awareness in Bank Angle – 360 turn level flight.

In Continuous training group in Digital cockpit lower scores in ATI scale gave lower Digital Overall Workload and therefore predict higher response times (worse performance, worse situation awareness) in Total time Climbing leg.

These results are very important as they are evidence that in demanding circumstances (such as the difficult conditions of climbing leg and 360 Turn) ATI scores served as moderator to workload to predict performance. It is interesting to consider that in the aviation field we need trainees in digital cockpits that score higher in ATI in order to have better performances (in addition we expect high workload) . Additionally, study results indicate that experts should not be lower in ATI as there is evidence that they do not engage enough to have higher score in overall workload, as to have the best performance (lower response time). Findings from other research (Schaarschmidt, M., Ivens, S., Homscheid, D., & Bilo, P. 2015;

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. 2011) detected same effects (ATI acted as moderator).

Nevertheless, it should be noted that the need to use statistical measures which handle small samples must be highlighted. Furthermore, in future studies the thought of use mixed-effects models is of great interest, as they have several advantages compared to classic statistical measures, such as handling designs, offering much greater flexibility in choosing covariates and better statistical power (Gueorguieva, and Krystal, 2004). The Bayesian estimation via Markov-Chain MonteCarlo sampling, using the MCMCglmm program from the correspondent package as supplied in the R system for scientific computing is recommended in several studies with small population samples (Rens Van de S., and Milica, M., 2020).

Discussion

The association between pilot cognition and accident risk is supported by annual reports linking pilot-related error to 75% of GA accidents, while mechanical failures account for just 8% (Geske, 2018). Therefore, the ability to predict pilot's performance is of great value. Considering that aircraft and aircraft technology systems increase in complexity, aviation academies need to explore predictors of flight performance. Enhancing program attributes at flight training academies is essential to the success of prospective students (Hankins, 2007). It seems essential that an effective feedback-data mechanism needs to be established from the training and operations departments to ensure that over time the right people are identified by the airline companies or military aviation with an ever-increasing reliability. Flight instructors may be able to judge specific performances of a candidate but due to the variable nature of factors, a standardized assessment is very difficult. Moreover, frequent changes in instructors, and insufficient experience and education of the instructors in the field of aptitude testing make it impossible to reliably diagnose important measuring dimensions such as personality traits, socio-interactive abilities and basic or composite mental abilities. This is the reason why some cadets may manage to get a license despite their weaknesses. As their flying experience increases, they may be able to compensate for their weaknesses in normal operational scenarios. However, in many cases, their deficiencies will not disappear and will resurface (several deficiencies even can compound or overlap) when encountering situations that demand high levels of performance (during times of fatigue, high operational complexity, unforeseen situations, emergencies/non-normal scenarios, etc.). Accordingly, the goal of this study was to contribute to improving the predictive capacity of early pilot selection, thus reducing training costs and improving operational capabilities. Indeed, when training is complex, costly, and dependent on specific abilities, such as is the case with aviation pilots, "Poor selection will result in increased training attrition, training requirements, and costs, and lead to poor job performance and poor organizational effectiveness" (Carretta & Ree, 2003, p. 359). Therefore, predicting pilot training success early has pragmatic and real-world benefits. Human performance is becoming even more relevant because of the accelerating innovation and technological advancements in the aviation industry. This process of continuous improvement induces changes in the job requirements and consequently drives the need for continuous adaptation of the pilot behavior and the airline training methodologies. The aviation industry must be able to produce safe pilots cost-effectively. Few empirical studies dealt with comparing pilot performance between digital and traditional cockpits (Whitehurst and Rantz, 2011; Smith, 2008; Wright, O' Hare, 2015). There is also limited aviation research literature regarding student performance predictors in university academic flight programs (Bell, 1998; Burrell, 1993; Kole, 2006). Flight performance predictors related to the practical operation of aircraft are vital to aviation safety and career success for the students in university flight programs. The importance of predictors of successful flight performance extends from university primary flight training to the selection and screening of commercial pilots in the aviation industry and up to the selection of military aviators who should follow the training in 4th generation helicopters and must succeed in high-risk special operations. Our study results can be used to calibrate and improve flight training to mental workload demands, and enhance effective use of simulated flight training, only to those who will have potentially a "poor" transition from analogue to digital cockpit. Moreover, training must go beyond simply providing pilots with facts about the digital display, as sometimes pilots possess knowledge in the sense of being able to recite facts, but that they are unable to apply the knowledge successfully in an actual flight context. This is called the problem of inert knowledge. Training must conditionalize knowledge to the contexts where it is utilized. Pilots need to learn not simply how the automated system works, but also how to work the system. This requires scenarios and instruction designed around managing the transitions between different modes of instrumentation. Pilots do learn a subset of methods to be able to make the system work under routine conditions, so situations that challenge their current understanding may arise relatively infrequently (or go unnoticed as such due in part to lack of feedback about the state and behavior of the digital system). This means that ongoing learning programs may need to be devised to help even experienced digital cockpit pilots discover and correct subtle bugs in their mental models or to elaborate their understanding of how the automation works situations in a risk-free environment. Performance feedback and continuous improvement loops are indispensable modules of the aptitude testing processes. This holds the potential to benefit all levels in the aviation industry, from individual student pilots and their instructors to airlines and agencies. Our study's objective goal was important because the relationship between cognitive criterion predictors and successful performance in transition from analogue to digital instrumentation could be key factors in future preadmission practices. Confronting the relationship that exists, it may help flight academy program administrators to provide assistance and accommodations, help students meet challenges, and improve learning environment for the students (Kole, 2006; Olson, 2002).

4.1 Limitations

A limitation concern was validity threats because of practice and ceiling effect. Practice effect is repeated practice of a task that over a period there is improvement in performance (Cozby, 2006, Shuttleworth, 2008; Shuttleworth, 2009). In the study, this limitation was minimized since it was not a repeated measures design. Instruction and practice over the allotted flight time among the sample was minimal and insignificant.

It is common practice to use human expert ratings as objective data, as we did in this study. Even though, the main advantage of the observer-rating technique, is its non-intrusive nature as it has no impact on the task being executed, it is doubtful that any observer can accurately rate the internal process of Situation Awareness (SA). A superior performance may not really equate to good SA. Some observable behaviors may suggest implication of SA, but the actual internal SA level cannot be precisely assessed by observation alone and therefore may be subject to bias. Consequently, the reliability of the performance's measure (response times) may also have been a limitation in that it was subjective and may have involved bias from the flight instructors' rating of the student flight performance. In future studies the Process indices procedure may be able to record, analyze and rate the processes that each subject follows to establish SA during the task performance of the experiment, by measuring the subject's eye movements during task execution. Eye-tracking devices can be employed to determine which situational elements the subject has fixated upon and evaluate how the subject's attention is allocated. Nevertheless, the use of an eye-tracking device outside of laboratory settings is not convenient. Furthermore, process indices have the indirect nature, i.e., the 'look but do not see' phenomenon by which the subject may fixate upon a certain environmental element but does not accurately perceive it. Maybe in the future when using this kind of procedure, researchers may be able to proceed in the verification of the correct answers from the indications of the instruments during the actual flight along with a flight instructor to avoid the look but no see phenomenon.

Another limitation of our study was the small sample of pilots, nevertheless the convenience sample and the nonrandom sampling of the participants. A convenience sample may potentially lead to bias in research results (Heiman, 2002). A convenience sample or a non-probability sampling is limited because the results cannot be applied to other groups (Gay & Afrasian, 2000). As a result, a caution in the study is that findings may be limited and should not be generalized to larger and more diversified aviation cohorts. Range restriction of the sample also led to limitations. A limitation in range restriction affects the correlation results with probable lower correlations as it was evidenced in our study (Bobko, 2001). In addition, predictive validity requires reliable criteria of job performance and data from a reasonably large sample of pilots. As a result, many of the non-significant findings may be due to a lack of statistical power. Small samples have the potential to produce statistical artifacts that artificially inflate correlations (Burke, Hobson, & Linsky, 1997) and are not robust to the violations of normality which larger sample sizes afford (Tabachnick & Fidell, 2007). Consequently, the results may not be an accurate representation of the associations in the population, and these relations should be re-examined in follow-on research with a larger sample size. Nevertheless, the small sample of flight students in this study was representative of this special population. However, to compensate for the small sample size, future research should involve concurrent validity studies with trained, experienced pilots, as well as longitudinal studies, which follow the pilot candidates past the training stage and into their actual flying careers. The challenge is to standardize the criteria across the various stages of a pilot's career (initial training, type rating training, command training), as these stages are controlled by different entities.

4.2 Ethical Dimensions

The ethical considerations in the study were the confidentiality of the participants and compliance regarding the transfer of private data of the students. Prior to data collection members of Global Aviation S.A reviewed and gave permission and approval for the research. For the study, a designated scorer entered the data results. Participant anonymity was protected by use of a coded identification of the demographic questionnaire, flight performance scores and COMPASS test battery data results. Encrypted passwords of the COMPASS computer data software program were also used to promote confidentiality for the participants. Prior to data collection, participants were given an informed consent form to identify that participation in the study was voluntary. The information within the form included a confirmation that participants could withdraw at any time and that data results would not affect their flight evaluations or grades if they did not participate in the research. Students were advised that once the research began there would be no further opportunities for them to participate in the testing. In addition, the Information Statement included information about testing procedures, debriefing options, and researcher contact information.

APPENDIX A: Information Statement

Information Statement

Principal Investigator:
Post – graduated Researcher:

Background information and invitation to participate
This project is aimed at pilots who are trained to fly airplanes. This project will examine the cognitive processes of pilots in an analogue and digital cockpit, during visual flight (VFR), in flight simulators.

Project and researcher interests
This experiment is being conducted to meet the requirements of the research project.

What participation will involve – time, effort, resources, costs, compensatory payments, etc
You will be required to complete flights in a simulator. The total experiment time will be approximately 30 minutes. The flight will require you to fly first as a pilot non-flying (PNF) in an analogue and digital cockpit and after as a pilot flying (PF) in an analogue and digital cockpit again. Additionally, you will be asked some questions during those flights. Flight Academy will provide the use of simulator at no cost.

Participant rights and interests – Risks & Benefits/Contingencies/Back-up Support
There are minimal risks associated with participating in this project. A potential benefit of participation is a greater understanding of your abilities and limitations, and experience while flying a plane with an analogue and a digital cockpit in a simulator.

Participant rights and interests – Free Consent/Withdrawal from Participation
Participation in this project is completely at your free will, and you may withdraw from participation at any time. There is no risk of penalty or repercussion from your decision to withdraw, and any recorded data will be removed at your request. Your consent to participate is acknowledged by completing and signing the attached “Demographic Questionnaire” form.

Participant rights and interests – Privacy & Confidentiality
All steps have, and will be, taken to ensure your privacy and confidentiality. Your signed consent form will be retained on file, while your background information questionnaire will be transposed (without identity information) into electronic/printed format. Additionally, all notes and results from the simulator session will be matched only by number with background data, to ensure you cannot be matched with your simulator outcome. All data will be password protected or stored in a locked filing cabinet.

Research output
The research data and conclusions reached will form part of the postgraduate research project for the above-named researcher. The data may also be used for publication in an applicable journal. In both possible outcomes no identifiable data or personal details will be published without your express written consent.

APPENDIX B: Demographic Questionnaire

DEMOGRAPHIC QUESTIONNAIRE

Age	
Country	**Language**
Gender	
Flight Experience (Yes or No)	**If any, how many hours**
Digital or analogue cockpit	
What license do you currently hold	**Include any additional ratings**

……/..…/…2022

APPENDIX C: Pilot Performance Questionnaire – PPQ

Pilot Performance Questionnaire – PPQ

PNF/PF

Pilot: ……………………………………………………………………………………
Evaluator: ……………………………………………………………………………………
Date: ……………………………………………………………………………………

CLIMBING LEG

1st Measure (Analog Cockpit) – S1

	Correct answer	Time	Record Check
Airspeed			
Climbing rate			
Heading			

2st Measure (Digital Cockpit) – S1'

	Correct answer	Time	Record Check
Airspeed			
Climbing rate			
Heading			

LEVEL FLIGHT LEG

1st Measure (Analog Cockpit) – S3

	Correct answer	Time	Record Check
Airspeed			
Altitude			
Heading			

2st Measure (Digital Cockpit) – S3'

	Correct answer	Time	Record Check
Airspeed			
Altitude			
Heading			

LEVEL FLIGHT LEG (360 Turn)

1st Measure (Analog Cockpit) – S4

	Correct answer	Time	Record Check
Airspeed			
Altitude			
Angle of Bank			

2st Measure (Digital Cockpit) – S4'

	Correct answer	Time	Record Check
Airspeed			
Altitude			
Angle of Bank			

DESCENTING LEG

1st Measure (Analog Cockpit) – S2

	Correct answer	Time	Record Check
Airspeed			
Rate of Descent			
Heading			

2st Measure (Digital Cockpit) – S2'

	Correct answer	Time	Record Check
Airspeed			
Rate of Descent			
Heading			

APPENDIX D: NASA TLX

Figure 8.6

NASA Task Load Index

Hart and Staveland's NASA Task Load Index (TLX) method assesses work load on five 7-point scales. Increments of high, medium and low estimates for each point result in 21 gradations on the scales.

Name	Task	Date

Mental Demand — How mentally demanding was the task?

Very Low — Very High

Physical Demand — How physically demanding was the task?

Very Low — Very High

Temporal Demand — How hurried or rushed was the pace of the task?

Very Low — Very High

Performance — How successful were you in accomplishing what you were asked to do?

Perfect — Failure

Effort — How hard did you have to work to accomplish your level of performance?

Very Low — Very High

Frustration — How insecure, discouraged, irritated, stressed, and annoyed wereyou?

Very Low — Very High

APPENDIX E: Flight Pattern

FLIGHT PATTERN

MEGARA AIRPORT (LGMG)

TRAFFIC PATTERN

ΥΠΟΜΝΗΜΑ
Σκέλος Ανόδου
Σκέλος ΕΟΠ
Σκέλος Καθόδου
Στροφή 360
Σημεία μετρήσεων

APPENDIX F: Affinity for Technology Interaction (ATI) Scale

Affinity for Technology Interaction (ATI) Scale

Franke, Attig, & Wessel (2019)

In the following questionnaire, we will ask you about your interaction with technical systems. The term "technical systems" refers to apps and other software applications, as well as entire digital devices (e.g., mobile phone, computer, TV, car navigation).

	Please indicate the degree to which you agree/disagree with the following statements.	completely disagree	largely disagree	slightly disagree	slightly agree	largely agree	completely agree
01	I like to occupy myself in greater detail with technical systems.	☐	☐	☐	☐	☐	☐
02	I like testing the functions of new technical systems.	☐	☐	☐	☐	☐	☐
03	I predominantly deal with technical systems because I have to.	☐	☐	☐	☐	☐	☐
04	When I have a new technical system in front of me, I try it out intensively.	☐	☐	☐	☐	☐	☐
05	I enjoy spending time becoming acquainted with a new technical system.	☐	☐	☐	☐	☐	☐
06	It is enough for me that a technical system works; I don't care how or why.	☐	☐	☐	☐	☐	☐
07	I try to understand how a technical system exactly works.	☐	☐	☐	☐	☐	☐
08	It is enough for me to know the basic functions of a technical system.	☐	☐	☐	☐	☐	☐
09	I try to make full use of the capabilities of a technical system.	☐	☐	☐	☐	☐	☐

References

Attig C.; Wessel D.; Franke T. (2017). Assessing Personality Differences in Human-Technology Interaction: An Overview of Key Self-report Scales to Predict Successful Interaction. *Springer International Publishing* AG 2017. http://dx.doi.org/ 10.1007/978-3-319-58750-9_3

Bailey, N. R.; Scerbo, M. W.; Freeman, F. G.; Mikulka, P. J.; Scott, L. A. (2003). A brain-based adaptive automation system and situation awareness: The role of complacency potential, *Proceedings of the Human Factors and Ergonomics Society 47*: 1048–1052. http://dx.doi.org/10.1177/154193120304700901

Bailey, N. R.; Scerbo, M. W.; Freeman, F. G.; Mikulka, P. J.; Scott, L. A. (2006). A comparison of a brain-based adaptive system and a manual adaptable system for invoking automation, *Human Factors 48(4)*: 693–709. http://dx.doi.org/10.1518/001872006779166280

Bruder, C., and Hasse, C. (2019). Differences between experts and novices in the monitoring of automated systems. *International Journal of Industrial Ergonomics 72 (2019)* 1–11 German Aerospace Center (DLR), Department of Aviation and Space Psychology, Sportallee 54, 22335. https://doi.org/10.1177/1541931213571065

Cabeza, G., Molesworh, B., Good M., Caponecchia, C., Steffensen, R. (2021). Investigating the predictive validity of the COMPASS pilot selection test. *The International Journal of Aerospace Psychology.* http://dx.doi.org/10.1080/24721840.2021.1885297

Cacioppo, J., T. and Petty, R., E. (1982). The Need for Cognition. *Journal of Personality and Social Psychology Copyright 1982 by the American Psychological Association, Inc. 1982, Vol. 42*, No. 1, 116-131

Coelho, L.de H.; Hanel, P.H.P & Wolf, L.J. (2018). The very efficient assessment of need for cognition: developing a six-item version *Journals.sagepub.com/home/asm.* http://dx.doi.org/ 10.1177/1073191118793208

Cook, R I; Woods, D D; Mc Colhgan E & Howle, MB. (1990). Cognitive consequences of clumsy automation on high workload, high consequence human performance' Fourth *Annual Workshop on Space Operations, Applications and Research '90 NASA, Washington DC, pp 543-546*, 199 (NASA Report CP-3103).

Dafermos, B., (2011). Social statistics and research methodology with SPSS, Ziti Publications, Athens.

Endsley, M. R. (1989). Pilot situation awareness: The challenge for the training community. *In Proceedings of the Interservice/Industry Training Systems Conference* (pp. 111–117). Fort Worth, TX: American Defense Preparedness Association.

Endsley, M. R. (1993). A survey of situation awareness requirements in air-to-air combat fighters. *International Journal of Aviation Psychology, 3*, 157–168.

Endsley, M. R. (1995a). Measurement of situation awareness in dynamic systems. *Human Factors, 37*, 65–84.

Endsley, M. R. (1995b). Toward a theory of situation awareness in dynamic systems. *Human Factors, 37*, 32–64.

Endsley, M. R., & Garland, D. G. (2000). Pilot situation awareness training in general aviation. *In Proceedings of the 14th triennial congress of the International Ergonomics Association and the 44th annual meeting of the Human Factors and Ergonomics Society* (pp. 2/357–2/360). Santa Monica, CA: Human Factors and Ergonomics Society.

Endsley, M. R., Garland, D. J., Shook, R. W. C., Coello, J., & Bandiero, M. (2000). Situation awareness problems in general aviation (SA Tech. Rep. No. 00-01). *Marietta, GA: SA Technologies. Federal Aviation Administration. (2005). General aviation (Chapter V).* Retrieved from http://www.faa.gov/data_statistics/aviation/aerospace_forecasts/2005–2016/media/ch5.pdf

Evans, C.J.; Kirby J.R. & Fabrigar, L.R. (2003). Approaches to learning, need for cognition, and strategic flexibility among university students. *British Journal of Educational Psychology, 73*, 507–528.

Feltovich, P. J., Spiro, R. J. & Coulson, R.L. (1997). Issues of expert flexibility in contexts characterized by complexity and change. *Chapter 5. Human and machine (pp.125-146).* Retrieved from http://www.researchgate.net/publication/232465540

Field, A. (2013). Discovering statistics using IBM SPSS Statistics, 4th Edition handbook. SAGE Publications Ltd.

Fink, M., R., W., Ma, X., Traue H., C. (2015). Trust in digital technology : Reliability and Validity. *International Symposium on Companion Technology, September 23*-25, 2015.

Franke, T., Attig, C., & Wessel D., (2018): A Personal Resource for Technology Interaction: Development and Validation of the Affinity for Technology Interaction (ATI) Scale. *International Journal of Human–Computer Interaction*. http://dx.doi.org/10.1080/10447318.2018.1456150

Gawron, V. J. (2000). Human Performance Measures handbook. Mahwaw, NJ*: Lawrence Erlbaum Associates.*

Glaholt, M. G. (2014). Eye tracking in the cockpit: A review of the relationships between eye movements and the aviator's cognitive state. *Ottawa, Canada: Defence Research and Development Canada.*

Gnardellis, C. (2013). Data analysis with IBM SPSS Statistics 21. Papazisis Publicatios.

Hamblin, C. J., Gilmore, C., & Chaparro, A. (2006). Learning to fly glass cockpits requires a new cognitive model. *In Proceedings of the Human Factors and Ergonomics Society Annual*

Meeting (Vol. 50, No. 17, pp. 1977-1981). Los Angeles, CA: Sage Publications.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.),*Human mental workload* (pp. 139–183). North-Holland. https://doi.org/10.1016/S0166-4115(08)62386-9

Hart, S. G. (2006)."NASA-Task Load Index (NASA-TLX); 20 Years Later". *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 50 (9): 904–908. http://dx.doi:10.1177/154193120605000909.S2CID6292200

Jones, R. E., Milton, J. L., & Fitts, P. M. (1949). Eye fixations of aircraft pilots, I. A review of prior eye-movement studies and a description of a technique for recording the frequency, duration and sequences of eye-fixations during instrument flight. (USAF Tech. Rep, 5837). Dayton, OH: Wright Patterson Air Force Base.

Kalavsky, P., Rozenberg, R., Mikula, B., Zgodavona, Z. (2018). Pilots' Performance in Changing from Analogue to Glass Cockpits. *Proceedings of 22nd International Scientific Conference. Transport Means.*

Lewin, K. (1939). Field theory and experiment in social psychology: Concepts and methods. *American Journal of Sociology, 44*, 868–896. http://dx.doi.org/10.1086/218177

Li, W. C., Chiu, F. C., Kuo, Y. S., & Wu, K. J. (2013). The investigation of visual attention and workload by experts and novices in the cockpit. In D. Harris (ed.), Engineering Psychology and Cognitive Ergonomics: *Applications and Services. Springer: Berlin Heidelberg,* pp. 167-176.

Martins, A. P. G. (2016). A review of important cognitive concepts in aviation. *Aviation 20,* 65–84. http://dx.doi.org/10.3846/16487788.2016.1196559

Mumaw, R.J., Sarter, N.B., Wickens, C.D. (2001). Analysis of pilots' monitoring and performance on an automated flight deck. *11th International Symposium on Aviation Psychology. Columbus,* Ohio State University.

NASA. (1986). NASA task load index (TLX): Paper-and-pencil version. Retrieved from https://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20000021488.pdf

Palmer, C. R. (2007). Applying Human Factors Principles In Aviation Displays: A Transition From Analog to Digital Cockpit Displays In The CP140 Aurora Aircraft. *Master's Thesis, University of Tennessee.* https://trace.tennessee.edu/utk_gradthes/185

Pingali, S., Mc Mahon, T.W., Newman D. (2016). Comparison of pilot visual scan patterns in glass vs analogue cockpits. *87th Annual Scientific Meeting AsMA.*

Rens Van de S., and Milica, M. (2020). Small Sample Size Solutions. A Guide for Applied Researchers and Practitioners. *Published by Routledge of the Taylor & Francis Group.*

Sarter, N.B., Mumaw, R.J., Wickens, C.D. (2007). Pilots' Monitoring Strategies and Performance on

Automated Flight Decks: An Empirical Study Combining Behavioral and Eye-Tracking Data. *Human Factors*, Vol. 49, No. 3, June 2007, pp. 347–357. http://dx.10.1518/001872007X196685

Socha, V., Socha, L., Hanakova, L., Valenta, V., Kusmirek, S., and Lalis, A. (2020). Pilots' Performance and Workload Assessment: Transition from Analogue to Glass-Cockpit. *Appl. Sci. 2020,* 10, 5211.
http://dx.doi.org/10.3390/app10155211

Smith, C.E. (2008). Glass cockpit transition training in collegiate aviation: Analog to digital. *Msc Graduate College of Bowling Green State University*.

Thevaki, K., Yuhan, L., De Roza, C., Loi En Qi, J., and Yip Kam Luen, G. (2021). Validation of Computerised Aptitude Selection System (Compass) In Predicting Success of Uav Applicants in the Republic of Singapore Air Force (Rsaf). *88th International Symposium on Aviation Psychology, 316-321*. https://corescholar.libraries.wright.edu/isap_2021/53

Woods, D.D. (1993). Price of flexibility in intelligent interfaces. Cognitive Systems. *Engineering Laboratory, Knowledge-Based Systems, Volume 6*, 1-8.

Woods, D.D. & Sarter, N.B. (1998). Learning from automation surprises and "going sour" accidents: Progress on human-centered automation. *Cognitive Engineering in Aerospace Applications* (Cooperative research agreement NCC 2-592). Moffett Field, CA: NASA-Ames Research Center.

Wiener, E.L. (1989). Human factors of advanced technology ("glass cockpit") transport aircraft (*Tech Report 177528). Moffet Field, CA: NASA Ames Research Center*.

Whitehurst, G., & Rantz, W. (2011). Digital training to analog flying: Assessing the risks of a stark transition. *Journal of Aviation/Aerospace Education & Research, 20(3),* 13-16.

Whitehurst, G., & Rantz, W. (2012). The digital to analog risk: Should we teach new dogs old tricks? *Journal of Aviation/Aerospace Education & Research, 21(3)*, 17-22.

Wright, S. & O'Hare, D. (2015). Can a glass cockpit display help (or hinder) performance of novices in simulated flight training? *Applied Ergonomics 47 (2015)* 292 - 299.