

# Examining the Use of Model-Based Work Instructions in the Aviation Maintenance Environment

Nathan W. Hartman<sup>1</sup> and Timothy D. Ropp<sup>2</sup>

<sup>1</sup> Purdue University, West Lafayette, IN USA  
nhartman@purdue.edu

<sup>2</sup> Purdue University, West Lafayette, IN USA  
tropp@purdue.edu

**Abstract.** A fundamental tenet of product lifecycle management (PLM) environments is the use of high-fidelity, 3D product models. The capability to create models with high degrees of fidelity to the physical world has driven companies to extract as much benefit and use from these digital assets as possible throughout the design, production, and support stages of the lifecycle. This is particularly apparent in the aviation industry where aircraft lifecycles routinely reach 80 years or longer. As the aviation industry migrates to the use of 3D model-based communications mechanisms in lieu of 2D drawings, multiple factors will impact the use of digital model-based work instructions, including the device, the form of the product model data, and levels of detail in geometry and interactivity. This paper will present a series of short studies conducted over the last three years using novice university students and expert university staff aircraft mechanics to evaluate the use of model-based work instructions in a general aviation maintenance environment. The results indicate that varying levels of detail and levels of interactivity have an effect on number of errors, time on task, and mental workload.

**Keywords:** product lifecycle management, aviation maintenance, model-based product definition, model-based work instructions, user experience, mobile computing, undergraduate research.

## 1 Introduction

The U.S. Federal Aviation Administration's (FAA) Next Generation Air Transportation System (NextGen), which transforms the U.S. national airspace system [1,2], has rapidly advanced the evolution of intelligent, networked aircraft and their sophisticated sustainment and support systems. This technology shift places tremendous pressure on aircraft maintenance and engineering organizations, as well as the individual technician, who require more access to detailed product information pipelines to work on more advanced and integrated air vehicle systems, minimize downtime, and meet unyielding air worthiness and quality requirements. This requires innovative data support systems integrated with product data definitions delivered to the point of maintenance. While electronic signoff and networked maintenance data capabilities currently exist within the aircraft maintenance industry [3,4], model based work instructions are neither widespread

nor standardized; however, they are being realized more and more as key tools for the 21<sup>st</sup> century technician's use. Despite the tremendous computerized capabilities of modern aircraft and existing electronic maintenance networks containing aircraft technical manuals and diagrams, maintenance job tasks are still largely accomplished and tracked using manual methods such as job task "signoff" on paper-based work instructions in larger proportion than complete electronic systems [5].

It has been noted that properly applied visualization and component presentation of technical or complex systems is critical for improving daily maintenance tasks in both efficiency and accuracy [6]. Specifically, use of high fidelity model images on demand at the point of maintenance (for example a 3D image of a wing tip fairing installation) has been noted to result in less rework or missed steps. By leveraging a model-based product definition, MRO technicians would have access to the most relevant product geometry and accompanying metadata [7,8,9,10]. Moreover, in an estimate of the financial cost of maintenance errors to industry, Markou & Kalimat [11] concluded the worldwide maintenance expenditure to be \$45.2B in 2008. Maintenance expenses represent approximately 10%-15% of an airline's operational cost.

This paper describes a series of experiments conducted in the context of a senior-level undergraduate research course at Purdue University. These studies involve the use of undergraduate students in an Aeronautical Engineering Technology program (novices) and University staff mechanics (experts) charged with the maintenance and support of the University's fleet of aircraft. All product data was acquired through the use of scanning and modeling technologies, and the accompanying technical documentation for the University's Boeing 727-200 records.

## 2 Experiment Frameworks

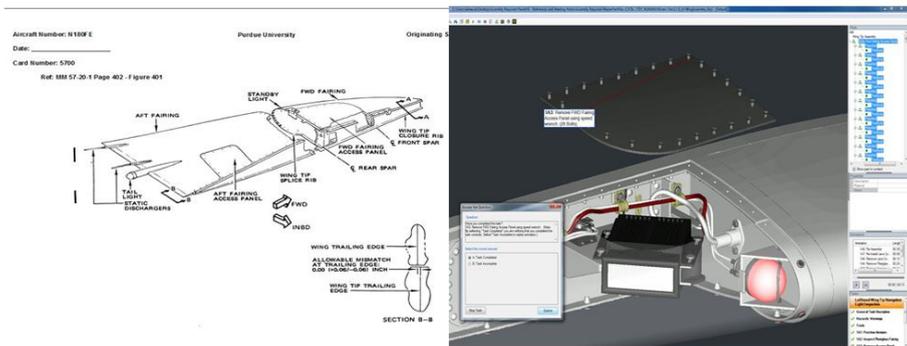
These pilot studies examined three critical characteristics associated with aircraft Maintenance, Repair and Overhaul (MRO) environments – time on task, error rate, and mental workload. As aerospace vehicle developers and maintenance firms leverage 3D assets beyond design and into assembly and service of the product it is important to decipher how these variables impact safety, airworthiness, and stability of the air vehicle.

Research frameworks in this study evolved on a number of dimensions. The first study examined differences between typical paper-based work instructions versus a non-interactive 3D animation of the same procedure. The study progressed to a comparison between non-interactive and interactive 3D graphics (as well as paper-based instructions). The third study used on 3D graphics (interactive and non-interactive), included variance in devices used to display the work instructions. The fourth study addressed varying levels of detail within the 3D model-based presentations as a way to influence the experimental variables and the computing capability on the devices used. Mobile devices and tablets have become increasingly popular, in the design, manufacturing and MRO space to disseminate technical data. Much work remains to assess their effectiveness in this environment as compared to traditional computers and flat-panel monitors used in the same way. Studies 3 and 4 began to assess these issues. A summary of the studies is included in Table 1.

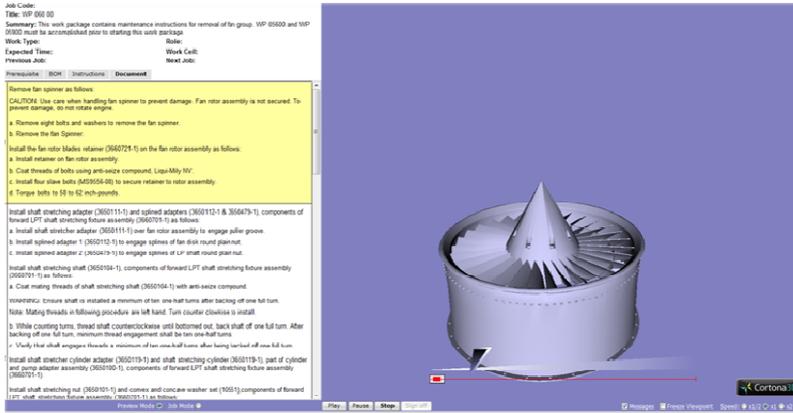
As is the case with many tasks, differences in experts and novices data interpretation is essential to understand if aviation maintenance is to become more efficient and more accurate. Experts and novices in Purdue University’s Aviation Technology department participated in the experiments. Study participants with more experience began to combine and omit assembly steps in the test procedure, indicative of expert behavior in a specific domain [12]. Experts in the study were FAA certified Purdue Aviation Technician staff with at least ten years of experience. According to Ericsson [13], ten years of experience is accepted in many industries. The novices consisted of Aeronautical Engineering Technology students with classroom and laboratory experience, but have yet to acquire an FAA certification. The NASA-TLX (Task Load Index) was used in the post-questionnaire in an effort to make connections between modality of the work instructions and the mental work load exerted by the participant [14]. Figures 1-4 show the various aircraft subassemblies used in these studies, with a brief description of how each one was used.

**Table 1.** Summary of Research Study Frameworks

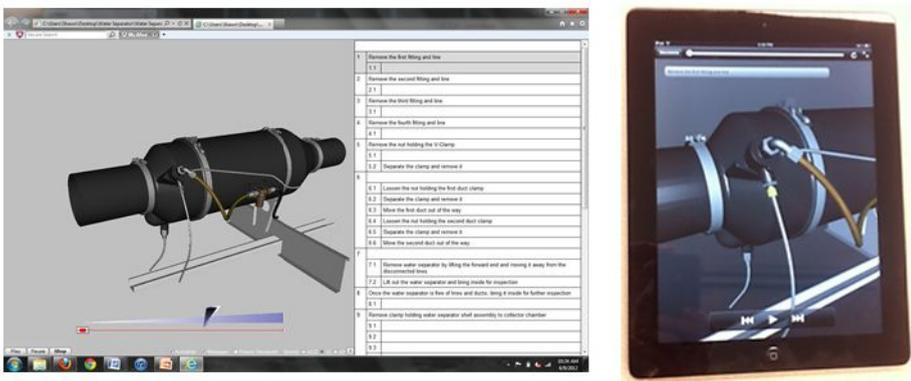
Research Study Frameworks					
	Experience Level	Group size	Device Type	Mode (2D or 3D)	Interactivity
Study 1	Novice Expert	8 8	Laptop, tablet	2D paper, 3D digital	2D paper, continuous loop video
Study 2	Novice Expert	9 7	Laptop, tablet	2D paper, 3D digital	2D paper, 3D non-interactive, 3D interactive
Study 3	Novice	28	Desktop computer, iPad	3D digital	3D non-interactive, 3D interactive
Study 4	Novice	35	Desktop	3D digital	High LOD vs. Low, LOD; continuous loop presentation vs. step-wise presentation



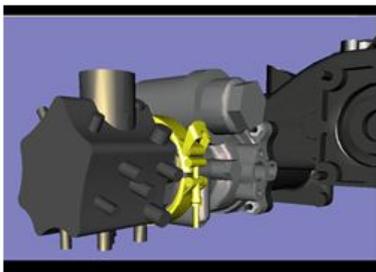
**Fig. 1.** Left wing-tip fairing position light from Boeing 727-200 Assembly instructions from 2D and 3D job task card



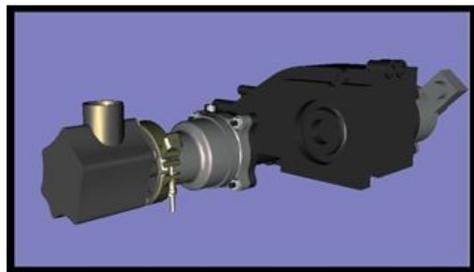
**Fig. 2.** Front Fan Assembly of F-109 Engine. Comparing 3D interactive (NGRAIN), 3D static (Cortona 3D), and 2D (paper) Job Task Cards.



**Fig. 3.** Water separator unit on Boeing 727-200. Comparing Mobile and Stationary Job Task Cards



High Level of Detail



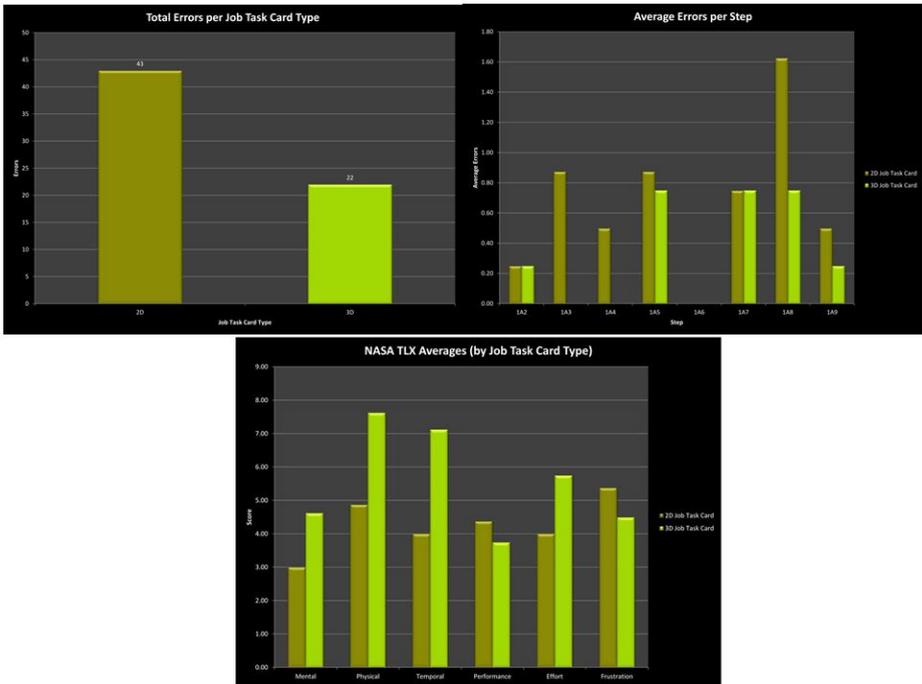
Low Level of Detail

**Fig. 4.** Fuel Pump and Gear Box from F-109 Engine. Comparing High and Low Levels of Detail and Interactivity of the CAD Assembly Work Instructions.

### 3 Results and Discussion

#### *Study 1: Comparing 2D and 3D Job Task Cards*

Leveraging the capabilities of personal computing devices and 3D graphics data could help reduce common errors potentially impacting air vehicle safety and improve efficiency of aviation maintenance technicians. Times to completion for each step and for the overall process between test groups showed no statistically significant difference. However, completion times of those subjects using the 3D job task card showed less variance than those using the 2D job task card. These results differed from those in an industrial case study [15]. After removing results determined to be outliers, range between the minimum and maximum completion times for the 2D job task card was three times that of the 3D job task card range. Mental workload, as determined by the NASA TLX, also showed no statistically significant difference between test groups. Statistically significant difference between test groups ( $p < .05$ ;  $p = .033$ ) was found in the total number of procedural and assembly errors made by participants. The group using the 2D job task card made twice as many errors as the group using the 3D job task card (Figure 5).

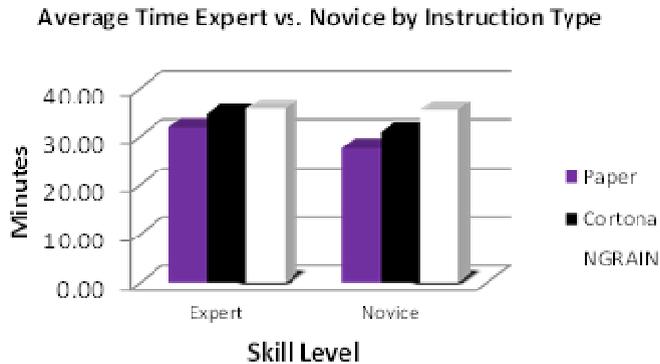


**Fig. 5.** Results from 2D vs. 3D Task Card Comparisons

#### *Study 2: Comparing 3D interactive (NGRAIN), 3D static (Cortona 3D), and 2D (paper) Job Task Cards*

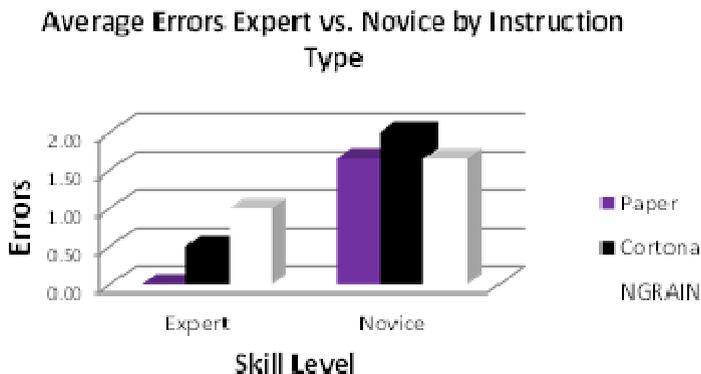
There were 16 participants total in the study, 7 of which were experts and 9 were novice. Due to the small sample size, results were analyzed using a non-parametric

method called the Kruskal-Wallis test. A p-value of .05 was used to evaluate statistical significance. The average time for experts was higher than that of the novices, although statistically insignificant. Between types of work instructions (paper, 3D interactive, 3D non-interactive), there was no significant difference in time; however, paper was found to be slightly faster than 3D static, and 3D static was marginally faster than 3D interactive in both experts and novices (Figure 6).



**Fig. 6.** Time on Task between Experts and Novices

As expected, the average number of errors for experts was lower than for novices. The experts made more errors on the 3D interactive mode than any other type of work instructions; whereas the average number of errors on the 3D static mode for novices was slightly higher than the other types of work instructions, although statistically insignificant. The majority of the NASA-TLX results were also found to be statistically insignificant. Modality of the work instructions significantly affected how the novice participants felt they had performed the task. Performance of novice participants using 3D static work instructions was significantly lower than those using 3D interactive and paper (Figure 7).

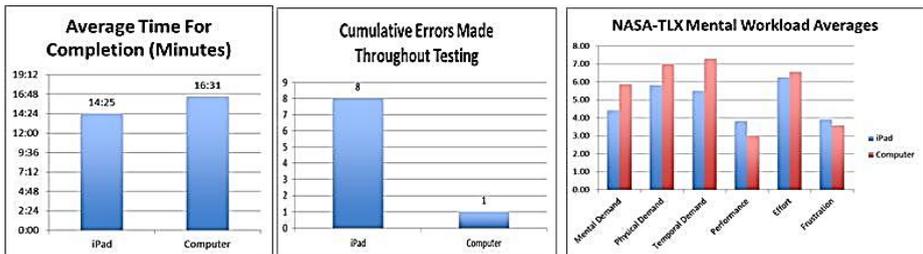


**Fig. 7.** Number of Errors between Experts and Novices

In the post questionnaire, participants were asked to comment on the method they were tested on. In summary, most of the participants felt the 3D work instructions were more effective than the 2D work instructions. One novice technician stated, “There was less room for error, as you have to verify completion of each and every step” in reference to the NGRAIN product. In contrast, one of the experts said that NGRAIN was “too slow for mind function” because he prefers to read through the entire task first rather than do it step by step. In addition, several of the participants, experts and novices alike, indicated graphics helped clarify the text instructions if they were unclear. One comment about the 2D paper work instructions was that it needed more pictures.

*Study 3: Comparing Mobile and Stationary versions of a 3D Job Task Card*

As with the previous studies, the 28 participants in this study were junior- or senior-level undergraduate students in the Aviation Technology program, with an emphasis in their studies of Maintenance, Repair and Overhaul (MRO) skills and system safety. The results of the study were analyzed using ANCOVA due to the selection methods and sample sizes. See Figure 10 for a comparison of the data between mobile and stationary platforms. A p-value of .05 was used to establish statistical significance. The average time for completion when using the iPad was found to be faster than using the computer by more than two minutes. The p-value found for this difference was .0565. This showed the difference in time for completion was just shy of statistical significance. Eight errors were made throughout the testing process for the iPad participants and one for those using the computer. The p-value for these differences was .028. This value showed the amount of errors made was statistically significant indicating iPad users had more errors. Finally, the difference in mental workload scores yielded a p-value of .3638. This value showed that there was no statistical significance in mental workload when comparing the computer and iPad task cards. See Figure 8.

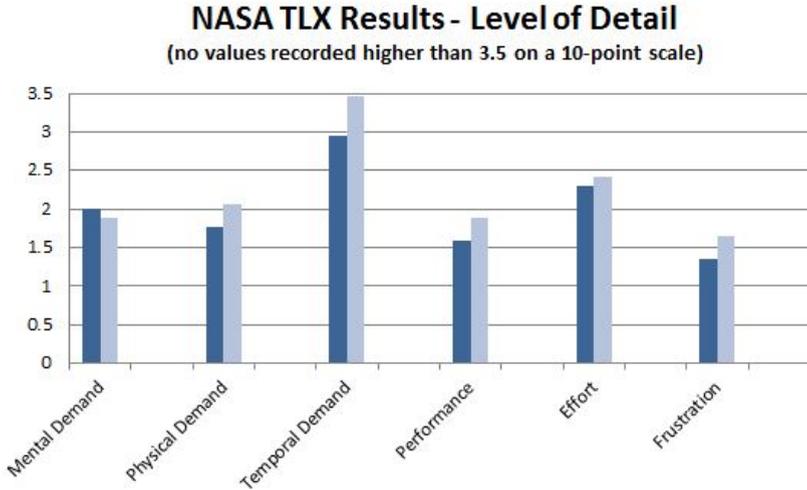


**Fig. 8.** Comparison of Data for Mobile and Stationary Platforms

*Study 4: Comparing Level of Detail and Presentation Mode for Graphics in the Work Instruction*

Results comparing levels of detail are presented first. The first was mental workload measured by assessing the NASA TLX test. The largest difference was in the temporal demand section with low level of detail being 5.3% higher than high level of

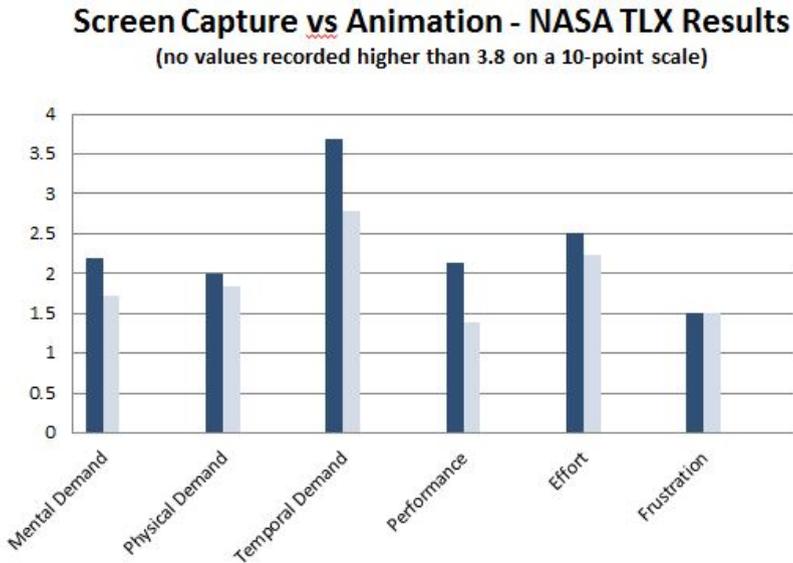
detail. While analyzing the data, it was predicted that providing a simple disassembly contributed to the low mental workload. Evaluation revealed a t-value of .318, indicating the difference in mental workload was not significant. Figure 9 shows mental workload data from the level of detail comparison.



**Fig. 9.** NASA-TLX Results for Level of Detail Comparison

The second variable tested was time on task, assessed by starting the timer as soon as test subjects began viewing the work instruction and stopped when test subjects stated they were finished. Subjects were required to verbally express when they were finished to confirm they fully understood the beginning and end of the work instruction. The difference for average time on task was minimal as high level of detail took 5:37.2 and low level of detail took 5:30.8. The t-value found for difference in time on task was .142, indicating no statistical significance and the null hypothesis cannot be rejected. The third variable tested was number of errors. After testing was complete, five errors for high level of detail and five errors for low level of detail were recorded. The t-value for difference in number of errors was .142. This indicated the difference in number of errors was not statistically significant, and the null hypothesis cannot be rejected.

The NASA TLX was utilized to measure mental workload variable for the delivery method similar to level of detail. The largest difference was in the temporal demand section with animation being 9.1% higher than screen capture. Figure 10 shows the results of this analysis. After analysis, it is likely a short disassembly process contributed to the low mental workload. Test subjects retained nearly all information by viewing only one time. The t-value was 1.985, indicating that the difference in mental workload (NASA TLX) was not significant. Number of errors was another variable tested. Four errors for the animation and six errors for the screen capture were recorded. The t-value was .442. The final variable tested was time on task. The procedure was done in the same manner, with little discrepancy noted between animation and screen capture. Dynamic animation averaged 5:41.1 and screen capture averaged 5:27.3. The t-value found for difference in time was .214, indicating once again no statistical significance and the null hypothesis cannot be rejected.



**Fig. 10.** NASA-TLX Results for Dynamic Presentation of Work Instructions

## 4 Summary

In Study 1, the original intent was to integrate a maintenance job task card into a digital 3D environment assessing performance impact on maintenance technicians. While test subject performance was measured by time to completion, sum of errors made, and mental workload, the only result acquired from experimentation with statistical significance was the total amount of errors made between the two job task card forms. Moreover, these errors are considered human error. According to Rankin & Allen [16], 20% - 30% of engine in flight shutdowns are caused by maintenance error and can cost an estimated \$500,000 per shutdown. Reducing the amount of human error in aviation maintenance processes can allow companies to realize tremendous cost savings. Considering that maintenance expenses make up 10%-15% of airline operational costs [11], a reduction in maintenance expenses would allow airlines to allocate resources elsewhere.

Study 2 involved the comparison of 3D interactive, 3D static, and 2D paper forms of the work instructions. Study 3 also involved the comparison of expert and novice mechanics. Novice technicians had a broader computer background than the experts, including CAD experience and familiarity with CAD interfaces. 3D interactive and non-interactive maintenance manuals were positively received by both expert and novice technicians, with Cortona being the most popular. They expressed the helpfulness of the 3D graphics and animation verbally and during the questionnaire. Despite this, there was one very puzzling result pertaining to 3D interactive work instructions and experts.

The number of errors committed by expert technicians was much higher than expected. It was believed that, since NGRAIN requires the user to stop and confirm

each step was completed, users would commit fewer errors than they would with the paper based or non-interactive manual. A possible explanation for this result could be the experts' use of chunking [12]. Since experts have been shown to gather task information together and cluster it as they go, it is likely many of them read the steps provided in NGRAIN, skipping the completion checks, and clustered the tasks together in a way they saw fit. In most instances of expertise, experts display a high ability to perform tasks almost without thinking about them by mixing procedural and declarative knowledge in strategic ways. If these maintenance experts had used the paper-based methods for a long time, it is highly likely that a level of automaticity had developed in their mental processing of the information. Anything that interrupted that level of performance (i.e., the introduction of a 3D mode for viewing work instructions) would likely result in a performance decrement.

The two alternatives used 3D graphics and animations providing visual communication of the task at hand to accompany the basic written instructions. The two products developed included an NGRAIN animation and a 3D PDF document. Participant feedback showed preference for 3D graphics-based support documentation of the newly developed job task cards. Participants agreed that task visualization was a tremendous instructional help. In Study 3, mobile and stationary model-based instructions were investigated. The iPad is still a fairly new technology, and a number of subjects mentioned more familiarity using laptops for productivity and iPads for entertainment. This may explain why so many errors were made. Portions of the NASA-TLX mental workload suggested this as well.

Study 4 examined multiple areas –delivery platform, form of the product model data, and levels of detail in geometry and interactivity – and gathered data on number of errors, time on task, and mental workload. If integration of model-based work instructions in aviation maintenance is to be useful, industry must address challenges with infrastructure and computing architecture, the level of graphics preparation needed, human information processing, and deployment platforms. There were no significant statistical differences between the high and low level of detail work instructions. These results are contrary to the initial hypothesis and research. The research indicated that removing unnecessary visual information provided the potential for improved learner efficiency and reduced cognitive load [17].

Similarly, animation based work instructions yielded no statistically significant results when compared to screen captures. These results were also contrary to initial predictions [18]. Most test subjects had experience working with similar assemblies resulting in high mechanical reasoning abilities. Subjective comments from test subjects imply both delivery methods have positives and negatives. Some subjects felt step by step clicking was an advantage, allowing for slower comprehension and feeling “more in control” of the work instruction. A negative expressed was the inability to see the parts come apart. This is where the animation work instruction gained more approval. Test subjects in favor of animation favored ability to see dynamic movement of parts occur. They said, “This improved understanding of how the parts went together and came apart.” It is evident that both methods have advantages and disadvantages. The challenge for the future is developing a process that effectively uses the correct method in each situation.

Study 4 did not find statistically significant differences in regards to a specific delivery method or level of detail. In order to test as many subjects as possible, the

work instructions were condensed. Allowing assessment of all 34 available test subjects, but results were very similar between subjects. This could suggest that simple maintenance procedures and tasks are minimally affected by the delivery method and level of detail. Industry professionals and data suggest that component complexity and assembly frequency are critical factors impacting work instructions. Additionally, certain regulatory guidelines require assembly technicians to have both hands available to be engaged in work while performing a task, thereby rendering them unable to hold a tablet or other mobile device while working. Future work should concentrate on interaction between variables associated with geometric level of detail and specific display modalities, as well as the positioning of the computing device in the assembly work environment.

**Acknowledgements.** The authors would like to acknowledge the following research student team members involved in this study: Bret Angel, Gilbert Bracey, Anthony Butcher, Zachary Carnahan, Nathan Christopher, Kelsey Crowe, Nick Delcore, Bret Gastineau, Brandon Hall, Matt Harris, Andrew Klaassen, Patrick McGuire, Ian Pack, Matthew Pizzatto, John Pourcho, Josh Rathke, Nick Rohe, Shawn Ruemler, Joshua Schliessman, Samantha Schreiber, Jacob Seeley, Stephen Trimboli, and Athan Valaskatgis,

## References

1. FAA – JPDO, Federal Aviation Administration – Joint Planning and Development Office (FAA-JPDO). “Safety Culture Improvement Guide” V.1.0. Paper number 08-010. p. 22 (2008)
2. Federal Aviation Administration, FAA NextGen Implementation Plan (2010)
3. ATE&M, Online maintenance tools: Online help for aircraft operations and maintenance staff. Aircraft Technology and Engineering Maintenance (ATE&M), (102), p. 70 (October-November 2009)
4. Boeing (2010). Boeing Commercial Aviation Services website. Maintenance Services page. Airplane Health Management: Managing Information to Improve Operational Decision Making (2010), [http://www.boeing.com/commercial/ams/mss/brochures/airplane\\_health.html](http://www.boeing.com/commercial/ams/mss/brochures/airplane_health.html) (retrieved October 18)
5. Block, E.E., Ropp, T.D., Specca, T.P.: Implementing digital technologies and network systems into aircraft maintenance. In: Proceedings of the 1st International Conference on Aviation Information Technology in Engineering and Management, New Orleans, LA, March 22-24 (2010)
6. Kim, S., Woo, I., Maciejewski, R., Ebert, D.S., Ropp, T.D., Thomas, K.: Evaluating the effectiveness of visualization techniques for schematic diagrams in maintenance tasks. In: Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization, Session: displays and visualization, Los Angeles, California, pp. 33–40 (2010) ISBN:978-1-4503-0248-7
7. Hartman, N., Rosche, P., Fischer, K.: A framework for evaluating collaborative product representations in product lifecycle workflows. In: Rivest, L., Bouras, A., Louhichi, B. (eds.) PLM 2012. IFIP AICT, vol. 388, pp. 424–434. Springer, Heidelberg (2012)

8. Lubell, J., Hartman, N.W., Cheney, D.: Model-based engineering standardization and validation. NISTIR 7796, National Institute of Standards and Technology (2011), [http://www.nist.gov/manuscript-publication-search.cfm?pub\\_id=908106](http://www.nist.gov/manuscript-publication-search.cfm?pub_id=908106)
9. Versprille, K.: Model-based Definition for the Masses, Collaborative Product Development Associates (2008), [http://www.righthemisphere.com/\\_base/static/img/whitepapers/2\\_cpda\\_rh\\_wp.pdf](http://www.righthemisphere.com/_base/static/img/whitepapers/2_cpda_rh_wp.pdf) (retrieved January 27, 2013)
10. Quintana, V., Rivest, L., Pellerin, R., Venne, F., Kheddouci, F.: Will model-based definition replace engineering drawings throughout the product lifecycle? A global perspective from aerospace industry. *Computers in Industry* 61(5), 497–508 (2010)
11. Markou, C., Kalimat, T.: IATA Airline Maintenance Costs An Executive Commentary (May 2010), [http://www.iata.org/workgroups/Documents/MCTF/Form41\\_2010Report\\_2010.10.27.pdf](http://www.iata.org/workgroups/Documents/MCTF/Form41_2010Report_2010.10.27.pdf) (retrieved)
12. Egan, D.E., Schwartz, B.J.: Chunking in recall of symbolic drawings. *Memory and Cognition* 7(2), 149–158 (1979)
13. Ericsson, A.K., Prietula, M.J., Cokely, E.T.: The making of an expert. *Harvard Business Review* (June 7, 2007), <http://www.coachingmanagement.nl/The%20Making%20of%20an%20Expert.pdf> (retrieved)
14. NASA, NASA TLX: Task Load Index (2006), <http://human-factors.arc.nasa.gov/groups/TLX/paperpencil.html> (retrieved)
15. NGRAIN, Field Case study: Leveraging ngrain-enabled IETMs for Performance Support (2009), [http://www.ngrain.com/pdfs/CaseStudy-20LeveragingNGRAIN-enabledIETMsforPerformanceSupport\(CanadianAirForce\).pdf](http://www.ngrain.com/pdfs/CaseStudy-20LeveragingNGRAIN-enabledIETMsforPerformanceSupport(CanadianAirForce).pdf) (retrieved)
16. Rankin, W., Hibit, R., Allen, J., Sargent, R.: Development and evaluation of maintenance error decision aid (MEDA) process. *International Journal of Industrial Ergonomics* 26, 261–276 (2000)
17. Allen, C.: The effects of visual complexity on cognitive load as influenced by field dependency and spatial ability. New York University (2011), <http://gradworks.umi.com/3478262.pdf>
18. Boucheix, J., Schneider, E.: Static and animated presentations in learning dynamic mechanical systems. *Learning and Instruction* 19, 112–127 (2009), <http://elsevier.com/locate/learninstruc> (retrieved)