SAND2019-0996C

Human performance differences between drawing-based and model-based reference materials

Siobhan M. Heiden[i] & Eric M. Moyer[1]

[1] Sandia National Laboratories, Albuquerque, New Mexico, USA
{smheide, emmoyer}@sandia.gov

Abstract. The Sandia National Laboratories Human Factors team designed and executed an experiment to quantify the differences between 2D and 3D reference materials with task performance. A between-subjects design was used where 27 participants were randomly assigned a 2D or 3D reference material condition. The experimental tasks required participants to interpret, locate and report dimensions on their assigned reference material. Performance was measured by accuracy of task completion and time-to-complete. After all experimental tasks were completed, cognitive workload data were collected. Response times were longer in the 3D condition vice the 2D. However, no differences were found between conditions with respect to response accuracy and cognitive workload which may indicate no negative cognitive impacts concerning the sole use of 3D reference materials in the work-place. This paper concludes with possible future efforts to address the limitations of this experiment and to explore the mechanisms behind the findings of this work.

Keywords: Human performance · Reference materials · Model-based engineering · Human-systems Integration

1 Introduction

The Model Authorized Product Realization (MAP-R) project is a United States Department of Energy project designed to characterize and analyze the differences between a traditional drawing-based design and manufacturing process versus a model-based design and manufacturing process. Literature can be found suggesting significant cost and time savings can be realized when shifting design and manufacturing processes from drawing-based to model-based (Quintana et al. 2010). Its ultimate purpose is to identify exactly what and where, within the process flow, a model-based process could contribute and improve on the traditional, currently-used engineering methods. Given the advent of computer numeric control (CNC) machining and additive manufacturing, many claim that design and manufacturing processes would be better supported by 3D reference materials (models) compared to the 2D reference materials (drawings) relied on in the past (Hudspeth 2006; Lamit 2007). However, no published work has directly compared and examined the different reference materials with respect to operator performance, operator cognitive workload, and usability for the design and manufacture

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology & Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

of products. In order to advance National Security Enterprise's (NSE) ability to quantify the differences between the current drawing-based (2D reference materials) process and the future model-based (3D reference materials) process, the research team designed and executed an experiment.

The purpose of this experiment was to compare the reference materials used for drawing-based (2D reference materials) and model-based (3D reference materials) design and manufacturing processes. Specifically, this experiment aimed to quantify the differences between 2D and 3D reference materials with respect to operators performing tasks requiring these reference materials, and to identify opportunities for design improvements in usability. Factors that were measured to assess reference material use were performance (based on success of task completion and time-to-complete), cognitive (or mental) workload, and usability. The findings of this work inform several areas, including considerations for reference material design and composition, operator training requirements, and work task design.

2 Methods

This section describes the study design, participant demographics, measurement tools, the experimental procedure, and the data analysis methods used.

2.1 Study design

Participants were asked to complete a series of tasks requiring them to find and report dimensions from digital 2D or 3D reference materials for an object, specifically a flange connector. A between-subjects design was used, meaning that participants were randomly assigned to just a single condition: 2D or 3D reference materials. Both reference material types illustrated the same part (MAP-R); the 2D and 3D versions are shown in Figures 1 and 2, respectively. Half of the questions were solely text-based; the other half included a supporting image directing participants to the dimension of interest. The same questions were posed regardless of the condition, except to note different views for each reference material type. For example:

> ***2D reference***: Provide the distance between the top surface of the part (where T-Bird Logo is located) to the surface where the top holes are located. **(Refer to Section B-B)**
>
> *or*
>
> ***3D reference***: Provide the distance between the top surface of the part (where T-Bird Logo is located) to the surface where the top holes are located. **(Refer to View 7A_ENVELOPE)**

2.2 Participants

A total of 27 participants (10 female) completed the experiment; 14 (4 female) of which were given the 2D condition, and the remaining 13 participants (6 female) were given the 3D condition. Most participants (15 out of 27; 8 with 2D condition, 7 with 3D condition) had no classroom experience with spatially referenced materials, and 4 total participants (2 per condition) had used spatially referenced materials in five or more classes. 41% (11) of all participants had no work experience using spatially referenced materials and a disproportionate number of those participants were given the 3D condition (8 participants) versus the 2D condition (3 participants). Using the ranges from the demographic questionnaire, we can conservatively estimate that the total work experience of participants given the 2D condition is 11 years greater than those given the 3D condition.

2.3 Measurement tools and materials

The following measurement tools used in this experiment, and the data collected by each are listed below:

- ***Demographic questionnaire***: Collected basic demographic information about participants (e.g. age, gender) and their experience with reference materials in classroom and workplace settings.
- ***Mental Rotation Test (MRT)***: Assessed the participant's spatial abilities (Vandenberg & Kuse 1978); a shortened version was created for this experiment by ensuring that the subset of questions chosen had the same proportion of questions as the original test with respect to item difficulty and item characteristics (Caissie, Vigneau, & Bors 2009).
- ***National Aeronautics and Space Administration (NASA) Task Load Index (TLX)***: Assessed the mental workload (MWL) experienced by participants while completing the experimental tasks. Scores on the NASA TLX range from 1 to 100; scores greater than or equal to 75 indicate "very high" workload levels (Hart & Staveland 1988).
- ***Usability and Preference questionnaire***: Collected data about participants' perception of the usability of their assigned reference material (2D or 3D) and their preference for reference material format. The usability portion of the questionnaire was derived from the Computer System Usability Questionnaire (Lewis 1995). The questionnaire included a total of 23 items, 15 of which were survey items aimed at measuring satisfaction/agreement with using the assigned reference material. 5 were survey items aimed at obtaining preferences concerning 3D and 2D reference materials in general. Finally, comment fields were presented where participants were asked to list up to 3 negative aspects, 3 positive aspects and general comments about their assigned reference material.

Additional materials used in this experiment were:

- ***E-Prime 3.0 software***: Presented the training questions (6) and experiment questions (20), and collected participants' responses and response times (RTs) for each question (Psychology Software Tools, Pittsburgh, PA)
  - Experiment questions with images: 1, 4, 6, 9, 11, 13, 14, 16, 18, 19
  - Experiment questions without images: 2, 3, 5 7, 8, 10, 12, 15, 17, 20

- ***Training PowerPoint slides***: Presented information to the participant on how to read the respective reference material, including line types, notation, and tolerances; presented to them prior to training and experiment questions
- ***Notes sheet:*** Summarized reference material notation for participant to use during the experiment
- ***Navigation sheet:*** Designated the various views of the reference material for participant to use during the experiment

2.4 Experimental procedure

The experimental procedure was 1) introduction and consent; 2) demographic questionnaire administration; 3) shorted version of MRT administration; 4) facilitator-led training on reading reference materials (PowerPoint slides): 5) training question practice (participant permitted to ask facilitator questions); 6) experiment questions (participants not permitted to ask facilitator questions); 7) NASA TLX administration; 8) Usability and preference questionnaire administration; 9) summary and close-out.

2.5 Data analysis

Descriptive statistics (mean, median, standard deviation) were calculated for overall RTs, accuracy, MWL, and spatial skills (based on MRT score) and with respect to each condition (2D and 3D). For RT data, data points that were three standard deviations from the mean with respect to the experiment question and the assigned condition (7 data points) and any outliers based on experiment log notes (e.g. bathroom break, emergency evacuation practice) (7 additional data points) were excluded from the analyses. RT, accuracy, and MWL data were statistically analyzed by condition (2D or 3D) using one-tailed t-tests (α=0.05), specifically two-sample t-tests assuming unequal variances. The findings from the Usability and Preference questionnaire data are not reported here.

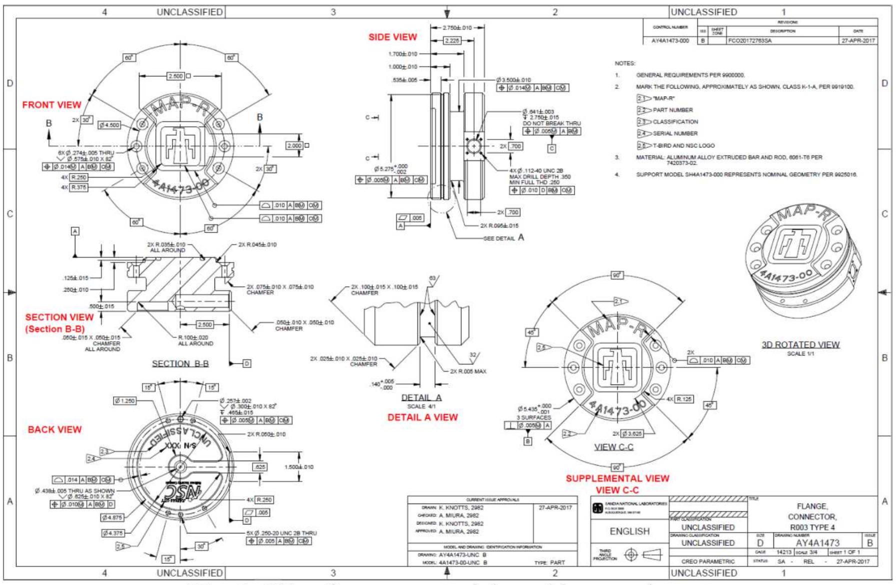

Fig. 1. 2D reference material used in experiment

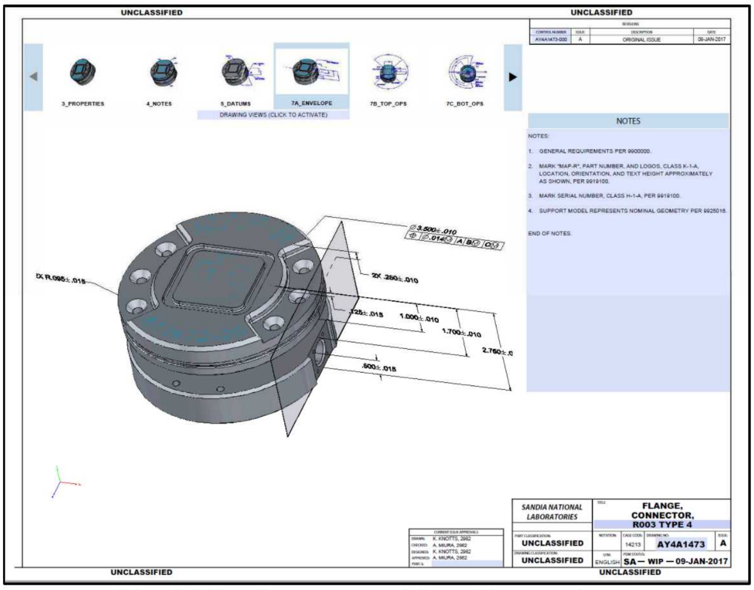

Fig. 2. 3D reference material used in experiment

3 Findings and results

This section contains the quantitative results and qualitative findings of the experiment and our analyses. The first section includes the descriptive statistics and t-test results for RT, accuracy, and MRT scores. The second section describes the MWL (NASA TLX scores) results.

3.1 Descriptive statistics and t-test results

Descriptive Statistics. The descriptive statistics (mean, median, and standard deviation) for RT, accuracy, and MRT scores are displayed in **Error! Reference source not found.**, **Error! Reference source not found.**, and **Error! Reference source not found.** respectively. RTs for the 3D condition had higher mean, median, and standard deviation than the 2D condition (**Error! Reference source not found.**). Participants in the 2D condition had higher mean and median accuracy rates (**Error! Reference source not found.**) and MRT scores (**Error! Reference source not found.**).

Table 1. Descriptive statistics for overall average response time in milliseconds.

Overall average response time (seconds)	Overall	2D	3D
Mean	41.842	32.658	51.025
Median	35.126	27.673	45.797
Standard deviation	24.637	17.344	27.704

Table 2. Descriptive statistics for accuracy rates (percentage of questions accurately answered).

Accuracy rate (%)	Overall	2D	3D
Mean	82.3%	82.9%	81.7%
Median	84.6%	85.7%	83.3%
Standard deviation	10.5%	14.2%	10.0%

Table 3. Descriptive statistics for Mental Rotation Test score.

Mental Rotation Test score	Overall	2D	3D
Mean	20.22	22.29	18.00
Median	24	24	19
Standard deviation	6.59	4.94	7.57

Note: Highest MRT score possible = 26; lowest MRT score possible = -26

T-test results. The higher mean MRT scores for participants with the 2D condition (as described in the above section) was statistically significant (p=0.050) in a one-tail t-test. However, no statistically significant differences were found in accuracy between conditions in a one-tail test (p=0.258).

Response times were analyzed further to determine whether there were statistically significant differences between the two conditions. One-tailed t-tests confirmed that the mean RT for the 3D condition was larger than the 2D condition (p=0.009) and for specific experiment questions (5, 6, 9, 12, 13, 18, 19, 20); -test statistics for each of these questions are presented in

Table 4. t-test statistics for experimental questions in which Response Times (in milliseconds) for 3D condition were longer than the 2D condition.

	Q5	Q6	Q9	Q12	Q13	Q18	Q19	Q20
2D mean	29242	26105	14752	18379	17437	15105	17043	35125
3D mean	151180	45547	28153	27092	34494	46047	33021	65784
t statistic	3.396	2.482	2.423	1.892	2.916	3.424	2.384	1.909
P(T≤t) one-tail	.003	.013	0.014	0.039	.005	.002	.015	.039
t Critical one-tail	1.796	1.753	1.753	1.753	1.740	1.761	1.746	1.761

Table 5. t-test statistics for experimental questions in which Response Times (in milliseconds) for 2D condition were longer than the 3D condition.

Mental Rotation Test score	Q15
2D mean	62080
3D mean	32981
t statistic	2.295
P(T≤t) one-tail	0.018
t Critical one-tail	1.746

. For one experiment question (15), the RT was longer for the 2D condition; test statistics are presented in **Error! Reference source not found.**.

Table 4. t-test statistics for experimental questions in which Response Times (in milliseconds) for 3D condition were longer than the 2D condition.

	Q5	Q6	Q9	Q12	Q13	Q18	Q19	Q20
2D mean	29242	26105	14752	18379	17437	15105	17043	35125
3D mean	151180	45547	28153	27092	34494	46047	33021	65784
t statistic	3.396	2.482	2.423	1.892	2.916	3.424	2.384	1.909
P(T≤t) one-tail	.003	.013	0.014	0.039	.005	.002	.015	.039
t Critical one-tail	1.796	1.753	1.753	1.753	1.740	1.761	1.746	1.761

Table 5. t-test statistics for experimental questions in which Response Times (in milliseconds) for 2D condition were longer than the 3D condition.

Mental Rotation Test score	Q15
2D mean	62080
3D mean	32981
t statistic	2.295
P(T≤t) one-tail	0.018
t Critical one-tail	1.746

3.2 Workload Results

As mentioned in Section 0, the researchers administered the NASA TLX to all participants after the experiment to uncover any workload challenges indicative of usability issues. **Table 6.** Descriptive statistics for mental workload based on weighted NASA TLX scale.

Mental Rotation Test score	Overall	2D	3D
Mean	56.0	56.4	55.5
Median	59	60	42
Standard deviation	12.9	11.2	15.3

below shows the TLX results. Note that while the mean and median were higher in the 2D condition, the difference was not statistically significant (p=0.515 and p=0.878, respectively) in a two-tailed t-test. Meaning participants experienced similar workload levels while working with each reference material. Additionally, the scores produced indicate that the levels of workload experienced were not "very high" (score of 75 or more).

Table 6. Descriptive statistics for mental workload based on weighted NASA TLX scale.

Mental Rotation Test score	Overall	2D	3D
Mean	56.0	56.4	55.5
Median	59	60	42
Standard deviation	12.9	11.2	15.3

4 Limitations and future work

This section notes key limitations of this experiment and describes potential next steps for future work.

4.1 Limitations

There are several limitations to this study that must be noted. First, the 2D and 3D reference materials (Figures 1 and 2, respectively) were original reference materials for a test part, meaning these materials were not modified for experimental use. This was done to maintain the authenticity of the materials and provide an analog to real-world use. However, the experimental tasks (i.e. finding and reporting on dimensions) may not have been valid representations of real-world use and/or may not have been difficult enough to detect differences in performance or MWL.

Another key limitation was the heterogeneous nature of our participants, who came from a variety of education and professional backgrounds, and with large variance in years of education and experience with reference materials. Additionally, experimental conditions were assigned at random, and demographic analysis revealed that the two experimental groups were not balanced with respect to education and experience with reference materials.

4.2. Future work

The results of this experiment have prompted several potential "next-steps" should opportunities arise in the future. First, more than half of our participants were non-engineers (17) and they were all selected from Sandia National Laboratories across various groups. This provided a glimpse into the performance and preferences of individuals at this organization in general but not those of who would actually be working with these materials. It would be a benefit to perform this study with participants who are real users (e.g. engineers working in parts design and inspection) and with groups at other Department of Energy (DOE) locations and laboratories. This would ensure we had an accurate picture of the performance and preferences of the individuals expected to work with the 3D reference materials.

Next, as mentioned in our Limitations section, our task was dependent on feature searching, rather than feature inspection. Experimental tasks more oriented to feature inspection would likely serve the advanced 3D capabilities of the 3D.pdf better and more clearly highlight its utility in that area. This would probably affect performance in favor of the 3D reference materials over the 2D one since our current task was slightly hampered by some usability challenges with finding information in the 3D condition. Beyond usability challenges, we hope to better understand the inherent benefits that 3D part viewing seems to provide inspectors.

Finally, we would like to re-test the 3D reference materials after any changes are made based on our usability recommendations (reported elsewhere) to determine whether those changes truly improve users' ability and efficiency in performing the tasks, and/or improve their perceptions of the 3D reference materials.

5 Conclusions

The experimental data clearly show two things in terms of performance for our feature searching and measurement finding task: 1) There is no significant difference between the 2D and 3D reference materials in terms of accuracy performance, but 2) there is a significant difference in response time for using the materials with 2D being "faster" than 3D. For the first conclusion, this bodes well for the use of 3D reference materials in the sense that the ability to find the information and use the 3D reference material will produce similarly accurate results. However, the discrepancy for response time is most likely due to how the 3D reference material is arranged and the controls provided. Since the 3D reference material is inherently more complicated than the 2D one, it does take a bit more time for the user to orient themselves to it and to subsequently find information. Future work could include redesigning experimental tasks to focus more on feature inspection rather than feature search, and running the experiment with more participants, particularly engineers and technicians, who better match our expected end users of these reference materials in design and manufacturing.

Acknowledgments. We thank the MAP-R project and members for providing the funding and impetus for this effort; Chris Forsythe for his expertise and input on measuring participants' spatial skills; Aaron Jones and Paul Schutte for their participation and review of our pilot experiment runs and experimental materials; Danielle Sanchez and

Paul Schutte for help in facilitating the experiment runs; and Megan Greenwald-Yarnell for her thorough and speedy peer review of the final document.

References

1. Quintana, V. et al. (2010). Will Model-based Definition replace engineering drawings throughout the product lifecycle? A global perspective from aerospace industry. *Computers in Industry*, 61 (5), pp. 497-508.
2. Hudspeth, M. (2006). Transitioning from 2D to 3D: the benefits of switching to 3D far outweigh what 2D offers. *Cadalyst*, 23 (9), pp. 39-41
3. Lamit, L. (2007). Moving from 2D to 3D CAD for Engineering Design: Challenges and Opportunities. BookSurge Publishing, pp. 5.
4. Caissie, A. F., Vigneau, F., & Bors, D. A. (2009). What does the Mental Rotation Test Measure? An Analysis of Item Difficulty and Item Characteristics. *The Open Psychology Journal, 2*(1), 94-102. doi:10.2174/1874350100902010094
5. Hart, S. G. & Staveland, L. E. (1988) Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati (Eds.) Human Mental Workload. Amsterdam: North Holland Press.
6. Lewis, J.R. (1995) IBM Computer Usability Satisfaction Questionnaires. *International Journal of Human-Computer Interaction*, 7(1), 57-78.
7. Psychology Software Tools, Inc. [E-Prime 3.0]. (2016). Retrieved from http://www.pstnet.com.
8. Vandenberg, S. G., & Kuse, A. R. (1978). Mental Rotations, a Group Test of Three-Dimensional Spatial Visualization. *Perceptual and Motor Skills, 47*(2), 599-604. doi:10.2466/pms.1978.47.2.599