

HOW PIPEFITTERS OBTAIN VISUAL INFORMATION FROM CONSTRUCTION ASSEMBLY DRAWINGS

SUBMITTED: July 2021

REVISED: January 2022

PUBLISHED: April 2022

EDITOR: Robert Amor

DOI: [10.36680/j.itcon.2022.015](https://doi.org/10.36680/j.itcon.2022.015)

Matthew Sears, PhD Candidate

Department of Civil, Environmental and Architectural Engineering, University of Colorado Boulder, USA

email: Matthew.H.Sears@colorado.edu

ORCID: <https://orcid.org/0000-0002-4629-7300>

Omar Alruwaythi, Assistant Professor

Department of Civil Engineering, Taibah University, Saudi Arabia

email: oruwaythi@taibahu.edu.sa

ORCID: <https://orcid.org/0000-0002-9587-0580>

Paul M. Goodrum, Professor and Department Head

Department of Construction Management, Colorado State University, USA

email: Paul.Goodrum@colostate.edu

ORCID: <https://orcid.org/0000-0002-5656-1240>

SUMMARY: *An extensive framework has been developed for studying the behavior of motor vehicle drivers using eye tracking technology. Previous work has revealed strong relationships between driver eye movements and performance, which has resulted in widely accepted guidance within the transportation industry. In this work, the same eye tracking analysis methods were applied to investigate 20 professional pipefitters' interactions with traditional isometric assembly drawings during a construction pipe model assembly task, in order to begin to understand the strategies that construction craft professionals use to gather visual information from engineering deliverables. A custom web application was developed to quantify and compare the pipefitters' interactions with the assembly drawings through several visit metrics. Results indicated that the pipefitters' interactions with the assembly drawings were associated with their performance and spatial cognition; however, the results did not suggest that the pipefitters were adhering to any particular visual information gathering strategies. The authors also investigated whether age or industry experience were associated with differences in visual information gathering strategies, but no significant relationships were observed. The primary contribution of this work is a demonstration of how existing eye tracking analysis methods can be applied to investigate how construction craft professionals extract visual information from engineering deliverables.*

KEYWORDS: *Eye tracking; construction; productivity*

REFERENCE: *Matthew Sears, Omar Alruwaythi, Paul M. Goodrum (2022). How pipefitters obtain visual information from construction assembly drawings. Journal of Information Technology in Construction (ITcon), Vol. 27, pg. 290-311, DOI: [10.36680/j.itcon.2022.015](https://doi.org/10.36680/j.itcon.2022.015)*

COPYRIGHT: © 2022 The author(s). This is an open access article distributed under the terms of the Creative Commons Attribution 4.0 International (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.



1. INTRODUCTION

Several recent construction model assembly studies have demonstrated that individuals obtain information from engineering deliverables differently. Spatial cognition and information format are associated with the way that individuals interact with engineering information (Dadi et al. 2014a; b; c; d). Additionally, construction industry experience and information formats are associated with construction craft professional performance (Alruwaythi and Goodrum 2019; Goodrum et al. 2016; Sweany et al. 2016). The authors propose that craft professionals have underlying strategies that they use to obtain visual information from engineering deliverables, and that those strategies may differ based upon industry experience, or other demographic factors. The authors also propose that certain aspects of visual information gathering strategies may be associated with performance. Thus, in the present work, existing eye tracking analysis methods have been applied and expanded upon to gain a better understanding of the strategies that craft professionals use to extract visual information from construction assembly drawings.

1.1 Previous construction model assembly studies

Several recent studies have examined the influence of construction engineering information during scale model assembly tasks. Significant findings are presented below in chronological order and are summarized in Table 1.

TABLE 1. Summary of previous construction model assembly task studies

Year	Study	Significant Findings
2014	Building assembly task (Dadi et al.)	• Engineering Information format associated with cognitive workload • Cognitive workload associated with performance • No single information format was superior for all participants • Workers had preferences for specific information formats • Information format preferences were based upon participants' previous experiences
2016	Building assembly task (Sweany et al.)	• Engineering information format associated with performance, even after controlling for differences in spatial cognition
2016	Pipe spool assembly task (Goodrum et al.)	• Performance deficiencies of below average spatial cognition workers were eliminated when 2D information was supplemented with 3D information
2018	Mixed Reality for electrical construction (Chalhoub and Ayer)	• Mixed reality information improved assembly time, rework, and errors made assembling electrical conduit, compared to traditional 2D paper drawings • Individuals with most experience using paper plans performed best with paper, while individuals with least experience using paper plans performed best with mixed reality • Most of the participants preferred the mixed reality information format and thought that mixed reality made the task easier than using traditional 2D paper plans
2019	Pipe spool assembly task with eye tracking (Alruwaythi and Goodrum)	• When only 2D information was provided, below average spatial cognition workers utilized the drawings more • When 2D information was supplemented with 3D information, below average spatial cognition workers utilized the 3D information more than the above average spatial cognition workers

1.1.1 2014 Building assembly task

Dadi et al. (2014a, d) tasked construction practitioners with assembling a small, plastic model from various information formats, including: traditional 2D drawings, a 3D CAD model, and a 3D physical model. After assembling the model, participant cognitive workload for the task was measured using the NASA raw task load index (NASA-rTLX). The NASA-rTLX is a widely used, subjective self-assessment questionnaire used to assess perceived workload. It consists of six factors, which combine to produce a single composite workload score. No overall relationship was found between information format and the six factors of the NASA-rTLX. However, the composite workload score was significantly lower for the 3D physical model than for the 2D drawings or the 3D

CAD model (Dadi et al. 2014d; b). Several demographic factors were also collected during the study that had a significant relationship with the outcomes of the task. Participant years of work experience, CAD experience, and frequency of use of drawings in their job were all found to influence the performance, mental demand, and time demand of the participants. Lower mental-workload scores also led to better productivity outcomes.

A post-test questionnaire was also presented to the study participants to assess their engineering information format preferences (Dadi et al. 2014c). Interestingly, while the 3D physical model led to the lowest composite cognitive workload and best performance, it was not the most preferred information format. The 2D drawings were preferred by 46% of the study participants, while 39% preferred the 3D physical model, and only 15% preferred the 3D CAD model. When asked why they preferred one format over another, most of the participants stated that their preference was based upon their familiarity with a format, or the ease of use of a format.

1.1.2 2016 Building assembly task

The assembly task study was repeated with a different model and the addition of tests to measure participant spatial cognition (Sweany et al. 2016). 2D spatial cognition was measured with a card rotation test, and 3D spatial cognition was measured with a cube comparison test (Ekstrom et al. 1976). 2D drawings were again found to be associated with the worst performance, even after controlling for differences in 2D and 3D spatial cognition. While spatial cognition measures were to control comparisons across individuals with similar levels of spatial cognition, the authors did not attempt to determine if spatial cognition was associated with performance due to limits in sample size.

1.1.3 2016 Pipe spool assembly task

Another model assembly task was conducted by (Goodrum et al. 2016). In this study, professional pipefitters assembled a ½" diameter PVC pipe spool assembly. Spatial cognition was measured (using both card rotation and cube comparison tests described in section 2.3.1 herein), and participants received engineering information in three different formats: 2D drawings, 2D drawings supplemented with 3D images, and 2D drawings supplemented with a 3D printed model. Participants with below average levels of spatial cognition were found to perform significantly worse than individuals with above average levels of spatial cognition when using the 2D drawings. However, when the 2D drawings were supplemented with 3D information, the individuals with below average spatial cognition performed just as well as individuals with above average spatial cognition.

1.1.4 2018 Mixed Reality for electrical construction

Eighteen electrical practitioners participated in an electrical conduit assembly task (Chalhoub and Ayer 2018). In this task, each participant assembled two similar electrical conduit assemblies, one using a Microsoft HoloLens mixed reality headset, and the other using traditional 2D paper drawings. The participants had to assemble the conduit components, as well as place the entire assemblies in the correct location and orientation within a room. On average, participants using the mixed reality headset were able to complete an assembly in less than half of the time required by the participants using the paper plans; the difference in assembly time was highly significant. Participants using mixed reality also made fewer mistakes during assembly (rework), and they also made fewer final errors.

Prior experience with paper plans was also important for assembly time. The participants with >10 years of industry experience were the fastest performers with paper plans, and the participants with <1 year of industry experience were the fastest performers using mixed reality.

The researchers also had the study participants complete a post-session questionnaire. All of the participants indicated that they agreed with, strongly agreed with, or had a neutral stance on the following statements: 1) "With Mixed Reality, I can effectively build electrical conduit without the need for traditional paper documentation," 2) "It is easier to build conduit using Mixed Reality than Paper Plans," 3) "It would be easier for inexperienced individuals to build electrical conduit with mixed reality than with paper plans." Additionally, only one participant disagreed with the statement, "I would rather use Mixed Reality than use Paper plans for assembling pre-fabricated electrical conduit."

1.1.5 2019 Pipe spool assembly task with eye tracking

A derivative of the 2016 pipe spool assembly study was developed by Alruwaythi and Goodrum (2019) and included one of the first efforts to investigate *how* craftworkers interact with 2D engineering information. The pipe

spool model used by Goodrum et al. (2016) was relatively simple, so a more complex model was developed by Alruwaythi and Goodrum (2019), in order to more closely resemble real-world conditions and increase the potential for confusion, rework, and errors during model assembly. Relationships between performance and spatial cognition in Alruwaythi's study were consistent with the previous findings by Goodrum et al (2016). Alruwaythi also fitted participants with eye tracking glasses to record precisely what they looked at while assembling the more complex model. With the recorded eye tracking data, Alruwaythi was able to determine the amount of time that participants utilized 2D information and 3D information.

When only 2D information was used, participants with below average spatial cognition looked at the drawings more than the participants with above average spatial cognition. This is unsurprising because the below average spatial cognition participants required more time to complete the task. When the 2D drawings were supplemented with 3D information, there were no differences in the amount of time that the participants looked at the 2D drawings. However, the below average spatial cognition participants looked at the 3D information significantly more than the above average spatial cognition participants. This was true for each of the ten 3D images, and for the 3D printed model. Alruwaythi also claimed that participants with below average spatial cognition consistently looked at the 3D printed model more than participants with above average spatial cognition for each of the ten 2D drawings. However, the 3D printed model was a single object, so it's unclear how fixations on the 3D printed model were associated with the ten individual 2D drawings. Nevertheless, the results were consistent with the previous findings by Goodrum et al. (2016) which suggested that below average spatial cognition participants benefited from 3D information used not as a substitute to 2D information, but as a supplement to 2D information since 3D information was provided to help augment 2D information provided to them.

1.2 Use of Eye Tracking to Monitor Participant Behavior with Complex Tasks in Other Industries

Alruwaythi and Goodrum (2019) assessed the fixation counts, total fixation durations, and revisits captured in the eye tracking data collected during the pipe spool assembly task, metrics that are found frequently throughout eye tracking literature. However, these three metrics are very similar to one another, and they represent some of the highest-level metrics for eye tracking data analyses. More advanced eye tracking analysis methods have been developed, including international standard ISO 15007, which provides guidance for the "Measurement of driver visual behavior with respect to transport information and control systems" (ISO 2014a; b). The methods described in the ISO 15007 standard are primarily concerned with the number of times that a motor vehicle driver "glances" away from the "road scene ahead," to look at a side mirror, roadway sign, cell phone, etc., and the duration of each glance. The number of glances and duration of glances form a glance duration distribution, and many motor vehicle driver behavior studies have analyzed glance duration distributions, resulting in widely accepted industry guidance for driver safety and performance. For example, a driver should not glance away from the road ahead for more than two seconds while operating a motor vehicle, because this behavior drastically increases the likelihood of an accident (Driver Focus-Telematics Working Group 2006; Horrey and Wickens 2007; Liang et al. 2012; NHTSA 2016; Tivesten and Dozza 2014; Zhang et al. 2006).

2 METHODS

2.1 Existing eye tracking analysis methods from Other Industries

ISO 15007 as well as previous research has developed a number specific terms and metrics that are commonly used in eye tracking analyses. These terms and metrics are used throughout the paper, including its analyses.

Stimulus image- In experiments with mobile eye tracking glasses, objects of interest are typically physical objects, for example, a printed construction assembly drawing. A stimulus image is a digital representation of a physical object and eye tracking data points are mapped from recorded fixation points on physical objects to corresponding points on stimulus images. The physical assembly drawings used in the pipe spool assembly study were printed from ten digital stimulus images, such as Figure 1.

Area of Interest (AOI) - An area of interest is a specific region within a stimulus image, for example, the title block of a construction assembly drawing. An area of interest can be any shape. In these experiments, each stimulus image in the pipe spool assembly task included a single AOI that covered the entire drawing, so the study had a total of 10 AOI's. As such, the terms stimulus image and AOI may be used interchangeably for this particular study.

Gaze point -Eye tracking data collection devices record where a participant is looking periodically. A common recording frequency for eye tracking devices is 60Hz, which records data approximately every 17ms. When recorded data is mapped back to a point on a stimulus image, the points on the stimulus image are known as gaze points. At a minimum, a gaze point includes an X coordinate, a Y coordinate, and a timestamp.

Fixation - Eye tracking devices generally record gaze points faster than a participant is able to look from one location to another. As such, raw eye tracking data will often include many sequential gaze points clustered in a single location. Eye tracking software algorithms detect when sequential gaze points are located at a single location and lump them together as a single fixation point. A single fixation typically has a duration of approximately 300ms.

Saccade - Huey (1900) stated that, “Eye movements typically do not occur as smooth movements, but rather ‘quick jerks’ from one fixation location to the next.” These “quick jerks,” or rapid eye movements, are known as saccades.

Glance - ISO 15007-1:2014 defines a glance as, “maintaining of visual gaze within an area of interest, bounded by the perimeter of the area of interest; may be comprised of more than one fixation and saccades to and from it. Its duration is measured as glance duration.”

Glance duration - ISO 15007-1:2014 defines glance duration as, “Time from the moment at which the direction of gaze moves towards an area of interest to the moment it moves away from it.” In other words, glance duration includes the time from the beginning of the saccade leading into the AOI until the beginning of the saccade trailing out of the AOI.

Visit - A visit can be defined as all eye tracking data that is collected starting with the initial fixation inside a specific AOI to the last fixation within the same AOI (Tobii Pro AB 2019). Note that a visit is simply a glance that does not include the initial saccade leading into the area of interest.

2.2 Pipe spool assembly task

For the experiments described herein, a total of 20 industrial pipefitters from the Colorado Front Range participated in the pipe spool assembly task study. The participating pipefitters ranged in age of 20 to 60 years of age, and their industry experience ranged from 1 to 39 years of experience (Table 2). Further, 45% of the pipefitters had received trainings on interpretation of engineering drawings. It is noted that both the Skewness and Kurtosis statistics, which are measures of a population’s normality, indicates that the observed pipefitters are positively skewed towards older and more experienced workers (based on a Skewness value greater than 0), but the Kurtosis values indicates that the age and years of industry experience can still be considered a normal univariate distribution, with Kurtosis values being between -2 and +2 (George and Mallery 2010). In this study, the pipefitters were provided with ten isometric pipe spool drawings, such as the drawing shown in Figure 1. The primary reason why pipefitters were chosen for this study was the inherent complexity of the drawings they use in the form of isometric drawings, which are not drawn to scale (for example, diameters of pipe are not differentiated by line thickness) and understanding an overall assembly involves referencing one sheet to another (for example, continuation notes on an individual sheet to other sheets). The isometric drawings were developed to represent the typical complexity of the information that pipefitters use to complete their assemblies. All ten isometric drawings can be viewed at Alruwaythi (2017). The pipefitters were fitted with eye tracking glasses that recorded what they looked at as they assembled pipe components, and they were free to look at each of the drawings as many times as needed, for as long as needed.

TABLE 2: Demographic metric descriptive statistics, all 20 pipefitters

	Minimum	Mean	Median	Maximum	Std Dev	Skewness	Kurtosis
Age (years)	20	34.3	31	60	11.8	0.753	-0.502
Years of Industry Experience	1	11.4	7.5	39	10.6	1.301	1.163
Engineering Drawing Training 0=no, 1=yes		0.45					

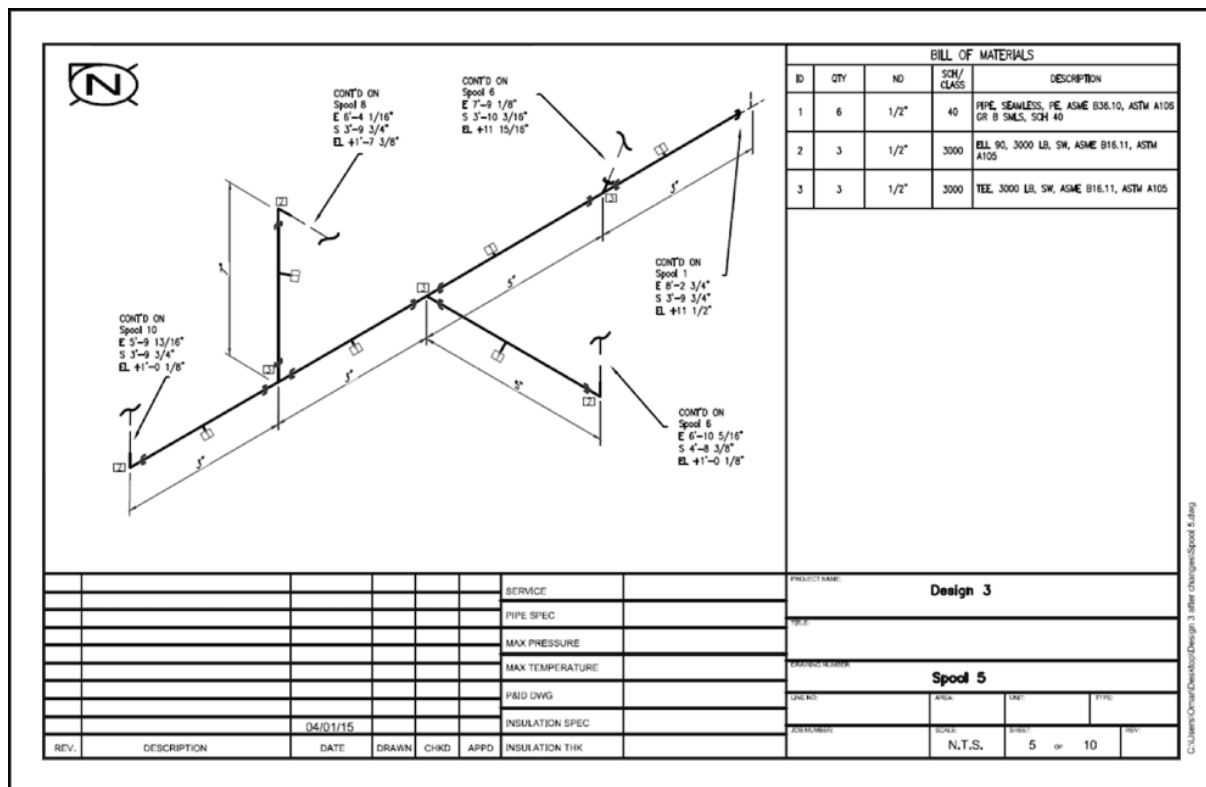


FIG. 1: One of the ten isometric assembly drawings

As mentioned, three metrics were used to quantify the level of experience of the pipefitters: years of industry experience, age, and whether or not they had received training for reading engineering drawings. Presumably, as pipefitters gain work experience, they become more proficient in their work, and as they become more proficient in their work, it is reasonable that their visual information gathering strategies may also evolve. Therefore, the authors hypothesized that the visit metrics for the novice pipefitters would be somehow different from the visit metrics of the more experienced pipefitters. No specific hypotheses were made of how the pipefitters' visit metrics might differ.

While years of work experience measured the pipefitters' level of industry-specific experience, age served as a proxy for their life experiences. Several studies have shown that Americans are spending more time consuming digital media, such as the internet, television, and social media, and less time consuming traditional print materials, such as books, magazines, and newspapers, than in the past (Bureau of Labor Statistics 2019; Chen and Adler 2019; Twenge et al. 2018). Older individuals are likely to have more experience with print materials than younger individuals, and traditional, paper engineering drawings are print materials. Therefore, the authors hypothesized that the older pipefitters would interact with the assembly drawings differently than the younger pipefitters during the pipe spool assembly task. Again, no specific hypotheses were made regarding how the pipefitters' interactions with the assembly drawings would differ, but simply that their visit metrics would differ by age.

The third and final experience metric collected from pipefitters was whether or not they had ever received any training for reading engineering drawings. Presumably, providing such training to a pipefitter should in some way alter the manner in which that pipefitter obtains information from engineering drawings. Therefore, the authors also hypothesized that the visit metrics for the pipefitters that had received training for reading engineering drawings would differ from the visit metrics for the pipefitters that had not received training.

Three performance metrics were recorded for each of the 20 pipefitters in the pipe spool assembly task, which included assemble time (the duration of time required to complete the pipe spool assembly), number of errors (the number of errors present in the completed pipe spool assembly), and rework % (the proportion of time that a participant spent disassembling and reassembling components that they assembled erroneously). The observed pipefitters took an average 40 minutes to complete the assembly, with a maximum of 71 minutes and a minimum

of 22 minutes. Furthermore, the pipefitters completed the assembly with an observed maximum of 4 errors and other pipefitters completing the assembly with zero errors. Next, the observed average proportion of time the pipefitters worked correcting rework was 8.87% (Table 3). It is observed that there is a correlation with increases in assembly time and the percentages of rework; there was zero rework and error with the minimum assembly time and the greatest percentage of rework and errors and the maximum assembly time. What was also observed but not evident by Table 3's data is that as the number of errors and rework increased so did the variability and standard deviation in the assembly time; some mistakes were fairly easy to correct but others required more effort.

TABLE 3: Performance metric descriptive statistics, all 20 participants

	Min	Mean	Median	Max	Std Dev	Skewness	Kurtosis
Assembly Time (mm:ss)	22:14	40:33	37:59	71:54	13:11	0.701	-0.005
Number of Errors	0	1.2	1.5	4	1.24	0.501	-0.667
Rework %	0%	8.87%	8.78%	19.28%	5.44%	0.152	-0.865

2.2.1 Spatial cognition

As part of the experiments, each pipefitter completed two paper-based tests to measure their “ability to see differences in figures” (Ekstrom et al. 1976), which is a component of their spatial cognition. These tests included the card rotation and cube comparison test.

2.2.2 Card rotation test

In each problem of this test, participants were given an image, and then asked whether eight other images were the same or different from the original image. Three examples problems are shown in Figure 2 below, and the first row has been answered correctly. The test included five problems, for a total of 40 possible responses. Participants had 1.5 minutes to answer as many items correctly as possible. Test scores were computed by subtracting the number of incorrect responses from the number of correct responses, so the maximum possible score was 40 and the minimum possible score was -40.

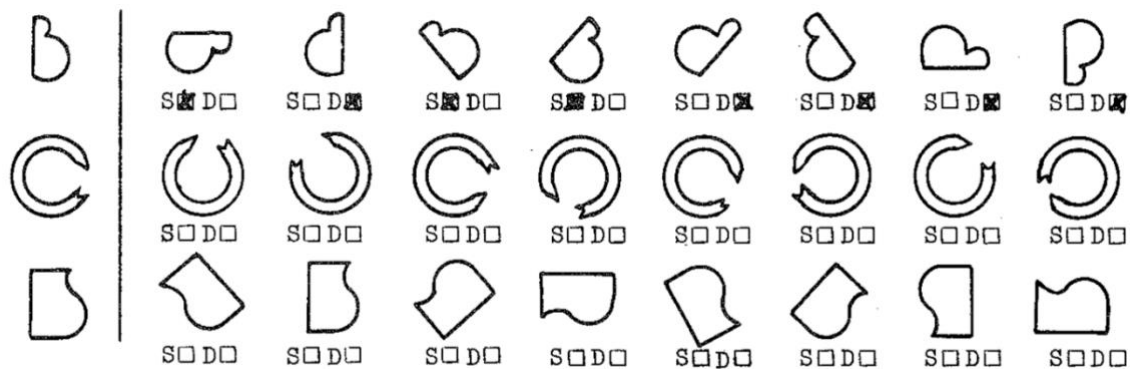


FIG. 2: Examples of card rotation test problems, reproduced from (Ekstrom et al. 1976)

2.2.3 Cube comparison test

This test is very similar to the card rotation test, but three-dimensional cubes are used instead of two-dimensional shapes. In each problem of this test, the participant compares two cubes and indicates whether the cubes are the same or different. Two example problems are shown in Figure 3. The participants had 2 minutes to answer as many items correctly as possible. Test scores were computed by subtracting the number of incorrect responses from the number of correct responses, so the maximum possible score was 14 and the minimum possible score was -14.



FIG. 3: Example of cube comparison test problems, reproduced from (Ekstrom et al. 1976)

Based on the observed population, the average Card Rotation Test Score was 30.2, which had a maximum score of 40 and a minimum of 15. The average Cube Comparison Test Score was 7.0, which had a maximum score of 14 and a minimum of -2 (Table 4).

TABLE 4: Spatial cognition metric descriptive statistics, all 20 pipefitters

	Min	Mean	Median	Max	Std Dev	Skewness	Kurtosis
Card Rotation Test Score	15	30.2	30	40	6.8	-0.327	-0.427
Cube Comparison Test Score	-2	7.0	8.5	14	5.2	-0.406	-1.252

2.2.4 Spatial cognition and visual information gathering strategy hypotheses

Previous studies have shown that spatial cognition has a profound impact on performance during construction model assembly tasks (Alruwaythi and Goodrum 2019; Goodrum et al. 2016; Sweany et al. 2016). Therefore, in the present work, the authors also hypothesized that the pipefitters' visit metrics would differ by their level of spatial cognition.

2.3 Visit metrics

Utilizing eye tracking data, a number of metrics were used to quantify and compare the pipefitters' visual information gathering strategies. The metrics have been defined in the context of the pipe spool assembly task, and descriptive statistics for these metrics are provided in Table 5.

Table 5: Visit metric descriptive statistics, all 20 pipefitters

	Min	Mean	Median	Max	Std Dev	Skewness	Kurtosis
Visit Count (n)	110	252.6	249	359	77.6	-0.253	-1.13
Mean Visit Duration (s)	2.24	3.42	3.34	4.67	0.72	0.226	-0.889
Count of Visits Longer than 7.5 seconds (90 th percentile)	6	25.9	24	63	14.6	0.910	0.716
Page Turn Count	12	48.0	41	114	28.7	0.722	-0.321

Visit count (n) - This is the total number of visits made by a pipefitter to the assembly drawings.

Mean visit duration - The mean visit duration is a measure of the typical amount of time that a participant spent searching a drawing for a piece of information.

Maximum visit duration - A pipefitters' maximum visit duration was the duration of their single longest visit to the assembly drawings.

Visit duration standard deviation - The standard deviation is a measure of how consistently the durations of the pipefitters visits to the assembly drawings were.

Count of visits longer than 7.5 seconds (90th percentile duration) - One of the most significant and widely accepted findings from historic motor vehicle driver behavior eye tracking studies is that drivers should not glance away from the road ahead for longer than 2 seconds. Each glance with a duration longer than 2 seconds significantly increases a driver's crash risk, i.e. reduces their performance. In the context of driving a motor vehicle, a 2 second glance is a relatively long duration glance. However, in the pipe spool assembly task, the average visit duration was 3.4 seconds, so a 2 second visit to the assembly drawings would actually have been a very short duration visit. In order to count the number of long duration visits made by each of the pipefitters, we counted the number of visits with a duration greater than the 90th percentile of all of the visit durations. The 90th percentile visit duration was 7.5 seconds.

Page turn count - Page turn count is the number of times that a pipefitter turned the page of the assembly drawings. Sequential visits to the same drawing did not constitute a page turn. If a pipefitter looked at each of the ten drawings without turning the page back to a drawing that they had looked at previously, then their page turn count would equal nine.

The three visit duration metrics above (Mean visit duration, Maximum visit duration, and Visit duration standard deviation) are derived from the pipefitters' visit duration distributions, and they can be used to characterize the visual information gathering strategies used by the pipefitters (Table 5). Meanwhile, page turn count is derived from the sequence in which the pipefitters visited the assembly drawings, which provides additional insight into the information gathering strategies used by the pipefitters. While the pipefitters were not asked to describe the strategies that they used to obtain information from the assembly drawings, it is reasonable that some of the pipefitters' strategies may have been to "complete" each assembly drawing before moving on to the next drawing. In other words, some of the pipefitters may have intended to assemble every component present on a single drawing before turning the page and assembling every component present on the next drawing. Pipefitters using such a strategy should have had lower page turn counts than pipefitters that did not follow this strategy.

Similar to a motor vehicle driver behavior study, duration distributions were first developed for the instances in which the pipefitters looked at the assembly drawings. Figure 4 displays a "visit" duration distribution, while ISO 15007 pertains to the analysis of "glances." In the context of an eye tracking study, a visit is very similar to, but distinct from a glance. Therefore, these terms, as well as other key eye tracking terms are defined in the following section.

Note that the maximum recorded visit duration in the study was 74.58 seconds (Table 6), but the abscissa (x-axis) of Figure 4 has been truncated at 20 seconds in order to better visualize the distribution. Twenty seconds was the 99.2 percentile of the recorded visit durations, so 99.2% of the recorded data is shown in Figure 4. The remaining 0.8% of the recorded visits had a duration longer than 20 seconds.

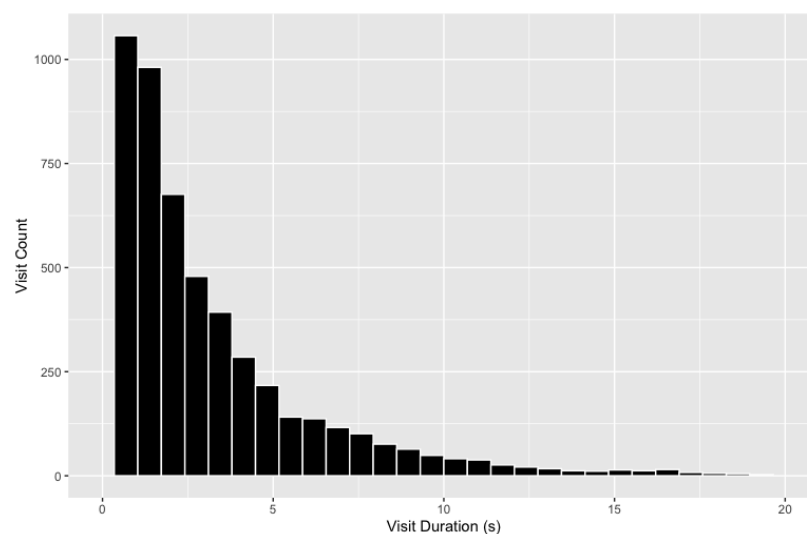


FIG. 4: Visit duration distribution, all 20 pipefitters, $n=5,052$ total visits

TABLE 6: Visit duration descriptive statistics, all 20 pipefitters

Metric	
Count (n)	5,052 visits
Minimum	0.33 seconds
Maximum	74.58 seconds
Mean	3.44 seconds
Median	2.16 seconds
Standard Deviation	3.85 seconds
Skewness	4.12
Kurtosis	37.34
Visits Longer than 7.5 seconds (90 th percentile duration)	518 visits

2.3.1 Visit duration distributions

Visit duration is the time elapsed between the beginning of the initial fixation within an area of interest and the end of the last fixation with the same area of interest. A total of 5,052 visits to the assembly drawings were recorded, and the experiment observed a relatively large variation in visit durations, with the longest visit lasting 74.6 seconds and the shortest visit lasting 0.3 seconds. The visit durations were inherently highly positively skewed (D'Agostino skewness=4.12, $p < 0.001$).

The primary critique of most eye tracking metrics is that they are not easily interpretable or that they do not provide practical insights. For example, fixation counts, and cumulative fixation durations are frequently encountered in eye tracking literature. These metrics provide the relative durations of time that participants spent looking at particular Areas of Interest (AOI's), but that information alone provides limited insight into the strategies that individuals use to obtain visual information. Alternatively, visit duration distributions can be visualized and qualitatively compared through density plots, and statistics can be derived from visit duration distributions to quantitatively compare the visual information gathering strategies used by individuals. Visit duration density plots for all 20 pipefitters are shown in Figure 5. This figure illustrates the overall variability of the pipefitters' visit duration distributions, but because of the large number of distributions shown, it is not very useful for comparing particular distributions. Two individual duration distributions are compared later in section 3.1.

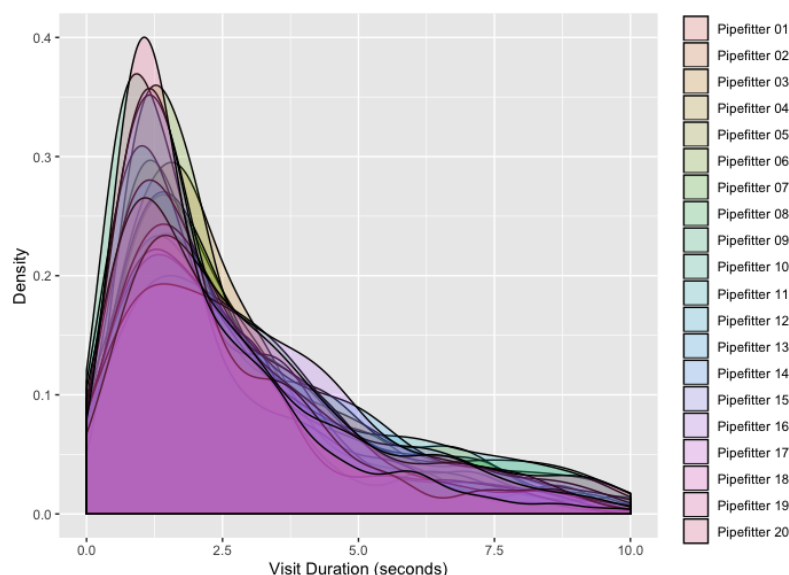


FIG. 5: Density plot of assembly drawing visit durations, all 20 pipefitters

2.4 Visit generation

A custom JavaScript web application was developed for the purpose of generating and analyzing visits from eye tracking data. The application is named “Visual Eyes” and the source code is available in the project repository on GitHub (Sears 2020).

Raw eye tracking data was collected using SensoMotoric Instruments (SMI) Eye Tracking Glasses 2.0, and the data was imported into the SMI BeGaze v3.4 eye tracking analysis software. The default SMI BeGaze event detection algorithm was used to classify the raw data into fixation, saccade, and blink events, and then each fixation point was manually mapped to the appropriate assembly drawing. Next, the eye event data was exported from SMI BeGaze as text files, which were then imported into the Visual Eyes application to generate visits. The process that the Visual Eyes application used for generating visits from fixations and saccades is provided by (Sears et al. 2020a). The visit generation process was completed once for each of the 20 pipefitters.

2.5 Minimum fixation duration (MFD)

Completing the primary task of driving a motor vehicle while completing secondary tasks of looking at mirrors, roadway signs, or a cell phone to gain visual information, is similar in many ways to completing the primary task of assembling construction components while completing secondary tasks of looking at assembly drawings to gain visual information. Therefore, in the present work, we applied the ISO 15007 glance analysis methods to eye tracking data collected during a pipe spool assembly task, in order to investigate how professional pipefitters interacted visually with construction assembly drawings. Initial results from this eye tracking assembly task were published previously by (Alruwaythi et al. 2017). The present work differs from the work published previously in that the previous analysis was concerned with a handful of relatively high-level eye tracking metrics, while the present work is a deeper analysis of each of the individual instances in which a pipefitter looked at an assembly drawing.

As discussed in Sears et al. (2020a), the eye tracking community has not adopted a single, standard minimum fixation duration value. However, ISO Standard 15007-1:2014 stipulates that, “Fixations to an area of interest $\leq 120\text{ms}$ are physically not possible” (ISO 2014a). Therefore, the authors elected to use a minimum fixation duration of 120ms for this work.

2.6 Maximum off-stimulus fixations (MOSF)

The concept of maximum off-stimulus fixations was introduced as a noise filtering mechanism in (Sears et al. 2020a). However, since maximum off-stimulus fixations have not been used in any other eye tracking research, and an optimal value has not yet been determined for this parameter, an MOSF value of zero was used in this study. An MOSF value of zero provides no noise filtering of the collected eye tracking data, but it is the most consistent with historically published eye tracking studies.

2.7 Visit metric calculations

Each pipefitter generated 100+ visits to the assembly drawings, which produced a distribution of visit durations for each pipefitter. The visit duration distributions were used by the Visual Eyes application to derive all of the visit metrics except for the page turn count. These metrics included: visit count, mean visit duration, maximum visit duration, visit duration standard deviation, and count of visits longer than 7.5 seconds. The Visual Eyes application also computed the page turn count for each pipefitter by first recording the name of the stimulus image associated with each visit, in chronological order, and then simply counting the number of times that the stimulus image changed.

2.8 Visit, performance, and demographic metric relationships

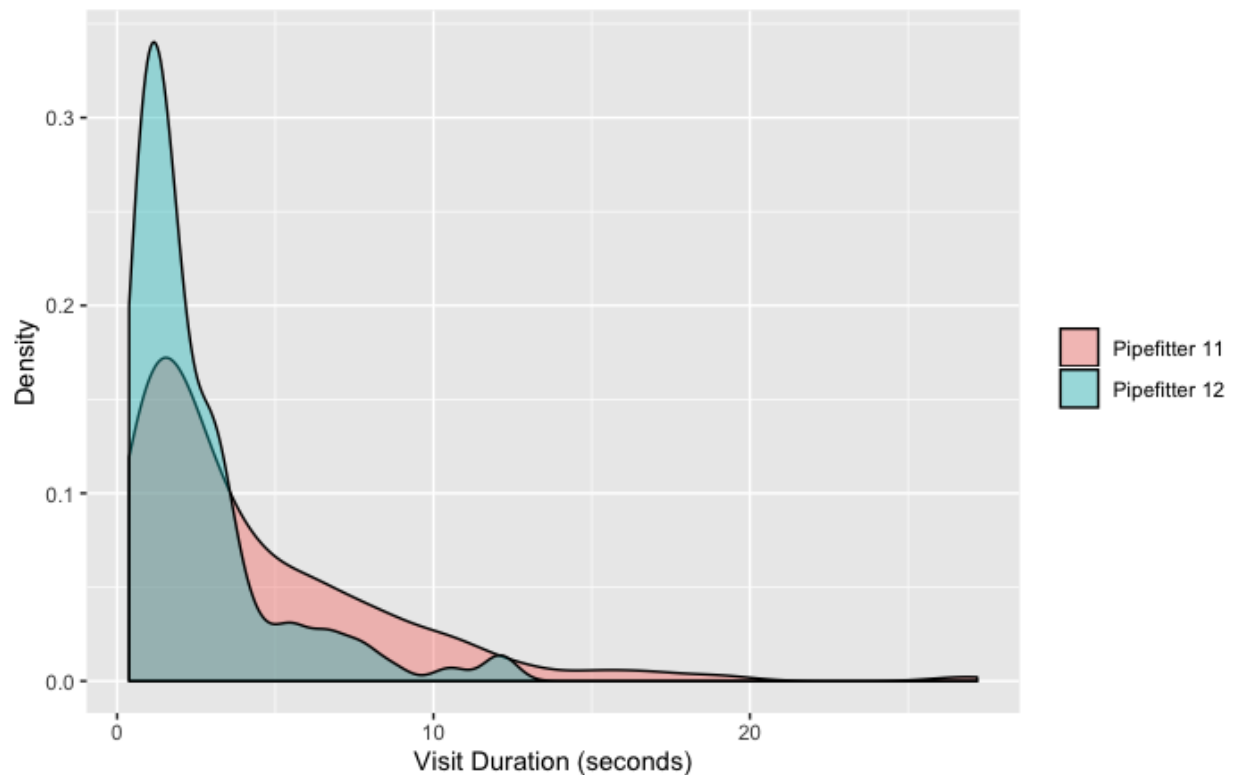
In order to investigate relationships between the visit, performance, and demographic metrics, results were first compared between the best performing pipefitter and the worst performing pipefitter in the study, Pipefitter 12 and Pipefitter 11, respectively (Section 3.1). The comparison of these two pipefitters also serves as an illustration of how individual visit duration distributions can be compared, both qualitatively and quantitatively. The visit duration distributions were inherently highly positively skewed, as indicated by the long tails to the right. Since a log-transformation is a very transformation method, especially with right skewed distributions (Wonnacott and Wonnacott 1986), a log base-10 transformation was applied to the visit durations to improve normality.

After comparing the best and worst performers, the analysis was broadened to the full sample of 20 pipefitters. In order to perform an analysis of whether a linear relation exists between a series of quantitative variables, Pearson correlation tests were performed. The Pearson correlation coefficient measures both the direction (positive or negative relation) as well as the strength of the association (larger values indicate a greater strength). First, Pearson correlation tests were conducted between the visit metrics to determine whether relationships between the visit metrics suggested that the pipefitters had followed any particular, overarching patterns in the strategies that they used for obtaining visual information from the assembly drawings. Then Pearson correlation tests were conducted between the visit and performance metrics and the visit and demographic metrics to search for relationships.

3 RESULTS

3.1 Best and worst performers

Figure 5 illustrated the overall variability of the pipefitters' visit duration distributions by plotting all 20 of the pipefitters' duration distributions. However, it is much easier to compare individual visit duration distributions when fewer distributions are plotted. Figure 6 shows the visit duration distributions for Pipefitter 11 and Pipefitter 12, the worst and best performers in the assembly task, respectively. Pipefitter 12 completed the assembly task in the least amount of time, 22.23 minutes (best performance), while Pipefitter 11 required the greatest amount of time, 71.9 minutes (worst performance).



Pipefitter 11 = worst performance, Pipefitter 12 = best performance

Fig 6: Visit duration density plot of two pipefitters,

The visit, performance, and demographic metrics for Pipefitters 11 and 12 are listed in Table 7. In addition to having the fastest assembly time, Pipefitter 12 made no errors, and had the third lowest rework % of all of the pipefitters (time Pipefitter 12 spent doing rework completed all errors in the final assembly). Meanwhile, Pipefitter 11 made three errors, which was the second highest of all of the pipefitters, and spent 12.7% of their time on rework, which was the seventh highest amount of all 20 pipefitters. Both pipefitters were approximately the same age, had less than two years of industry experience, and had not received any prior training for reading engineering drawings. Pipefitter 12 scored the maximum possible score on both of the spatial cognition tests, while Pipefitter

11 had the third lowest card rotation test score (23), and the eighth lowest cube comparison test score (5). Pipefitter 12 made the fewest number of visits to the assembly drawings (110) and turned the page of the drawings fewer times than any other pipefitter (12 page turns). Meanwhile, Pipefitter 11 had the third highest number of visits to the assembly drawings (348), and the sixth highest number of page turns (71).

As noted in section 2.9., the visit duration distributions were highly positively skewed, so a log base-10 transformation was applied to improve normality before testing for differences between Pipefitter 11 and Pipefitter 12's visit duration distribution metrics. The transformed values are listed in square brackets in Table 7. Pipefitter 12 had the third lowest mean visit duration (2.50s), and the second lowest maximum visit duration (12.18s), while Pipefitter 11 had the fourth highest mean visit duration, and the eighth highest maximum visit duration (27.18s). Pipefitter 12 also had the third lowest visit duration standard deviation (2.35s), while Pipefitter 11 had the fifth highest visit duration standard deviation (4.14s).

What is perhaps most notable in the comparison between these two pipefitter is their differences in spatial cognition and the relation this has in terms of how both fitters were able to use their drawings. Pipe fitter 12 required 68% fewer total visits to the plans along with 83% fewer page turns. Furthermore when both pipe fitters were looking specific drawings, pipe fitter was able to obtain the information needed in a 41.5% shorter duration based on the different in the average visit duration. Finally, pipefitter 11's longest visit duration was 55% longer than the longest duration visit of pipefitter 12.

TABLE 7: Performance, demographic, and visit metrics for the best and worst performers (metric ranks in parentheses, 1=lowest rank, 20=highest rank)

Metric	Pipefitter 11, Worst		Pipefitter 12, Best	
	[Log ₁₀] (rank)		[Log ₁₀] (rank)	
Assembly Time (mm:ss)	71:54	(20)	22:14	(1)
Number of Errors	3	(19)	0	(1)
Rework %	12.70%	(14)	2.22%	(3)
Age	24	(5)	23	(3)
Years of Industry Experience	1	(1)	2	(4)
Engineering Drawing Training	No		No	
Card Rotation Test Score	23	(17)	40	(1)
Cube Comparison Test Score	5	(12)	14	(1)
Visit Count (n)	348	(18)	110	(1)
Page Turn Count	71	(15)	12	(1)
Mean Visit Duration (s)	4.28	[0.631] (17)	2.50	[0.399] (3)
Maximum Visit Duration (s)	27.18	[1.434] (13)	12.18	[1.086] (2)
Visit Duration Standard Deviation	4.14		2.35	
Visit Duration Skewness	2.005		2.166	
Visit Duration Kurtosis	5.638		5.128	
Count of Visits Longer than 7.5 Seconds	37	(16)	19	(8)

As further evidence that significant differences occurred in the effective use the pipefitters' engineering drawings, statistical was performance on visit durations. An F-test was conducted to compare the visit duration variance between Pipefitter 11 and Pipefitter 12. There was a significant difference in the visit duration variance between Pipefitter 11 and Pipefitter 12 ($F(109, 347)=1.417$, $p=0.032$). In addition, A two-sample t-test for means with unequal variance (Welch) was conducted to compare the visit durations between Pipefitter 11 and Pipefitter 12. There was a significant difference in the visit durations for Pipefitter 11 ($M=0.445$, $SD=0.416$) and Pipefitter 12 ($M=0.251$, $SD=0.350$); ($t(215.0)=4.83$, $p<0.001$). Finally, a two-sample proportion z-test was conducted to compare the proportion of visits longer than 7.5 seconds between Pipefitter 11 ($19/110 = 0.173$) and Pipefitter 12 ($37/348 = 0.106$). There was a significant difference in the proportion of visits longer than 7.5 seconds between the Pipefitters ($z=-1.853$, $p=0.064$).

As a whole, these differences suggest that the difference the pipefitter's spatial cognition is related to each pipefitter's relative effectiveness in using the information required to complete the assembly, since there was little to no difference between their industry experience and having received engineering drawing training.

3.2 Relationships between visit, performance, and demographic metrics, all 20 pipefitters

After comparing the best and worst performing individuals, relationships between the visit, performance, and demographics metrics for the full sample of 20 pipefitters were investigated through Pearson correlation tests (Table 8).

TABLE 8: Pearson correlation test results – visit, performance, and demographic metrics

			Visit Metrics						
			Visit Count	Mean Visit Duration	Page Count	Turn	Visit Duration Std Dev	Maximum Visit Duration	Visits Longer than 7.5s
Visit Metrics	Visit Count		—						
	Mean Visit Duration		r=0.067 p=0.779	—					
	Page Turn Count		r=0.627 p=0.003**	r=-0.181 p=0.444	—				
	Visit Duration Std Dev		r=0.118 p=0.620	r=0.897 p<0.001**	r=-0.271 p=0.248	—			
	Maximum Visit Duration		r=0.217 p=0.358	r=0.730 p<0.000**	r=-0.185 p=0.434	r=0.929 p<0.001**	—		
	Visits Longer than 7.5s		r=0.055 p=0.817	r=-0.014 p=0.953	r=0.083 p=0.727	r=-0.142 p=0.552	r=-0.134 p=0.573	—	
Performance Metrics	Assembly Time		r=0.759 p<0.001**	r=0.436 p=0.055*	r=0.266 p=0.257	r=0.458 p=0.042**	r=0.446 p=0.049**	r=0.187 p=0.430	
	Number of Errors		r=0.573 p=0.008**	r=0.172 p=0.468	r=0.271 p=0.247	r=0.095 p=0.690	r=0.150 p=0.529	r=0.126 p=0.596	
	Rework %		r=0.398 p=0.082*	r=0.156 p=0.512	r=0.063 p=0.792	r=0.176 p=0.459	r=0.180 p=0.447	r=-0.084 p=0.725	
Demographic Metrics	Age		r=-0.082 p=0.731	r=-0.109 p=0.647	r=-0.077 p=0.747	r=-0.140 p=0.557	r=-0.255 p=0.279	r=-0.260 p=0.269	
	Years of Experience		r=-0.274 p=0.242	r=-0.146 p=0.540	r=0.020 p=0.934	r=-0.230 p=0.328	r=-0.220 p=0.352	r=-0.094 p=0.695	
	Card Rotation Test Score		r=-0.511 p=0.021**	r=-0.342 p=0.140	r=0.012 p=0.960	r=-0.447 p=0.048**	r=-0.375 p=0.103	r=-0.212 p=0.370	
	Cube Comparison Test Score		r=-0.473 p=0.035**	r=-0.056 p=0.815	r=0.181 p=0.444	r=-0.220 p=0.352	r=-0.269 p=0.252	r=0.072 p=0.764	
	Drawing Training		r=0.039 p=0.869	r=-0.193 p=0.416	r=0.469 p=0.037**	r=-0.150 p=0.527	r=-0.110 p=0.645	r=0.020 p=0.932	

******Correlation is significant at the 0.05 level; *****Correlation is significant at the 0.10 level

The purpose using the Pearson correlation tests was to identify significant relationships, as determined with a p-value of 0.10 or smaller, to identify which relationships were justified to be explored further through regression. It should be noted that conducting correlation tests between several variables increases the likelihood of Type I errors (false positives). A Bonferroni adjustment, or similar adjustment, could have been applied to account for the fact that several statistical tests were conducted in this work; however, a Bonferroni adjustment increases the likelihood of Type II errors (false negatives) (Perneger 1998). As this work was exploratory in nature, and the objective was to identify any possible relationships between the visit, performance, and demographic metrics, a Bonferroni adjustment was not applied.

Several of the correlations between the visit and performance metrics and visit and demographic metrics were significant at the $p=0.10$ level, so regressions were calculated to determine the relationship coefficients. Results of the visit metric correlations did not suggest to the authors that the pipefitters had adhered to any particular, overarching information gathering strategies, so no further analysis of the visit metric relationships was conducted. Relationships between the visit metrics are discussed further in Section 4.1. The drawing training metric was a dummy coded binomial variable, so a t-test was conducted for this variable instead of a regression. A summary of the regression results is provided in Table 9.

TABLE 9: Regression results – visit, performance, and demographic metrics

Eq.	Dependent Variable	Independent Variable	Constant (p)	β (p)	$F(1,18)$	R^2	Adj R^2
A	Assembly Time	Visit Count	479.356 (0.260)	7.736 (0.000)	24.53	0.577	0.553
B	Assembly Time	Mean Visit Duration	789.2 (0.347)	480.5 (0.055)	4.217	0.190	0.145
C	Assembly Time	Visit Duration Std Dev	1281.18 (0.032)	321.37 (0.042)	4.768	0.209	0.165
D	Assembly Time	Max Visit duration	1780.87 (0.000)	24.43	4.461	0.199	0.154
E	# Errors	Visit Count	-1.112 (0.188)	0.009 (0.008)	8.801	0.328	0.291
F	Rework %	Visit Count	1.821 (0.654)	0.028 (0.082)	3.385	0.158	0.111
G	Visit Count	Card Rot. Score	429.172 (0.000)	-5.847 (0.021)	6.351	0.261	0.220
H	Visit Duration Std Dev	Card Rot. Score	5.829 (0.000)	-0.074 (0.048)	4.507	0.200	0.156
I	Visit Count	Cube Comp. Score	301.461 (0.000)	-7.030 (0.035)	5.199	0.224	0.181

3.3 Simple linear regression results – performance and visit metrics

3.3.1 Relation between assembly time and visits count

A number of the regressions confirmed a positive relation between the time required to complete the assembly and different visit metrics. Specifically, Equations A, B, and D confirmed that among the observed pipefitters their time to complete the assembly increased as the total visits counts, mean visit duration, and the maximum visit duration increased as we, which suggests that the pipefitters that completed the assembly faster were partly able to do so based on their ability to obtain the information from the drawings to their works. Assembly time also increased the standard deviation of the observed visit durations, which suggests that the observed pipefitters that completed the assembly faster were also more consistent in the time required to find the information they needed.

3.3.2 Relation between errors, rework and visit counts. Number of errors and visit count (Eq E)

Equations E and F confirmed the relation between the visit counts and both the number of final errors in the assembly (Equation E) and the observed rework percentage (Equation F). The observed pipefitters with fewer visit counts were involved in less rework. Also, the observed pipefitters that had greater number of visit counts also

had a greater number of final errors in their assembly. If this had been an actual construction operation, the ramifications of the final errors would undoubtedly impact the assembly time since the error would have to be corrected (this was not required of the participating pipefitters) and potentially a system that has been assembled incorrectly may also have significant impact on both the safety of construction and operations.

3.4 Simple linear regression and t-test results – visit and demographic metrics

Of all the different demographic measures used in the experiments (e.g. industry experience, prior drawing training, and spatial cognition, the measures related to spatial cognition were observed to be the strongest predictor of the number of visit counts an observed pipefitter used during their task assembly. Pipefitters with higher card rotation and cube comparison scores required fewer visit counts (Equations G and I). Finally, the observed pipefitters with a higher card rotation score also experienced with a smaller standard deviation in terms of their visit counts (Equation H). However, the same was not observed with a cube rotation score, so while Equation H suggests that a greater spatial cognition is related to more consistent visit counts, this was not confirmed across both of the studies' measures of spatial cognition among the observed pipefitters.

4 DISCUSSION

This work showed that differences in pipefitter spatial cognition were associated with differences in visit duration distributions, and differences in visit duration distributions were associated with differences in performance. Additionally, higher performance was achieved during the pipe spool assembly task when the pipefitters looked at the assembly drawings more consistently, and as little as possible.

4.1 Visual information gathering strategies

The purpose of computing and comparing the visit metrics used in this study was to characterize the pipefitters' interactions with the assembly drawings, in order to begin to develop an understanding of the strategies that craft professionals use to gather visual information from assembly drawings. Figure 3 illustrated the overall variability between the 20 pipefitters' visit duration distributions, and Figure 7 presented the visit duration distribution of Pipefitters 11 and 12, the best and worst performers in the study, respectively. As discussed in section 3.1, Pipefitters 11 and 12 consistently landed near the extreme and opposite ends of nearly every one of the visit metrics. Given the pronounced differences in their visit metrics, it's possible that Pipefitters 11 and 12 had different underlying strategies that they used to obtain visual information from the assembly drawings. Pipefitters 11 and 12 also landed near the extreme and opposite ends of nearly every performance metric recorded, so it is also possible that differences in their performance were influenced by differences in their visual information gathering strategies.

One aspect of visual information gathering strategies that may differ between craft professionals is the quantity of information that they attempt to gather before looking away from engineering information deliverables to complete a task. In the present work, the strategy used by some of the pipefitters may have been to look at the assembly drawings only long enough to identify a single component that hadn't been installed yet, and to gather the information necessary for installing that single component. Afterward, the pipefitter would look away from the drawings, locate the single component, install it, and then repeat the process for the next component until the assembly was complete. Alternatively, the strategy used by other pipefitters may have been to gather the information necessary for installing two, three, four, or more components all at once, before looking away from the assembly drawings and installing that batch of components. In the context of LEAN manufacturing, the latter strategy would be considered batch processing, while the former strategy would more closely resemble the LEAN principle of "continuous flow" or "one-piece flow" (Shah and Ward 2003). One-piece flow processing is widely regarded as more productive than batch processing for manufacturing (Brown et al. 2006). Therefore, if analogous differences exist in visual information gathering strategies, then it's possible that those differences are also associated with differences in performance.

The results of the visit metric correlations (Table 8) were reviewed to identify any patterns that might be present in the information gathering strategies used by the pipefitters. The authors anticipated that relationships between the visit metrics might suggest particular, overarching strategies that had been used by the pipefitters, which could have then been further investigated for relationships with performance or demographics. For example, the batch processing and one-piece flow processing strategies described above are purely speculative; however, if the pipefitters had in fact use batch processing and one-piece flow processing visual information gathering strategies,

then it should have been apparent from their visit metrics. Pipefitters using a batch processing strategy should have had lower visit counts and longer mean visit durations than pipefitters following a one-piece flow processing strategy. Therefore, if a correlation had been observed between visit count and mean visit duration, then that would have supported the theory of the existence of batch processing and one-piece flow processing strategies. However, no relationship was observed between visit count and mean visit duration, so the results did not support the existence of batch processing and one-piece flow processing information gathering strategies. In fact, the relationships observed between the visit metrics did not immediately suggest to the authors that the pipefitters were adhering to any particular information gathering strategies.

4.2 Visit metrics and performance

While the correlations between the visit metrics did not suggest that the pipefitters were following any particular, overarching information gathering strategies, it was still possible that particular aspects of the pipefitters' interactions with the assembly drawings—as measured by their visit metrics—were associated with their performance. As shown in Table 8, several significant relationships were observed between the visit and performance metrics.

4.2.1 Visit count and performance

Results of the visit count and performance metric correlations consistently indicated that fewer visits to the assembly drawings were associated with better performance. The pipefitters who made fewer visits to the assembly drawings required less time to complete the assembly task, made fewer errors, and spent a smaller percentage of their time on rework. The same relationship was observed between the best and worst performer in the study (Table 7). Pipefitter 12, the best performer, made the fewest number of visits to the assembly drawings, and Pipefitter 11, the worst performer, had one of the highest number of visits to the assembly drawings.

4.2.2 Mean visit duration and performance

There was no significant relationship between mean visit duration and number of errors made, or mean visit duration and rework %. However, pipefitters with a higher mean visit duration required more time to complete the assembly task, which was consistent with the comparison between the best and worst performers in the study. Pipefitter 12, the best performer, had a significantly lower mean visit duration than Pipefitter 11, the worst performer. Collectively, these results indicate that the pipefitters who spent less time looking at the assembly drawings during each visit performed better.

4.2.3 Visit duration standard deviation and performance

Similar to the mean visit duration metric, there were no significant relationships observed between visit duration standard deviation and number of errors made, or visit duration standard deviation and rework %, but there was a significant relationship between visit duration standard deviation and assembly time. Pipefitters with higher visit duration standard deviations required more time to complete the assembly task. This finding is also consistent with the comparison of the best and worst performers; the best and worst performers had significantly different visit duration standard deviations. These findings indicate that the pipefitters' performance increased as the consistency in which they looked at the assembly drawings increased. That is, pipefitters' performance increased as the duration of their visits to the assembly drawings became more consistent.

4.2.4 Maximum visit duration and performance

There was also a significant relationship observed between maximum visit duration and assembly time. A pipefitter's maximum visit duration is the duration of their single longest visit, and the pipefitters' performance decreased as their longest visit duration increased. The presence of a long duration visit could indicate that a pipefitter was having difficulty searching for a piece of information during a particular visit. Search difficulty is discussed more in Section 4.4 below.

4.2.5 Visits longer than 7.5 seconds and performance

Historic motor vehicle driver behavior studies have shown that long duration glances away from the road while operating a motor vehicle are strongly related to driver performance. Therefore, in an effort to identify an analogous relationship in the manner that pipefitters interact with assembly drawings, we counted the number of long duration visits made by each pipefitter to the assembly drawings during the pipe spool assembly task. No

significant relationships were observed between the number of visits longer than 7.5 seconds and any of the performance metrics when all 20 pipefitters were considered. However, the proportion of visits lasting longer than 7.5 seconds was significantly different between the best performer in the study, Pipefitter 12, and the worst performer, Pipefitter 11. Given this finding, it's possible that a larger sample size could have revealed a significant correlation between the performance metrics and the number of visits lasting longer than 7.5 seconds.

4.2.6 Performance summary

As a whole, the visit and performance relationships indicate that the best performance was achieved when the pipefitters looked at the drawings the fewest number of times, and for the shortest durations. Or, more simply stated, the best performance was achieved when the pipefitters looked at the drawings as little as possible. Unfortunately, this finding provides relatively little insight into the visual information gathering strategies used by the pipefitters. However, if future efforts were to develop a taxonomy of the visual information gathering strategies that craft professionals use when interacting with engineering deliverables, then the visit metrics considered in the present work could be used to determine whether particular information gathering strategies are associated with differences in performance.

4.3 Visit metrics and demographics

4.3.1 Age and work experience

Interestingly, no significant relationships were found between age or years of construction experience and any of the visit metrics. (Dadi et al. 2014c) found that craft professionals had preferences for information formats which were based upon past experiences, and (Chalhoub and Ayer 2018) found performance differences between information formats based upon industry experience. Therefore, the authors hypothesized that older individuals would look at the assembly drawings in some way that was different than younger individuals, based upon the older individuals' propensity for being more familiar with print materials. However, no differences in visit patterns were observed based upon the age of the pipefitters. For the same reasons, the authors also hypothesized that pipefitters with more industry experience would interact with the assembly drawings differently than less experienced individuals. However, no significant relationships based upon industry experience were observed either.

4.3.2 Spatial cognition

In addition to hypothesizing differences based upon age and experience, the authors also hypothesized that the pipefitters' interactions with the assembly drawings would differ based upon their level of spatial cognition. Multiple studies have shown that individuals with lower spatial cognition utilize engineering information deliverables differently than individuals with higher spatial cognition (Alruwaythi and Goodrum 2019; Goodrum et al. 2016). Therefore, the visit metrics were anticipated to differ based upon card rotation test scores and cube comparison test scores. Pipefitters that scored higher on the spatial cognition tests did make fewer visits to the assembly drawings, and higher card rotation test scores were associated with lower maximum visit durations, so the pipefitters' interactions with the assembly drawings were associated with their level of spatial cognition.

4.3.3 Engineering drawing training

Finally, it was observed that the pipefitters that had previously received engineering drawing training turned the pages of the assembly drawings significantly more times than the pipefitters that had not previously received engineering drawing training. It's possible that the pipefitters that had received training felt more confident in their ability to navigate through the assembly drawings, and that is why they had larger page turn counts. However, the pipefitters were not asked to describe their information gathering strategies or to explain their actions, so it is only possible to speculate about this relationship.

4.4 Information complexity

Presumably, when a pipefitter made a visit to an assembly drawing, they did so to obtain a piece of information necessary for completing the assembly task. Section 4.1 introduced two hypothetical information gathering strategies to described how the quantity of information that pipefitters attempt to obtain during visits to assembly drawings could influence the duration of their visits. Greater quantities of information gathering should result in longer visit durations. Additionally, previous motor vehicle driver behavior studies have shown that the complexity of visual information can also influence visit duration distributions. (Horrey and Wickens 2007) found that

“complex, demanding, or compelling in-vehicle tasks” can result in a higher number of relatively long duration visits, and (Beijer et al. 2004) found that “complex, visually captivating advertising signs” can result in a higher number of fixations and a higher number of long duration visits.

This study did not consider the complexity of information presented within the assembly drawings. However, the study included 10 assembly drawings, and each drawing would have inherently had a varying degree of visual complexity. Thus, it is reasonable that the complexity of information presented within the assembly drawings could have also influenced how the pipefitters interacted with the assembly drawings. More complex information within the drawings could have resulted in longer visit durations and greater visit counts. If future efforts were to consider the complexity of information presented within construction assembly drawings, then insights might be gleaned regarding how information complexity influences interactions with assembly drawings, and how information complexity influences the performance of craft professionals.

4.5 Search efficiency

This study observed that differences in spatial cognition were associated with differences in visit duration distributions, that differences in engineering drawing training were associated with differences in interactions with the assembly drawings, and that differences in visit duration distributions were associated with differences in performance. Previous studies have also shown that information complexity is associated with differences in visit duration distributions, and section 4.1 speculated that the quantity of information that the pipefitters attempted to obtain from the drawings may have also been influenced their visit duration distributions. As such, the authors propose that each of these findings and theories may be explained through a concept of search efficiency, where search efficiency is a measure of the amount of time required to successfully obtain visual information from an assembly drawing.

By definition, the pipefitters with higher levels of spatial cognition should have been more efficient at searching for information within the assembly drawings, and the purpose of providing pipefitters with training to read engineering drawings is to make them more efficient at obtaining information from assembly drawings. Logically, pipefitters should also be able to obtain small quantities of information faster than large quantities of information, and they should be able to obtain information from less complex drawings faster than information from more complex drawings. Thus, the authors propose that the pipefitters’ search efficiency during the pipe spool assembly task was a function of their spatial cognition, previous engineering drawing training, the quantity of information being gathered, and the complexity of the information presented in the assembly drawings (Eq 1).

$$\text{Search Efficiency} = f(SC, EDT, Q, C) \quad (1)$$

Where SC = spatial cognition, EDT = engineering drawing training, Q = quantity of information gathered, and C = complexity of assembly drawing information

A definition for search efficiency was proposed above as the amount of time required to successfully obtain visual information from an assembly drawing. It would not be possible with the information collected during this study to determine whether the pipefitters were successful or unsuccessful at obtaining information during any given visit. However, repeated visits to the same assembly drawing or same area within an assembly drawing would suggest that a pipefitter was unsuccessful at obtaining information during their initial visit(s) to that particular drawing or area of the drawing. Therefore, the authors propose that the efficiency in which the pipefitters were able to successfully search for information must have been a function of the number of visits and the duration of visits that they made to each of the assembly drawings (Eq 2). That is, the pipefitters’ search efficiency was a function of their visit counts and visit duration distributions for each assembly drawing (Eq 2). Visit counts and other visit duration distribution metrics were found to be associated with performance in the present work, so it is possible that the pipefitters who performed best during the pipe spool assembly task did so because they were most efficient at searching the assembly drawings for information.

$$\text{Search Efficiency} = f(\text{Visit Counts}, \text{Visit Duration Distributions}) \quad (2)$$

The relationship between search efficiency and visit duration distributions, as proposed herein, is theoretical, and based upon the findings in this study. However, search efficiency has been considered in eye tracking literature for some time. Various metrics such as scanpath length and convex hull areas have been proposed for measuring search efficiency (Blascheck et al. 2014; Goldberg and Kotval 1999; Poole and Ball 2005). The authors have

recently incorporated these additional metrics into the Visual Eyes eye tracking analysis application and are currently conducting further investigations related to search efficiency.

4.6 Limitations

This study was very limited by sample size. Twenty pipefitters participated in the study, and most of the visit, performance, and demographic metric relationships tested were not significant. If the study had included a larger sample size, then it's possible that more of the metric relationships could have been significant. Additional relationships between the visit metrics could have suggested particular strategies that the pipefitters were using to obtain information from the assembly drawings, and additional relationships between the visit and performance metrics and the visit and demographic metrics could have led to insights for improving craftworker performance.

It could also be said that this work was limited by the fact that the visual information gathering strategies used by the pipefitters were unknown. This work characterized the pipefitters' information gathering strategies through visit metrics; however, an alternative approach to investigating information gathering strategies would have been to interview the pipefitters. If the pipefitters had been asked to describe their visual information gathering strategies, then it's possible that a taxonomy of information gathering strategies could have been developed and the resulting strategies could have been investigated for associations with performance and demographics.

5 CONCLUSION

In the present work, existing eye tracking analysis methods initially developed for motor vehicle driver behavior studies were applied to quantify and compare the strategies used by professional pipefitters in obtaining visual information from assembly drawings during a construction model assembly task. No particular information gathering strategies were suggested by the results of the visit metric correlations; however, several of the visit metrics were associated with the performance and demographics of the pipefitters. Results indicated that the best performers in the study were the individuals who looked at the assembly drawings the least amount and were also more consistent in their search. While these findings are noteworthy, they are not enough to identify a singular search strategy would maximize performance.

This work was an exploratory effort to begin considering how construction productivity might be improved through incremental modifications to construction drawings or craftworker training for reading construction drawings. Future efforts could build upon this work by introducing drawing modification interventions or training interventions, and then determine how the interventions influenced craftworker interactions with construction drawings, search efficiency, and performance. For example, a drawing modification interventions might include the use of color or callout notes to improve search efficiency.

6 ACKNOWLEDGEMENTS

The presented work has been supported by the National Science Foundation (NSF) through grant award number 1928398. The authors gratefully acknowledge NSF's support. Any opinions, findings, conclusions, and recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

- Alruwaythi, O. F. (2017). "Influence of Engineering Information Format, Demographics and Spatial Cognition on Craft Worker Performance." Doctoral Dissertation. University of Colorado Boulder. p. 121.
- Alruwaythi, O. F., and Goodrum, P. M. (2019). "A Difference in Perspective: Impact of Different Formats of Engineering Information and Spatial Cognition on Craft-Worker Eye-Gaze Patterns." *Journal of Construction Engineering and Management*, 145(11), 1–14.
- Alruwaythi, O. F., Sears, M. H., and Goodrum, P. M. (2017). "The Impact of Engineering Information Formats on Craft Worker Eye Gaze Patterns." *ASCE International Workshop on Computing in Civil Engineering 2017*, American Society of Civil Engineers, Seattle, WA.
- Beijer, D., Smiley, A., and Eizenman, M. (2004). "Observed driver glance behavior at roadside advertising signs." *Transportation Research Record*, (1899), 96–103.

- Blascheck, T., Kurzhals, K., Raschke, M., Burch, M., Weiskopf, D., and Ertl, T. (2014). "State-of-the-Art of Visualization for Eye Tracking Data."
- Brown, C. B., Collins, T. R., and McCombs, E. L. (2006). "Transformation from batch to lean manufacturing: The performance issues." *EMJ - Engineering Management Journal*, 18(2), 3–14.
- Bureau of Labor Statistics. (2019). "American Time Use Survey - 2018 Results." (202), 1–26.
- Chalhoub, J., and Ayer, S. K. (2018). "Using Mixed Reality for electrical construction design communication." *Automation in Construction*, Elsevier, 86(October 2017), 1–10.
- Chen, W., and Adler, J. L. (2019). "Assessment of Screen Exposure in Young Children, 1997 to 2014." *JAMA Pediatrics*, 173(4), 391–393.
- Dadi, G. B., Goodrum, P. M., Taylor, T. R. B., and Carswell, C. M. (2014a). "Cognitive Workload Demands Using 2D and 3D Spatial Engineering Information Formats." *Journal of Construction Engineering and Management*, 140(5), 04014001.
- Dadi, G. B., Goodrum, P. M., Taylor, T. R. B., and Maloney, W. F. (2014b). "Effectiveness of communication of spatial engineering information through 3D CAD and 3D printed models." *Visualization in Engineering*, 2(1), 9.
- Dadi, G. B., Taylor, T. R. B., Goodrum, P. M., and Maloney, W. F. (2014c). "Cognitive Demands of Craft Professionals Based on Differing Engineering Information Delivery Formats." 767–776.
- Dadi, G. B., Taylor, T. R. B., Goodrum, P. M., and Maloney, W. F. (2014d). "Performance of 3D computers and 3D printed models as a fundamental means for spatial engineering information visualization." *Canadian Journal of Civil Engineering*, 41(10), 869–877.
- Driver Focus-Telematics Working Group. (2006). Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems Including 2006 Updated Sections.
- Ekstrom, R. B., French, J. W., Harman, Ha. H., and Dermen, D. (1976). *Manual for Kit of Factor-Referenced Cognitive Tests*. Princeton, New Jersey.
- George, D., & Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference*, 17.0 update (10a ed.) Boston: Pearson.
- Goldberg, J. H., and Kotval, X. P. (1999). "Computer interface evaluation using eye movements: Methods and constructs." *International Journal of Industrial Ergonomics*, 24(6), 631–645.
- Goodrum, P. M., Miller, J., Sweany, J., and Alruwaythi, O. F. (2016). "Influence of the Format of Engineering Information and Spatial Cognition on Craft-Worker Performance." *Journal of Construction Engineering and Management*, 142(9), 04016043.
- Horrey, W. J., and Wickens, C. D. (2007). "In-Vehicle Glance Duration." *Transportation Research Record: Journal of the Transportation Research Board*, 2018(1), 22–28.
- Huey, E. B. (1900). "On the Psychology and Physiology of Reading." *The American Journal of Psychology*, 11(3), 283–302.
- ISO. (2014a). "ISO 15007-1:2014 Road Vehicles - Measurement Of Driver Visual Behaviour With Respect To Transport Information And Control Systems - Part 1: Definitions and parameters." International Organization for Standardization.
- ISO. (2014b). "ISO 15007-2:2014 Road vehicles — Measurement of driver visual behaviour with respect to transport information and control systems - Part 2: Equipment and procedures." International Organization for Standardization.
- Liang, Y., Lee, J. D., and Yekhshatyan, L. (2012). "How dangerous is looking away from the road? Algorithms predict crash risk from glance patterns in naturalistic driving." *Human Factors*, 54(6), 1104–1116.

- NHTSA. (2016). "Visual-Manual NHTSA Driver Distraction Guidelines for Portable and Aftermarket Devices." *Federal Register- Rules and Regulations*.
- Perneger, T. V. (1998). "What's wrong with Bonferroni adjustments." *British Medical Journal*, 316(7139), 1236–1238.
- Poole, A., and Ball, L. J. (2005). "Eye Tracking in Human-Computer Interaction and Usability Research : Current Status and Future Prospects." *Psychology*, 211–219.
- Sears, M., Alruwaythi, O., and Goodrum, P. (2020). "Current Eye Tracking Glance and Visit Analysis Standards can Produce Inconsistent Results." *Under review, submitted to Automation in Construction*.
- Sears, M. H. (2020). "Visual Eyes." <<https://github.com/mattsears18/visual-eyes>> (Mar. 26, 2020).
- Shah, R., and Ward, P. T. (2003). "Lean manufacturing: Context, practice bundles, and performance." *Journal of Operations Management*, 21(2), 129–149.
- Sweany, J., Goodrum, P. M., and Miller, J. (2016). "Analysis of empirical data on the effects of the format of engineering deliverables on craft performance." *Automation in Construction*, Elsevier B.V., 69, 59–67.
- Tivesten, E., and Dozza, M. (2014). "Driving context and visual-manual phone tasks influence glance behavior in naturalistic driving." *Transportation Research Part F: Traffic Psychology and Behaviour*, Elsevier Ltd, 26(PA), 258–272.
- Tobii Pro AB. (2019). "Tobii Pro Lab User's Manual."
- Twenge, J. M., Martin, G. N., and Spitzberg, B. H. (2018). "Trends in U.S. Adolescents' Media Use, 1976-2016: The Rise of Digital Media, the Decline of TV, and the (Near) Demise of Print." *Psychology of Popular Media Culture*, 8(4), 329–345.
- Wonnacott, T. and Wonnacott, R. (1986). Regression: A Second Course in Statistics. Krieger Publishing Company. ISBN: 0898749700
- Zhang, H., Smith, M. R. H., and Witt, G. J. (2006). "Identification of real-time diagnostic measures of visual distraction with an automatic eye-tracking system." *Human Factors*, 48(4), 805–821.