The Use of Virtual Welding Simulators to Evaluate Experienced Welders

The ability of virtual reality welding simulations to evaluate existing skills in experienced and trained novice welders was assessed

BY A. P. BYRD, R. T. STONE, R. G. ANDERSON, AND K. WOLTJER

ABSTRACT

Virtual reality welding simulations have been, and continue to be, a trend in welding training programs. The goal of this study was to examine the use of virtual reality simulations as an assessment tool for existing welders. This study used a virtual reality welding simulator, VRTEX® 360, to assess the existing skills of experienced and trained novice welders. This study also used the shielded metal arc welding (SMAW) process to perform simple and complex welds. Performance was evaluated through a quality score, which was based on the following five welding parameters: arc length, position, work angle, travel angle, and travel speed. The virtual reality welding simulator was able to evaluate performance, but it could not distinguish between experienced and trained novice welders. On average, experienced welders as a group scored 10 quality points higher than trained novice welders. Welding experience also had a large to very large effect on the quality score for each weld type. One identified trend for both experienced and trained novice welders was as weld difficulty increased, the quality score decreased. It is recommended that industries use virtual reality simulators to evaluate welders for ensuring high-quality welding in production practices.

KEYWORDS

- Virtual Reality Simulations
- Shielded Metal Arc Welding (SMAW)
- Experienced Welders
- Trained Novice Welders

Introduction

An emerging method of assessment within several industries is the immersion of individuals into a virtual reality (VR) simulation (Ref. 1). Since the conception of virtual reality welding simulations, many strides have been made to enhance the virtual reality experience. One major stride has been creating the most realistic weld pool. The realistic 3D weld pool along with the creation of human welder response models have led to many other advancements in robotic welding and welding training through the years. Human welder response models were created by observing the reactions of a welder to the appearance of the weld pool during a welding process (Ref. 2). Understanding how a welder reacts to the weld pool may lead to why novice welders perform at a lower level than advanced welders (Ref. 3).

In the realm of robotic welding, human welder response models have been used to bridge the gap between programmed automated welding machines and automated welding machines with a human welder’s intelligence (Ref. 4). This means the automated welding machines can change welding parameters during a welding process based upon the appearance of the weld pool. The ability to do this comes from the creation of neuro-fuzzy logic based upon the human welder response models (Ref. 5). These abilities of the automated welding machines with a human welder’s intelligence have been verified by comparing the reactions of an experienced welder to the recreation of an automated welding machine (Ref. 6). This has led to the machine-human cooperative control scheme, which uses machine algorithms to help less experienced welders produce better quality welds. This new control scheme has helped generate the avenue for developing today’s welder training systems (Ref. 7).

Enhancements in welding training have stemmed from the creation of neuro-fuzzy logic for robotic welding machines. From this advancement, visual feedback devices have been created for virtual reality trainers to help novice welders correct themselves (Ref. 8). Another feedback system that has been created incorporates vibration sensors on the welding helmet to give feedback to novice welders. This method has been found to be easier to integrate into a training program than the visual feedback system (Ref. 9).

From these feedback devices came virtual reality welding simulation trainers that fully immerse a student into a virtual reality environment. Some of the virtual reality welding simulators are ARC+, the Fronius virtual welding system, and VRTEX® 360 (Ref. 10). Industries have used VR simulations to allow trainees to learn basic skills in a safer environment (Ref. 11). One study found that full and partial VR integration into a welding training program was appropriate,
but depended on the level of task difficulty (Ref. 12). However, the ability of VR simulations to evaluate existing skill has received only limited attention.

Virtual reality (VR) environments can be used to train workers to acquire the basic skills to perform the tasks required for a job in a technical field. However, performance in a VR environment could be used as an indicator to hire an individual. Training within a VR environment can prepare a trainee to anticipate and recognize when situations go awry, as well as to test an individual’s decision-making skills under normal and stressful conditions. Such evaluation is possible through dynamic and continuously changing VR environments. Training in such environments can lead to increased memory retention, reduced human error, and a deeper understanding of the complexities of a work environment. Evaluating the critical thinking skills is also possible by evaluating how an individual adapts to changing conditions within a VR environment (Ref. 13).

A study conducted at Yale University trained surgical residents in postgraduate years 1–4 utilizing virtual reality surgical simulation to train skills and reduce surgical error in the operating room. It was concluded that residents trained on the VR surgical simulator were 20% more efficient at performing gallbladder surgery. The residents who were not trained on the simulator were five times more likely to injure the gallbladder and nontarget tissue. The study found that using a VR surgical simulator helped the residents hone their surgical skills in a safe environment, thus reducing surgical error in the operating room. It was concluded that residents trained on the VR surgical simulator were 20% more efficient at performing gallbladder surgery. The residents who were not trained on the simulator were five times more likely to injure the gallbladder and nontarget tissue. The study found that using a VR surgical simulator helped the residents hone their surgical skills in a safe environment, thus reducing surgical error in the operating room.

Virtual reality offers the potential to act as an evaluation tool because every simulation within the virtual reality environment can be recorded and produce quantitative feedback that could facilitate assessment tasks (Ref. 16). Research has found that simulations can be used with younger and older individuals. Furthermore, the focus of incorporating virtual simulation is changing from educating novices and interns to the importance of continued training for experienced personnel. One example of experienced personnel using VR is found in aviation. Aviation simulations are applied regularly to individuals and to groups regardless of participant seniority (Ref. 17).

Another example is the utilization of medical simulators. One study that used virtual reality medical simulations found that on average medical residents \((M = 64.9)\) performed better than medical students \((M = 57.1)\) (Ref. 10). The study suggested that VR simulations could play a future role in continuing medical education and re-certification (Ref. 18).

A VR simulation in certifying medical personnel is an important consideration for medical programs. Several associations and training programs are considering the use of simulators for health care professionals’ skills certification. For example, the Joint Commission on Allied Health Personnel in Ophthalmology (JCAHPO) uses simulator-based evaluations to certify certified ophthalmic technician skills. Simulator-based evaluations have replaced hands-on skills used in JCAHPO evaluations in 250 test centers nationwide. Research has found simulations...
that look and feel like actual procedures help clinicians develop skills and maintain those skills throughout their professional practice (Ref. 19).

Theoretical Framework

The theory of individual differences in task and contextual performance guided this study. Individual performance is described as behavioral, episodic, evaluative, and multidimensional. Furthermore, performance is defined as an aggregate of behavioral episodes that add value to an organization. The theory of individual differences uses the distinction between task and contextual performance to identify and define behavioral episodes to describe an individual’s performance.

Contextual performance refers to behaviors that influence the psychological, social, and organizational environments of an organization. Task performance refers to an individual’s effect on the technical core of an organization, which assembles the products of that organization. Task performance is divided into two types of tasks. The first type of tasks includes the transformation of raw material into a finished product or service. The second includes service and maintenance of the technical core by helping restock raw materials, move finished products, plan, and supervise or coordinate the first type of task performance (Ref. 20).

For this study, the researchers focused on the first type of task performance that deals with the construction of finished products. Task skill is affected directly by an individual’s cognitive abilities, such as prior experiences. Additionally, cognitive ability can have positive or negative effects on performance depending on the nature of prior experiences.

Specifically, cognitive ability can affect three intervening variables of task performance: task knowledge, task skill, and task habit. Task knowledge, which is largely affected by an individual’s cognitive ability, allows for a higher rate of retention and mastery of procedures, principles, and facts. Task skill refers to an individual’s ability to use technical information, make judgments, and perform technical skills related to an organization’s core functions. Task habits refer to patterns of learned behavior that contribute to or impede the execution of organizational tasks (Ref. 21).

Within a task performance, the researchers focused specifically on the effect of an individual’s cognitive ability for task knowledge and task skill when completing welding-related tasks.

Research Goal

The goal of the study was to examine the ability of VR as an effective assessment tool for existing skill among welders. The researchers addressed this goal by comparing experienced welders to trained novice welders in terms of participants’ VR performance. Performance was defined in terms of a quality score based on five welding parameters.

For this study, the VRTEX® 360 welding simulator was selected because it is capable of providing realistic simulations that were appropriate for this study. The researchers hypothesized that a VR simulator would be able to indicate the difference between experienced welders and trained novice welders.

Methods

Experimental Materials

This portion of the study used one VRTEX® 360 virtual reality arc welding training machine with a SMAW stinger, helmet, and plastic coupons (see Figs. 1 and 2). This trainer was chosen because it was the highest fidelity VR simulator on the market at the time of this study. Also, the machine allowed users to be fully absorbed in a VR welding environment. Participants wore a welding helmet with integrated stereoscopic VR screens (see Fig. 3).

Location

This study took place at several locations to obtain a sufficient number of participants during the fall of 2013. The research sites varied between nine locations ranging from classrooms to union halls. The researchers moved the VRTEX® 360 virtual reality arc welding trainer from one facility to the next when invited. The day, time, and number of participants at each location varied because of the specific times requested by each location.

Participants

The population of this study consisted of 49 male participants of vary-
ing ages. In addition, all the participants of this study had completed a formal welding training program and were employed as welders. They were categorized as either experienced welders or trained novice welders. Experienced welders were categorized as either having at least 10 years of welding experience or being a certified welder. Trained novice welders were individuals who had no more than one year of experience. This study included 18 experienced and 31 trained novice welders.

Independent and Dependent Variables

The dependent variable was the quality score for the SMAW test welds (2F, 1G, 3F, 3G). The two independent variables were the experience level of participants and weld type in order of increasing difficulty as follows: 2F (horizontal fillet weld), 1G (flat groove weld), 3F (vertical fillet weld), and 3G (vertical groove weld).

Participant performance was evaluated using a quality score generated by the VR simulator. The quality score was determined by averaging the individual scores of five welding parameters: weld position, arc length, work angle, travel angle, and travel speed. Weld position refers to where the weld bead is formed in relation to the center of the weld joint. The appropriate distance from the tip of the electrode should be at the weld coupon, which is termed arc length. Work and travel angles are the appropriate horizontal and vertical angles the welder should keep between the electrode and weld coupon. Travel speed is the appropriate horizontal speed that a welder should move the electrode across a weld joint.

After the completion of each test weld, the VRTEX® 360 welding simulator calculated a score that ranged from 0 to 100 for each of the five welding parameters. The welding simulator also averaged the five parameter scores into an overall quality score, which was recorded for each participant. For a participant to achieve a high-quality score, the participant had to maintain each welding parameter within a certain range. For this study, the VRTEX® 360 welding simulator was using AWS D1.1, Structural Welding Code — Steel, welding specifications as the basis for evaluating the welds.

The researchers specifically wanted to know how the VR simulator would be used if brought in as an assessment tool in a real-world setting. As a result, the study was designed in a manner consistent with an independent evaluator used to assess a worker’s welding skills. Therefore, participants were not trained to use all of the features available on the VRTEX® 360 welding simulator. Allowing participants to fully learn how to use the simulator would not have aided in answering the research questions of this study.

Experimental Procedure

At each research site prior to experimentation, participants were given informed consent forms followed by a questionnaire. The researchers then trained all participants to use the VRTEX® 360. To minimize usability of the VR system as a factor, all participants were allowed enough time to become acclimated with the system until they were comfortable with using the VR simulator. This included participants being able to correctly adjust the stereoscopic VR screens and produce a proficient bead on a practice plate. The participants then were instructed to perform test welds as they would in a real-world welding environment.

The sequence of the tests performed were 2F (horizontal fillet weld), 1G (flat groove weld), 3F (vertical fillet weld), and 3G (vertical groove weld). The tests were done in sequence with no breaks in between. It should be noted that participants had only one chance to complete the test welds, and the settings for each weld were identical for each participant.

The score for each test weld was recorded from the live action student evaluation report screen (Ref. 22). This protocol was followed with every participant for this portion of the study at all locations.

Limitations

The skilled worker shortage in the welding industry (Ref. 23) presented this study with some limitations. The first limitation was the short amount of time we had with the participants of this study. The limited time frame only allowed for a specific amount of data to be collected. This time frame also did not allow for the collection of real-world welds to compare to the VR welds completed.

Results

The performance measures for this study were welder’s experience and test weld quality scores; the relationship between these variables was also examined.

Test Weld Quality Score

Once participants completed each test weld, a quality score was recorded (see Table 1). Differences were identified between experienced and trained novice welders in several instances. A major difference was found in the range in quality scores, where trained novices had a wider range of scores than did experienced welders. Another difference was identified when examining the minimum quality scores.

Experienced welders maintained an average minimum score in the low 70s for all weld types; the 3G weld yielded the lowest quality score at a 49. Trained novice welders’ minimum scores fluctuated between 20 (3G) and
61 (2F). Examining the standard deviation, experienced welders were more consistent in ability than were trained novice welders for each weld type. Consistency amongst experienced welders is very evident with the 3F weld type, in which the standard deviation was 4.43.

When comparing the minimum score of the experienced welders and the maximum score of the trained novice welders, it is evident that there is overlap. This data illustrates that the VR simulator can evaluate current skill, but it cannot accurately identify an individual as an experienced welder or a novice welder.

The test weld quality scores were averaged for each weld type by welder experience. An overall average was also calculated by averaging all scores from each experience level. In Fig. 4, the scores for the experienced welder group are plotted to show the average with error bars. The novice trained welder group scores are plotted in Fig. 5.

Based on the data, the experienced welders outperformed the trained novice welders by an average of 10 quality points. As a descriptive trend, the experienced welders outperformed the trained novice welders on all weld types. The separation in quality scores grew progressively as the weld difficulty increased. The separations in average quality scores were 2F (7.39), 1G (10.21), 3F (10.53), and 3G (15.04).

Another identified trend was that the 2F weld was the easiest weld performed by welders of both experience levels. The scores then decreased progressively as the weld difficulty increased. Overall performance was highest among the experienced welder group with an average of 83, which was higher than the trained novice welder group by 12 quality points.

To determine whether the groups were significantly different, it was necessary to determine whether the data were normal and had homogeneity of variance. Homogeneity of variance was examined by conducting Levene’s Test (Ref. 24) (see Table 2). Each test was conducted using $\alpha = 0.05$ to indicate significance. Significance of the Levene’s Test indicates that equal variances are not assumed. All weld types, except for 3F, had equal variances; 3F has a $p$-value of 0.015, indicating which line to use to interpret the following $T$ test.

After homogeneity of variance was calculated, a $t$ test was calculated for the levels of experience on all four weld types. The $t$ test was conducted for each weld using $\alpha = 0.05$ and 47 deg of freedom (see Table 3). The results identified statistical significance for all weld types with $p$ values ranging from 0.000 to 0.007. These findings indicated that a significant difference existed between the experienced welders’ and trained novice welders’ average quality score for each weld.

To examine the effect that experience had on the average quality score, Cohen’s $d$ (representing the effect size of a statistical significant relationship found with a $t$ test) was calculated, which describes the effect size on the dependent variables, and is interpreted as small (0.2), medium (0.5), and large (0.8) (Ref. 14). According to the results of the Cohen’s $d$, experience had a large effect on the 2F (0.830) and 1G (0.977). The 3F (1.582) and 3G (1.024) weld types exhibited a very large effect as Cohen’s $d$ was above 1.00.

**Discussion**

For the effectiveness of VR as an assessment tool to evaluate existing ability in welders, it was hypothesized that VR simulations would indicate a difference between experienced and trained novice welders. A discussion in light of this hypothesis follows for each of the weld types.

---

**Table 2 — Levene’s Test for Equality of Variance**

<table>
<thead>
<tr>
<th>Weld Type</th>
<th>F Ratio</th>
<th>$P$ Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>2F</td>
<td>0.657</td>
<td>0.422</td>
</tr>
<tr>
<td>1G</td>
<td>1.402</td>
<td>0.242</td>
</tr>
<tr>
<td>3F</td>
<td>6.315</td>
<td>0.015</td>
</tr>
<tr>
<td>3G</td>
<td>2.458</td>
<td>0.124</td>
</tr>
</tbody>
</table>

**Table 3 — T-Test Results for Welders’ Experience**

<table>
<thead>
<tr>
<th>Weld Type</th>
<th>$T$</th>
<th>$P$ Value</th>
<th>Cohen’s $d$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2F</td>
<td>2.846</td>
<td>0.007</td>
<td>0.830</td>
</tr>
<tr>
<td>1G</td>
<td>3.352</td>
<td>0.002</td>
<td>0.977</td>
</tr>
<tr>
<td>3F</td>
<td>5.426</td>
<td>0.000</td>
<td>1.582</td>
</tr>
<tr>
<td>3G</td>
<td>3.512</td>
<td>0.001</td>
<td>1.024</td>
</tr>
</tbody>
</table>
2F Weld Type

In this study, the 2F weld type was the simplest to complete. This was seen in both groups as the 2F weld had the highest average quality scores of all weld types tested. This finding indicates that the welders, both experienced and trained novices, are, on average, most competent to complete the 2F weld type. However, the VR simulator was not able to distinguish between an experienced and trained novice welder. From the results obtained, it was determined that experience level had a large effect on the weld quality score for each group of participants, experienced and trained novice.

1G Weld Type

The 1G weld type was the second-easiest weld for both levels of welders, experienced and trained novices. The difficulty was identified in the results as the average quality scores were second highest when performing the 1G weld. The 1G average quality score decreased from the average quality scores of the 2F weld type.

The relationship between experience and quality score was apparent in both weld types using the t-test. Specifically, the findings revealed a large effect of experience on the 1G weld type quality score. The results of the VR simulator continued to show a difference between the experienced and trained novice welders with the 1G weld type.

3F Weld Type

The vertical fillet weld, 3F, was the second-most-difficult weld performed in this study by both experienced and trained novice welders. The average quality scores continued to decrease as the 3F weld type was identified as more complicated to perform than the 2F and 1G weld types. A distinct difference was identified between the quality scores of experienced and trained novice welder groups for the 3F vertical fillet weld. The results also revealed that the experience level had a very large effect on the quality score for the 3F weld type.

3F performance showed the influence of experience more than any other weld type in this study. Experienced welders also demonstrated more consistent ability when examining the standard deviation of the quality scores on the 3F weld type compared to trained novice welders. The VR simulations were able to distinguish between both experienced and trained novice welders for the 3F weld type.

3G Weld Type

The most difficult weld performed by the experienced and trained novice welders in this study was the 3G weld type. Experienced and trained novice welders both scored the lowest quality scores on this weld type. A distinct difference was also apparent using the VR simulation between experienced and trained novice welders for the 3G weld type. Experience was also identified as having a large effect on quality scores.

Conclusions

The results of this study suggest that VR simulations can be used as assessment tools to assess existing skill levels in welders. The differences between the experienced and trained novice welders were distinct for all weld types examined in this study. These findings support the theory of individual differences because individuals with more experiences were able to perform tasks at a higher level of success due to increased task knowledge and skill.

It can also be concluded that experienced welders as a group were able to perform significantly better than the trained novice welder group. Both groups showed a trend of decreasing quality scores as the weld difficulty increased. The 2F and 1G welds tended to be easier for both the experienced and trained novice welders. The 3F and 3G welds were more complex in nature, and yielded a lower quality score from both experienced and trained novice welders. This conclusion supports the studies based on the theoretical framework that experience allows welders to perform better because of prior task knowledge and task skill of each weld type. However, the quality scores were able to identify which weld types the welders were most and least competent in.

Because VR has demonstrated the ability to assess existing welding skills, this method could be used to track a welder’s skills over a period of time. This use of longitudinal data could be used in quarterly worker assessments for promotional purposes as well as for identifying when novice welders are ready to test for certification. In a setting where VR simulation would be used to assess existing skills, a system could be put into place for routine assessment to ensure high-quality welds in a production setting. The results of this study demonstrated the ability of VR simulation, specifically VRTEX® 360, to assess the welder’s existing skills in terms of a quality score based on five welding parameters (position, arc length, work angle, travel angle, and travel speed).

Because of the limitations this study faced, it is recommended to conduct a follow-up study that will be able to validate these findings by comparing participants’ VR welds to real-world welds. Furthermore, it is recommended to evaluate the participants on each individual welding parameter as well as the quality score.

References


The Welding Journal can now be enjoyed by AWS members for free using their computers or Internet-ready mobile phones or tablets, including iOS (iPad® and iPhone®), Android®, Windows 7®, and HP Web OS® devices.

Presented below are three key areas with details for reading the magazine in various ways.

• For desktops/laptops, the Welding Journal can be read online, or downloaded for offline reading and archiving. You can also do custom searches within a single issue or throughout all the issues in the archive (dating back to December 2011), and print individual pages or whole issues. Active links add convenience.

• For mobile devices, the Welding Journal can be used on Internet-capable tablets or phones. A web app runs in your mobile browser and automatically recognizes the device, optimizing presentation and functionality. It is the best choice for those who prefer reading content online vs. downloading to their mobile device.

• For iPad and iPhone users, go to iTunes, type in “Welding Journal” in the search window, and download the app to retrieve the magazine online or download for immediate access. This is the best choice for those who prefer to store issues in their iOS devices for offline reading. In addition, build your Welding Journal catalog with all the available issues in the archive.

The digital edition will be sent to you automatically every month, but make sure your e-mail address is up to date by logging in at www.aws.org.