# Virtual Reality

## *Does it have a role in certification testing?*

**NCCCO FOUNDATION**

*An NCCCO Foundation Report*

# Virtual Reality

## *Does it have a role in certification testing?*

Wallace Judd, PhD,
President, Authentic Testing

March 2020

NCCCO
FOUNDATION

This report contains the results of a study commissioned by the National Commission for the Certification of Crane Operators (NCCCO) following a demonstration of virtual reality (VR) technology at NCCCO's Lift Safety Zone at CONEXPO 2017. An ad hoc group of experienced crane operators was subsequently convened at which generally favorable reactions to the technology were recorded.

Following consultation with the American National Standards Institute (ANSI), NCCCO commissioned a study to try to determine if VR could play a role in high-stakes testing such as that conducted by NCCCO to the point where VR might even substitute for some or all of a conventional practical test on a real crane.

A pilot study was followed by a larger study, the results of which are published here by the NCCCO Foundation through its SIREN information network.

Thanks and acknowledgements are due the following:

**Wallace Judd**, PhD, President, Authentic Testing, who designed and conducted the study and authored this report. wjudd@AuthenticTesting.com

**Zack Parnell**, President & CEO, Industrial Training International (ITI), who provided all of the VR equipment and handled the logistics of testing including the scheduling of candidates.

**Roy Swift**, PhD, who as Senior Director of ANSI Personnel Credentialing Accreditation Programs, provided the inspiration for the study.



The **NCCCO Foundation** is a 501(c)3 charitable organization formed to promote industry safety by conducting research, educating about–and advocating for–personnel safety and certification, and facilitating access by youth, military personnel, and underserved communities to education, training and certification through scholarships and grants.

**NCCCO Foundation**
2750 Prosperity Ave., Suite 520
Fairfax, VA 22031, USA.
info@ncccofoundation.org
www.ncccofoundation.org

**S**afety
**I**nformation
**R**esearch
**E**ducation
**NETWORK**

NCCCO FOUNDATION

# CONTENTS

# Can Virtual Reality Play a Role in High-Stakes Certification Testing?

## Goal of the Study

The goal of this study was to determine the extent to which a Virtual Reality (VR) crane operator test could predict the results of a test conducted on an actual crane. Training and testing sessions were conducted on ITI VR Crane Simulators developed in partnership by Industrial Training International (ITI) and Serious Labs.

A preliminary study had previously been conducted, but variations in location, crane types, VR simulators and preparation limited the reliability of the conclusions. The study described below resolved many of these issues by requiring minimum VR practice, limiting testing to one location, and reducing the number of crane types.

## Overview

From May to October 2019, 40 candidates for NCCCO mobile crane operator certification in three categories (swing cab telescopic, fixed cab telescopic, and lattice boom crawler crane) were tested both via Virtual Reality and on an actual crane. The results indicated that the VR test is a highly reliable measure for predicting a passing score on an actual crane. The likelihood of a candidate passing the VR test but failing the subsequent parallel test on an actual crane was 5.6%. The likelihood that a candidate will get the same pass/fail score on both the VR test and on the actual crane is 0.87.

The remainder of this report describes the prerequisites, the testing conditions and the outcomes of the tests.

## Candidate Demographics

All candidates were male and had an average age of 40.4 years.

| Age | | |
|---|---|---|
| 18-24 | 2 | 5% |
| 25-34 | 13 | 33% |
| 35-44 | 11 | 28% |
| 45-54 | 8 | 20% |
| 55-64 | 4 | 10% |
| 65+ | 1 | 3% |
| NA | 1 | 3% |
| **Total** | 40 | 100% |

Candidates had an average of 6.3 years’ experience. Rows in the table below show the number of candidates above the >= number, and equal to or below the middle column.

| **Years** | | |
|---|---|---|
| **>=** | **<=** | **N** |
| | 0 | 9 |
| 1 | 3 | 11 |
| 4 | 5 | 5 |
| 6 | 10 | 5 |
| 11 | 20 | 7 |
| 21 | | 3 |
| | **Total** | **40** |

Candidates had an average of 2,366 hours of crane experience.

| **Hours** | | |
|---|---|---|
| **>=** | **<=** | **N** |
| | 0 | 9 |
| 1 | 100 | 4 |
| 101 | 200 | 3 |
| 201 | 500 | 5 |
| 501 | 1,000 | 3 |
| 1,001 | 10,000 | 14 |
| 10,001 | 30,000 | 2 |
| | **Total** | **40** |

Candidates’ crane experience in years and hours correlated as follows:

| Years | | | Hours | | |
|---|---|---|---|---|---|
| >= | <= | N | >= | <= | N |
| 0 | 0 | 9 | | 0 | 8 |
| | | | 1 | 200 | 1 |
| 1 | 5 | 16 | | 0 | 1 |
| | | | 1 | 200 | 6 |
| | | | 201 | 1,000 | 6 |
| | | | 1,001 | 10,000 | 3 |
| 6 | 10 | 5 | 201 | 1,000 | 2 |
| | | | 1,001 | 10,000 | 2 |
| | | | 10,001 | 100,000 | 1 |
| 11 | 20 | 7 | 1,001 | 10,000 | 6 |
| | | | 10,001 | 100,000 | 1 |
| 21 | 50 | 3 | 1,001 | 10,000 | 3 |
| | **Total** | **40** | | **Total** | **40** |

## Testing Location

Instruction, VR study and testing, and actual crane testing reported in this study were conducted at the ITI Houston training center over the course of several months in 2019.

## Candidate Training

The candidates received 24 hours of classroom instruction in preparation for the actual crane test and written examinations, and an additional four (4) hours of crane practice. Prior to the VR testing, candidates received 45 minutes of VR practice on a VR desktop simulator with an Oculus Rift headset.

Immediately following the practice session, candidates were given a simulated test which followed the structure and sequence of the actual NCCCO certification test. Two tasks, the pre-operational inspection and the crane shutdown and securing procedure were evaluated using multiple choice items instead of open-ended questions. The other test tasks – ball in barrel, chain in stop circle, hand signals, zigzag corridor forward and backward – were all done with active simulations.

Candidates were also given the opportunity in the simulation for a 15-minute pre-test familiarization period before the operational portion of the practical exam, and a 5-minute pre- task familiarization period before the zigzag corridor tasks.

## Models of VR Simulators Used

The candidates took the exam initially on the VR desktop simulator. The simulator was developed by Industrial Training International, Inc. in partnership with Serious Labs, Inc.

The simulator displayed a full 360° simulation on an Oculus Rift CV1 headset. The display in the headset was duplicated on a monitor so a proctor could follow the progress of the candidate.

Hand controls and the simulation in the VR headset were specific to the actual crane being tested and matched the live controls as closely as possible. The cranes simulated were the Tadano GR-1000XL Rough Terrain Crane, the Link-Belt 218-HSL Lattice Boom Crawler Crane, and the Broderson IC-80 Carry Deck Crane. The hand controls for both hands were not visible to the candidate, which was an additional challenge of the VR examination.

***Rough Terrain Mobile Crane, Tadano 1000XL***

***Lattice Boom Crawler, Modeled after Link-Belt 218 HSL***

***Carry Deck Crane controls, Modeled after Broderson IC 80***


The simulator allowed candidates to use the appropriate controls and to watch the simulation as they completed the test. The simulator had both hand controls that allowed them to rotate the crane on its vertical axis, to manipulate the boom and raise or lower the headache ball.

As a practical matter, the desktop simulator mirrored only the hand controls of a crane. Below is the cab interior for the Link-Belt crane.

### Candidate Tasks

Candidates were tested on VR simulators of Tadano, Link-Belt and Broderson cranes. While the number of candidates tested was 40, several candidates completed multiple tests on VR and actual crane pairs. Each VR-Actual Crane pair is a test session. The number of test sessions included in the study was 53.

| VR Crane | N |
|---|---|
| TLL (Tadano) | 27 |
| LBC (Link-Belt) | 15 |
| TSS (Broderson) | 11 |
| **Total Sessions** | 53 |

After VR testing, candidates were then tested on an actual crane, either a Tadano, Link-Belt, Broderson or P&H crane.

| Actual Crane | N |
|---|---|
| TLL (Tadano) | 26 |
| LBC (Link-Belt) | 14 |
| LAT (P&H) | 1 |
| TSS (Broderson) | 12 |
| **Total Sessions** | 53 |

In some situations, because of crane availability at testing locations, candidates were tested on one crane model in the VR simulator and then tested on a different crane model for the live exam. These mismatches are noted in the table below.

| VR Type | N | CR Type | N | VR=CR Pass | VR=CR Fail | Agree | P0 |
|---|---|---|---|---|---|---|---|
| TLL (Tadano) | 27 | | | | | | 78% |
| | | TLL (Tadano) | 26 | 5 | 15 | 20 | |
| | | TSS (Broderson) | 1 | 0 | 1 | 1 | |
| LBC (Link-Belt) | 15 | | | | | | 73% |
| | | LBC (Link-Belt) | 14 | 6 | 4 | 10 | |
| | | LAT (P&H) | 1 | 1 | 0 | 1 | |
| TSS (Broderson) | 11 | | | | | | 82% |
| | | TSS (Broderson) | 11 | 7 | 2 | 9 | |
| **Total** | **53** | | | | | **Overall** | **77%** |

The P0 measure is called classification accuracy. It is the degree to which the VR test and the actual crane test agree on the classification of the candidate, either pass or fail.

**Types of Cranes Used**

*Tadano 1000XL*


*Link-Belt 218 HSL*


*Broderson IC 80*

**Test Layout**

The test layout in the VR simulator was identical to the layout of the actual NCCCO test. Below is the VR view of the layout of the simulated **zigzag corridor** as seen from the opposite side to the crane.

Note that the test weight has to be kept off the floor of the corridor, and hitting any of the poles will knock a tennis ball off the top and reduce points. The candidate must begin with the weight in the circle at one end of the zigzag corridor and negotiate through to the circle at the other end of the zigzag corridor.

Seen from the interior of the simulated crane, this is what the **ball-in-the-barrel** test looks like to the candidate. Note that, in this example, the ball is partially in the barrel.

**Hand signals** were also part of the test.  Here, the examiner is signaling the operator to move the boom to the left.

**Practice Effects**

The possibility of practice effects was considered, since most candidates took the VR test prior to the actual crane test. Practice effects had been measured in a preliminary study and found to be not significant. The likely reason for this is that the content of the actual crane test is published on the NCCCO web site, and many of the candidates had likely practiced either on portions of the test setup or simulated testing maneuvers as they practiced on their employers' cranes.

In addition, the study was designed to evaluate whether a VR simulation could be used to predict the score on an actual crane and not vice-versa.

**Results**

The data will first be presented as raw data directly from performance on the VR and actual crane. The results will then be presented as calculated using an Artificial Intelligence (AI) program to weight scoring components.

Raw Data

| | | Raw Data | |
|---|---|---|---|
| | | Crane | |
| | | Pass | Fail |
| VR | Pass | 19 | 8 |
| | Fail | 4 | 22 |

There were 53 testing sessions.

A measure called Classification Accuracy calculates the ratio of those candidates who were correctly classified (VR pass, Crane pass or VR fail, Crane fail) on the two measurement types to the total of all candidates tested.

Classification Accuracy = P0 = ( 19 + 22 ) / ( 53 ) = ( 41 ) / ( 53 ) = 0.77

The type one error is for many purposes the most critical measure. It is a measure of how many candidates failed the actual crane test after passing the VR test. This is potentially the most dangerous type of error, since it puts candidates of uncertain competence in the seat of a working crane.

Type 1 Error = 8 / ( 19 + 8 ) = 8 / 27 = 29.6%

Revised Analysis

After reviewing the test results it was clear that some subtests were more predictive of the total test score than others. For example, manipulating the headache ball through a zigzag corridor was a much better predictor than the walkaround segment of the test, even though both were scored.

To sophisticate this observation, an AI program was used to optimize the effectiveness of each subtest, weighting those that were highly predictive more heavily than those which were less predictive, and revising the cut score to minimize Type 1 Error. The algorithm used is called an evolutionary algorithm.

The results are shown in the table below. Note that in the raw analysis above, candidates in 27 sessions passed the VR test, whereas after application of the algorithm only 19 sessions resulted in a pass, as shown in the table below.

| | | Revised Data | |
|---|---|---|---|
| | | **Crane** | |
| | | **Pass** | **Fail** |
| **VR + AI** | **Pass** | 17 | 2 |
| | **Fail** | 6 | 28 |

However, the Classification Accuracy has improved from 0.77 to 0.85, a substantial improvement.

Classification Accuracy = P0 = ( 17 + 28 ) / ( 53 ) = ( 45 ) / ( 53 ) = 0.85

And the Type 1 Error has improved from 29.6% to 10.5%.

Type 1 Error = 2 / ( 17 + 2 ) = 2 / 19 = 10.5%

However, one of the candidates who passed the VR test but failed the actual crane test had zero hours of documented crane experience. Consequently, we felt it was appropriate to consider his data point an outlier and eliminate his data from the analysis. (This operator was clearly a video gamer with extensive VR experience but no documented crane experience. He actually scored a 98 on the VR test and 46 on the actual crane test.)

## Final Analysis

After removing that spurious data point, the results are as shown in this revised table:

| | | Revised Data | |
|---|---|---|---|
| | | **Crane** | |
| | | **Pass** | **Fail** |
| **VR + AI** | **Pass** | 17 | 1 |
| | **Fail** | 6 | 28 |

There are now 52 sessions in the analysis.

Classification Accuracy = P0 = ( 17 + 28 ) / 52 = 0.87

Type 1 Error = 1 / ( 17 + 1 ) = 1 / 18 = 5.6%

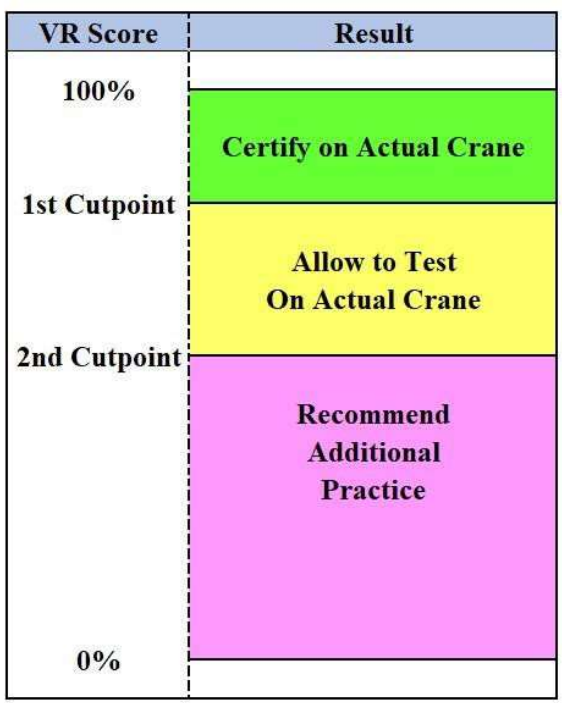
As a consequence, we can state that using the revised analysis, the likelihood of a Type 1 Error using a VR simulator to predict Pass/Fail performance on the same test is 5.6%.

Going Forward

Going forward, a likely strategy would be to set two cut scores for the VR test: Passing the $1^{st}$ Cutpoint would indicate that the candidate will likely pass the actual crane test. This could be used to certify the candidate or possibly administer an abbreviated confirmatory test on an actual crane.

Failing the $1^{st}$ Cutpoint but passing the $2^{nd}$ Cutpoint might entitle the candidate to a second test on the actual crane. Thus the candidate would have the ability to demonstrate the necessary skills at an additional cost.

Failing the $2^{nd}$ Cutpoint would indicate that the candidate is highly unlikely to pass the actual crane test and should practice more on an actual crane.



## Conclusions

While the results are encouraging, this study was too small to determine conclusively whether it is appropriate to certify a candidate using only VR technology as a replacement for an actual crane. Additional studies will be needed to evaluate VR certification as this technology becomes further refined.

Moreover, the conclusions of this study only generalize to:

- candidates who have a minimum number of documented hours of prior crane experience;
- sessions in which candidates have a minimum of 45 minutes directed VR practice; and
- test parameters which have been specified clearly and meticulously adhered to in the administration of the VR and actual crane tests.

The heavy equipment industry in general, and the crane industry in particular, can look forward to increased use of VR as more studies are conducted.

FOUNDATION

NCCCO
FOUNDATION