Developing a Data Strategy for Immersive Training Scenarios

Jennifer Solberg, Ph.D.
Quantum Improvements Consulting
Orlando, FL, USA

QUANTUM IMPROVEMENTS CONSULTING

VR is becoming commonplace in military training…but how do we know it works?

- Usability
- Behavioral measurement
- Physiological measurement

There is no single “right way” to do this.

Usability

Evaluates the “User Experience”

Negative reactions to AR/VR are frequently reported

- Physiological issues (headaches, nausea)
- Lack of situational awareness (outside of headset)
- Appropriate level of fidelity
- Occlusion of visual field

If users find the experience unpleasant, the solution will sit on a shelf.

The Five “Fs”

(Abich & Sikorski, 2002)

Fit	Ergonomics of the solution to meet user specifications
Feel	Physical sensations, perceptions, and emotional responses elicited from interaction with the solution
Form	The shape, size, dimensions, or other parameters of the solution
Function	Validity and reliability of the solution to perform as intended
Future Use	The extent to which users foresee a training benefit based on their interaction with the solution

Measuring User Experience

- Combination of qualitative and quantitative measures
- Think-aloud protocol
- Standardized and custom measures
- Feedback questionnaires
- Eye tracking and physiological measures

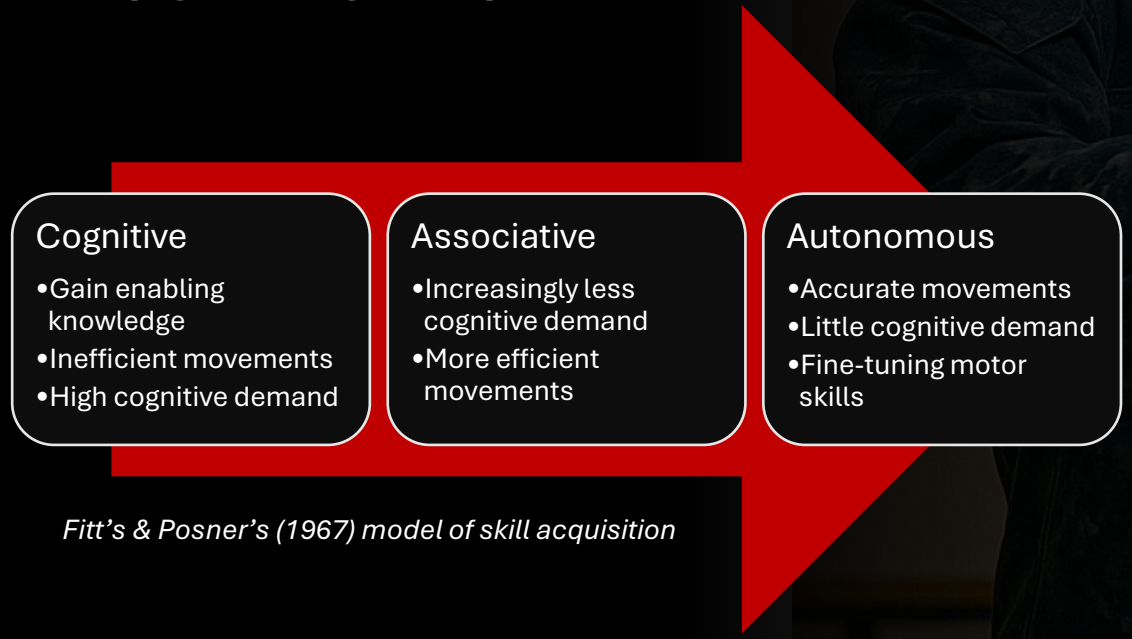
Resources:

- https://Usability.gov
- https://www.nngroup.com/topic/user-testing/
- https://www.interaction-design.org/literature/topics/usability

Behavioral Measurement in VR

Why is this important?

- Often, VR solutions just report on scenario completions, time spent in the scenario, or other summary measures.
- More granular data is needed to inform effective remediation.
- Allows real-time training content adaptation.
- VR is best used for skills development, and ***skills develop gradually with practice***.



Fitt's & Posner's (1967) model of skill acquisition

Novice

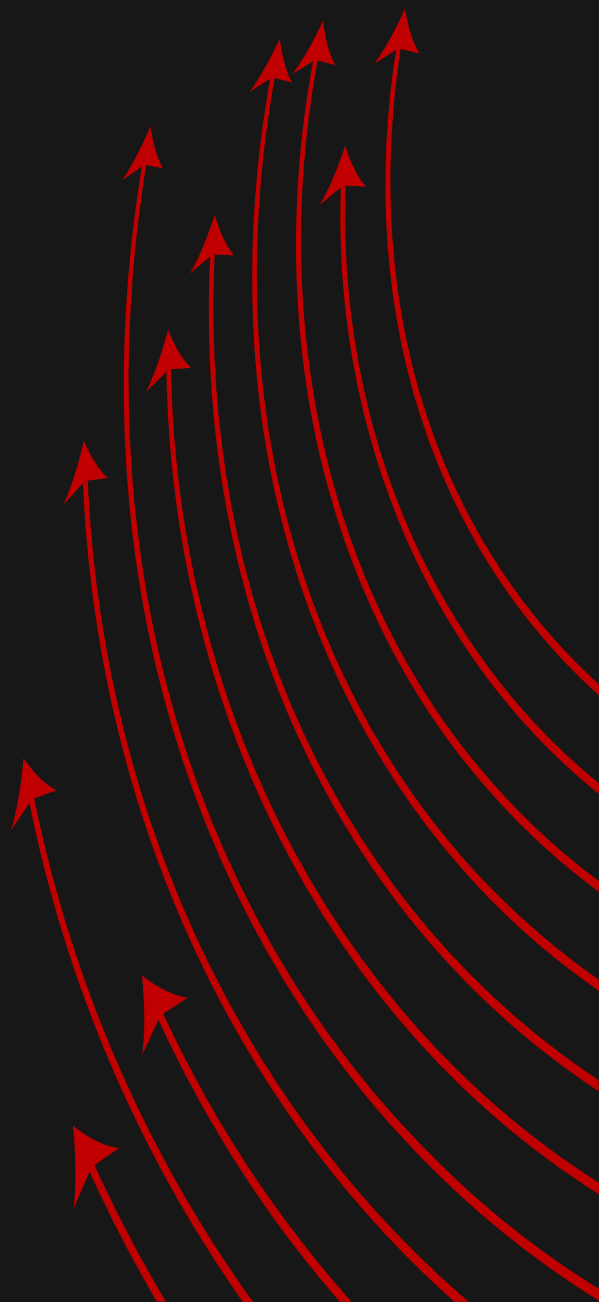
- Plays mostly correct notes
- Keeps time
- Holds guitar correctly
- Plays from memory

Expert

- Plays mostly correct notes
- Keeps time ***with the band***
- Holds guitar correctly
- Plays from memory ***for 3.5 hours***
    - In front of large crowds
    - While wearing a sparkly costume and heels
    - With choreography

Development Process

Front-end analysis	Understand current training environment, gaps, and user needs
Develop learning objectives	Identify what the trainee should be learning during the scenarios
Design VR scenarios	Design activities to train each learning objective
Develop performance measures	Map scenario activities to measurable outcomes
Establish performance standards	Determine what “right looks like” for different skill levels



Measure Development

1. Different criteria for novices and experts

Learning Objective → Scenario Measure → Novice criteria
- *Lower standard*
- *Not measured*

Expert criteria
- *Higher standard*

Example:

Trainee should order medical device → 1 = orders device
0 = fails to order device → Novice = 0-1

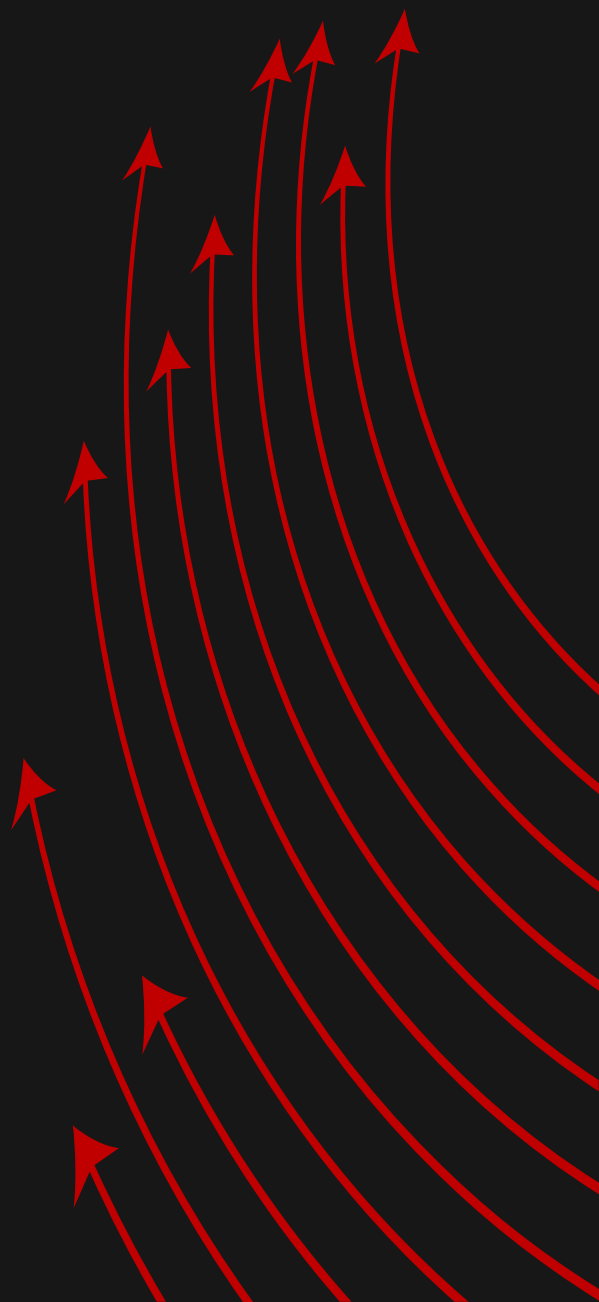
Expert = 1

2. Different measures for experts

Learning Objective → Expert Only Scenario Measure → Expert only criteria

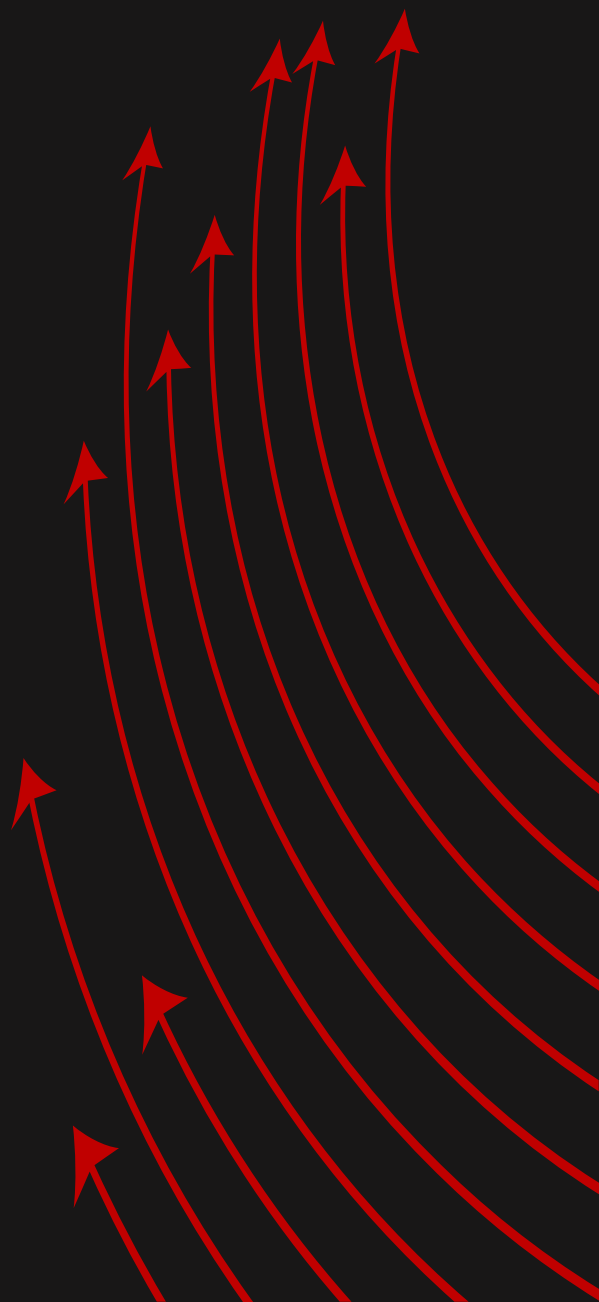
Example:

Trainee should choose best medical device → 1 = correct
0 = incorrect → Expert = 1



Measures to Consider

Accuracy	Novices make more mistakes
Speed	Novices are slower
Consistency	Novices are less consistent across trials or scenarios
Task Complexity	Novices succeed in simple, structured tasks; experts can succeed in complex tasks
Scaffoding and Support Needs	Novices rely on real-time assistance; experts may ignore it
Emotional Regulation	Experts are resilient in stressful situations; novices are overwhelmed
Adaptability	Experts can change approach to problem solving quickly; novices require time to reflect



Physiological Measurement with VR

Why are we doing this?

- Traditional performance metrics (e.g., time and accuracy) miss internal cognitive and emotional states.
- Physiological sensors (HRV, GSR, eye tracking) expose cognitive load, stress, and attention.
- Allows real-time training content adaptation, keeps trainees in the optimal learning zone.
- Critical when training under pressure matters more than rote skills.
- Can detect onset of adverse states (e.g., simulator sickness, excessive workload)

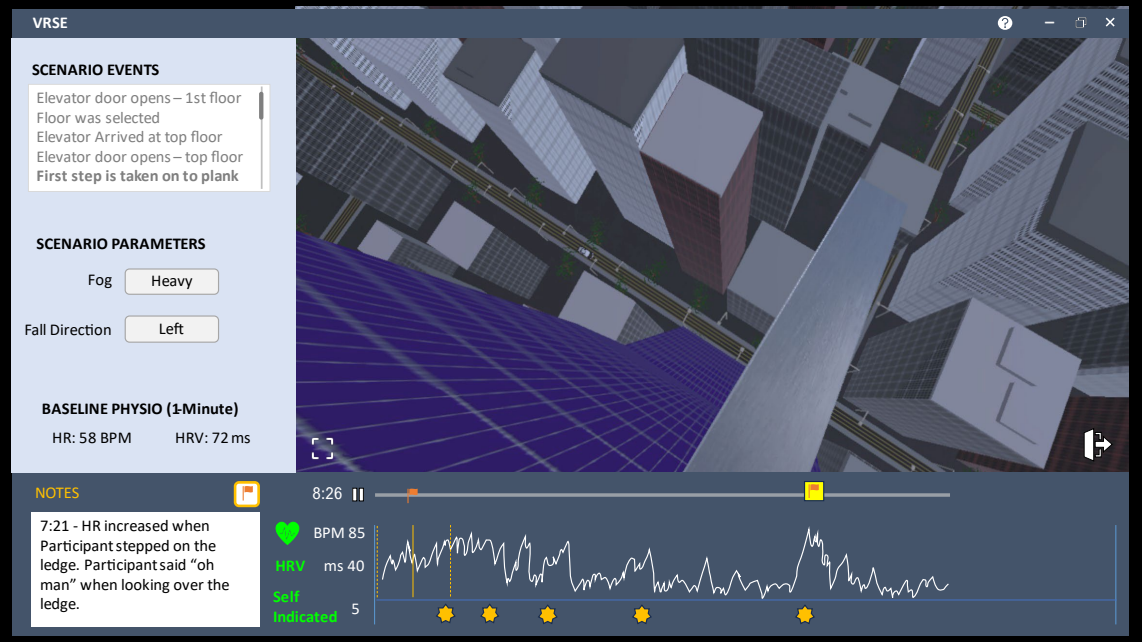
Current Sensors: Pros and Cons

Sensor	Pros	Cons
Heart Rate Monitor (HR/HRV)	• Inexpensive • Reliable	• Less effective for rapid state changes
Galvanic Skin Response	• Inexpensive • Lightweight	• Temperature and humidity can skew data
Eye Tracking	• Headset probably already has one	• Sensitive to lighting changes
Respiration	• Inexpensive • Non-intrusive	• Requires proper placement/calibration
EEG/fNIRS	• Monitors brain activity	• Expensive • Prone to artifacts and data loss • Challenging to analyze

Example: Stress Management Training

Problem: Service members are trained to deal with operational stress, but not low-level, daily annoyances.

Solution: Stress exposure in VR with real time physiological monitoring of stress responses, paired with practitioner intervention and guidance



Mockup of Practitioner Interface

Designing the Solution

- Needs analysis to understand existing hardware infrastructure, stakeholder requirements
  - Determined headset, sensors, tablet, room size, etc.
- Scenario design based on SME and stakeholder interviews
  - Introductory scenario to capture physio baseline and familiarize users with VR
  - Adaptive scenario difficulty to elicit stress (built-in and practitioner-driven)
- Practitioner interface requirements determined through stakeholder conversations

User Testing

Measures

- Demographics: experience with VR
- Thrill-seeking scale (Zuckerman et al., 1964)
- Simulator Sickness Questionnaire (Kennedy et al, 1993)
- Short Stress-State Questionnaire (Helton, 2004)
- Feedback Questionnaire (custom)
- HR/HRV in each scenario compared to baseline

Do they already know how to use the system?

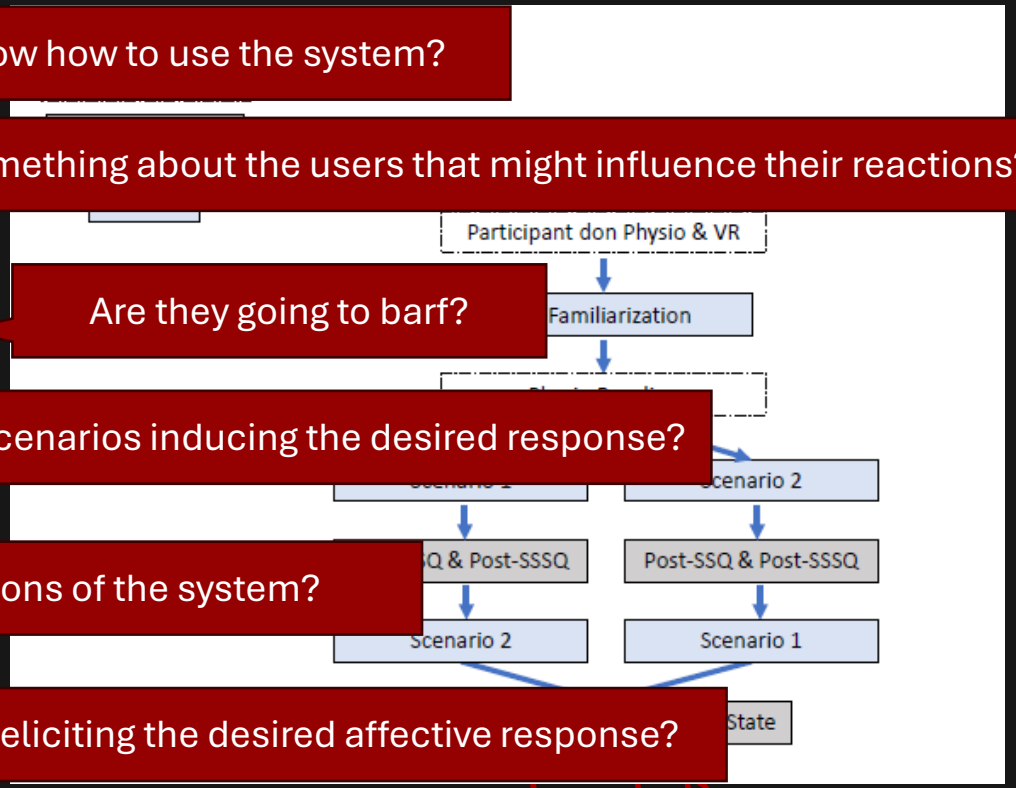
Is there something about the users that might influence their reactions?

Are they going to barf?

Are the scenarios inducing the desired response?

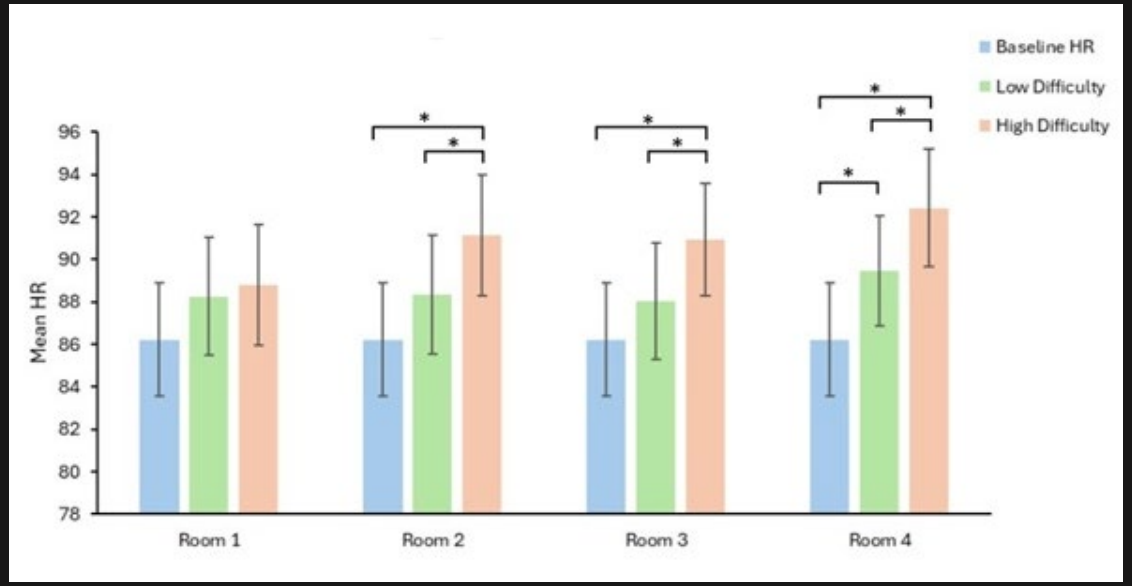
What are their opinions of the system?

Are you eliciting the desired affective response?



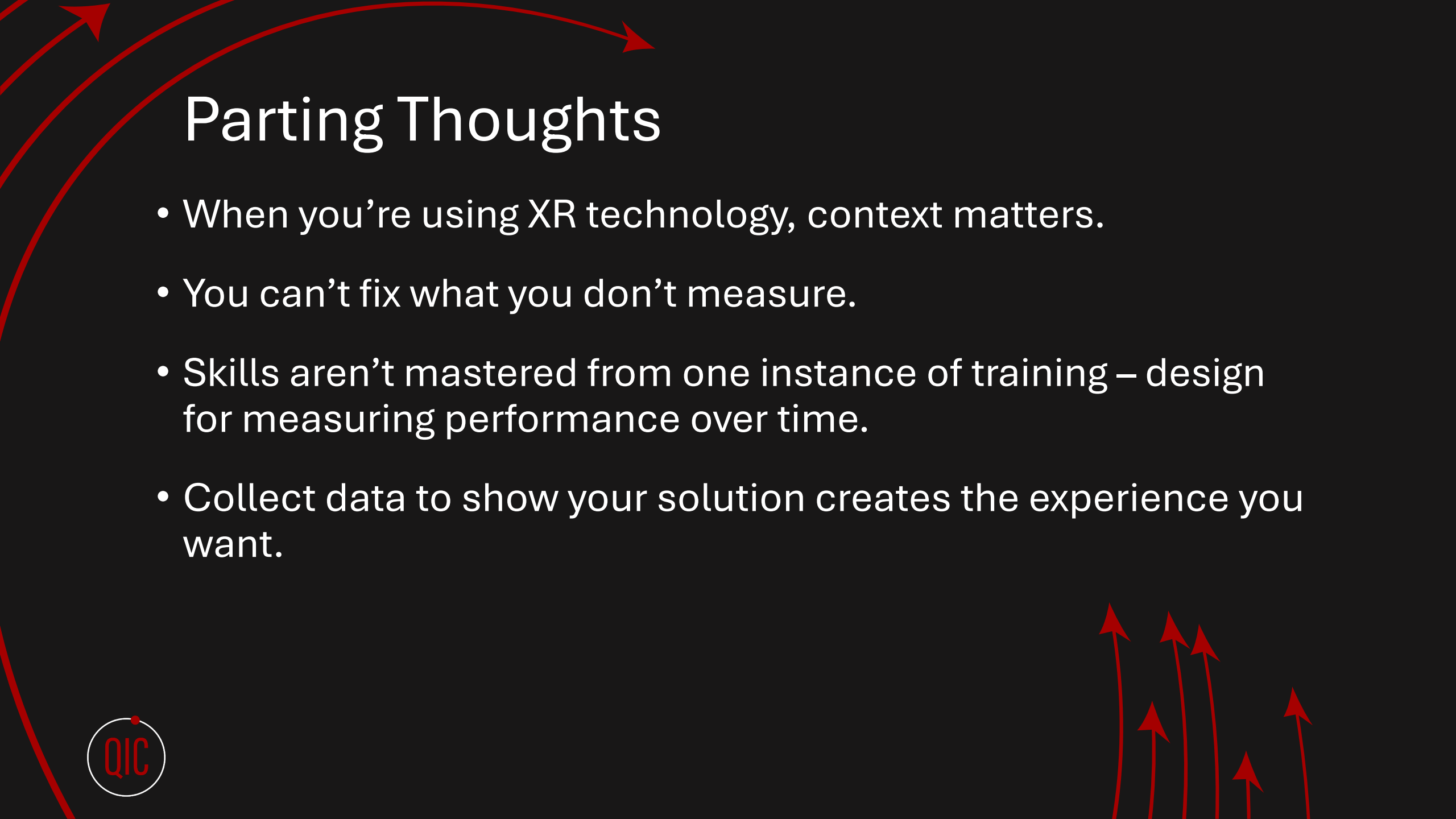
Results

Finding	So What?
Most of our participants had experience with VR (88%)	Naïve participants might have shown higher baseline levels of stress.
Our participants were not particularly “thrill-seeking” (*M*=6.24)	They weren’t predisposed to enjoy the scenarios.
SSQ ratings did not change during the scenarios	The scenarios are not inherently barf-inducing.
SSSQ scenario ratings increased from baseline (*M*= 5.35) in some scenarios (*M* = 8.82; *t*(16) = 26.5, *p* = .03).	The scenarios that were designed to be annoying and irritating were experienced as such.
HR increased during more stressful scenarios.	The scenarios induced a measurable physiological change.



Parting Thoughts

- When you're using XR technology, context matters.
- You can't fix what you don't measure.
- Skills aren't mastered from one instance of training – design for measuring performance over time.
- Collect data to show your solution creates the experience you want.



Contact

jsolberg@quantumimprovements.net

www.quantumimprovements.net

QUANTUM IMPROVEMENTS
CONSULTING