QUIVRR **- Quality of Education in Virtual Reality Rubric**

Focus on Active and Embodied Learning

Dr. Mina C. Johnson-Glenberg
Arizona State University and
Embodied Games, LLC

Beta – Preliminary Data Gathering Version 1.1.8 (Nov. 2019)

Send forms and questions to
Contact@Embodied-games.com
All feedback appreciated!

QUIVRR.
Designed to assess the quality of a learning experience in a VR Head Mounted Display (HMD).
Three sections are included - A: for Pedagogy/Content, B: for Mechanics in VR, and C: Bonus items (not used in the denominator of final score). Each line item is described in the Appendix. The score range is 0-3: generally, 0 = none, 1 = a little, 2 = medium, 3 = high/a lot. (Note: Two items are reverse coded.)
You may use .5, but it is not recommended to use intervals smaller (e.g., .25). An asterisk means an item is known to be "double barreled"*. The denominator is 60. Multiple the final score by 100.
With bonuses, it is possible, but highly unlikely, to score 106%.

***QUIVRR* v 1.1.8** **Name of Module**_____

Your name (Optional) _____**Your position** _____**date**_____

A: Pedagogy/Content	Notes- if any	Score (0-3)
1. Clear critical thinking Learning Goals (several are higher order per Bloom's Taxonomy)		
2. Content Suited to Immersive VR (is enhanced by 3D and Presence)		
3. Does VR move the content towards a redefinition of the technology and towards more transformative learning? (via the SAMR model)		
4. Scaffolding (building up to complexity)		
5. Active Learning (includes user-driven choices and body movements; agency well used)		
6. Actions Congruent to/Reinforcing of Content		
7. Guided Exploration (not all free exploration)		
8. Opportunities to Practice Learning Goals		
9. Corrective Feedback		
10. Prompts for Metacognition (Think about thinking; chance to be reflective, also include outside of headset prompts in lesson plan)		
Subtotal	**Subtotal**	
B: Mechanics		
11. Amount of Text and Eye Strain Reverse: 0 = too much text/eye strain, 1 = a lot of text, 2= good usage of text, 3 = very minimal amount of text.		
12. Induction of Cybersickness Reverse: 0 = Highly likely to lead to nausea and disorientation, 1 = more than somewhat likely, 2 = somewhat likely over extended time, 3 = Highly unlikely to lead to nausea and disorientation.		
13. Frequency and How Much of Content is Manipulable?		
14. Avatar - Aspects under user design		
15. Avatar - Movement Match		
16. Interface and Ease of Use (intuitivity)		
17. Sound and Its Quality		
18. Haptics/Other Modalities (Beyond visual and auditory stimuli)		
19. Engaging (holds attention, not repetitive)		
20. Overall Module Quality (design aspects, avoids gender/racial biases, low aliasing, etc.)		
Subtotal	**Subtotal**	
C: Bonus (Optional, not included in denominator)		
21. **Assessment** Included (and not simply true/false)		
22. **Adaptive** (scales with learner's performance) and/or **Collaborative** in real time. Both of these are expensive; they will become more common with time and no longer be bonuses.		
Subtotal of Bonuses	**Optional Subtotal**	
Sum of All Subtotals		
Final Score - Denominator is 60	**Sum subtotals, divide by 60, multiple by 100.**	

General Notes.

If there is a bias in this rubric, it favors STEM content over the humanities which rely more on POV, editing savvy, narrative, and empathy. The rubric honors the 4E's of cognition, i.e., embodied, embedded, extended and enactive [1], with a strong emphasis on enaction and gesture.

Double Barreled Responses*. The primary end-user for this rubric is a teacher trying to decide between, or to justify to others, the choice of an educational module. Teachers are busy. We acknowledge it is cleaner to avoid queries with more than one construct in a single item response, but that would result in more than 35 items, making this rubric more burdensome. Thus, users are sometimes asked to make judgments about **both quantity and the quality with a single score. The more complex, multi-faceted queries are marked with an asterisk.

The Module. The term "module" refers to the learning experience inside the headset. Educational content may reside outside the headset, but unless noted in this Appendix (as in item 10), the scores are given for *in-headset experiences*.

Whole Integers. Reviewers can add .50 anywhere in between .00 and 3.0, but adding .25 will probably not gain you much in precision or sensitivity. *QUIVRR* is meant to be a fast rubric and it is not supposed to be agonized over. It is similar to a 20-item rubric for cognitive thinking skills in videogames published by Rice [2].

Bonus - Multi-user Collaboration. Collaboration and cooperation are well-researched and well-regarded constructs in education [3]. Multiplayer in VR and is still expensive to create and maintain, so it seems too early to penalize the majority of modules for not having something that is too expensive for independent studios to create. The collaboration variable may move out of the bonus section in the next iterations. But we note that it is not a simple, purely positive-valenced construct. Per Graesser, Kuo [4], three negatives are associated with Collaborative Problem Solving including a) diffusion of responsibility for completing tasks *(which affects uni-level or single student assessments*), b) disagreements that can paralyze progress, and c) learners getting sidetracked by irrelevant discussions.

Appendix

A: <u>Pedagogy/Content</u>

1. Learning Goals. The stated learning outcome goals can be embedded in the beginning of the module. But, it is best to minimize text in a headset, so this is content that can reside in a separate document. The goals need to be accessible and attainable, and the learner should be exposed to the goals before the module begins for priming. The focus should be on the higher levels of Bloom's Revised Taxonomy. As a reminder, the lower levels are: remembering, understanding, and applying, and the higher levels are: analyzing, evaluating, and creating.

Example 0 = no stated goals, 1 = goals incomplete, only listed at end, only lower level, 2 =goals adequate, but mainly middle level, 3 = goals clear, easy to access and higher level.

2. Content Suited to VR.* This decision can be based on two constructs called the profound affordances of VR [5] or using several other criteria from other VR researchers like Dalgarno and Lee [6] or Bailenson [7]. Many agree the first important affordance is **3D.** Is the concept well-suited to the 3D effects of VR? The second affordance is **Presence** and a more detailed score on the emotionality of presence is in the next item. For the first affordance, ask if the learning would be enhanced by more dimensions? Can the topic be better explored in 3D and in 360°? E.g., Protein folding and magnetic fields are good examples. I VR is also good for fantastical and making the unseen be seen.

Presence is 'the sense of being there'. Ask whether high presence is induced and if it facilitates the learning. Presence can heighten emotions, but overstimulation will lead to a reduction in learning. A well-crafted narrative can also increase presence and 'buy in'. Attendant with deep presence is a feeling of empathy. For example, seeing your virtual body in a mirror with a different skin color, or gender, has strong effects on implicit biases [8]. Is the Point of View (POV) that immersive 3D VR allows creatively utilized, and is it conducive to cognitive change?

Example 0 = content could be learned just as well in 2D, 1 = content is suited to VR, but design does not take advantage of this (i.e., most artwork is in 2D), 2) = content is well-suited, and the module makes adequate use of 3D and Presence, 3 = content is well-suited and takes advantage of VR's many affordances including 3D and Presence.

3. Does VR move the content towards a redefinition of the technology/towards more transformative learning? (via the SAMR model)**. SAMR** ranks the levels that a technology might serve**.** The lowest is as a Substitute, higher is when the technology Augments the learning, higher still, does it positively Modify what has been learned, or at the highest does it Transform, (redefine or create brand new tasks). If the lesson is a word list that could be presented *Powerpoint*, then it should receive a low

score because VR has not augmented the learning task. **SAMR -** http://www.hippasus.com/rrpweblog/archives/2014/06/29/LearningTechnologySAMRModel.pdf

Example 0 = no real technological modification, merely serves as a substitute, 1 = low of augmentation and/or level of modification,2 = moderate level of modification, 3 = high level of modification and/or redefinition.

4. Scaffolding. This means that the module systematically builds up to the more complex concepts over time. The designers have added more complex components at an appropriate learning pace. As an example, in a virtual biology lab, learners first view a simulation passively, afterwards they are able to begin manipulating and exploring the dynamic system [9]. Another route involves in-process (real time) assessing, as learners show mastery of one concept, a new concept is added. The module should display some evidence of leveling up in difficulty with time or mastery. This is part of the second "E" in Kolb's [10] *Triple E Framework* for learning with technology (i.e., Engage, Enhance and Extend).

Example 0 =none, 1 = some leveling up of difficulty, but too many variables, 2= attempts to level and scaffold, but poorly paced/designed at times, 3 = appropriate, well-paced, and helpful scaffolding.

5. Quality of Active Learning*. This means learners are actively making choices in the module and, hopefully, moving their bodies to engage with content (beyond moving eyes). The example below favors the use of the hand controls in creative and kinesthetic ways. You may adhere to a more traditional 'generative' definition of the term active, and perhaps compare lecturing versus more constructive learning like small group discussions. A meta-analysis by Freeman, Eddy [11] shows that students in STEM classes with traditional lecturing were 1.5 times more likely to fail than students in classes with active learning. This item focuses on the QUALITY:

Example: 0= primarily passive presentations of content or written word, 1 = very small amount of activity by learner (clicking forward), 2 = some active learning, 3 = multiple instances of high quality active learning.

6. Actions Congruent to the Content*. This means the learners' actions map well to the learning of the topic. E.g., stroking a virtual brush down your hair at a faster rate will pick up more electrons. This is one reason why learning goals are important, if the goal is to learn about friction and electron induction from hair, then virtual brush stroking would be a highly congruent action or gesture (as opposed to pushing a button labeled "move brush"). Not every movement must be isomorphic to real world movements. If the learning goal is to construct an entire car engine in a short time and picking a proper sized screw is what is important, then it may not be critical for a learner to twist a wrist multiple times; it may simply be enough to collide the screw to the hole. (This example is an expansion of lesson 3, number 4, by Schell [12].)

Example: 0 = no instances, 1 = few instances of congruency and of low quality, 2 = several instances and of fair quality, 3 = multiple instances of congruency that are creative and further the learning goals.

7. Guided Exploration. This is different from free exploration where learners do as they please. In educational games/sims it is appropriate to allow from some discovery and exploration early in the game, but it is important that at some point learners receive cognitive and navigational guidance. Otherwise learners, in STEM especially, tend to make spurious hypotheses [13]. Several methods to guide in multimedia are visual (lighting, arrows, text, etc.) or with audio.

Example: 0 =none, 1=mainly free exploration, 2= some guidance, 3= well-paced and well-designed guidance.

8. Quality/Quantity of Opportunities to Practice Learning Goals*. The learning goal(s) could be instantiated by either an action or application of a learned rule. An example of an action that shows learning is when learners begin to capture only certain patterned butterflies to demonstrate a comprehension of Batesian mimicry (from the VR game "Catch a Mimic", 2019). Another example might be when learners grasp a virtual moon and turn it to examine its dark side (see *Oculus Rift* version of "Titans of Space", 2019). This item addresses two constructs: a) **Quality** – were the opportunities for practice creative (varying and engaging), versus rote (like drills, e.g., summing a small set of numbers over again - similar to *Kumon* math training), and b) **Quantity** – the number of times the learning goals were reinforced (*some noted overlap with item 13*). This is related to the learner interaction section in Dalgarno and Lee [6] which includes "Construction/scripting of objects and behavior". Construction usually means moving/manipulating content and this construct assesses whether appropriate and meaningful opportunities for construction have been made available. It is understood that creativity is, by its nature, a subjective judgment.

Example. 0 = no practice, 1 = low practice/non-creative, 2 = mid-level practice/somewhat creative, 3 = high and appropriate level of creative practice.

9. Useful Corrective Feedback. Feedback is not simplistic in formative and mediated education [14]. When feedback is constant it can be distracting, when it lags too long after the incident, it is not useful. When formative evaluative feedback is included it should not be a simplistic pop-up of "great try" after a failure. When feedback is evaluative, it should give meaningful hints as

well. A hint box callout could appear with more in depth information if an incorrect answer has been submitted three times in a row.

Example. 0 = no feedback; 1 = minimum feedback, or at end of module only, 2 = mid-level; 3 = Useful and actionable feedback with proper pacing.

10. Prompts for Metacognition. Does the content encourage learners to "think about their thinking"? Is there a space for learners to be reflective? For this item you may take into account "outside of the headset" situations. E.g., if the written lesson plan has during-during or post-module prompts, or does it include dyads with screen-casting where another student can ask reflective questions? Some examples of evidence-based metacognition prompts are: Asking learners if they want to pause to think through anything, asking them to pretend they are teaching to another student - what would they say, asking them to make a prediction about what might happen next, or, create a question you might ask on a test[1].

Example. 0 = no prompts; 1 = one or two prompts, 2 = three or more prompts, of varying quality; 3 = three or more consistently effective prompts.

Mechanics

11. Amount of Text and Eye Strain. As of 2019, all commercial VR headsets force the muscles attached to the lenses in the eyes to strain after a while (also called vergence-accomodation strain). Certain tasks induce this strain faster, e.g., tracking disproportionately sized objects in different planes. A common educational task conducive to eye strain is text reading. Designers should **minimize the amount of text**. Black background, white text preferred. Decoding is not an interesting affordance of 3D. As you quantify "text in the module", consider the entire length of the module, then consider how much of the visual display is dedicated to reading text or tracking minute objects. Note that having a voice-over does not mean players will *not* also read. (Do **not** tally the title or 'optional viewable content', e.g., the credits or answers to hints – multimedia learners rarely read those). Just remember "less is better".

Example Reverse: 0 = too much text/eye strain, 1 = a lot of text, 2= good usage of text, 3 = very minimal amount of text.

12. Induction of Cybersickness. Cybersickness is polysymptomatic and polygenetic (affects each individual differently). The Simulator Sickness Questionnaire (SSQ) from Kennedy, Lane [15] is often cited and it includes three categories of 1) nausea (e.g., stomach awareness, nausea, etc.), 2) oculomotor issues (e.g., headache, eyestrain[2],) and 3) disorientation (e.g., vertigo, dizziness, etc.). With experience, many of these effects attenuate and players become habituated. Approximately 50% of discomfort is gone by the 10th session [16]. Students are encouraged to leave a module as soon as they feel 'weird'.

It is commonly agreed that the negative effects are caused by a visual/vestibular disconnect. Rebenitsch and Owen list several VR design fixes including "… partially limiting the degrees of freedom in control when navigating. Inclusion of the real world[3] and increased tactile feedback appear to decrease symptoms." (p.122). The ideal speed of navigation is still not known, but cybersickness tends to increase with increasing speed [17]. There are known tricks to decrease cybersickness like using ramps rather than stairs, and making learners more agentic, that is, in control of where they go and what they interact with. The scoring below is another instance where less is better, the anchors range from *very unlikely* to lead to nausea and disorientation, to *highly likely* to lead to nausea and disorientation.

Example Reverse. 0 = Highly likely to lead to nausea and disorientation (e.g., includes instances of accelerating navigation, excess yaw axis movement, little agentic control – a "roller coaster" is an example, 1 = more than somewhat likely, 2 = somewhat likely over extended time, 3 = Very unlikely to lead to nausea and disorientation (e.g., POV does not change often, user has large degree of agentic control).

13 How often, and how much of the content is manipulable?*
(*Related to active learning and congruency in Pedagogy/Content*.) Throughout the entire module, how often is the learner encouraged to interact with or manipulate the content? This item attempts to quantify the instances of 'interactive objects', and the frequency of interactivity under user control. The term manipulable means **actionable and movable content** in the virtual world, it does not include multiple instances where the user merely pushes a button to start a more complex sequence (see

[1] Several good prompts come from the Reciprocal Teaching (see Palincsar and Brown). The VR research community is still working out what it means to break immersion and if that is worth the putative positive effects on learning. More to come on this item in the future as the education research is published.
[2] Note. This rubric pulls eyestrain out from the SSQ and treats it as an individual line item because so much text is often included in educational modules. We leave headache in the definition of Cybersickness; eyestrain is only one pathway to a headache.
[3] We do NOT recommend adding real world in the periphery – because – well, that would turn the lesson into "AR" module and AR will soon have its own rubric. This rubric is for VR modules.

"brush hair" example in item 5 for congruency). Navigation across a room should only be counted as one instance and not counted over and over.

The "Catch a Mimic" game is an unusual example. There is no navigation via locomotion, and no forced change in POV, but there is a large amount of actionable content on the screen at once.[4] In contrast, in the *Oculus Go* Version of "Titans of Space", the user can navigate to visit a specific planet, but once at the planet the only interaction is clicking on a text panel. We agree with Schell's (2015) advice in lesson 3, point 1, "You are wiser to create a small game with rich object interaction than a big game with weak ones."

Example: 0 = no manipulation or very low level of manipulable content, 1 = some manipulable content, but few chances to interact, 2 = more instances and more chances, 3 = a high amount of manipulable content and high frequency for interaction.

14. **Avatar: User choice.** Can the learner choose an avatar? How many components of the avatar should be customized? Research in 2D supports that some customization is valuable [18]. Recent work in VR shows that after participants inhabited Einstein's virtual body, they showed a decrease in implicit biases towards older people. Body type carries meaning…"this meaning has implications for the perceptual processing, attitudes and behaviors of the person experiencing" the avatar [19]. When users can customize their avatars, is there an inflection point where too much choice leads to wasted time? These questions await research. Hopefully, the gross categories below will suffice for now. If the module only allows for humanoid avatars, this item should not go above a 2.

Example: 0 = no avatar, 1 = yes, hand or body present, but no avatar choice - preassigned, 2 = yes, a body and 1 to 3 components (e.g., clothes, hair, etc.), only humanoid options, 3 = yes, body and more than four components customizable.

15. Avatar: Extent of Body and Movement Match.* Gonzalez-Franco and Peck [20] gathered questions from 30 relevant avatar studies and culled them down to the six main categories pertaining to type of embodiment and ownership of an avatar. Several of those categories are already dispersed throughout this rubric, but their items 2, 4, and 5 - issues of co-location and external appearance of avatar - are addressed in this # 15 QUIVRR item. This item depends on two dimensions: A) **how much of the avatar body** is displayed, and B) **how aligned are the movements**. Both are important. Both constructs are generally correlated in bi-manual headsets, such that, if a body (from neck down) is visible, then the upper part will generally move veridically with users' hand controls.

Example: 0=none, 1=hand only or poor movement match, 2=more of body shown, adequate movement match, 3=full body shown and high quality match.

16. Intuitive Interface.* Modules and interfaces should be designed with first-time users in mind. Intuitive means the " the users' unconscious application of pre-existing knowledge leads to effective interactions" [21]. There should not be a strong dependence on complicated button sequences for navigation[5], nor to answer questions. Actionable items should not be overly spread out throughout the interface, i.e., the learner should never be forced to spin rapidly around in a lesson (see Cybersickness note). Entire books are written on interface design, you will have to use your instincts. A percentage of the population is colorblind, so critical elements should not rely on red/green distinctions. Less clutter is better. *Remember* not all youth are gamers.

Example: 0 = highly cluttered and unnavigable, 1 = somewhat cluttered and not easy to navigate, 2 = clean interface, but not easy to navigate, 3 clean interface and easy to navigate.

17. Sound Quality.* Sound in VR elevates all experiences. Nonetheless, a unidirectional tune can play in a loop and become distracting. Even very nuanced, omnidirectional sounds, when overused, can become overkill. Quality also depends on whether the sound furthers the educational goals. Creative sounds used in feedback count. If the sound is of an extremely low quality, or is distracting and irritating, it should be given a reduced rating.

Example: 0 = silent, 1 = low quality and poor mapping, 2 = acceptable quality and acceptable mapping, 3 = high quality, creative and synced omnidirectionally.

18. Haptics/Other Modalities.* To date, the common modalities in VR are visual and auditory. Several higher-end hand controls are able to vibrate. Even though haptics are not yet common in most VR modules (e.g., *Google Cardboard*), we predict that vibrotactile feedback will become more common. Tactile and haptic add-ons are evolving rapidly, they should be included *where*

[4] In each of the six levels there are 60 butterflies, only half of those should be caught with the virtual net. In the *Oculus Go* version of "Catch a Mimic" the upper body and controller arm are very active throughout the module.
[5] If the learner cannot figure out how to move forward in two tries, the interface is at fault.

they further learning (they should not be distractors). Olfaction is being used, and several vestibular-altering rigging systems have been researched [22], though more research is needed in all these fields.

Example: 0 = no stimuli beyond audio and visual, 1 = one extra modality, adequately integrated, 2 = one extra modality, integrated very well, 3 = two or more modalities (beyond audio and visual) are included and well integrated.

19. Engaging*. The next two items are highly subjective and touch on *both* pedagogy and mechanics. It is possible to hit every line item above and create an experience that is not engaging or fun. The author was once part of a beautiful, mixed reality lesson that would have scored high on most constructs, but once in the field it was determined that few students could figure out how to walk with a 'negative acceleration'. Repeated gameplay led to expressions of frustration.[6] Every reviewer will come to this item with a bias as to what it is engaging. You may have a bias for a satisfying narrative. But, for STEM at least, the evidence is not conclusive that a deep narrative story line consistently significantly affects STEM learning. Several Randomized Control Trial studies in labs have not supported that narrative leads to significant learning gains [23, 24].

Trust your instincts. Engagement must be linked with educational payoff; the learning goals must be supported. You may also change any of these scores after you observe users engage with and discuss a module.

Example: 0 = not at all engaging *(more broccoli than chocolate*), 1 = somewhat engaging, 2 = engaging, 3 = very engaging, and evidences high educational worth.

20 Overall Quality* – This touches on *both* pedagogy and mechanics. Some reviewers dislike all low poly artwork, others expect perfect shading. Resolution versus framerate is a constant trade off in VR; ask if this has been addressed with creativity? There are certain touchstones that should result in points deleted, e.g., are there obvious stitch lines, is the content highly repetitive, is there Z fighting, is there aliasing on thin lines, are icons uninterpretable, is the module sensitive to gender and racial biases, was bad judgment shown for physiological comfort, e.g., letters that zoom towards you, or the POV includes flying around without agency, aka, a "magic carpet" effect, etc.

Example: 0 = very low quality, 1 = some quality attempted, 2 = good quality, 3 = high overall quality.

C: Bonuses: The following two items could be worth a total of six points; however, those points are *not* included in the denominator because these constructs are not yet commonplace in educational VR. Over time, assessment, adaptivity and collaboration are expected to become the norm and then they should be moved out of the bonus category.

21. Assessment Included*. Assessment comes in a multitude of forms. When it happens during the experience it is called formative, when at end - either in the VR headset or outside - it is called summative. It is possible to embed assessment during learning so that players do not even know they are being assessed. The literature on quality for assessments is extensive (for a crisp summary of quality see Darling-Hammond et al. [25], for evidence-centered design see Mislevy, Behrens [26]). The test could be old-fashioned paper and pencil, but it should always be "valid, reliable, and fair".

Example: 0 =no assessment, 1 = some assessment, 2 = assessment both within module (formative or in-process) and at the end, 3 = Assessment is high quality and occurs both within module (formative) and at end for summative reflection.

22. Adaptive/Collaborative* – Adaptive means that the experience in the module changes according to the learner's performance. The experience is dynamic and based on the choices made during encoding. Linear pathways cost less to create compared to dynamic pathways. Adaptivity is expensive, so at this stage in VR it is an aspirational construct. Note: having the choice over a series of *linear pathways* is only worth one point. Collaborative (more than one human player in the VR space) is also expensive, but it is powerful and when it becomes more commonplace, it will be moved out of the bonus category.

Example: 0 = no adaptivity, 1 = user control only over a series of predefined, linear pathways, 2 = user control and up to two instances of adaptivity, one other player in spaces, 3 = user control and three or more instances of high quality adaptivity and more than one player in space.

References

1. Rowlands, M., *The new science of the mind: From extended mind to embodied phenomenology*. 2010, Cambridge, MA: MIT Press.
2. Rice, J.W., *Assessing Higher Order Thinking in Video Games.* Journal of Technology and Teacher Education, 2007. **15**(1): p. 87-100.

[6] Probably inadequate scaffolding could be blamed, but we did not know that until we were in the field and could observe that the high school students did not possess the prior knowledge to be successful.

3. Johnson, D.W. and R.T. Johnson, *Cooperation and Competition: Theory and Research*. 1989, Edina, MN: Interaction Book Company.
4. Graesser, A., B. Kuo, and C. Liao, *Complex problem solving in assessments of collaborative problem solving.* Journal of Intelligence, 2017. **5**(10): p. 1-14.
5. Johnson-Glenberg, M.C., *Immersive VR and education: Embodied design principles that include gesture and hand controls.* Frontiers in Robotics and AI, 2018. **5**(81).
6. Dalgarno, B. and M.J.W. Lee, *What are the learning affordances of 3-D virtual environments?* British Journal of Educational Technology, 2010. **41**(1): p. 10-32.
7. Bailenson, J.N., *Experience on Demand: What Virtual Reality Is, How It Works, and What It Can Do*. 2018, New York, New York: W.W. Norton & Co.
8. Banakou, D., P.D. Hanumanthu, and M. Slater, *Virtual Embodiment of White People in a Black Virtual Body Leads to a Sustained Reduction in Their Implicit Racial Bias.* Frontiers in Human Neuroscience, 2016. **10**(601).
9. Hossain, Z., et al., *Authentic Science Inquiry Learning at Scale Enabled by an Interactive Biology Cloud Experimentation Lab*, in *Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale*. 2017, ACM: Cambridge, Massachusetts, USA. p. 237-240.
10. Kolb, L., *Learning First, Technology Second: The Educator's Guide to Designing Authentic Lessons*. 2017: ISTE.
11. Freeman, S., et al., *Active learning increases student performance in science, engineering, and mathematics.* Proceedings of the National Academy of Sciences, 2014. **111**(23): p. 8410-8415.
12. Schell, J., *Making Great VR: Six Lessons Learned From "I Expect You To Die"*, in *Gamasutra*, K. Graft, Editor. 2015, Gamasura: The Art and Business of Making Games.
13. Kirschner, P.A., J. Sweller, and R.E. Clark, *Why Minimal Guidance During Instruction Does Not Work: An Analysis of the Failure of Constructivist, Discovery, Problem-Based, Experiential, and Inquiry-Based Teaching.* Educational Psychologist, 2006. **41**(2): p. 75-86.
14. Shute, V., *Focus on formative feedback.* Review of Educational Research, 2008. **78**(1): p. 153-189.
15. Kennedy, R.S., et al., *Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness.* Int J Aviat Psychol, 1993. **3**(3): p. 203–220.
16. Rebenitsch, L. and C. Owen, *Review on cybersickness in applications and visual displays.* Virtual Reality, 2016. **20**(2): p. 101-125.
17. So, R.H.Y., A. Ho, and W.T. Lo, *A Metric to Quantify Virtual Scene Movement for the Study of Cybersickness: Definition, Implementation, and Verification.* Presence: Teleoperators and Virtual Environments, 2001. **10**(2): p. 193-215.
18. Lin, L., et al., *How character customization affects learning in computational thinking*, in *Proceedings of the ACM Symposium on Applied Perception*. 2017, ACM: Cottbus, Germany. p. 1-8.
19. Banakou, D., S. Kishore, and M. Slater, *Virtually Being Einstein Results in an Improvement in Cognitive Task Performance and a Decrease in Age Bias.* Frontiers in Psychology, 2018. **9**(917).
20. Gonzalez-Franco, M. and T.C. Peck, *Avatar Embodiment. Towards a Standardized Questionnaire.* Frontiers in Robotics and AI, 2018. **5**(74).
21. Israel, J., et al., *On intuitive use, physicality and tangible user interfaces.* International Journal of Arts and Technology, 2009. **2**(4): p. 348-366.
22. Tamaddon, K. and D. Stiefs, *Embodied experiment of levitation in microgravity in a simulated virtual reality environment for science learning.* IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual Augmented Reality (KELVAR) 2017: p. 1-5.
23. Adams, D.M., et al., *Narrative games for learning: Testing the discovery and narrative hypotheses.* Journal of Educational Psychology, 2011. **104**(1): p. 235-249
24. Johnson-Glenberg, M.C. and M.C. Megowan-Romanowicz, *Embodied science and mixed reality: How gesture and motion capture affect physics education.* Cognitive Research: Principles and Implications, 2017. **2**(24).
25. Darling-Hammond, L. and etal., *Criteria for high-quality assessment*. 2013, Stanford Center for Opportunity Policy in Education, Stanford University: Stanford, CA.
26. Mislevy, R.J., et al., *Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining.* JEDM-Journal of Educational Data Mining, 2012. **4**(1): p. 11-48.

Referenced VR Games- Available in *Oculus* Store for both Go and Rift platforms as of 7/2019
https://www.oculus.com/experiences/go
"Catch a Mimic" v1.0.7
"Titans of Space" v2.5.5