Quality of Education in Virtual Reality Rubric **(QUIVRR)**

Dr. Mina C. Johnson-Glenberg
Department of Psychology
Arizona State University
and
Embodied Games, LLC
www.embodied-games.com

To appear in, S. L. Macrine and J. Fugate (Eds.) *Movement Matters: How Embodied Cognition Informs Teaching and Learning*. Cambridge, MA: The MIT Press

This is shortened version for the Web.

The entire chapter gives the pedagogical reasoning behind each item. After the rubric is a more in depth description of the item and some examples for the numbering system.

As of 2020, this is a work-in-progress and the author appreciates any feedback at:

Mina.Johnson@asu.edu

The Catch a Mimic game can be found at https://www.embodied-games.com/games/natural-selection-catch-a-mimic/

Notes before using QUIVRR:

Double Barreled Responses. The primary 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 30 items, making this rubric more burdensome. Thus, users are sometimes asked to make judgments about *both* quantity and the quality with a single score. (Survey design hawks just need to take a deep breath.)

Whole Integers. Users can add .50 if they choose. We do not want you to overthink precision though. This is modeled on a short and easy 20-item rubric for "Cognitive thinking skills in videogames" published by Rice (2007).

QUIVRR: Quality of Immersive VR in education Rubric

Module Title: Hardware

M.C. Johnson-Glenberg, In general: 0 = not present, 1 = low, 3 = moderate, 5 = high

A: Pedagogy/Content	Notes	Score (0-5)
1.Clear **learning goals** stated: Does module align well with learning objectives and critical thinking standards?		
2. **Suited to immersive VR**: Is content enhanced by 3D and/or increase in presence?		
3. Does module supporting higher level **"transformative learning**"? See notes on the SAMR model and using technology optimally		
4. Is **scaffolding** present? Does module build up in complexity		
5. **Active learning**: **Quality Only** Could include user-driven choices and body movements; agency is included, learners kinesthetically practice learning goals		
6. **Actions congruent to, or reinforcing of the content**? Is there an authentic match between actions and agency and learning		
7. **Guided exploration:** Is there a beginning tutorial? Module should not be totally free exploration, some guidance included		
8. Prompts for **metacognition:** Think about thinking; are there chances for reflection built in (could also include outside of HMD prompts, working with a partner? Etc.)		
9. **Corrective feedback**: Given appropriately during activity		
10 Assessment: Either in headset or afterwards – more sophisticated than simple t/f		
B: Mechanics		
11. **Designed for comfort**: Amount of Text and Eye Strain. 0 = too much text/small objects to track, 3 = moderate amount of text, efforts made to keep comfortable, 5 = very minimal amount of text		
12. Induction of **cybersickness**: 0 = Highly **likely** to lead to nausea & disorientation, e.g., rollercoaster; 1 = more than somewhat likely, 3 = perhaps over extended time, 5 = Highly **unlikely** to lead to nausea & disorientation		
13. Content is **interactive and manipulable: Quantity** Frequency and type of manipulation (relates to item 5)		
14. **Avatar creation**: Has multiple aspects under user design		
15. **Avatar in play**: Movement match to user gestures, control and ease over gestures. (If module is all 3rd person POV, but hand controls well-mapped=2)		
16. **Overall interface** and ease of use: Interface intuitive and easy to navigate		
17. **Sound** and its quality: e.g., ambisonic, not distracting		
18. **Haptics/Other modalities**: Beyond visual and auditory stimuli, e.g., vibrotactics well-integrated		
19. **Engagement**: Holds attention, not repetitive		
20. **Overall module quality/Other**: What you care about that is not listed - design aspects, does it avoid gender/racial biases, creative use of low poly, etc. Write in notes.		
Subtotal	**Subtotal**	
C: Bonuses *(Optional - These should be the norm in several years.)*		
21. **Adaptive**: Scales with learner's performance		
22. **Collaborative**: Multiple users in same synchronous space		
Final Score	**TOTAL**	

QUIVRR Items Descriptions

Scores can range from 0 to 5. QUIVRR is weighted towards the embodied and active experience (i.e., more points are awarded when the content is designed to use hand controls and includes multiple instances of well-designed body-based metaphors for learning). It may be more relevant for STEM content than the Humanities.

QUIVRR is broken into three sections: Pedagogy/Content, Mechanics, and bonus items. The first 50 points relate to pedagogy and content, whereas the last 50 points relate to mechanics. At the end, there are two additional "bonuses" , it is unlikely a module will ever score 110.

1. Learning Goals. The stated learning outcome goals can be embedded in the beginning of the module. The term module here means what is experienced "inside the head mounted display (HMD)". Since it is wise to minimize text in a headset, however, the learning goals could also 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 module should align well with the learning objectives and critical thinking standards. Ideally, the goals would include higher levels of Bloom's Revised Taxonomy. Those lower levels are: remembering, understanding, and applying, and the higher levels in Bloom's are: analyzing, evaluating, and creating.

Example 0 = no stated goals, 1 = goals incomplete, only listed at end, only lower level, 3 =goals adequate, but mainly middle level, 5 = goals clear, easy to access and higher level.

2. Content Suited to Immersive VR. These points can be allotted based on two profound affordances of VR (highlighted above; see J(Johnson-Glenberg, 2018), or based on several other criteria from other VR researchers (see e.g., Dalgarno and Lee (2010)or Bailenson, 2018). These are: 1) **Presence** - does the module do a good job of eliciting presence and transporting the learner in a way that reaffirms the learning?, and 2) **Use of the multiple dimensions**. Some good examples of manipulation in 3D are protein folding and understanding the flux in magnetic fields around the Earth over time, etc.

Example: 0 = content could be learned just as well in 2D, 1 = content is suited to VR, but design does not take advantage (i.e., most artwork is 2D, fairly static), 3) = content is well-suited, and the module makes adequate use of 3D and presence, 5 = content is well-suited and takes full advantage of VR's many affordances including presence and manipulation in 3D.

3. Is the technology moving the content towards transformative learning? There is a marriage of product (the new technology) and process (the new way of learning). We use the Substitution, Augmented, Modified or Redefined (SAMR) model (Hamilton, Rosenberg, & Akcaoglu, 2016) for technology to clarify levels. At the **S**ubstitution level, the technology makes no difference from the old 2D method of learning. At the **A**ugmented level the technology is "adding something new but not transformative", an example would be instead of a teacher reading a story out loud now students use a tablet to both listen and read the text. At the higher **M**odification level, technology integration involves a meaningful modification in learning or assessment. An example might be instead of having a student pick a multiple choice answer regarding the concept of acceleration, a student could move a finger across a large tablet to show comprehension of acceleration in an embodied manner that the technology facilitates (see Johnson-Glenberg and Megowan-Romanowicz (2017). At the highest level, **R**edefinition, learning is transformed into a new novel task, into something previously inconceivable. An example might be a co-located mixed reality lesson where teams of students use hand-held trackers to manipulate the frequency of light waves and alter the projected, digitized colors on the floor (for examples of transformative mixed reality lessons see SMALLab Learning[1]).

[1] For examples of SMALLab Learning see https://www.smallablearning.com/videos, for an original article on SAMR see http://www.hippasus.com/rrpweblog/archives/2014/06/29/LearningTechnologySAMRModel.pdf

Example: 0 = no real technological modification, merely serves as a substitute, 1 = low level of augmentation and/or low level of modification, 3 = moderate level of modification, 5 = high level of redefinition for learning, this experience could only happen in immersive 3D VR.

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 might first view a complicated simulation, in which afterwards they are able to begin manipulating and exploring the dynamic system (Hossain, Bumbacher, Blikstein, & Riedel-Kruse, 2017). 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, which is related to Kolb's (2017) *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 at once, 3= attempts to level and scaffold, but not well paced/designed at times, 5 = appropriate, well-paced, and helpful scaffolding.

5. Quality of Active Learning. Active learning means learners are actively making choices in the module and kinesthetically moving their bodies to engage with content (beyond moving eyes). It implies they have agency. For example, in a mixed reality astronomy simulation, learners ran to show the path of a meteor (Lindgren et al., 2016), that is high quality, but limited in occurrence (see: Mechanics question). You may adhere to a more traditional 'generative' definition of the term active and constructive learning. A meta-analysis by Freeman et al. (2014) showed that students in STEM classes with traditional lecturing were 1.5 times more likely to fail than active learning students.

Example: 0= primarily passive presentations of content, 1 = very small amount of activity by learner (clicking forward), 3 = some active learning, 5 = high quality, creative active learning.

6. Actions Congruent to the Content. This means the learners' actions map well to the learning of the topic. Not every movement must be isomorphic to real world movements, but there should be overlap. In the *Catch a Mimic* game, the location and velocity of the VR hand controller maps to the location and velocity of the virtual butterfly net swoop. In another example: If the learning goal is to construct an entire car engine in a short time and picking a properly sized screw is important, then it may not be so important for the learner to spend many minutes twisting the hand controller multiple times to simulate screw-turning. (see lesson 3, number 4, by Schell (2015).)

Example: 0 = no instances, 1 = few instances of congruency and of low quality, 2 = several instances and of fair quality, 3 = a fair amount of congruency and the actions further a learning goal, 5 = multiple instances of congruency that are creative and further several learning goals.

7. Guided Exploration. In educational games/sims it is appropriate to allow from some discovery and free exploration early in the game, but it is important that *eventually* learners receive cognitive and perceptual navigational guidance. Otherwise learners, in STEM especially, tend to make spurious hypotheses (Kirschner, Sweller, & Clark, 2006). Several methods to guide in multimedia are visual (lighting, arrows, text, etc.). This can also be done via audio or haptic cues.

Example: 0 =none, 1=mainly free exploration, 2= some guidance, 3= more guidance but not properly paced, 5 = well-paced and well-designed guidance.

8. 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. Does the written lesson plan include in HMD prompts, as well as post-module prompts? Some examples of evidence-based metacognition prompts are: Asking learners if they want to pause to think through anything, and asking them to pretend they are teaching to another student. This latter question might include what would they say, or asking them

to make a prediction (Palincsar & Brown, 1984) about what might happen next, or create a question they might ask on a test.

Example. 0 = no prompts; 1 = one prompt, fairly low level, 3 = several prompts of varying levels of quality, 5 = multiple prompts of high quality.

9. Useful Corrective Feedback. Feedback is not simplistic in formative and mediated education (Shute, 2008). When feedback is constant, it can be distracting especially if it lags too long after the incident. 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 several times in a row.

Example. 0 = no feedback; 1 = minimum feedback, or at the end of module only, 3 = mid-quality feedback; 5 = Useful and actionable feedback with proper pacing.

10. Assessment Included. Assessment comes in a multitude of forms. When it happens during the experience it is called formative. When feedback happens 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. (2013) (for evidence-centered design, see Mislevy, Behrens, DiCerbo, and Levy (2012). The test could be old-fashioned paper and pencil, but it should always be "valid, reliable, and fair". If only simplistic true/false questions are asked the module should be scored low.

Example: 0 = no assessment, 1 = some assessment, low level, 3 = some assessment both within module (formative or in-process) and at the end, 5 = assessment is high quality and occurs both within module (formative) and at end for summative reflection.

11. Designed for Comfort. As of this writing (2020), all commercial VR headsets force the muscles attached to the lenses in the eyes to fatigue after a while. This effect is also called vergence-accommodation strain. Certain tasks induce this strain faster (e.g., tracking disproportionately sized objects in different planes). Text reading is straining. Designers should minimize the amount of text, and use a black background with white text. As you quantify comfort and amount of 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 with text does not mean players will *not* also read. Just remember "less is better".

Example: 0 = too much text/small object tracking leads to eye strain, 3 = a moderate amount of text, efforts have been made to keep the experience comfortable, 5 = very minimal amount of text, overall the composition and experience is very comfortable. (Do **not** tally the title or 'optional viewable content', e.g., the credits or answers to hints).

12. Reduction of Cybersickness. Cybersickness is polysymptomatic and polygenetic, and affects each individual differently. The Simulator Sickness Questionnaire (SSQ) (Kennedy, Lane, Berbaum, and Lilienthal (1993) includes three categories: 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. In one study, approximately 50% of discomfort was gone by the 10th session (Rebenitsch & Owen, 2016). It is commonly agreed that the negative effects are caused by a visual/vestibular disconnect. They list several VR design fixes including "… partially limiting the degrees of freedom in control when navigating…and increased tactile feedback..." (p.122). The ideal speed of avatar navigation is

[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.

still not known, but cybersickness tends to increase with increasing speed (So (So, Ho, & Lo, 2001). There are known tricks to decrease cybersickness, including using ramps rather than stairs, and making learners more agentic - in control of where they go and with what they interact.

Example: 0 = experience is **highly likely** to lead to nausea and disorientation (e.g., includes instances of accelerating navigation, excessive yaw axis movement, little agentic control – e.g., a roller coaster, 1 = more than somewhat likely, 3 = may lead to cybersickness over extended time, e.g., the POV changes often, 5 = **very unlikely** to lead to nausea and disorientation, e.g., there is large degree of agentic control and nothing should induce vertigo, etc.

13 How often, and how much of the content is manipulable? 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 (i.e., push the "T" key for a screw to be turned). 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. Navigation across a room should only be counted as one instance and not counted repeatedly. 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, only passive viewing. 1 = very low level of manipulable content, 2 = some manipulable content, but few chances to interact, 3 = more instances and more chances, but repetitive, 5 = a high amount of manipulable content and high frequency for interaction, novel instances throughout.

14. **Avatar Creation.** Is there an avatar? Can the learner choose an avatar? How many components of the avatar are customizable? Research in 2D supports that some customization is valuable (Lin et al., 2017). But, when users can fully customize their avatars, might there be an inflection point where too much choice leads to wasted time? Additionally, what are the pros and cons of inhabiting non-human forms? More research is needed in this domain[3]

Example: 0 = no avatar, 1 = yes, hands or body present, but no avatar choice - preassigned, 3 = yes, a body and two or three components can be chosen (e.g., clothes, hair, etc.), but only humanoid options, 5 = yes, a body and more than four components are customizable, and non-human options are available.

15. Avatar In Play: Specifically, this item depends on the two dimensions of: a) **Amount of avatar displayed**, and b) **Alignment of movements** paired with ease of control (i.e., there is not a complex button sequence to be memorized). For more on avatars see Gonzalez-Franco and Peck[4].

---

[3] The *Catch a Mimic* game does not show the body beyond hands, yet it was an effective learning tool. We must await further VR avatar research, for now points are awarded along a continuum of allowing users more freedom and creativity in avatar construction.

[4] The avatar section was based, in part, on Gonzalez-Franco and Peck's (2018) thoughtful review of over 30 embodiment experiments that have used questionnaires since 1998. They chose 25 most indicative questions regarding embodiment in VR and ran a principal-components analysis (PCA). They ended up with six question "types" that are present for embodiment of avatars (Gonzalez-Franco & Peck, 2018, p.3).
Our proposed QUIVRR rubric addresses types 1, 2, 3, and 5.

1. **Body ownership**. Present whenever there is a substitute body or body part. …
2. **Agency and motor control** of the body. Present whenever there is motion tracking and the participant can move parts or all of the virtual body.
3. **Tactile sensations**. Present whenever there is tactile or haptic stimulation to enhance the embodiment illusion.
4. **Location of the body**. Present whenever there is a substitute body or body part that is either collocated or not collocated with the participant. Participants must feel that their body is in the same location as the virtual body in order to experience an embodiment illusion.

Example: 0 = no avatar, no control, 1 = hand(s) only or poor movement match, 3 = more of the body is shown, and there is adequate movement match, 5 = full body shown and high-quality match, e.g., lips sync well**.**

16. Intuitive Interface. Intuitive means the "the users' unconscious application of pre-existing knowledge leads to effective interactions" (Israel et al., 2009). There should not be a dependence on complicated button sequences for navigation, nor to get questions in-game answered. 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 also Cybersickness item). Modules and interfaces should be designed with first-time users in mind. Remember, not all youth are gamers. A percentage of the population is colorblind, so critical elements and feedback should not rely on red/green distinctions.

Example: 0 = highly cluttered and unnavigable, 1 = somewhat cluttered and not very easy to navigate, 3 = clean interface, but not easy to navigate, 5 = clean interface and very 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.

Example: 0 = silent, 1 = low quality and/or poor directional mapping, 3 = acceptable quality and acceptable mapping, 5 = high quality, creative and synced omnidirectionally. If the sound is of an extremely low quality, or is distracting and irritating, it should be given a reduced rating.

18. Haptics/Other Modalities. Even though haptics are not yet common in many VR modules, vibrotactile feedback should become more common. Tactile and haptic add-ons are evolving rapidly, they should be included *when they further learning*, and not be distractors.

Example: 0 = only visual, 1 = auditory is included, adequately integrated, 2 = one extra modality, integrated well, 3 = two modalities (beyond audio and visual) are integrated well, 5 = multi-modal and very well integrated such that the inclusion creatively furthers learning.

19. Engagement. 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 mixed reality lesson that would have scored high on most items, but once in the field it was determined that few high school students could figure out how to walk with a 'negative acceleration'. Repeated gameplay led to expressions of frustration.[5] Several Randomized Control Trial (RCT) studies have not supported that more narrative leads to significant multi-media learning gains (Adams, Mayer, MacNamara, Koenig, & Wainess, 2011; Johnson-Glenberg & Megowan-Romanowicz, 2017). Trust your instincts. Engagement must be linked with educational payoff and learning goals must be supported.

Example: 0 = not at all engaging *(more broccoli than chocolate*), 1 = somewhat engaging, 3 = engaging, 5 = very engaging, and evidences high educational worth. You may also change any of these scores after you observe users engage with and discuss a module.

---

5. **External appearance**. Present when the self-avatar is a lookalike avatar or as control questions when there are shape, gender, race, clothing, or other visual modifications … different from the self.

6. **Response to external stimuli**. In many occasions during the experiment there is an event that modifies or threatens the body or body parts of the self-avatar (p.3).

[5] Probably inadequate scaffolding could be blamed, but we did not realize until we were in the field that the high school students did not possess the prior knowledge of the term "acceleration".

20. Overall Quality/Other. This question allows you to make a summative decision or can include an item that you feel strongly about. If race and gender are not well represented then ding the module and write that under Notes. Some reviewers dislike very 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. Certain elements should result in points deleted, e.g.,) are there obvious stitch lines?,is the content highly repetitive?, is there obvious aliasing on thin lines, 4) are icons uninterpretable, and 5) was very poor 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 or missed something salient to the reviewer, 1 = some quality was attempted, 3 = good quality, 5 = high overall quality.

BONUSES

The following two items could be worth a total of 10 points; however, I have never experienced a module that would rate 110. The hope is these two constructs will be the norm in the not too distant future and no longer considered bonuses.

21. **Adaptive.** Adaptive means that the experience in the module changes according to the learner's performance. The experience is "dynamic" and based on the players' choices made during encoding. Linear pathways cost less to create compared to dynamic pathways, but we should certainly all be striving for dynamic experiences.

Example: 0 = no adaptivity, 2 = user control only over a series of predefined, linear pathways, 3= up to three pathways of adaptivity, 5 = four or more pathways feel seamless, rigorous research has been done to optimize pathway selection. Note: Having the choice over a series of *linear pathways* is only worth 2 to 3 points.

22. **Multi-user Collaboration.** Collaboration and cooperation are well-researched and well-regarded constructs in education (Johnson & Johnson, 1989). Currently, this item is in the "bonus section" because it is expensive and complex to create multiplayer experiences in VR. We also note , there are negatives associated with Collaborative Problem Solving (CPS) including: a) diffusion of responsibility for completing tasks (which affects single student assessments), b) disagreements that can paralyze progress, and c) learners getting sidetracked by irrelevant discussions (Graesser, Kuo, and Liao (2017).

Example: 0 = single user only, 1 = a screencast option is included – but only one learner is in the experience, 3 = multi-user for small group only, 5 = multi-user for larger group with synchronous teacher dashboard support.

Acknowledgments

The author wishes to thank James Comstock, Hannah Bartolomea, Man Su, Elena Kalina, Ricardo Nieland Zavela, and Vanessa Ly.

Full References from Chapter

Adams, D. M., Mayer, R. E., MacNamara, A., Koenig, A., & Wainess, R. (2011). Narrative games for learning: Testing the discovery and narrative hypotheses. *Journal of Educational Psychology, 104*(1), 235-249

Antle, A. A., Corness, G., & Droumeva, M. (2009). What the body knows: Exploring the benefits of embodied metaphors in hybrid physical digital environments. *Journal of Interactive Computing, 21*(1-2), 66-75.

Banakou, D., Kishore, S., & Slater, M. (2018). Virtually Being Einstein Results in an Improvement in Cognitive Task Performance and a Decrease in Age Bias. *Frontiers in Psychology, 9*(917). doi:10.3389/fpsyg.2018.00917

Dalgarno, B., & Lee, M. J. W. (2010). What are the learning affordances of 3-D virtual environments? *British Journal of Educational Technology, 41*(1), 10-32. doi:10.1111/j.1467-8535.2009.01038.x

Darling-Hammond, L., & etal. (2013). *Criteria for high-quality assessment*. Retrieved from Stanford, CA: https://edpolicy.stanford.edu/sites/default/files/publications/criteria-higher-quality-assessment_2.pdf

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410-8415. doi:10.1073/pnas.1319030111

Goldin-Meadow, S. (2011). Learning through gesture. *WIREs Cognitive Science, 2*, 595–607. doi:10.1002/wcs.132

Goldin-Meadow, S. (2014). Widening the lens: What the manual modality reveals about learning, language and cognition. *Philosophical Transactions of the Royal Society, Biological Sciences, 369*(20130295). doi:10.1098/rstb.2013.0295

Gonzalez-Franco, M., & Peck, T. C. (2018). Avatar Embodiment. Towards a Standardized Questionnaire. *Frontiers in Robotics and AI, 5*(74). doi:10.3389/frobt.2018.00074

Graesser, A., Kuo, B., & Liao, C. (2017). Complex problem solving in assessments of collaborative problem solving. *Journal of Intelligence, 5*(10), 1-14. doi:10.3390/jintelligence5020010

Greenbaum, Z. (2020). Going high tech in academia. *Monitor on Psychology, 51*(1).

Hamilton, E. R., Rosenberg, J. M., & Akcaoglu, M. (2016). The Substitution Augmentation Modification Redefinition (SAMR) Model: a Critical Review and Suggestions for its Use. *TechTrends, 60*(5), 433-441. doi:10.1007/s11528-016-0091-y

Hossain, Z., Bumbacher, E., Blikstein, P., & Riedel-Kruse, I. (2017). *Authentic Science Inquiry Learning at Scale Enabled by an Interactive Biology Cloud Experimentation Lab*. Paper presented at the Proceedings of the Fourth (2017) ACM Conference on Learning @ Scale, Cambridge, Massachusetts, USA.

Israel, J., Hurtienne, J., Pohlmeyer, A. E., Mohs, C., Kindsmuller, M., & Naumann, A. (2009). On intuitive use, physicality and tangible user interfaces. *International Journal of Arts and Technology, 2*(4), 348-366.

Johnson-Glenberg, M. C. (2018). Immersive VR and education: Embodied design principles that include gesture and hand controls. *Frontiers in Robotics and AI, 5*(81). doi:10.3389/frobt.2018.00081

Johnson-Glenberg, M. C. (2019). The Necessary Nine: Design Principles for Embodied VR and Active Stem Education. In P. Díaz, A. Ioannou, K. K. Bhagat, & J. M. Spector (Eds.), *Learning in a Digital World: Perspective on Interactive Technologies for Formal and Informal Education* (pp. 83-112). Singapore: Springer Singapore.

Johnson-Glenberg, M. C., Birchfield, D., Megowan-Romanowicz, M. C., & Snow, E. L. (2015). If the gear fits, spin it! Embodied education and in-game assessments. *International Journal of Gaming and Computer-based Simulations, 7*(7), 40-65. doi:DOI: 10.4018/IJGCMS.2015100103

Johnson-Glenberg, M. C., & Megowan-Romanowicz, M. C. (2017). Embodied science and mixed reality: How gesture and motion capture affect physics education. *Cognitive Research: Principles and Implications, 2*(24). doi:10.1186/s41235-017-0060-9

Johnson-Glenberg, M. C., Megowan-Romanowicz, M. C., Birchfield, D., & Savio-Ramos, C. (2016). Effects of embodied learning and digital platform on the retention of physics content: Centripetal force. *Frontiers in Psychology, 7:1819*. doi:doi: 10.3389/fpsyg.2016.01819

Johnson, D. W., & Johnson, R. T. (1989). *Cooperation and Competition: Theory and Research*. Edina, MN: Interaction Book Company.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness. *Int J Aviat Psychol, 3*(3), 203–220.

Kirschner, P. A., Sweller, J., & Clark, R. E. (2006). 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, 41*(2), 75-86. doi:10.1207/s15326985ep4102_1

Lamb, R., Antonenko, P., Etopio, E., & Seccia, A. (2018). Comparison of virtual reality and hands on activities in science education via functional near infrared spectroscopy. *Computers & Education, 124*, 14-26. doi:https://doi.org/10.1016/j.compedu.2018.05.014

Lin, L., Parmar, D., Babu, S., Leonard, A. E., Daily, S. B., & Joerg, S. (2017). *How character customization affects learning in computational thinking*. Paper presented at the Proceedings of the ACM Symposium on Applied Perception, Cottbus, Germany.

Lindgren, R., Tscholl, M., Wang, S., & Johnson, E. (2016). Enhancing learning and engagement through embodied interaction within a mixed reality simulation. *Computers & Education, 95*, 174-183.

Mislevy, R. J., Behrens, J. T., DiCerbo, K. E., & Levy, R. (2012). Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. *JEDM-Journal of Educational Data Mining, 4*(1), 11-48.

Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cognition and Instruction, 1*(2), 117-175. doi:10.1207/s1532690xci0102_1

Radianti, J., Majchrzak, T. A., Fromm, J., & Wohlgenannt, I. (2020). A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. *Computers & Education, 147*, 103778. doi:https://doi.org/10.1016/j.compedu.2019.103778

Rebenitsch, L., & Owen, C. (2016). Review on cybersickness in applications and visual displays. *Virtual Reality, 20*(2), 101-125. doi:10.1007/s10055-016-0285-9

Rice, J. W. (2007). Assessing Higher Order Thinking in Video Games. *Journal of Technology and Teacher Education, 15*(1), 87-100.

Schell, J. (2015). Making Great VR: Six Lessons Learned From "I Expect You To Die". Retrieved from https://www.gamasutra.com/blogs/JesseSchell/20150626/247113/Making_Great_VR_Six_Lessons_Learned_From_I_Expect_You_To_Die.php

Shute, V. (2008). Focus on formative feedback. *Review of Educational Research, 78*(1), 153-189. doi:10.3102/0034654307313795

Slater, M., & Wilbur, S. (1997). A framework for immersive virtual environments (FIVE): speculations on the role of presence in virtual environments *Presence: Teleoperators and virtual environments, 6*, 603–616.

So, R. H. Y., Ho, A., & Lo, W. T. (2001). A Metric to Quantify Virtual Scene Movement for the Study of Cybersickness: Definition, Implementation, and Verification. *Presence: Teleoperators and virtual environments, 10*(2), 193-215. doi:10.1162/105474601750216803

Yoon, S., & Wang, J. (2014). Making the invisible visible in science museums through augmented reality devices. *Tech Trends, 58*(1), 49–55.