A situation that we had never imagined: post-Fukushima virtual collaborations for determining robot task metrics

Michael Vallance, Stewart Martin, Catherine Naamani

Abstract: There is no consensus regarding a common set of metrics for robot task complexity in associated human-robot interactions. This paper is an attempt to address this issue by proposing a new metric so that the educational potential when using robots can be further developed. Tasks in which students in Japan and UK interact in a 3D virtual space to collaboratively program robots to navigate mazes have resulted in quantitative data of immersion, circuit task complexity and robot task complexity. The data has subsequently been collated to create a proposed new metric for tasks involving robots, which we have termed task fidelity. The paper proposes that task fidelity is a quantitative measure of a set robot task in relation to a learner's solution. By quantifying task fidelity educators utilising robots in schools and in higher education will be able to provide tasks commensurate with the expected successful outcomes achieved by the learners.

Keywords: architectures for educational technology system; improving classroom teaching; robotics; programming; simulations; virtual reality.

To reference this paper: Vallance. M., Martin, S., & Naamani, C. (2015). A situation that we had never imagined: post-Fukushima virtual collaborations for determining robot task metrics. *International Journal of Learning Technology*, Vol. 10, No. 1, pp.30-49.

Biographical notes: Michael Vallance is a Professor in the Department of Media Architecture, Future University Hakodate, Japan. He has been involved in educational technology design, implementation, research and consultancy for over 15 years, working closely with Higher Education Institutes, schools and media companies in UK, Singapore, Malaysia and Japan. He was awarded second place in the Distributed Learning category of the United States Army's 2012 Federal Virtual Worlds Challenge for his research in virtual collaboration and robots. His website is at http://www.mvallance.net.

Stewart Martin is Professor of Education and Head of Department at the Centre for Educational Studies at the University of Hull, UK. He is a National Teaching Fellow of the UK Higher Education Academy, a keen advocate of research-led teaching and of the application of experiment, new pedagogy and theory in the use of technology, where creating unique, effective educational applications that do not just replicate established computer-free approaches is often seen as 'difficult'. His research interests include the educational use of computer technology; citizenship; images of the self; psychometrics; cognition, learning and technology; educational leadership and organisational change. http://www.iverg.com.

Catherine Naamani is Head of Technology Enhanced Learning based in the Centre for Excellence in Learning & Teaching at the University of South Wales, UK. She has worked

with local schools for the past six years to promote the use of technology to enhance learning and teaching. During this time, she was also seconded to the Welsh Government to advise on developing an ICT Strategy for schools. In her current role, she leads the university's strategy to research and develop innovative approaches to technology enhanced learning, teaching and assessment. Her website is http://staff.southwales.ac.uk/users/1338-cnaamani.

Introduction

The educational potential of using robots in schools and higher education can only be fully realised through understanding the specific categories of knowledge being employed when learners are engaged in robot specific tasks (Gaudiello and Zibetti, 2013). In any task design it is also important to consider its difficulty for the intended learners. Therefore, task designers such as teachers and higher education practitioners need to provide tasks commensurate with the expected successful outcomes which, it is anticipated, will be developed by the learners; although achieving this is not straightforward. In this paper we demonstrate how tasks can be matched and quantified with learner outcomes within the particular context of communicating the programming of robots.

The motivation for programming robots in a simulated 3D virtual world was the Fukushima Daiichi nuclear power plant disaster of March 2011, which revealed much about Japan's lack of preparedness for nuclear accidents. Despite the brave efforts of its labour force leading up to and in the aftermath of the reactor explosions it became apparent that coordination and communication were disorganised. One of the most surprising technology related episodes during the post-disaster efforts was Japan's lack of robots to assist with the recovery operations. Despite Japan being a robotics-friendly nation with the world's highest levels of automation, it had to count on foreign assistance in the form of disaster recovery robots donated by iRobot USA (Vallance et al., 2013) and it took two weeks of training before the robots were allowed to inspect the damaged reactors. Japan may be a nation whose image is that of advanced technology and creative media, but the actual uptake of technology in Japan's education system 'remains comparatively low' (UNESCO, http://www.unescobkk.org/index.php?id=1381). As described by Mima (2003), the Japanese educational curricula are largely designed for fact-based, exam-oriented learning and the pedagogy is founded upon a hierarchical flow of information from 'knowers to non-knowers' (ibid ., p.266). This has a significant impact on the nature of university education in Japan.

Japanese undergraduates arrive at university ill-equipped with computer literacy skills and are incapable of applying studied theoretical concepts; for instance, they are often unable to progress from static declarative knowledge to active procedural knowledge. Moreover, Japan's assessment-dominated educational culture dismisses collaboration as a supplementary activity, rather than a core learning experience to be valued. In addition, science and engineering courses remain unattractive to Japanese teenagers. This is likely to have a negative impact upon Japan's future 'human capital' if not reversed. Our solution to this is to engage undergraduate students to actively participate in international tele-collaboration tasks involving basic robots in both the real world and a virtual world simulation.

Robot-mediated interaction (RMI) of this kind is a novel but effective form of communication, promising significant benefits in remote collaboration. Systems that afford RMI provide remote users with the ability to navigate in a local environment and communicate with individuals in that environment (Rae et al., 2013). LEGO robot programming components have been shown to be useful in quantifying task complexity and for iteratively increasing the challenges given to students (Vallance and Martin, 2012).

3D virtual simulations provide interesting, engaging, realistic yet safe contexts where robots are ordinarily utilised; such as in disaster recovery situations. A virtual world simulation allows remotely located students to enter as avatars to communicate and collaborate with other students, often across different continents. Cross cultural collaboration is seen as increasingly valuable and necessary in many fields of science, technology, engineering, research and education and international teams increasingly rely on synchronous, asynchronous and virtual technologies in such work. Although many of these intellectual domains may share procedural and content knowledge, this is not always the case with educational experiences and habits, where cognitive approaches to learning and thinking may draw on different cultural pedagogical traditions. Bringing individuals together who have very different pedagogical experiences may also be a valuable way of creating discussion and facilitating new thinking. This is important when trying to encourage students to think about which approaches are most helpful in problem solving and as an approach to promoting metacognitive learning. The use of students from different cultures also may allow for the exchange and modification of acquired learning strategies and could also be useful in testing some of the implicit assumptions in taxonomies such as Bloom’s.

Higher education academics, researchers and school educators therefore need to find ways to support the experience of collaboration in virtual spaces. To facilitate such work we developed an OpenSim 3D virtual space (Figure 1) and a simulation of the Fukushima nuclear power plant in a Unity3D virtual space (Figure 2) for Japanese students to collaborate with UK students. Both groups of students were highly motivated by this setting: Japanese students because of its national proximity and immediate impact on themselves, their families or their friends; and UK students because of the international interest the accident generated, especially in light of similar incidents in recent history and also because of the UK’s current review of its strategy towards replacing its ageing nuclear power infrastructure and concerns about meeting future energy needs for a growing economy.

Figure 1 Student as an avatar manoeuvring a robot in our 3D virtual world

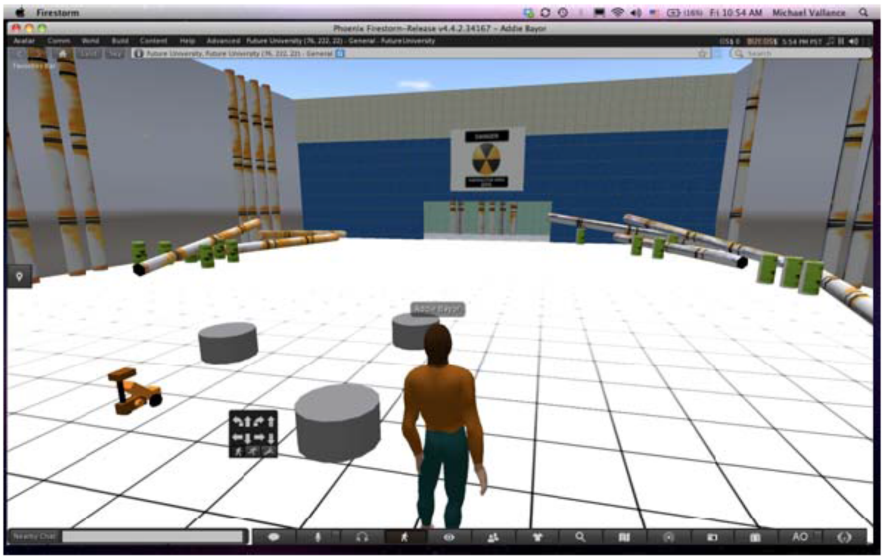


Figure 2 Instructor as an avatar in our virtual Fukushima nuclear plant

We begin with a brief summary of task complexity in the disciplines of robotics and human-robot interaction (HRI) then set the context of programming robots in 3D worlds within a new interpretation of complexity. After that the development of ‘circuit task complexity (CTC)’ and ‘robot task complexity (RTC)’ are explained within the specified context. In order to give educational value to these terms so that transfer can be made across various robot and HRI contexts, the term task fidelity (TF) is proposed. Data is collated and analysed from 29 RMI s involving students in the UK and Japan collaborating in constructing and programming robots. Finally we discuss the value of a common metric such as TF for educators in schools and higher education when teaching with robots. An illustrative video is available to view on our companion website at http://www.mvallance.net.

Robot task complexity

Common metrics are valuable for benchmarking within many domains. An example is road transportation, where cars, motorcycles and trucks can be compared on important objective features such as top speed, acceleration, engine capacity, fuel economy, transmission and price. However, although there are a number of measures which can be applied to robot-related tasks and the complexity of a given domain where robots are utilised, common metrics do not exist in the same way because, as Steinfeld et al. (2005) state, “the primary difficulty in defining common metrics is the incredibly diverse range of human-robot applications” (p.33).

An attempt at common metrics has ben provided via the USUS Evaluation Framework for HRI by Weiss et al. (2009), which focuses on usability, user experience, social acceptance and social impact. Usability is the extent to which a robot can be used by specified users to achieve specific goals and produce effectiveness, efficiency and satisfaction in a particular context of use. Metrics include effectiveness (i.e., task completion rate), efficiency (i.e., speed at which a task is completed), learnability (i.e., how easily the system can be learned by human users), flexibility (i.e., the number of different ways users can communicate with the system), robustness (i.e., the level of support provided) and utility (i.e., the number of tasks the interface is designed to perform). However, although useful, the USUS outcome is descriptive and does not produce objective numeric comparators.

In another study Murphy and Schreckenghost (2013) conducted a meta-analysis of 29 papers that between them proposed 42 different metrics for HRI. These metrics were categorised according to the object being directly measured; such as the human ($N = 7$), the robot ($N = 6$), or the system ($N = 29$). The metrics for the system were subdivided into productivity, efficiency, reliability, safety and coactivity. However, metrics were often not measured directly but were instead inferred through observation and the authors concluded that they “have no functional, or generalizable, mechanism for measuring that feature” (ibid). Although many attempts have been made to develop a taxonomy of metrics, the research community has yet to develop a standard framework and many metrics remain highly task-specific.

Perhaps because of this complex situation, when discussing robots that undertake specific manoeuvres, some researchers adopt task complexity as a common metric, where tasks are defined as physical action units that are undertaken by a robot and the designation ‘complexity’ is used to characterise the task that consists of different parts in potentially intricate arrangements. In an example of robots which manoeuvre around obstacles and follow distinct circuits (or mazes), Barker and Ansorge (2007) derive task complexity as $TC = \Sigma S + time$, where ΣS is the number of portions or turns of a maze. Olsen and Goodrich (http://icie.cs.byu.edu/Papers/RAD.pdf) define task complexity as $TC = TE + IE$, where task effectiveness (TE) reflects the number of commands successfully programmed into the robot and interaction effort (IE) the amount of time required to interact with the robot (to take into account mistakes).
These metrics (Barker and Ansorge, 2007; Olsen and Goodrich http://icie.cs.byu.edu/Papers/RAD.pdf) have been successfully adapted in the development of our research (Vallance and Martin, 2012) although due to the absence of a common set of metrics, we

feel it appropriate to develop our own task complexity value specific to the context of our RMIs discussed in this paper.

Robot tasks in 3D virtual worlds

Our research collated data from students collaborating in a 3D virtual world to program a LEGO robot to successfully navigate mazes from start to completion in both the physical world and within our 3D virtual space (see Figure 1). This was undertaken by:
1 designing circuits which necessitate the use of robot manoeuvres and sensors
2 students in Japan and UK experiencing collaboration in virtual worlds.
These experiences led to the development of personal strategies for teamwork, planning, organising, applying, analysing, creating and reflection. Complex problems were presented which necessitated the use of programming skills, design, cross-cultural collaboration and RMIs.

We divided task complexity into CTC and RTC because the task focuses upon the robot and what the human has to do to manipulate that robot. We call this the 'product' of a robot task. We appreciate that HRI is the 'intelligent interaction' between a human and a robot but the word 'interaction' assumes that the human and the robot are engaged in two-way communication. We call this the 'process' of a robot task. Although there is feedback from the robot in our tasks we would not necessarily claim that there is any intelligent interaction; it is simply feedback. Therefore, we consider the collaboration to be an example of RMI where humans act upon feedback provided by both robot and other humans.

Method

Task fidelity

In this section we explain our use of TF, which is the value resulting from the complexity of the circuit compared with the complexity of the program to complete that circuit. We will provide an example problem and a possible solution as illustrated in Figures 3 and 4, respectively.

Figure 3 Example of CTC

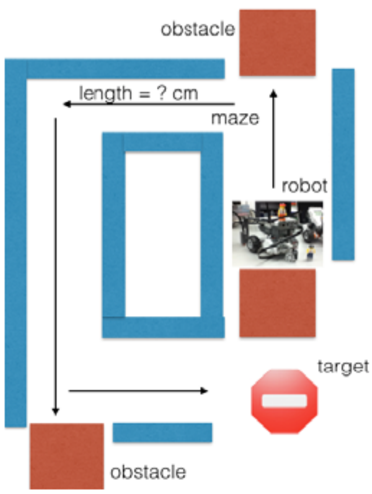
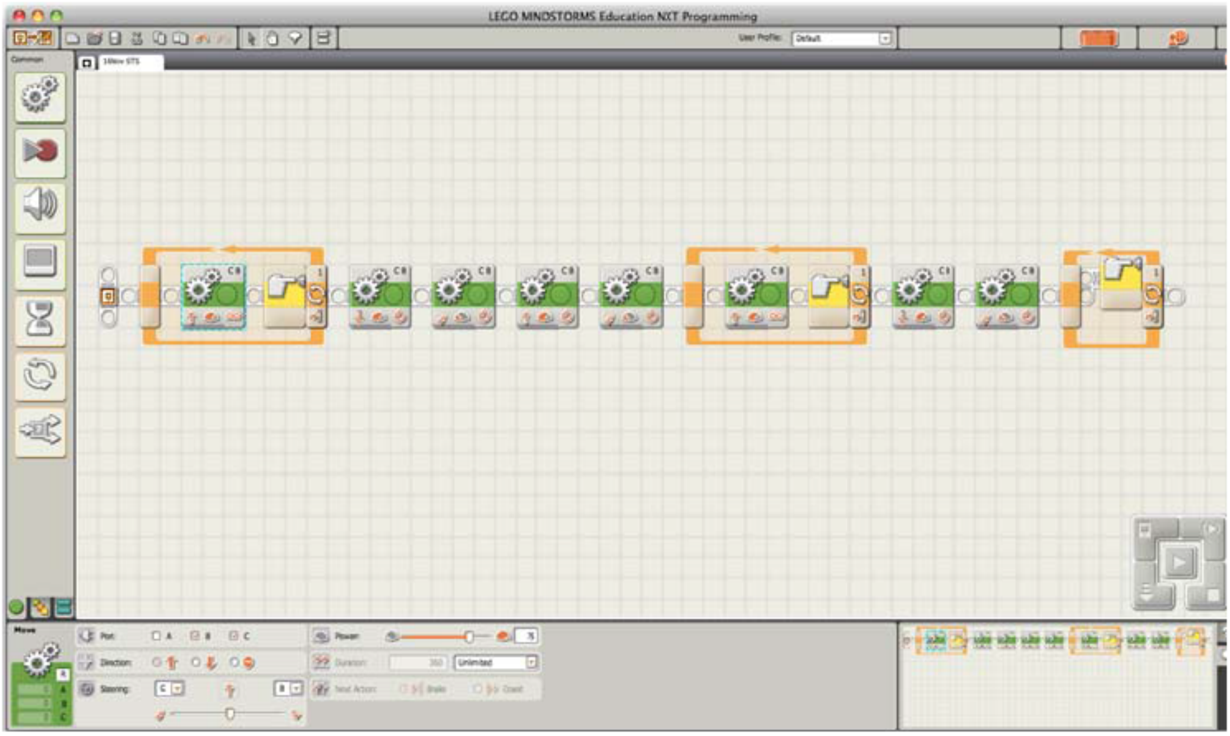


Figure 4 LEGO NXT program solution



In the first iteration of our research, quantifying each task complexity the programming of the LEGO robot required a determination of an action and a vector (Vallance and Martin, 2012). Given the specific purposes of the robot in our research, we utilised the work of Barker and Ansorge (2007) and Olson and Goodrich (http://icie.cs.byu.edu/Papers/RAD.pdf); where task complexity is calculated according to the number of portions that make up a given maze. We called this CTC which is composed of the number of directions *(d)* + number of manoeuvres *(m)* + number of sensors *(s)* + number of obstacles (*o*), which may be written as:

$$CTC = \Sigma \, (d + m + s + o)$$

For example, in Figure 3 the robot must manoeuvre around at least two obstacles in order to reach its target. The number of directions to be programmed is 4, the number of manoeuvres is 3 and the number of sensors is 2 (i.e., two touch sensors), so this can be expressed as:

$$\mathrm{CTC} = \Sigma \ (d+m+s+o)$$

$$\mathrm{CTC} = \Sigma \ (4+3+3+2) = 11$$

However, we found that the logic we were using to assign task complexity to circuits was inadequate. Initially we assigned complexity values to distinct manoeuvres such as forward – turn – back, but we found over the course of our research that as circuits became more challenging, the Mindstorms NXT programming became more complex.

This was especially the case when we needed to add sensors to manoeuvre around and over obstacles. Simply adding the number of obstacles to the CTC was insufficient because the programming required to manoeuvre over a bridge using touch sensors, for instance, was far more complex than that required to manoeuvre around a box using touch sensors. Consequently, we modified our task complexity so as to be determined by the NXT program solution rather than the circuit to be navigated. We now call this RTC, which is measured as:

$$RTC = Mv_1 + Sv_2 + SW + Lv_3$$

where

M number of moves (direction and turn)

S number of sensors

SW number of switches

L number of loops.

Also, where v = number of decisions required by the user for each programmable block so that, as explained below, v_1 = 6, v_2 = 5 and v_3 = 2.

In the NXT Mindstorms software, the 'Move' instruction block controls the direction and turns that the LEGO robot will take. There are six variables that need to be considered: NXT processor port, direction, steering, power, duration and next action. In other words, students have to make six specific decisions about the values that make up the programmable block and so we assign v_1 a value of 6. There are eight common sensors which are used in our tasks (timer, light, ultrasonic, colour, touch, sound, distance, wait) with each sensor's capabilities determined by five variables (so we assign v_2 = 5). Although some sensors have six decisions built in and some have five, the difference is that the extra decision is simply cosmetic as in 'speak an alert' so does not impact on the robot's performance or capability to complete the task. All sensors are tagged as S . A loop has only two variables (if/else) to consider so we assign v_3 = 2.

Given the circuit shown in Figure 3, the robot has to be programmed to move in four directions, with three turns and two touch sensors. A possible NXT program solution such as shown in Figure 4 can then be used to calculate the RTC, which can be written as:

$$RTC = Mv_1 + Sv_2 + SW + Lv_3$$

There are eight move blocks, three sensors and three switches.

$$RTC = (8x6) + (3x5) + 3 + 0$$
$$RTC = 66$$

It is acknowledged that other possible programming solutions could produce different RTC values.

TF is the resulting value of the complexity of the circuit compared with the complexity of the program generated to complete the circuit.

In order to develop a measure that includes the most relevant variables the collated data will need to include values for total challenge and skill, CTC and RTC. In order to compare the data from all tasks it is therefore useful to scale the students' challenge and skill values between 0 and 1. To do this, in each task we divide the sum of the challenge values of the students by the maximum score possible. Similarly, we divide the sum of the skill values of the students by the maximum score possible. For the CTC values we take the maximum CTC value and divide it into each CTC value. Similarly, for the RTC values we take the maximum RTC value and divide it into each RTC value. Measures are thus converted to values between 0 and 1. This allows us to represent comparative data graphically and thus the immersion in the case of challenge and skills and TF (see below) in the case of CTC and RTC.

As a result, the complexity of the task can now be quantified by the new metric, TF, which is calculated as:

Task Fidelity = Circuit Task Complexity - Robot Task Complexity or

TF = CTC - RTC

Immersion

When linked to students' immersion in tasks, TF is a useful indicator of the complexity of a task. To record 'immersion' [a cognitive phenomena also referred to as 'flow' (Csikszentmihalyi and Nakamura, 2010)], data can be collected from the students during and after each task, using questions developed from research in immersivity by Pearce et al. (2005). The assumption is that with optimal parameters for challenge and for skill relationship, students become 'immersed' in the RMI tasks. In order to determine how deeply immersed students were in each task we asked them about the challenge and their skills during and after all tasks. In our initial development we had a virtual iPad appear in front of the avatars where the two questions were displayed for the avatars to answer. The data automatically transferred to a database but this system proved unreliable so we have resorted to pen and paper until this is resolved.

To calculate immersion we utilised Pearce et al.'s flow criteria of task challenge and skill: "Amongst the various studies researching flow, an ongoing issue has been to find a method for measuring flow independently from the positive states of consciousness (such as enjoyment, concentration, control, lack of self-consciousness, lack of distraction). One solution has been to use a measure of the balance between the challenge of an activity and the participant's perception of their skill to carry out that activity" (ibid , p.250). In order to capture this data immediately after the completion of a task and while still in communication with their virtual collaborators in the virtual world, the students reported on the task's challenge and their skill in attempting the task. For 'challenge' they had to report whether they considered the task difficult, demanding, manageable or easy. For 'skill' they had to report whether they considered their ability to undertake the task as hopeless, reasonable, competent or masterful. Once the task had been completed, students logged out of the virtual world and a general discussion of the task process and its outcome was held locally with the researchers.

To calculate an immersion value, the challenge and skill metrics were assigned scores of 1 to 4: for challenge, difficult = 4, demanding = 3, manageable = 2 and easy =1; for skill, hopeless = 1, reasonable = 2, competent = 3 and masterful = 4. Then the number of participants was used to determine a task's maximum challenge and skill score. An example is given in the results section.

We acknowledge that we have applied our modified RTC metric only to the LEGO Mindstorms robot, but argue that this provides a useful indicator of experiential learning during collaborative tasks and so the next section will demonstrate the development of TF from collected data.

Participants

Prior to entering the virtual spaces undergraduate students in Japan studying media architecture (N = 6) and A-level students in UK studying science-based subjects (N = 10) have been undertaking robot related tasks in a learner-designed, OpenSim 3D virtual space. Of the 16 participant students two are female, 14 are male, all are aged between 17 and 19 and none have experienced working with LEGO Mindstorms or in OpenSim prior to this project.

Tasks

To date we have conducted and recorded reliable data from 39 tasks conducted in our OpenSim 3D virtual space and our Unity 3D virtual Fukushima nuclear power plant space. Task 1 was an introductory task while tasks 13 to 15 and 22 to 27 were remote robot manoeuvring. These tasks were removed from the data set as no programming was involved. All tasks are summarised in Table 1. We posit that this data can be used in conjunction with previous data of learning and communication to develop a framework for virtual world learning. Some tasks have involved Japanese students collaborating with other remotely located Japanese students and some with Japanese students collaborating with UK students. Tasks have included manoeuvring around obstacles using distance and turn commands, using touch sensors to find ways around obstacles, constructing a bridge and using touch sensors to move over obstacles, using light sensors to avoid obstacles, using RGB sensors to locate items and manipulating the telerobotic controls to virtually manoeuvre our LEGO robot within the virtual Fukushima space as part of 'search and rescue' simulations. Communication between students required the use of virtual world tools such as text panes, voice, live video streaming of respective real-world labs and 3D presentation boards where NXT program images could be deposited. The use of avatars in-world enabled students to remotely manoeuvre a real-world robot (tele-robot communication). The 3D space represented a disaster-area simulation in order to engage students in a contextualised task challenge.

Table 1 Tasks conducted in 3D virtual spaces

Tasks 1 ~ 13		*Tasks 14 ~ 26*		*Tasks 27 ~ 39*	
T1	Assemble LEGO robots. JPN + UK students introductions	T14	Programming LabView for remote control.	T27	Remote control for search and rescue circuit.
T2	NXT program + circuit. JPN teaching UK	T15	Programming LabView for remote control.	T28	Move to black line, stop and throw ball to hit over obstacle. UK teaching Japan.
T3	NXT program + circuit (90 degree turns + measured length). UK teaching JPN	T16	UK teaching Japan. Robot construction + NXT program + stop and swing arm to hit ball.	T29	Assemble LEGO robots. NXT program + circuit. Move fwd + touch sensor + stop + move back. JPN teaching UK students.
T4	Circuit + NXT program. Move. Touch sensor. Turn 90 degrees. JPN teaching JPN.	T17	Suika robot. Rotate + follow line+ sensor + chop down. Japan preparation 1.	T30	NXT program + circuit. Move fwd + sound sensor + stop + turn 90 degrees + move JPN teaching JPN
T5	Circuit + NXT program. Around obstacles. JPN teaching JPN.	T18	Suika robot. Rotate + follow line+ sensor + chop down. Japan preparation 2.	T31	NXT program + circuit. Program robot to race fwd + touch + return. Telerobot robot to race fwd + touch + return. Compare average times.
T6	Circuit + NXT program. Around obstacles. JPN teaching JPN.	T19	Suika robot. Rotate + follow line+ sensor + chop down. Japan preparation 3.	T32	Assemble LEGO robots. NXT program + circuit. Program robot to race fwd + touch + return using gears.
T7	NXT program + touch sensors + circuit. Locate and press switch off. JPN teaching JPN.	T20	Robot construction + NXT program + obstacles + sensors.	T33	Telerobot controlling. Robot to move fwd + turn left + move fwd + turn left + move fwd + turn left + move fwd + turn right + move fwd + turn right + move fwd + stop (pathfinder circuit)
T8	Over an obstacle. NXT program + sensors + bridge building (cardboard boxes). JPN teaching JPN.	T21	Suika robot. Rotate + follow line+ sensor + chop down. Japan teach UK.	T34	Assemble LEGO robot. NXT program + circuit. Robot to move fwd + turn left + move fwd + turn left + move fwd + turn left + move fwd + turn right + move fwd + turn right + move fwd + stop (Pathfinder circuit)

Table 1 Tasks conducted in 3D virtual spaces (continued)

Tasks 1 ~ 13		*Tasks 14 ~ 26*		*Tasks 27 ~ 39*	
T9	Over an obstacle. NXT program + sensors + bridge building (wood). JPN teaching JPN.	T22	Programming LabView for remote control.	T35	NXT program + circuit. Robot to move fwd + turn left + move fwd + turn left + move fwd + turn left + move fwd + turn right + move fwd + turn right + move fwd + stop (pathfinder circuit)
T10	Robot arm + scoop. UK teaching JPN	T23	Remote control	T36	NXT program + circuit. Robot to move fwd + turn left + move fwd + turn left + move fwd + turn left + move fwd + turn right + move fwd + turn right + move fwd + stop (pathfinder circuit). Use of ultrasonic + sound sensors.
T11	Robot arm + NXT program. JPN preparation	T24	Remote control for search and rescue circuit.	T37	NXT program + circuit. Robot to move fwd + turn left + move fwd + turn left + move fwd + turn left + move fwd + turn right + move fwd + turn right + move fwd + stop (pathfinder circuit). Use of colour sensor.
T12	Robot arm + scoop + NXT program. Streaming video. JPN teaching UK.	T25	Remote control for search and rescue circuit.	T38	NXT program + circuit. Robot to move from start point + a colour sensor for 'follow the line' + a touch sensor + a move to goal.
T13	Programming LabView for remote control.	T26	Remote control for search and rescue circuit.	T39	Assemble LEGO robot. NXT program + circuit. Robot to move from start point + a sensor + a touch sensor + a move to goal.

Results

We conducted tasks of various specifications over three semesters and for each task we measured CTC and RTC. In order to compare data from all the tasks for the CTC values we first took the maximum CTC value and divided it into each task's CTC value. Then we looked at the program solutions by our students and calculated the RTC. Similarly, for the RTC values we took the maximum RTC value and divided it into each task's RTC value. All values could thus be represented between 0 and 1. Finally, we could calculate TF as explained (see Table 2).

Taking task 2 (T2) as an example, this consisted of constructing and programming a LEGO robot to move in a maze in order to reach a specific target, as illustrated in Figure 3. The task was designed by the UK students who then had to 'teach' the circuit and its solution to the students in Japan with all communication taking place in the 3D virtual world.

As explained above,

$$\text{Circuit Task Complexity } (CTC) = \Sigma (d + m + s + o)$$

CTC for T2 was determined to equal 11.

In this series of tasks with the same students in the same lab configuration in UK and Japan and with the same technologies (MacBook Pros and LEGO NXT 2.0), the maximum CTC of all the tasks was determined as 20. For comparison, the CTC of T2 was calculated as 11/20 = 0.55.

Similarly for RTC, as explained above,

$$(RTC) = \Sigma M v_1 + \Sigma S v_2 + \Sigma SW + \Sigma L v_3$$

RTC for T2 was determined to equal 66.

The maximum RTC of the tasks was 300. For comparison, the RTC of T2 was calculated as 66/300 = 0.22.

$$\text{Task Fidelity} = \text{Circuit Task Complexity} - \text{Robot Task Complexity}$$
$$\text{Task Fidelity} = CTC - RTC - 0.55 - 0.22 = 0.33$$

All CTC, RTC and TF values for all the tasks were calculated and tabulated in Table 2.

Table 2 CTC, RTC and TF

Task	*CTC*	*RTC*	*TF*	*Task*	*CTC*	*RTC*	*TF*
T2	0.55	0.22	0.33	T20	0.70	0.48	0.18
T3	0.50	0.42	0.08	T21	0.30	0.65	–0.35
T4	0.80	0.22	0.58	T28	0.25	0.17	0.08
T5	0.80	0.57	0.23	T29	0.45	0.31	0.14
T6	1.00	0.85	0.15	T30	0.55	0.31	0.24
T7	0.70	1.00	–0.30	T31	0.40	0.93	–0.53
T8	0.25	0.39	–0.14	T32	0.40	1	–0.6
T9	0.30	0.33	–0.03	T33	0.6	1	–0.4
T10	0.20	0.20	0	T34	0.6	0.46	0.14
T11	0.65	0.76	–0.11	T35	0.6	0.70	–0.1
T12	0.65	0.84	–0.19	T36	0.7	0.78	–0.08
T16	0.55	0.83	–0.28	T37	0.7	0.29	0.41
T17	0.25	0.22	0.03	T38	1	0.82	0.18
T18	0.30	0.65	–0.35	T39	1	0.89	0.11
T19	0.30	0.65	–0.35				

To calculate an immersion value, the challenge and skill metrics were assigned scores of 1 to 4: for challenge, difficult = 4, demanding = 3, manageable = 2 and easy =1; for skill, hopeless = 1, reasonable = 2, competent = 3 and masterful = 4. Then the number of student participants were used to determine a task's maximum challenge and skill score. In order to compare the data from all tasks it was necessary to mathematically translate and scale challenge and skill to values between 0 and 1. In each task we therefore divided the sum of the challenge values provided by the students by the maximum score possible. Similarly, we divided the sum of the skill values provided by the students by the maximum score possible.

For example, in T2 there were six students so maximum challenge and skill score equals 4 x 6 = 24. The Task challenge score was then calculated by dividing the total task challenge score indicated by the participants by the maximum task challenge score.

- Total T2 challenge score = 12.
- Maximum T2 challenge score = 24.

Therefore, T2 challenge score = 12/24 = 0.5

This process was repeated for the task skill score. Calculations were repeated for all the tasks and results tabulated in Table 3. The values obtained in this way for task challenges will always come to a value between zero and 1. This allows our data to be comparable across different experimental settings, even with different numbers of students. The advantage of this approach is that this metric therefore allows cross-study and interdisciplinary comparisons.

Table 3 Table of task challenge and skill

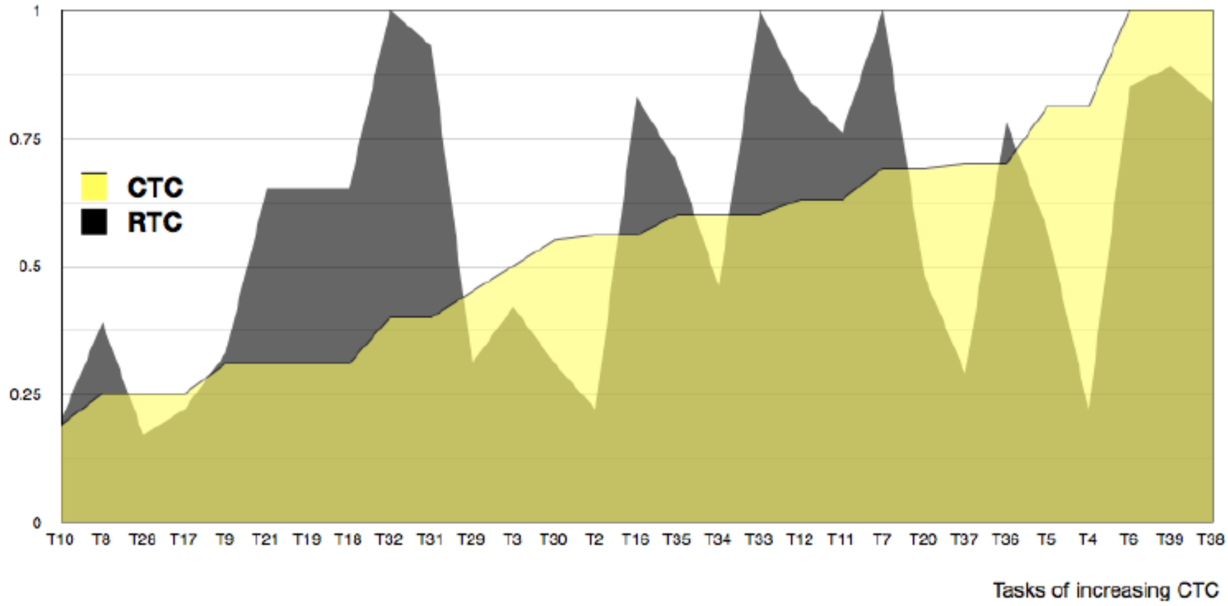
Task	*Challenge*	*Skill*	*Task*	*Challenge*	*Skill*
T1	0.5	0.75	T19	0.25	1
T2	0.5	1	T20	0.94	0.5
T3	0.75	0.5	T21	0.75	0.75
T4	0.5	0.75	T28	0.75	0.58
T5	1	0.67	T29	0.25	0.83
T6	0.8	0.67	T31	0.33	0.92
T7	0.67	0.8	T31	0.66	0.83
T8	0.67	0.5	T32	0.83	0.75
T9	0.42	0.92	T33	0.92	0.58
T10	0.42	0.5	T34	0.4	0.58
T11	0.8	0.5	T35	1	0.5
T12	0.58	0.58	T36	1	0.5
T16	0.8	0.45	T37	0.92	0.58
T17	0.25	1	T38	0.3	1
T18	0.7	0.7	T39	0.3	1

According to Pearce et al. (2005) these values are a valid indicator of flow or immersivity. The 'optimal line' of immersivity is shown in the graph of Figure 7 where boredom and anxiety are indicated at the two extremes of the graph (ibid). However, boredom is not a particularly accurate descriptor as the students were very positive in their reflection reports in all the tasks. A more appropriate suggested descriptor is therefore 'disengaged' although the students admitted to much anxiety when tasks were deemed very challenging.

The collated data included total challenge and skill values, CTC values and RTC values. To re-iterate, in order to compare the data from all tasks it was necessary to scale challenge and skill to values between 0 and 1. In each task we divided the sum of the challenge values provided by the students by the maximum score possible. Similarly, we divided the sum of the skill values provided by the students by the maximum score possible. For the CTC values we took the maximum CTC value and divided it into each CTC value. Similarly, for the RTC values we took the maximum RTC value and divided it into each RTC value. All values are thus represented between 0 and 1. This allows us to represent the data graphically and thereby determine the immersion in the case of challenge and skills and TF (see below) in the case of CTC and RTC values.

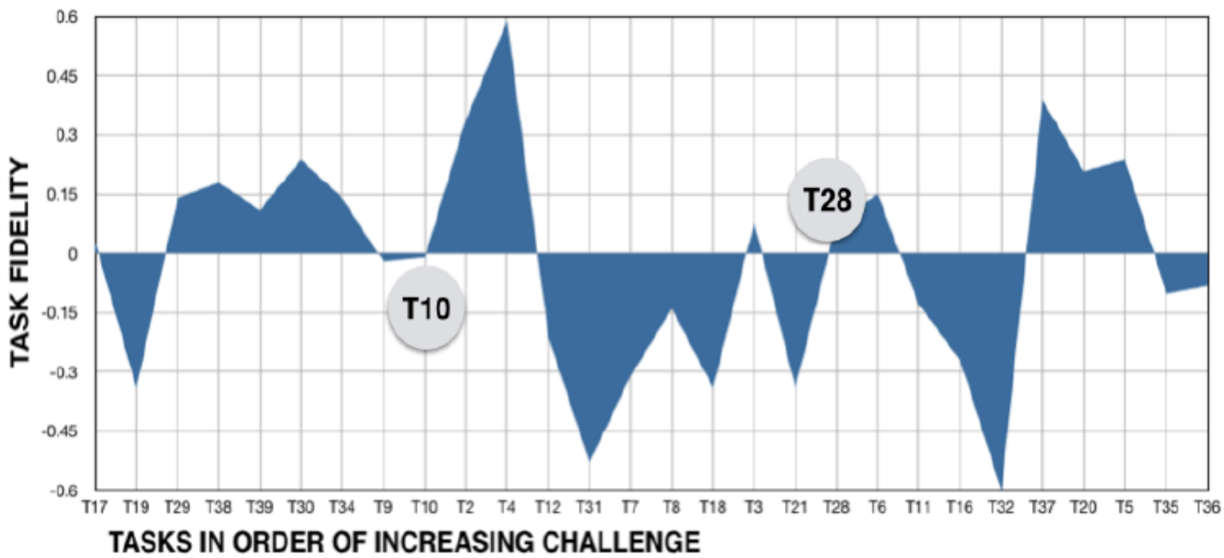
As a result, the complexity of the task could be quantified by a new metric which we term TF. For example, from the data discussed below, the graph in Figure 5 of CTC versus RTC reveals the plotted differences in the researcher's (in the role of instructor or teacher) expected level of complexity (i.e., the CTC) and the students' achievement (i.e., the RTC). As CTC is increased (see Figure 5), one might expect the two plotted areas to merge; in other words, the researcher (or teacher) has provided a task commensurate with the expected successful outcome that is likely to be developed by the learners. We can also represent numerically the differences between anticipated task complexity and successful accomplishment, which we refer to as TF.

Figure 5 Graph of CTC – RTC in order of task challenge



TF is defined as an indicator of the complexity of the circuit compared with the complexity of the program created to complete the circuit. The Y axis in Figure 6 indicates TF; or task complexity where zero is the ideal state. Data above the zero line indicate that the robot program was more complex than the circuit the robot had to manoeuvre. Data below the zero line indicate that the circuit was more complex than the optimum robot program required to successfully navigate it.

Figure 6 TF in order of increasing challenge

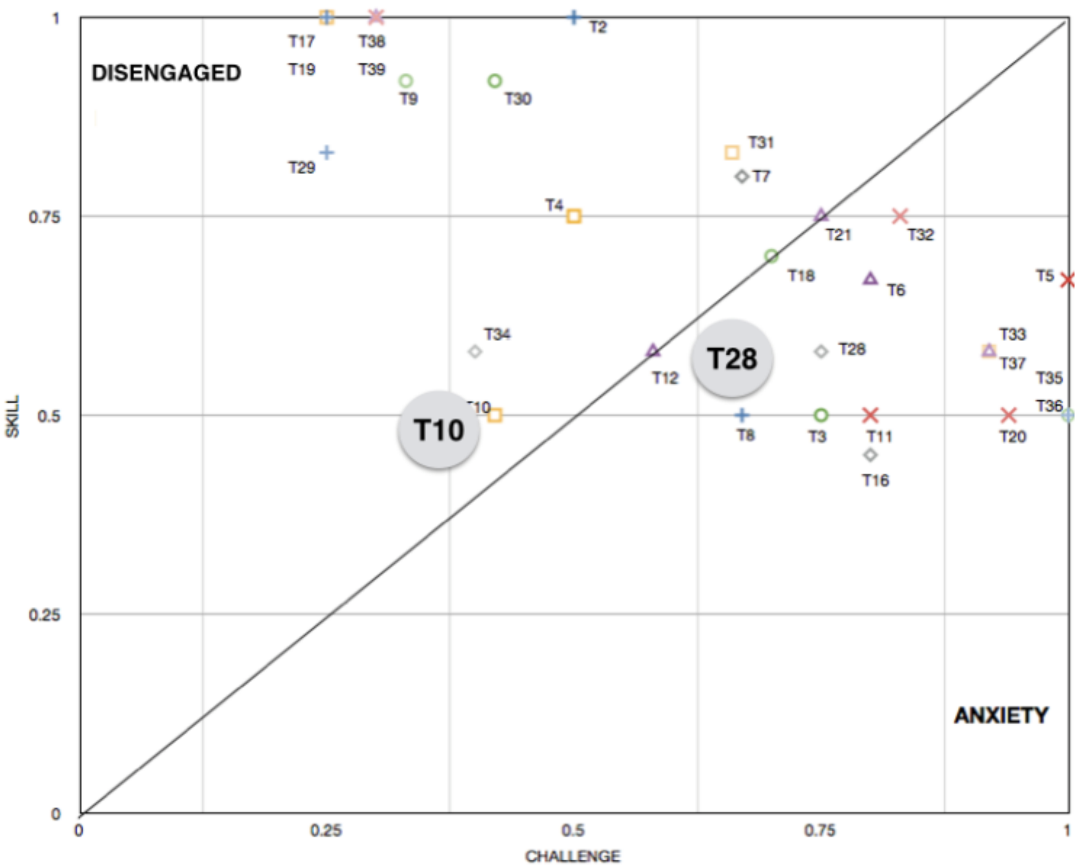


Discussion

In our research numerical values for task complexity (as determined by the 'teacher') and task solution (as determined by the 'student') have been calculated. The difference between teacher task complexity and student task solution has been calculated as TF. TF was calculated by subtracting the RTC from the CTC. We plotted TF against the order of increasing challenge as determined by the students. Ideally, one might anticipate that TF should be zero. The tasks above the zero line indicate that the teacher's task value (of CTC) was higher than the students' solution value (RTC). In other words, the parts of the graph above the Y-axis zero value in Figure 6 reveal that the task was difficult or that the students did not meet an anticipated solution. The tasks below the zero line indicate that the students' solutions exceeded the teacher's expectations of task complexity. Ideally, the teacher should provide a task commensurate with the expected successful outcome to be developed by the learners. If that was the case for all our tasks then TF would be zero and a horizontal line plotted in Figure 6.

The data represented in Figure 6 reveals that for most tasks the programming required to complete the circuits was less than the considered complexity of the circuit (i.e., most data points are below the zero line of TF). This appears to be the case across the range of challenges faced by the students. This reveals that students mostly found the tasks easier than expected. The tasks above the zero line (T2, T4, T5, T20, T30 and T37) all have one factor in common: they all necessitated the use of sensors. For example, TF value for T28 was only + 0.08; slightly above the optimal TF line and slightly below the optimal immersivity line; similarly for T10, with immersivity slightly above optimal path and TF at +0.01. Students found tasks that involved sensors most difficult and they were also more anxious during these tasks as we will see from our immersion data below. T10, T28 and T36 had small TF values so these would be considered the ideal tasks of those undertaken. These tasks also involved the use of sensors. It is acknowledged that the use of sensors contributes to the proficiency of the program, but it needs to be recognised by educators that, for the students, the inclusion of sensors represents additional demands. Combining the observations of data of Figure 6 and Figure 7, the data for TF and immersivity suggest that T10 and T28 could be considered the most successful tasks when students are engaged in robot mediated interactions because the TF value for T28 was only + 0.08; slightly above the optimal TF line and slightly below the optimal immersivity line. Similarly for T10, with immersivity slightly above the optimal path and TF at +0.01.

Figure 7 Students' immersivity



So even though sensors were used in the task and even though students reported that they found sensor related tasks difficult, being immersed in a task with sensors led to greater student success. An examination of the screen capture videos of communication by avatars in-world revealed that the UK students used more procedural language and confirmation questions, whereas the Japanese students offered only instructional language with no checking for understanding. This might suggest that the UK students were more adept at giving instructions than the Japanese students and is a feature reported in Vallance and Martin (2012). This may have more to do with normative cultural communication strategies than English fluency and is a factor which all international collaborations must consider.

The challenge for researchers and teachers is to seek tasks similar to T10 and T28 where immersivity is close to or on its optimal path and task complexity is close to or on the optimal line of TF. Another major challenge is to seek ways to transfer our understandings of this to the creation of further tasks with different participants so that we may develop more reliable optimal learning tasks for RMIs. For instance, tasks could be designed around simulations such as 'search and rescue' where the students are required to collaboratively program robots to solve specified scenarios. The metric for optimal task complexity can be developed by experienced engineers working with educators. From the students' task solution the RTC value can be calculated. TF for all participants can then be calculated.

Limitations of the study

The number of participants (N = 16) is too low to generalise our findings at present, although we argue that it is sufficient to demonstrate proof of concept. This paper therefore presents a rationale plus an implementation of proposed metrics for tasks involving robots that will be of use to educators in schools and higher education. Although the skills versus challenge data is specific to the participants in this research, it has allowed us to determine how immersed our participants were in each task and concurrently associate their programming success in order to locate optimal task complexity. It is also acknowledged that to more fully contextualise and supplement our development of metrics a more effective method of collecting and collating cognitive data from educational interventions would be helpful, especially the use of psychometric data in order to measure skills, knowledge and achievement and this forms part of the next phase of our work, to be reported separately. As a result of developments and progress to date we will also be recruiting an increased number of participants that may permit more generalised solutions to be proposed.

Conclusions

For educators at schools and in higher education, the teacher seeks to provide a task commensurate with the expected successful outcome to be developed by the learners. To understand if this is happening in practice, the complexity of a task can be compared with the solution developed by the student and we have used robot programming and student collaboration to determine a metric which we have called TF to assist with this. By combining TF data with immersion data we can observe and quantify the usefulness of a task for promoting learning. For example, we found that the programming of robot sensors by the students proved to be more complex than manoeuvring a robot and this was also reflected in the immersion data mentioned above; students were most anxious when engaged in tasks requiring sensor programming and were thus less immersed in the challenge. However, as their skills in sensor programming increased, immersivity increased; as indicated by task 28 where Japanese students were taught by UK students within our 3D virtual space to program the robot's use of light and colour sensors to initiate specific actions. The TF value for T28 was + 0.08; only slightly above the optimal level. The challenge is to seek tasks similar to T28 where immersivity is close to or at its optimal value and task complexity is close to or on the optimal line of TF. Our evidence suggest that this will create better engagement with learning and a greater likelihood that students will succeed in reaching their learning objectives.

To sum up, this applied research is developing metrics for recognising the most effective learning when learners are engaged in collaborative virtual world tasks:

- the motivation to implement this research was the nuclear disaster of 3-11 in Japan: a situation that we had never imagined (Lochbaum et al., 2014)

- a virtual Unity 3D Fukushima nuclear plant and an OpenSim training space have been iteratively designed and constructed

- international collaboration by students as non-experts has highlighted the benefits and challenges posed when engaged in constructing RMIs within the context of distance-based communication in 3D spaces

- students' immersion, CTC, RTC and TF have been calculated
- optimal learning tasks have been identified.

The literature reveals that there is no common consensus about metrics for RTC and associated HRIs. Our proposals for CTC and RTC, alongside appropriate arrangements for immersion, are suggested as ways to determine a new metric for measuring tasks involving robots, which we have termed TF.

We are continuing with our work to implement these metrics in diverse robot scenarios within our 3D virtual space involving synchronous collaboration between students in Japan and UK. We will attempt to overcome the current limitations and will publish further findings in the future.

References

Barker, S.B. and Ansorge, J. (2007) 'Robotics as means to increase achievement scores in an informal learning environment', *Journal of Research in Technology and Education*, Vol. 39, No. 3, pp.229–243.

Csikszentmihalyi, M. and Nakamura, J. (2010) 'Effortless attention in everyday life: a systematic phenomenology', in Bruya, B. (Ed.): *Effortless Attention: A New Perspective in the Cognitive Science of Attention and Action,* pp.179–189, MIT Press, Cambridge MA.

Gaudiello, E. and Zibetti, E. (2013) 'Using control heuristics as a means to explore the educational potential of robotics kits', *Themes in Science & Technology Education*, Vol. 6, No. 1, pp.15–28.

Lochbaum, D., Lyman, E. and Stranahan, S.Q. (2014) *Fukushima. The Story of A Nuclear Disaster*, The New Press, New York.

Mima, N. (2003) 'Online technology is not enough: transforming the teacher-student learning process', in Bachnik, J.M. (Ed.): *Roadblocks on the Information Highway*, Lexington Books, Oxford.

Murphy, R. and Schreckenghost, D. (2013) 'Survey of metrics for human-robot interaction', *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, Tokyo, pp.197–198.

Olsen, D.R. and Goodrich, M.A., Metrics for Evaluating Human-Robot Interactions [online] http://icie.cs.byu.edu/Papers/RAD.pdf (accessed 14 March 2009).

Pearce, M., Ainley, M. and Howard, S. (2005) 'The ebb and flow of online learning', *Computers in Human Behavior*, Vol. 21, No. 5, pp.745–771.

Rae, L., Takayama, L. and Mutlu, B. (2013) 'The influence of height on robotic communication products', *Proceedings of the 8th ACM/IEEE Conference on Human-Robot Interaction (HRI 2013),* Tokyo, Japan.

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A. and Goodrich, M. (2005) 'Common metrics for human-robot interaction', *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-Robot Interaction*, pp.33–40.

A situation that we had never imagined, *UNESCO, ICT in Education*: Japan [online] http://www.unescobkk.org/index.php?id=1381 (accessed 5 August 2007).

Vallance, M. and Martin, S. (2012) 'Assessment and learning in the virtual world: tasks, taxonomies and teaching for real', *Journal of Virtual Worlds Research*, Vol. 5, No. 2, pp.1–13.

Vallance, M., Naamani, C., Thomas, M. and Thomas, J. (2013) 'Applied information science research in a virtual world simulation to support robot mediated interaction following the fukushima nuclear disaster', *Communications in Information Science and Management Engineering (CISME),* Vol. 3, No. 5, pp.222–232.

Weiss, A., Bernhaupt, R., Lankes, M. and Tscheligi, M. (2009) 'The USUS evaluation framework for human-robot interaction', *Proceedings of the Symposium on New Frontiers in Human-Robot Interaction*, pp.158–165 [online] http://www.aisb.org.uk/convention/aisb09/ Proceedings/NEWFRONTIERS/FILES/Proceedings.pdf (accessed 4 March 2014).