MASTER THESIS

EDUCATIONAL SCIENCE AND TECHNOLOGY

Self-led Written Debriefing as a Scaffold for Serious Games on Individual Learning Outcomes

Shahbaz Javed Qureshi

University of Twente

1st Supervisor: Dr. H.H. Leemkuil

2nd Supervisor: Dr. Bas Kollöffel

27th August 2022

Acknowledgement

I would like to extend my gratitude to Dr. Henny Leemkuil and Dr. Bas Kollöffel for their endless support, patience, and supervision in the making of this thesis. In addition, I would also like to thank Lalababa, my father and my mother for their continuous support, compassion and understanding in undertaking this research. Lastly, I would further like to thank David Goggins, Sam Harris and Marcus Aurelius (yes, the ancient Roman emperor) for being excellent role models.

.

Abstract

Serious Games refer to games that are used primarily to provide educational value. It is suggested that the potential of serious games can be increased when combined with scaffolds such as debriefing. Debriefing is a process of turning the game experience into an intended discussion and analysis to improve learning. This study addresses multiple ways through which debriefing can be occurred but narrows its scope to a self-led written form of debriefing. The serious game in question is A.I. for Oceans – a game about an A.I. bot that teaches the fundamentals of machine learning such as training data & bias, and its impact on society. The study seeks to investigate the effect of self-led written debriefing and serious games on the learning outcomes of learners when compared to those who rely solely on serious games (n=68). The primary conclusion of the study is that respondents within the self-led written debriefing condition showed substantially higher learning outcome than those who relied only on serious games.

Keywords: debriefing, written debriefing, self debriefing, serious games, learning outcomes, self led written debriefing, instructional support

Table of Contents

Introduction

In 2021 alone, the number of US citizens who continue to play videogames is 221 million – approximately two-thirds of the total population (Entertainment Software Association, 2021). The number has increased in the past two years, suggesting COVID and by extension working from home as a direct determinant of the sudden increase (ESA, 2021). Games have a capacity to engage and motivate people (Tobias, Fletcher, Dai and Wind, 2011) and are seen as having a positive effect on learning outcomes which suggests the need for further research in the field of educational science. Subsequently, a serious game is an idea that utilizes the principles of game design where learning is seen as the primary goal as opposed to recreation (Annetta, Minogue, Holmes, & Cheng, 2009; Leemkuil, 2011; Susi, Johannesson, & Backlund, 2007).

An underlying problem with current research is that playing serious games as a standalone tool is not enough to warrant high learning outcomes (Van der Meij, Leemkuil and Juo-Lan Li, 2013). Wouters et al. (2013) mentions that for learning to be made effective, individuals need to consciously analyze and review the events that occur during the gaming process, whereas simply playing the game does relatively little to enforce that. To circumvent this issue, instructional supports are often encouraged (Ke, 2009; Wouters & Van Oostendrop, 2013). There is not enough evidence to suggest that serious games transfer well to learning outcomes if not paired with some form of instructional support (Crookall, 2010; Van der Meij et al., 2013). In this paper, instructional support is defined as an external type of guidance, instruction or assistance that could help the individuals learn (Tobias & Fletcher, 2011). One such type of instructional support is debriefing (Crookall, 2010) that is expanded upon further in the text.

Debriefing is a type of instructional support that enables learners to reflect on the gaming experience (Van der Meij, Leemkuil, & Li, 2013). It can be defined as “a facilitated or guided reflection in the cycle of experiential learning” (Fanning and Gaba, 2007, p. 116). It serves as the processing of a game experience to turn it into a valid learning outcome. Furthermore, it changes the learner’s subjective experience to objective as misunderstandings and mistakes are cleared (Linn, Lee, Tinker, Husic & Chui, 2006; Van Ments, 1983). The debriefing sessions are further used to clarify participants’ knowledge and apply experiences to other subjects (Asakawa & Gilbert, 2003, p. 15).

While debriefing in a conventional setting is often led by an expert, as is very often the case with military and healthcare, it can also be self-led by answering a set of debriefing questions that are constructed by an expert to facilitate guidance and learning (Boet et al., 2011; Fanning & Gabba, 2007; Moreno-Ger Burgos et al., 2008). The aim of the study is to read into the effectiveness of self-led, written debriefing when combined with serious games in terms of learning outcome when compared to a group that relies solely on serious games. The instructional support in this study emphasizes self-debriefing as a means for learners to critically think and reflect on the gaming experience to improve learning outcomes.

Theoretical Framework

Serious Games

Games are entertaining and interactive environments that are based on underlying models in which challenging goals must be achieved based on specific rules and constraints (Zyda, 2005). It can also be defined as a system that allows players to engage in an artificial challenge, that is defined by rules and results in a quantifiable outcome (Salen & Zimmerman, 2004). While a unified agreement on the definition of games between theorists do not exist, but many agree on the characteristics of games: which are based on defined rules of gameplay; provide instant feedback based on players' actions; challenging; and events that build from preceding actions (Mayer, 2014). Plass, Homer & Kinzer (2015) also mention the significance of rewards, events rooted in storyline with rich visual and auditory aesthetics as elements that make a complete game.

Game-based learning is the process and practice of learning by using games or serious games (Wouters at al. 2012). The study further mentions that it can be used to increase engagement, transfer knowledge, learn new skills and gain abstract knowledge. What distinguishes it from serious games is that game-based learning is a learning methodology, and a serious game is a product in which game-based learning is achieved (Wouters et al., 2012). This line of reasoning is further perpetuated by the study done by Leemkuil (2011) that serious games are games that borrow elements from a typical game, but its core objective rests on learning as opposed to recreation. It can be a custom-built product that fits a specific learning objective and can hold various benefits for education such as higher motivational, immediate feedback based on actions, and the risk-free environment where individuals are free to experiment through trial and error (Bakker et al., 2015; Garrit et al., 2002; Malone, 1981). Sitzmann (2011) also proposes further benefits that include learning outcomes such as self-efficacy, declarative knowledge, procedural knowledge, and retention.

One of the problems of learning through serious games, especially among novice learners is its reduced ability to articulate knowledge beyond the complex gaming environment (Leemkuil & de Jong, 2011). As a result learners find themselves trying to only solve the game as opposed to truly learning the concepts presented in the game. It was also found that processing information through reactive means by trial and error in the absence of reflection produces shallow learning that is devoid of reflection (Hickey, Kindfield, Horwitz, & Christie, 2003; Koops & Hoevenaar, 2012). Furthermore, considering the amount of information present in the game, it can be easy for learners to be overwhelmed and focused on irrelevant information and processes (Wouters & Van Oostendorp, 2013). Wouters (2013) describes several types of support that can mitigate this issue with one of them being debriefing~~.~~

Debriefing

Debriefing involves a facilitated, systematic process of reflection where knowledge gained during a learning experience such as simulation or a game is synthesized (Eppich & Chang, 2015). It is a process that occurs after the experience through a purposive discussion of that experience which is mostly led by a facilitator (Lederman, 1992, p. 146). For the debriefing of educational gaming simulations, the learning is attributed largely in part of debriefing as opposed to the game itself (Crookall, 2010). Lederman (1992) regards debriefing as a postexperience analytical process which is the heart of learning experiences.

On top of debriefing being widely considered as a critical component in the process of experiential learning (Fanning & Gabba, 2007; Kolb, 1984; Koops & Hoevenaar, 2012; Kriz, 2010; Lederman, 1984, 1992), it allows students to make the intuitive knowledge gained through simulation or games more explicit (Leemkuil & de Jong, 2012). Consequently, it allows students to compare the simulated reality with their own frames of reality that entails applying said knowledge to the real world (Kriz, 2003; Peters & Vissers 2004). It is often characterized as a set of sequential stages with descriptions or example questions (El-Shamy, 2001; Lederman, 1992; Lennon, 2006; McGaghie, Issenberg, Petrusa, & Scalese, 2010; Petranek, Corey, & Black, 1992; Sims, 2002; Steinwachs, 1992; Thiagarajan, 1992; Van Ments 1983; Vollmeyer & Rheinberg, 1999) that guides the learner to revisit experience. Several authors claim that the root for debriefing comes from Kolb's (1984) experiential learning cycle (Fanning & Gaba, 2007) where four phases are identified and separated between concrete experiences, reflective observation, abstract conceptualization, and active experimentation. The motivation for debriefing in this study also holds weight from the experiential learning cycle of Kolb. In the experiential learning cycle, the learner engages with the experience and learns from it through reflecting on the experience (Gardner, 2013). Afterwards, the learner tries to apply the knowledge gained from the experience into real life and adapts it into different situations followed by a reiteration of the learning cycle (Kolb, 1984, as cited by Gardner, 2013).

In the case of simulations or serious games, debriefing is considered as one of the most important phases in learning (Crookall, 2011). The study also mentions that a major reason why is to draw out and explicitly articulate the connections between experiences that learners have gained from playing the game or simulation into real-life situations following from the previous example. Furthermore, as per the same study it can also be used to identify the learning gaps between the actual performance in game and the learning objectives. A possible way this can be done is by identifying mistakes or discussing alternative actions through intermediary debriefing (Peters & Vissers, 2004).

Lederman (1992) identifies several elements that make up the process of debriefing which are the debriefers, the participants who are expected to debrief, the experience, the effect of experience, revisiting experience, articulating thoughts from the events and the time spent. To further expand, the debriefers are the facilitators who prompt questions to participants who engage in the learning experience; the experience is the learning experience such as serious game or a virtual simulation; the effect of experience describes the experience itself that the game has had on the learner. After having played the game, learners are expected to revisit the experience through recollection and reflection followed by reporting of said experiences in the debriefing session. Conventionally, the reporting is often done orally and collaboratively to the facilitator or an expert present (Lederman, 1992; Petranek et al., 1992). However, debriefing can also be done in a self and written manner that provides distinct benefits (Petranek, 2000) in the absence of a facilitator.

Self Debriefing

In educational settings, the cost for having an expert lead the debriefing may cost a lot of resources which allows for self-debriefing as an alternative (Van der Meij et al., 2013; Verkuyl et al., 2018). Self-debriefing offers opportunities to enable outcomes of learning from playing serious games by decreasing the number of required faculty debriefers and collaborators (Verkuyl et al., 2010). There is evidence to suggest that self-debriefing strategies are equivalent to facilitator-led debriefing in some situations in terms of learning outcomes (Moreno-Ger et al., 2008). Van der Meij et al., 2013 further proposes that some of the benefits of self-debriefing over its collaborative counterpart includes convenience, and cheaper faculty and institutional resources. Furthermore, some studies also claim that self-debriefing improves individual's self-assessment and evaluation – a skill that is heavily beneficial for continuous learning in healthcare services (Boet et al., 2011; Oikawa et al., 2016). As far as its application goes, it also makes for a good fit with virtual single-player simulation where learning experience is uniquely personal and can be completed outside of the formal educational setting and as such holds considerable scope for this study.

One of the ways through which self-debriefing can be done in the absence of a facilitator is if the questions are constructed using the 3D model of debriefing that entails defusing, discovery and deepening of the experience (Zigmont, Kappus, & Sudikoff, 2011). The study further mentions that defusing is designed to help the individual articulate the experience for clarity, discovering focuses on ‘reflective observation’ and ‘abstract conceptualization’ to help learners develop mental models that can be tested during ‘active experimentation’ and deepening helps to move learners to potential changes in practice with greater context.

Written Debriefing

Petranek (2000) suggests debriefing to be employed in a written form. One of the arguments is that it allows more room and private time for learners to reflect upon their experiences. Moreover, it addresses learners own personal feelings and insights that is particularly favourable to participants who normally remain silent during group discussions (Vries, Van der Meij., Boersma & Pieters, 2005; Oertig, 2010). Furthermore, it involves learners to explicitly articulate that can be effective in revealing misconceptions and makes it easier for facilitators and participants to evaluate their writing in formative or summative way to assess what the participant has learned from the experience. Van der Meij et al. (2013) suggests that it can be structured with supports such as concepts, suggestions on or leading questions on paper which are then open for self-reflection, discussion, or feedback given more leverage on the time spent as opposed to oral forms in collaborative debriefing.

Written debriefing makes it convenient for the participants to revisit gaming experience on a level that is personal to examine their thought processes, feelings, and statements (Petranek, 1992). Author further mentions that it harnesses the energy spent on simulation and idea forming into a form of a structured, coherent essay. An added benefit of writing offers individuals a better approach to articulate information (Wollman-Bonilla, 1989). This is further supported by Hughes et al. (1997) which states that when having to deal with complex issues or concepts, writing enables students to sort, clarify, organize, and personalize their learning. Moreover, Irmscher

(1979) stresses that since education is concerned with learning and applying learned information to other contexts, then writing serves as the way forward in terms of learning and development.

Learning Outcomes

A study conducted by Wouters et al. (2009) proposes that learning outcomes entail four categories. These are cognitive, motor, affective and communicative skills. For this paper, learning outcomes narrows its scope only to the cognitive dimension of learning. Study further claims that cognitive learning outcomes is split into knowledge and cognitive skills. In this part, knowledge is related to the encoded knowledge that reflects text-oriented learning such as verbal knowledge and non-text oriented learning such as the type of knowledge that comes from an image. Moreover, the study claims that a cognitive skill is related to more complex cognitive processes in which learners reapply their knowledge from the learning experience into new and novel situations.

Wouters, Paas & Van Merrienboer (2008) state that active cognitive process of any educational material or a learning experience is a prerequisite for an effective and sustainable form of learning. This is also supported by other educational researchers make a similar claim that cognitive processing is mandatory to achieve genuine learning (Chi, de Leeuw, Chiu, & LaVancher, 1994; Mayer, 2001; Wittrock, 1974). Moreover, Bandura (1976) states that stronger learning effects are reported when learners engage in active coding. In this respect, a learning environment that fosters active procession of skills such as serious games may foster more effective learning when compared to passive forms of learning such as reading or listening to a lecture. Serious games as a form of learning experience that entails active cognitive procession to solve challenges also replicates application of solving challenges in real life (Tobias, Fletcher, Dai & Wind, 2011).

Research Question and Hypothesis

Question 1: Is there a difference in learning outcomes between learners who participate in written, self-debriefing after playing a serious game and learners who rely solely on serious game?

H_1: Learning outcome will be higher for learners who participate in written, self-debriefing after playing serious games than those who rely solely on serious games.

Methods

Research Design

The Quantitative research method was used to investigate the research question. The experiment was split into two conditions to include an experimental group and a control group. The experimental group included the act of playing a serious game, answering self-led written debriefing questions, and then answering a series of knowledge test questions. In contrast, the self-led debriefing questions was absent in the control group while maintaining the rest of the structure.

Participants

The respondents included a total of 68 anonymous participants distributed randomly between 46 Pakistani and 22 Bachelor of Psychology in the University of Twente students aged 16 and above with no experience in Machine Learning. Participants were assigned randomly to each condition using a randomizing tool present in the survey program Qualtrics split between the control group (N = 40) and the experiment group (N = 28). The Dutch students were recruited through the University of Twente's SONA platform whereas Pakistani students were recruited from a class of foundation medical students from LUMHS University, Jamshoro.

Instruments

A.I. for Oceans.

The serious game used for this study is called A.I. for Oceans. This serious game engages participants to learn about artificial intelligence (A.I), machine learning, training data and bias while exploring ethical issues and its implementation on global issues.

The game asks students to train a bot called ‘A.I’ to classify random objects as either ‘fish’ or ‘not fish’. It also allows participants to break A.I. by giving it bad data, which emphasizes the idea that machine learning is only as good as the data used to train it. For instance, if the player were to repeatedly classify fish as trash, the A.I. then comes to the same conclusion – similar to how the principles of training data apply to machine learning.

In an example of a given level, the user/participant is provided with images of fish where they are expected to provide A.I. data on what constitutes a ‘circular’ fish in terms of their shapes, colours, and bodies (see Figure 1.) The emphasis is drawn on the idea that the term ‘circular’ is purely subjective that leads A.I. to provide output that is based on human’s subjective perception leading to biasness. It provides further examples through videos and in-game text of how biasness may hold real-life negative implications when not corrected.

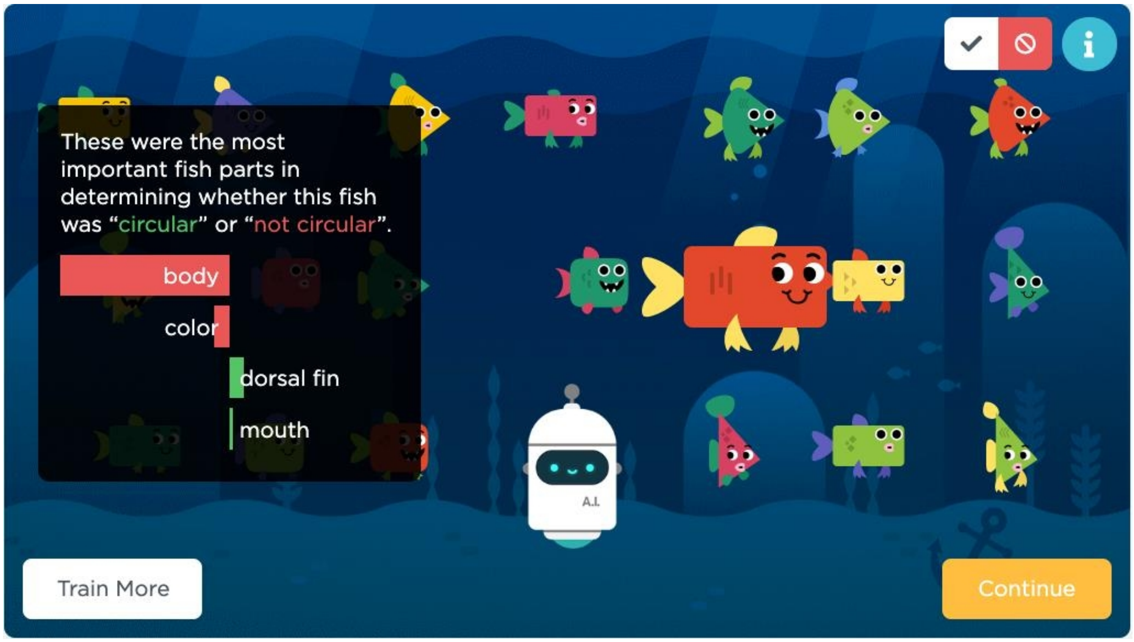


Figure 1: *AI for Oceans Training Data Characteristic View*

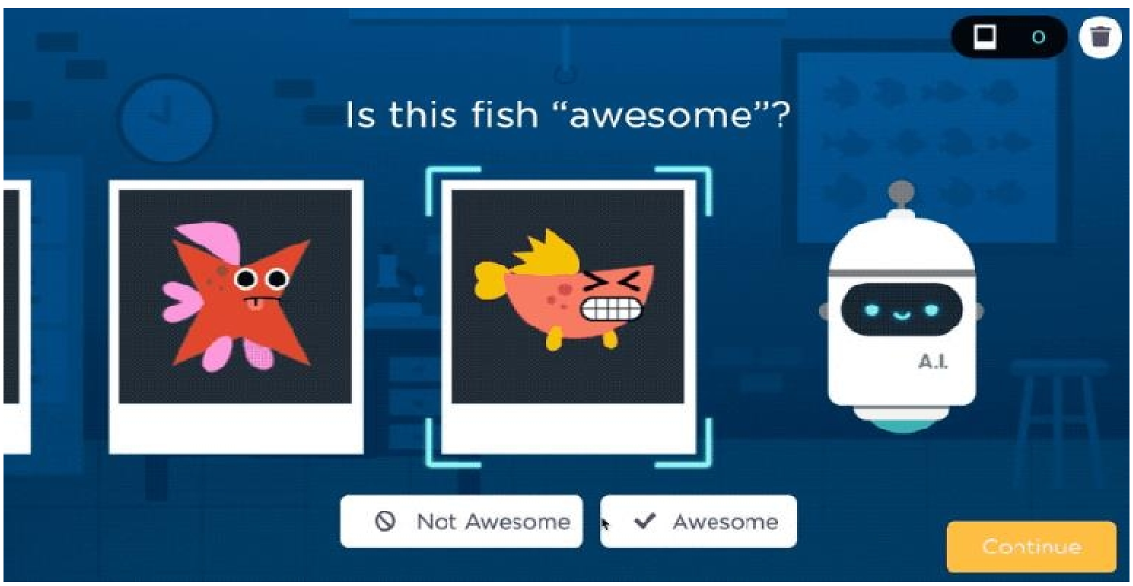


Figure 2: *AI for Oceans Training Data*

Knowledge Test

The knowledge test (n=10 questions) measured knowledge about game concepts, "What is Machine Learning, and how is it similar to the way humans learn?", principles "Consider a student who wants to train AI to recognize an angry fish. She wants to accomplish this by selecting any fish with eyebrows that point inwards and frowning mouths. Can AI figure out what the student is doing without being explicitly programmed? If so, how?", and heuristics "Imagine you are a fan of rock music. How would you get the YouTube algorithm to recommend you more rock music? What specific actions would you make and what effects are expected?" Conceptual questions ask for definitions or descriptions of phenomena; principle questions inquire about the ways in which events and actions influence each other, as well as influencing outcomes; and heuristics questions refer to the coherence between the various principles of the game (See Appendix C for all examples.)

The structure of the questions is open ended. In addition, question 7 and 8 are multiple choice with a section underneath for participants to justify their answer over the other option. For example, "Which is true about biases in the context of machine learning? Please explain why your choice is better than the other one. A. Its occurrence is dependent on how data is collected, who is doing the collecting, and how the data is fed, or B. Computer automatically creates biases which is not based on human input." and "When the objective of the AI is to detect an apple, which of the actions do you think is better? Please explain why your choice is better than the other one. A. Have users select apple from a list of random objects. B. Have users select objects that resemble apple in terms of their color, body, and shape."

Self-led Written Debriefing Guideline

The self-led written debriefing questions included a set of ten open-ended questions. It is preceded with an introductory text to explain participants how to answer by recollecting their experiences, examining their actions in game, and thinking about the possible strategies and future applications. Questions are structured and organized around the contents, actions and interactions present in the game. Moreover, it also prompted participants to think about possible interactions that they may have missed within the game and to describe what may have happened if they had chosen a different answer (See Appendix B.)

Qualtrics Survey Form

The Qualtrics Survey Form is an online data collection platform that enabled participants to input answers/data for researchers to analyze. The contents of this platform for this study includes five blocks that include two screening questions; one introductory/consent form; one external link to the serious game; one set of ten self-led debriefing questions; and one set of ten knowledge-test questions. The arrangement of blocks is structured sequentially as above with one key difference being that the control group does not receive the self-led debriefing questions which is possible through the randomization tool in the Qualtrics software.

Procedure

The online experiment was first published on SONA with an intention to recruit bachelor students from the BMS department of University of Twente for Dutch students. Once published, then the participants were expected to click on the study/link on SONA which redirected them to a Qualtrics Survey Form. For Pakistani students however, the link for the Qualtrics Survey Form experiment was passed manually to a class of first year medicine students through a professor at LUMHS University. The rest of the procedure is similar for students of both countries. After accessing the link, the participants were automatically placed in a random condition for the study. Both the conditions started with a set of two screening questions to confirm whether they were at least 16 years old or older, and whether they had any experience in machine learning. If the respondents answered no on ‘above 16’ or yes in ‘experience in machine learning’ questions, then Qualtrics booted them out of the survey as it did not match the criteria for the study.

Having passed the screener, the participants were then fronted with an introductory text to explain the motive for study, degree of user anonymity, introduction of the researcher and an option to consent. The game booted participants out of the survey if participants did not consent. Once consented, the participants were then instructed to play the game by clicking on an external link underneath the main text. After having played the game and depending on the condition that they had been randomly assigned, participants were then presented with either one of the two blocks – a set of knowledge test for the control group or a set of self-led written, debriefing questions in addition to a set of knowledge test for the experimental group.

Participants within the experimental group were then expected to fill out ten questions within the debriefing section and click next after to access the knowledge test questions. The survey ended after participants finished the knowledge test for both conditions.

Data Analysis

Answers from open-ended questions were measured against three criteria –namely: Excellent (2), Satisfactory (1) and Unsatisfactory (0). Excellent Criteria was coded using the numeric 2 that represents that response entail all required elements and including additional elements that add to the answer. Satisfactory Criteria was coded using the numeric 1 that represents that one/some required elements were missing but additional elements that added to the answer were present (e.g. thoughtful comments). Unsatisfactory Criteria was coded using the numeric 0 that represents several required elements were missing or heavily plagiarized – the former of which was analyzed through a quick google search. All responses within the Excellent, Satisfactory and Unsatisfactory Criteria entail 2, 1, and 0 points, respectively (see Appendix D). An inter-rater reliability between two raters showed a mean intercoder agreement of 90% for the first three knowledge test questions. This was after several differences were discussed and adjusted accordingly.

Internal reliability was checked by looking at Cronbach's alpha for the two conditions: knowledge-test control group ($\alpha = 0.747$) and knowledge-test experimental group ($\alpha = 0.678$). The difference of average scores between both the conditions were further compared using an independent sample t-test.

Before running the independent samples t-test, Levene's test was used to check for homogeneity in variances within each scale that showed a value of $F(66) = 0.302$, $p = 0.006$.

Results

An Independent samples t-test results showed that there was a statistically significant difference between the experimental and control group. On average, the debriefing condition scored higher ($M = 8.64$, $SD = 4.56$) than the control group ($M = 5.65$, $SD = 4.07$) in the knowledge test $t(66) = 2.84$, $p = .006$ (see Table 1).

Table 1

Mean and Standard Deviations of Knowledge Test Scores for Both Groups

	Experiment Group			Control Group		
	N	M	SD	N	*M*	*SD*
Knowledge Test	28	8.64	4.556	40	5.65	0.644

Discussion

This study investigated the effect of learning outcomes on individuals who participated in serious games, and the students who participated in self-led written debriefing in conjunction with serious games. The main research findings, implications and limitations of the study are discussed.

Based on the research findings, using self-led written debriefing as a scaffold for serious games showed better results when compared to its control group. These results were consistent with previous studies' claims that instructional support such as debriefing holds promise in individual learning outcomes (Crookall, 2010). The value of self-debriefing in having learning significant outcomes was also in line with study conducted by Verkuyl et al. (2019) albeit it falls short in being completely parallel to this study as the scope of this study does not include self-debriefing with small group and a bigger group. Furthermore, it was also consistent in line with study conducted by Boet et al., (2016) and Oikawa et al., (2016) that self-debriefing holds merit. In addition, studies that support debriefing as a strong instructional support for learning to be made effective was also consistent with several studies (Garris et al., 2002; Hays, 2005; Lederman 1992).

It is assumed that most of the benefits that individuals being able to reflect on the game by articulating thought processes through writing (Van der Meij et al., 2013), having enough time to think privately to make 'sense of it all' (Petranek, 2000), being able to draw connections from the game experience and being prompted to think about applying the learning objectives of the experience in real-life situations (Crookall, 2011). Students participating in the experimental group also benefited from learning by doing, in the sense that writing about their experience constantly kept them engaged which has its roots in merits from reflection on action (Schön,1992). The written portion of this experiment in particular encouraged individuals to harness abstract thoughts and emotions while playing the game into coherent sentences under the pretense that there were no right or wrong answers. This is further amplified by the assumption that debriefing the gaming experience was not a mentally strenuous exercise as opposed to answering questions in the knowledge-test which measured performance – latter of which holds some ground for stress (Zunhammer at al. 2013). Writing also seemed to have invited more articulation that helped in clearing out misconceptions, a claim supported by the studies of Van der Meij et al. (2013) and Petranek (2000). It is also assumed that individuals who were particularly shy also benefited from this method as it afforded them the privacy of typing out answers to debriefing guidelines which did not involve having to participate with others (Petranek, 2000). These findings for self-led, written debriefing in conjunction with serious games also provide good news in terms of letting individuals learn individually with no collaboration as the study done by Van der Meij et al. (2011) shows that communication with others was not needed to boost gameplay and learning. Another reason why could also be attributed to the claim that instructional support focuses on learning beyond just solving problems inside the gaming environment in terms of domain specific knowledge and skill. (Ke, 2009; Leutner, 1992).

With the advent of modern web-based technology, it increases the extends the scope of debriefing to go beyond conventional methods due to rapid progression in assessment delivery, assessment content, and in many different fields (Tippins, 2015). A cloud-based decentralized for instance server can allow students to submit responses from their debriefing guidelines present in-game into an open repository of responses. This holds implications for better learning as students are allowed to publish, view, interact, and compare their self-led, written, debriefing responses with other students' responses. In addition, this may also potentially bridge the benefits of both individual and collaborative debriefing as it allows individuals to bounce off other respondents' ideas virtually to strengthen their own understanding of the learning objective of the serious game. While some studies suggest that collaborative debriefing does not pose any significant advantage over self-debriefing while playing a serious game in terms of learning outcomes (Van der Meij et al., 2013), the former still adds different benefits that could as a whole benefit learning outcomes through different means such as fostering team-learning, and learning from other respondents' experiences (Kriz, 2010).

In addition to individual respondents benefitting from this technology, teachers, or researchers responsible for creating such self-led debriefing guidelines can also submit their debriefing questions to the repository and learn by studying how other teachers/researchers created their debriefing guidelines (Cohen et al., 2013). To further utilize debriefing within the serious game, game-developers may add a 'question-editor' executable file that comes with the game to allow teachers to tailor or personalize debriefing questions in game for one or more students to encourage unexplored means and modes of self-led debriefing under different frameworks.

Furthermore, psychologists and educational scientists may also develop certain metrics in theory within the serious game or simulation to collect data from students to evaluate and strengthen debriefing. For example, by developing an eye-tracking tool, it can help understand which objects in-game students focused most on (Argasinski et al., 2017). An algorithm may then allow the debriefing section to include AI generated questions like, "Why were you focused on 'x' object for so long as opposed to other objects?" Other similar tracking tools include time monitoring per question and recording keystrokes among more. Users can also have an option to upload this data to the open repository for multidisciplinary experts to learn from to improve upon debriefing. Ultimately, the field of knowledge that goes into creating such a system this extends the field of educational science from a stand-alone subject to one that becomes multi-disciplinary.

Another benefit this technology provides is a formation of an online platform for further discussion on self-led, written debriefing inside a serious game. A lot of videogames like Dark Souls, Diablo 2 and Witcher 3 have dedicated communities that center around players discussing key aspects of the game like tutorials, interactions between actions and specific walkthroughs on overcoming an obstacle on websites such as Reddit. Platforms like these are multifaceted and have rich environments with complex dynamics that induce feelings of belongness, cognitive challenges and peer tutoring (Gandolfi et al., 2021) Serious game developers who keep ingame debriefing in mind can then also leverage this idea to create a similar online structure for teachers, researchers, and students alike to discuss characteristics of the game – including debriefing guidelines. They can then further also be catered to specific disciplines such as educational and behavioural scientific research to encourage scientific discussion on different techniques and methods that can improve self-led, written debriefing and player experience on their learning outcomes.

All in all, having an online sphere also draws implications for schools to cut back on cost of physical and human resources by sharing digital copies of the serious game with self-led, written debriefing characteristics in it for students to work privately at their own leisure without the need of an external facilitator. This is because of the ubiquitous nature of online platforms, repositories, and digital softwares (Tranos E & Ioannides YM, 2021) such as a virtual serious game.

Moreover, this also prompts a discussion on what combination of debriefing when paired or supplemented with another mode or scaffold causes the highest learning outcome. This includes, but is not restricted to self-led, written debriefing versus small group, a large group across different modes such as oral or a variation of two including a long-term summative analysis to test retention. Furthermore, this can be leveraged through the addition of online tools such as ingame customizable self-led, written debriefing guidelines, online repository, data metrics and communities.

While the experimental group performed better in learning outcomes for Pakistani and Dutch subgroups, a clear distinction in the quality of answers was observed between the two. It reflects that better data collection methods can be used to gain more insight into future studies. Some responses showed plagiarized, repetitive and joke responses in the former subgroup that reflected a lack of incentive on their part to carry out the experiment. The plagiarism was detected through a quick google search and answers being unnecessarily detailed. Furthermore, the study being anonymous also may have contributed to it since there was no accountability suggesting the need for a non-anonymous and moderated research if the study were to be replicated for better results. As far as the distribution of the experiment goes, this study relied on Qualtrics through a randomization tool to equally distribute participants between experiment and control group. However, the numbers of respondents in both the groups skewed heavily to the control group suggesting a room for error in the randomizing process. Furthermore, when contacted by the person who distributed the experiment link to Pakistani students, it was found that a lot of participants experienced electrical and internet outage at the time of participation that might potentially explain incomplete answers. This goes against the principle of authentic debriefing sessions (Fanning & Gaba, 2007) that emphasizes on a conducive environment for debriefing. In comparison, students at the University of Twente were given incentive by gaining credit points upon completion of the experiment through based on the quality and completeness of their answers at the discretion of the researcher.

In addition, a further look into how self-led written debriefing in conjunction with serious games may affect students' learning outcome from different cultures may help get more insight into how different cultures – particularly those where independent thinking is encouraged versus those where teacher authority laid forefront. In addition, measuring student experience with learning from different platforms such as on a desktop computer vs a traditional pen and paper route may also be worth evaluating when conducting self-written debriefing.

Conclusion

Overall, when comparing the average scores of the experiment and control group showed that even with disparities in the quality of answers among Dutch and Pakistani subgroups, the self-led written debriefing condition still showed higher learning outcomes. Further points were included for future studies such as the potential of customizability and leveraging modes of self-led written debriefing in an online sphere to promote multidisciplinary collaboration for strengthening this scaffold to promote higher learning outcomes. Furthermore, a non-anonymous and moderated study was encouraged for better quality of answers – especially when no other incentives are provided.

References

Annetta, L. A., Minogue, J., Holmes, S. Y., & Cheng, M.-T. (2009). Investigating the impact of video games on high school students' engagement and learning about genetics. *Computers & Education, 53*, 74-85. doi: dx.doi.org/10.1016/j.compedu.2008.12.020

Argasinski, J.K., & Grabska-Gradzinska, I. (2017). Serious Games Evaluation Using eye-tracking and Affective Computing Techniques. Case Study of OSH Training Course. *ACHI 2017*.

Asakawa, T., & Gilbert, N. (2003). Synthesizing experiences: Lessons to be learned from internet-mediated simulation games. *Simulation & Gaming, 34(1)*, 10–22.

Boet, S., Bould, M. D., Bruppacher, H. R., Desjardins, F., Chandra, D. B., & Naik, V. N. (2011). Looking into the mirror: Self-debriefing versus instructor debriefing for simulated crises. *Critical Care Medicine*, *39(6),* 1377–1381. http://dx.doi.org/ 10.1097/CCM.0b013e31820eb8be.

Cohen, Anat & Kalimi, Sharon & Nachmias, Rafi. (2013). The Use of Digital Repositories for Enhancing Teacher Pedagogical Performance. *Interdisciplinary Journal of E-Learning and Learning Objects. 9*. 201-218. 10.28945/1926.

Crookall, D. (2010). Serious Games, Debriefing, and Simulation/Gaming as a Discipline. *Simulation & Gaming, 41(6), 898–920*. https://doi.org/10.1177/1046878110390784

Chiu, Y. h., Kao, C. w., & Reynolds, B. L. (2012). The relative effectiveness of digital game-based learning types in English as a foreign language setting: A meta-analysis. *British Journal of Educational Technology, 43(4).*

El-Shamy, S. (2001). Training games: Everything you need to know about using games to reinforce learning. *Stylus Publishing, LLC*.

Entertainment Software Association. (2021). Number of U.S. Video Gamers Hits 227 Million, and Most Say They've Played More During COVID: ESA Study. Variety. https://variety.com/2021/gaming/news/number-video-game-players-2021-esa-study-covid-1235016079/

Fanning, R. M., & Gaba, D. M. (2007). The role of debriefing in simulation-based learning. *Simulation in Healthcare, 2(2)*, 115–125. http://dx.doi.org/10.1097/SIH.0b013e3180315539.

Gandolfi E, Ferdig RE, Soyturk I (2021). Exploring the learning potential of online gaming communities: An application of the Game Communities of Inquiry Scale. *New Media & Society.* doi:10.1177/14614448211027171

Gardner, R. (2013). Introduction to debriefing. *Semin Perinatol, 37(3)*, 166-174. doi:10.1053/j.semperi.2013.02.008

Hughes, H. W., Koody, M., & Kanevsky, L. (1997). Dialogic refection and journaling. Clearing House, 70, 187-191.

Irmscher, W. F. (1979). Writing as a way of learning and developing. *College Composition and Communication, 30,* 240-244

Ke, F. (2009). A qualitative meta-analysis of computer games as learning tools. In F. R.E. (Ed.), *Handbook of research on effective electronic gaming in education (Vol. 1, pp. 1-32).* New York: IGI Global.

Kriz, W. C. (2010). A systemic-constructivist approach to the facilitation and debriefing of simulations and games. *Simulation and Gaming, 41(5),* 663–680. http://dx.doi.org/10.1177/1046878108319867.

Lederman, L. C. (1992). Debriefing: Toward a Systematic Assessment of Theory and Practice. Simulation & Gaming, 23(2), 145–160. https://doi.org/10.1177/1046878192232003

Leemkuil, H., & Jong, T. d. (2011). Instructional Support in Games. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 353–369).

Leemkuil, Henny & Jong, Ton. (2012). Adaptive Advice in Learning with a Computer-Based Knowledge Management Simulation Game. *Academy of Management Learning & Education. 11*. 653-665. 10.5465/amle.2010.0141.

Lennon, J. L. (2006). Reports & communications: Debriefings of web-based malaria games. *Simulation and Gaming, 37,* 350–356. http://dx.doi.org/10.1177/ 1046878106291661.

Linn, M. A., Lee, H. S., Tinker, R., Husic, F., & Chiu, J. L. (2006). Teaching and assessing knowledge integration in science. *Science, 313*, 1049–1050. http://dx.doi.org/ 10.1126/science.1131408.

McGaghie, W. C., Issenberg, S. B., Petrusa, E. R., & Scalese, R. J. (2010). A critical review of simulation-based medical educations research: 2003–2009. *Medical education, 44(1),* 50–63. http://dx.doi.org/10.1111/j.1365-2923.2009.03547.x.

Moreno-Ger, P., Burgos, D., Martínez-Ortiz, I., Sierra, J. L., & Fernández-Manjón, B. (2008). Educational game design for online education. *Computers in Human Behavior, 24(6),* 2530–2540. http://dx.doi.org/10.1016/j.chb.2008.03.012.

Oertig, M. (2010). Debriefing in Moodle: Written feedback on trust and knowledge sharing in a social dilemma game. *Simulation & Gaming, 41*(3), 374–89. https://doi.org/10.1177/1046878108325982

Oikawa, S., Berg, B., Turban, J., Vincent, D., Mandai, Y., & Birkmire Peters, D. (2016). Self-debriefing vs instructor debriefing in a preinternship simulation curriculum: night on call. *Hawaii Journal of Medicine & Public Health, 75(5)*, 127-132.

Papastergiou, M. (2009). Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. *Computers & Education, 52(1),* 1–12.

Petranek, C. F., Corey, S., & Black, R. (1992). Three levels of learning in simulations: Participating, debriefing, and journal writing. *Simulation Gaming, 23(2),* 174–185. http://dx.doi.org/10.1177/1046878192232005.

Petranek, C. F. (2000). Written Debriefing: The Next Vital Step in Learning with Simulations. *Simulation & Gaming, 31(1),* 108–118. https://doi.org/10.1177/104687810003100111

Peters, V. A., & Vissers, G. A. (2004). A simple classification model for debriefing simulation games. *Simulation & Gaming, 35*, 70–84.

Schön, D.A. (1992). The Reflective Practitioner: How Professionals Think in Action (1st ed.). *Routledge.* https://doi.org/10.4324/9781315237473

Sims, R. R. (2002). Debriefing experiential learning exercises in ethics education. *Teaching Business Ethics, 6(2),* 179–197. http://dx.doi.org/10.1023/ A:1015237928724.

Steinwachs, B. (1992). How to facilitate a debriefing. *Simulation Gaming, 23(2),* 186–195. http://dx.doi.org/10.1177/1046878192232006.

Susi, T., Johannesson, M., & Backlund, P. (2007). Serious games - an overview. *Skövde: University of Skövde (Technical Report HS-IKI-TR-07-001),* 1-28.

Thiagarajan, S. (1992). Using games for debriefing. *Simulation Gaming, 23(2),* 161–173. http://dx.doi.org/10.1177/1046878192232004.

Tippins, N. T. (2015). Technology and Assessment in Selection. *Annual Review of Organizational Psychology and Organizational Behavior, 2*(1), 551–582. doi:10.1146/annurev-orgpsych-031413-091317

Tobias, S., & Fletcher, J. D. (2011, October 19–21). Learning from computer games: A research review. Paper presented at the Serious Games: The challenge: ITEC/CIP/T 2011: *Joint Conference of the interdisciplinary research group of technology, education, communication, and the scientific network on critical and flexible thinking, Ghent, Belgium.*

Tobias, S., Fletcher, J. D., Dai, D. Y., & Wind, A. P. (2011). Review of research on computer games. In S. Tobias & J. D. Fletcher (Eds.), *Computer games and instruction* (pp. 127–221). Charlotte, NY: *Information Age Publishing Inc*

Tranos E, Ioannides YM (2021) Ubiquitous digital technologies and spatial structure; an update. *PLOS ONE 16(4): e0248982*. https://doi.org/10.1371/journal.pone.0248982

Plass, Jan & Homer, Bruce & Kinzer, Charles. (2015). Foundations of Game-Based Learning. Educational Psychologist. 50. 258-283. 10.1080/00461520.2015.1122533.

Salen, Katie & Zimmerman, Eric. (2004). Rules of Play : Game Design Fundamentals / K. Salen, E. Zimmerman ; pról. de Frank Lantz..

Van der Meij, H., Albers, E., & Leemkuil, H. (2011). Learning from games: Does collaboration help*? British Journal of Educational Technology, 42(4),* 655–664.

Van der Meij, H., Leemkuil, H., & Li, J. L. (2013). Does individual or collaborative self-debriefing better enhance learning from games?. *Computers in human behavior, 29(6),* 2471-2479.

Van Ments, M. (1983). The effective use of role-play: A handbook for teachers and trainers. *London, UK*: Kogan Page.

Verkuyl, M., Lapum, J. L., Hughes, M., McCulloch, T., Liu, L., Mastrilli, P., Romaniuk, D., & Betts, L. (2018). Virtual Gaming Simulation: Exploring Self-Debriefing, Virtual Debriefing, and In-person Debriefing. *Clinical Simulation in Nursing, 20*, 7–14. https://doi.org/10.1016/j.ecns.2018.04.006

Verkuyl, M., Hughes, M., Atack, L., McCulloch, T., Lapum, J. L., Romaniuk, D., & St-Amant, O. (2019). Comparison of Self-Debriefing Alone or in Combination With Group Debrief. *Clinical Simulation in Nursing, 37*, 32–39. https://doi.org/10.1016/j.ecns.2019.08.005

Vollmeyer, R., & Rheinberg, F. (1999). Motivation and metacognition when learning a complex system. *European Journal of Psychology of Education, 14(4),* 541–554. http://dx.doi.org/10.1007/bf03172978.

Vries, Bregje & Meij, Hans & Boersma, K.T. & Pieters, Jules. (2005). Embedding E-Mail in Primary Schools: Developing a Tool for Collective Reflection. *Journal of Educational Computing Research. 32*. 10.2190/AM6M-JACJ-PG51-C8UF.

Wollman-Bonilla, J. E. (1989, November). Reading journals: Invitations to participate in literature. *The Reading Teacher, pp. 112-120*.

Wouters, P., Van der Spek, E. D., & Van Oostendorp, H. (2009). Current practices in serious game research: A review from a learning outcomes perspective. *Games-based learning advancements for multi-sensory human computer interfaces: techniques and effective practices* (pp. 232-250). London: IGI Global.

Wouters, P., Van Nimwegen, C., Van Oostendorp, H., & Van Der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology.* doi:10.1037/a0031311

Wouters, P., & Van Oostendorp, H. (2013). A meta-analytic review of the role of instructional support in game-based learning. *Computers & Education, 60*, 412-425. doi: 10.1016/j.compedu.2012.07.018

Zigmont, J. J., Kappus, L. J., & Sudikoff, S. N. (2011). The 3D model of debriefing: defusing, discovering, and deepening. *Seminars in perinatology, 35(2),* 52–58. https://doi.org/10.1053/j.semperi.2011.01.003

Zunhammer M, Eberle H, Eichhammer P, Busch V (2013) Somatic Symptoms Evoked by Exam Stress in University Students: The Role of Alexithymia, Neuroticism, Anxiety and Depression. *PLOS ONE 8(12): e84911.* https://doi.org/10.1371/journal.pone.0084911

Zyda, M. (2005). From visual simulation to virtual reality to games. *Computer, 38(9),* 25-32.

Appendix

Appendix A - Informed Consent

Dear Participant, thank you for taking the time to participate in this study. The principal researcher in this study is Shahbaz Qureshi, an Educational Science & Technology master student. In this experiment, you will be going through a simple game explaining the foundations of Machine Learning and its social impacts, followed by a debriefing guideline and a knowledge test to test your learning outcome.

This experiment investigates the effect of self-led debriefing as a scaffold for serious games on learning outcome. The estimated time to complete the experiment is 30 minutes.

Please note that while completion of the experiment is highly appreciated, you are free to withdraw at any point. You can be assured that **none** of your personal details, including your name, email address and the results of your study will be forwarded to a third party.

For any questions regarding your rights as a research participant, you can contact the Secretary of the Ethics Committee of the Faculty of Behavioural, Management and Social Sciences at the University of Twente through the following email: bms@utwente.nl

If you have questions about the study itself, you can email them to s.j.qureshi@student.utwente.nl.

If you agree to participate, please complete the form below.

- I consent to participating in this research. (*Y/N*)
- Are you older than 16? (*Y/N*)
- Have you had any formal training on Machine Learning? (*Y/N*)

Appendix B - Self-led Written Debriefing Prompts

In this section you are expected to recollect your experience, examine your actions in game, and think about the possible strategies and future applications. Your response will not be assessed and is only used for your own reference and understanding.

1. What in your opinion was the learning objective of the game?

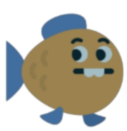
2. Did the A.I. bot identify some non-fish items as fish (and vice versa)? If so, why do you think that happened?

3. Where have you seen or experienced artificial intelligence in your lives?

4. What characteristic (in terms of bodyshapes & colours) of fish did you select for the AI and how did that influence your results?

5. What less obvious word did you select when identifying the fish?

6. What does this icon represent and what in your opinion is its significance? 

7. Did you find it fair to use AI to judge a fish by its looks? While AI may seem fair and neutral, its analysis comes from the training we provide. What unintended bias could this cause? 

8. Did you click on the information icon with the most important fish parts? If so, what did A.I Bot learn was the most and least important feature for making a decision? Do you agree with the bot's judgement?

9. How could you help A.I. Bot improve its decision?

10. How many images did you train with and how many did the A.I. Bot get correct? (You don't have to enter the exact number; an estimate would do.)

Appendix C - Knowledge Test Questions

1. What is Machine Learning,and how is it similar to the way that humans learn?

2. What types of data can machine learning learn from?

3. Why does bias often occur in training data?

4. How do you make sure that the computer labels/identifies the object based on your training?

5. What are some of the elements that make identification of data more concrete?

6. Consider a student who wants to train A.I. to recognize an angry fish. She wants to accomplish this by selecting any fish with eyebrows that point inwards and frowning mouths. Can A.I. figure out what the student is doing without being explicitly programmed?

7. Which is true about biases in the context of machine learning? Please explain why your choice is better than the other one.

   A. Its occurrence is dependent on how data is collected, who is doing the collecting, and how the data is fed.

   B. Computer automatically creates biases which is not based on human input.

8. When the objective of the A.I. is to detect an apple, which of the actions do you think is better? Please explain why your choice is better than the other one.

   A. Have users select apples from a list of random objects.

   B. Have users select objects that resemble apples in terms of their colours, bodies and shapes.

9. Imagine you are a fan of rock music. How would you get the YouTube algorithm to recommend you more rock music? What specific actions would you make and what effects are expected?

10. As a head engineer for the self-driving car company Tesla, your job is to ensure that your car does not hit a human. What steps will you take to teach Tesla to recognize humans from other objects?

Appendix D - Rubric with Examples of Answers

Table 2

Rubric used to Assign Grade Points with Examples for Each Knowledge-test Question.

	Codes		
	Unsatisfactory (0 points)	Satisfactory (1 point)	Excellent (2 points)
Description	**Several required elements are either missing or blatantly plagiarized.**	**One/some required elements are missing but additional elements that add to the answer are present (e.g., thoughtful comments)**	**All required elements are present, and additional elements that add to the answer.**
Example Q1 / What is Machine Learning and how is it similar to the way that humans learn?	Machines learn to predict from large data sizes	it is a computerised programme that can identify patterns without being explicitly programmed to do so. just like humans, it learns better with more data.	Machine learning is when a computer "learns" from training data provided. By showing a machine thousands of pictures of fish, it will eventually be able to spot fish in other images. It is similar to how we learn, because humans also study "training data" to be able to repoduce it on a test. And the result is only as good as the initial input of information.
Example Q2 / What types of data can	Humans provide machine learning	structured data, unstructured data, numbers, text	data from computers, provided training

machine learning learn from? Separate your answers with a comma (like, this,)			data by people, medical images,
Example Q3 / Why does bias often occur in training data?	because of the glitches in training	the sample might not be truly representative of what it actual is	When humans with certain biases train AI, the AI will also have the same biases when predicting outcomes.
Example Q4 / How do you make sure that the computer labels/identifies the object based on your training?	by testing your AI?	By training it more and more till it does	By making the machine practice more and checking whether it is correct or not and if not correct then practice more.
Example Q5 / What are some of the elements that make identification of data more concrete?	I have no idea.	no mistakes, longer training	Continue practice and training of the machine which give it experience
Example Q6 / Consider a student who wants to train AI to recognize an angry fish. She wants to accomplish this by selecting any fish with eyebrows that point inwards and frowning mouths. Can AI figure out what the student is doing without	No it needs to learn.	with enough data it can create its own reasoning	Yes, by detecting the shape of the eyebrows and mouth it can detect an angry fish but only in the way it was trained by the student

being explicitly programmed? If so, how?			
Example Q7 / Which is true about biases in the context of machine learning? Please explain why your choice is better than the other one. a. Its occurrence is dependent on how data is collected, who is doing the collecting, and how the data is fed; b. Computer automatically creates biases which is not based on human input.)	b. Computer automatically creates biases which is not based on human input.) Perhaps the algorythme could cause some biases at some point	a. Its occurrence is dependent on how data is collected, who is doing the collecting, and how the data is fed a is better because it includes human error B. does not	a. Its occurrence is dependent on how data is collected, who is doing the collecting, and how the data is fed. Computers do not create biases, it is created by humans who train the computer. Human biases are simply transferred to the computer depending on who trained them
Example Q8 / When the objective of the AI is to detect an apple, which of the actions do you think is better? Please explain why your choice is better than the other one. - a. Have users select apple from a list of random objects. b. Have users select objects	b. Have users select objects that resemble apple in terms of their color, body, and shape. Red colour round shape circle body	a. Have users select apple from a list of random objects. I think a lot of fruits share similar characteristics but having to select apple many times will convince the bot to identify better.	a.Have users select apple from a list of random objects. The second option leaves too much room for discrimination or interpretation, for example regarding the shape and body. (A small apple with holes because of warms is still an apple, even though the

that resemble apple in terms of their color, body, and shape.			shape is not the usual apple-shape).
Example Q9 / Imagine you are a fan of rock music. How would you get the YouTube algorithm to recommend you more rock music? What specific actions would you make and what effects are expected?	By don't using clickbaits. I will use great thmbnails.	By watching more and more videos of rock music them computer can know what to show or not.	By watching more rock music videos and engaging more with the rock videos by liking or commenting on them, AI will detect your interests and show you more options for rock music.
Example Q10 / As a head engineer for the self-driving car company Tesla, your job is to ensure that your car does not hit a human. What steps will you take to teach Tesla to recognize humans from other objects?	Sensor	Train AI in detecting humans by showing it multiple images of humans in different environments	Provide data that slowly teaches Tesla, the difference between more obvious things that are not human and humans and then work further towards perfecting this difference with more ambitious stimuli.