Assessing Learning Engagement Using Humanoid Robots in Higher Education

by

Nicholas Wong Wai Hong

A thesis submitted in fulfillment of the requirements for the degree of

MASTER OF PHILOSOPHY

supervised by

Dr. Anuja DHARMARATNE
Dr. Esyin CHEW
Dr. Lim Jen Nee JONES
Dr. Jojo WONG Sze-Meng

MONASH University
School of Information Technology

January 10, 2018

Copyright Notice

Abstract

A systematic review of educational robotics literature in the past 10 years has shown a lack of application and research of humanoid robots to support teaching as well as a lack of understanding on its impact in higher education. Hence, this study aims to compare and better understand the effects of robot tutors compared to human tutors on learning experience in the higher education context, as well as to suggest a possible framework or guidelines for an effective application of robots in the university learning environment. An exploratory case study through eight 15-minute tutorial sessions assessed with flow psychology was carried out with or without a robot tutor on undergraduate students in Monash University, Malaysia. The introduction of robot tutors in university tutorial classes was observed to positively affect concentration, perception of time, and feeling of reward; but imposes a more rigid interaction and lesson structure which loses sense of control, spontaneity of action and negatively impacts self-consciousness. There are limitations in classroom human-robot interaction which emphasizes the need for an integration framework incorporating automation, accessibility, control and emotion-based assessment.

Declaration

This thesis contains no material which has been accepted for the award of any other degree or diploma at any university or equivalent institution and that, to the best of my knowledge and belief, this thesis contains no material previously published or written by another person, except where due reference is made in the text of the thesis.

Signature: [redacted]

Print Name: NICHOLAS WONG WAI HONG

Date: January 10, 2018

Publications during Enrolment

Published Papers:

Wong, N. W. H., Chew, E. and Wong, J. S.-M. (2016). The review of educational robotics research and the need for real-world interaction analysis, in '2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)', IEEE, pp. 1-6. doi: 10.1109/ICARCV.2016.7838707.

(CORE Ranked-A conference; indexed by ISI Proceedings, EI Compendex and included in the IEEE Xplore)

[Parts of the extensive systematic literature review (Chapter 2) from this thesis were published in this paper]

Submitted for Review:

Wong, N. W. H., Chew, E., Wong, J. S.-M., Dharmaratne, A. and Jones, L. J. N. (under review). 'Investigating the differences in flow and learning experience of undergraduate students with and without the intervention of a humanoid robot tutor', *Computers & Education*.

(ISI-indexed, Q1 ranked journal based on Thompson Reuter's Journal Citation Reports 2016)

[Parts of the research methodology (Chapter 3), results (Chapter 4) and discussion (Chapter 5) from this thesis will be published in this paper]

Acknowledgment

The writing of this thesis would not have been possible without the support of many people around me, and I would like to express my gratitude and acknowledge their efforts in this section.

Firstly, I would like to thank Catherine Hiew Fui Chin, who was my lecturer during my time at SEGi University, but who is also now a colleague and fellow graduate researcher. If not for her dedication in her teaching, as well as her informative advice on scholarships and paths to education, I would have not enrolled in this Masters course in the first place. By extension, I would also like to express my appreciation for the financial support (merit scholarship) and top-class facilities provided by Monash University, Malaysia which made this entire research possible.

Next, I would like to express my deepest gratitude to Dr. Esyin Chew, who was my previous main supervisor when I first joined Monash University. Her expertise, devoted guidance and tremendous support has driven my research to explore and test the field of pedagogy in robotics in various ways I would have never thought of. She was actively involved in many parts of the research, and was also committed to ensure that I was able to complete the Masters course without any hiccups. Also, I would like to thank another fellow graduate researcher, Alena Tan, for her various insights on educational psychology, which this research has greatly benefited from.

Many thanks to my current main supervisor, Dr. Anuja Dharmaratne, who did an impeccable job of taking over Dr. Esyin's place as the main supervisor, providing useful suggestions to the research, as well as ensuring that every part of my candidature was progressing well. I am also very grateful for the support provided by my remaining supervisors, Dr. Jojo Wong Sze-Meng and Dr. Lim Jen Nee Jones for their contributions to the research, especially in research methods such as statistical or validity concerns.

Last but not least, I am greatly indebted to my parents, who despite their situation, have provided me unwavering support, guidance and love, which helped shaped my mindset, passion and motivation in life to move forward and pursue this research.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The usage of robots in education is not entirely new. Examples of dated research include the use of robotics as a collaborative tool for problem-based learning of 10-year-old children (Denis and Hubert, 2001); a programmable toy for learning by design in early childhood education of 3-year-old children (Bers et al., 2002); and a social partner or peer tutor to improve English of first-grade and sixth-grade students (Kanda et al., 2004). These studies suggest that educational robotics improve the quality of interaction, creative thinking and performance. However, Fagin and Merkle's (2002) quantitative analysis on the usage of LEGO Mindstorms robots in introductory college-level computer science education found no evidence in the improvement of student learning or retention. Moreover, a systematic review by Benitti (2012) found that empirical evidence on the effectiveness of educational robots in schools is still limited albeit painting a positive picture. It is mostly used in Science, Technology, Engineering and Mathematics (STEM) subjects with few studies on robot tutors for non-technology subjects. Teachers were also found uncomfortable with the high requirement of mastering the programming of robots. More importantly, Benitti's (2012) review excludes literature on higher education and focuses only on elementary, middle and high schools.

1.1 Problem Statement

The current literature seems to have inconsistencies in the evidence for the benefits of educational robotics in general. There is a lack of research addressing the higher education context as well as a lack of analysis and methods that can accurately describe the factors which affect the success of robot integration in classrooms. The impact of robots on learning experience in university classes is yet to be explored.

A systematic review of 62 articles about robotics in education and learning analytics from year 2008 to 2016 found that the usage of robotics to support teaching has been gaining attention recently and has seen positive results, but they are mostly focused on kindergarten to middle schools and there is a lack of such research in universities. Thus, the motivational and performance benefits observed in children may differ when it comes to the higher education context (Wong et al., 2016). There are also a lack of guidelines on integrating robots into classrooms and on how it can be used to enhance tertiary education. As such, colleges and universities may not see its role in enhancing education. The review identifies the problem of educational robotics research in that it does not focus on the conversations and social interactions that mediate the teaching-learning process in classrooms. These interactions need to be studied further to provide us a better understanding of student learning. The full literature review along with its systematic procedures and findings are presented in Chapter 2.

Therefore, an explorative case study conducted in this research attempts to fill the aforementioned research gap in exploring the application of a humanoid robot tutor in the Faculty of Information Technology (FIT) tutorial classes in Monash University, Malaysia and its effects on student experience compared to human tutors throughout 8 weekly sessions. By incorporating flow psychology with the application of a humanoid robot tutor, this case study attempts to make use of these theories in order to discover the extent of the robot's impact on student experience and flow in university classrooms, as well as the efforts needed to establish a conducive robot-integrated learning environment in the higher education context through the inquiry of the following research questions.

1.2 Research Questions

1. How does a humanoid robot tutor affect students' experiences of flow in university tutorial classes?
2. Are there significant differences in student learning experience with or without the introduction of a humanoid robot tutor in Monash University, Faculty of Information Technology (FIT) classes?
3. What would be a possible guideline or framework for the application of a robot tutor in university tutorial classes?

1.3 Significance of Research and Contribution to the Knowledge

As mentioned in Section 1.1, there is a lack of studies performed on robots used for teaching in the higher education context. Many of the studies in today's literature are focused on using robot tutors for childhood education and as a care robot for autistic patients or the elderly. While humanoid robots seem to grab a lot of attention in kindergarten and primary school classrooms, this may not be the case for university subjects.

This project is an explorative case study carried out on 63 undergraduate students in Monash University, Malaysia from 2 units offered by the Faculty of Information Technology on their experiences with or without the intervention of a robot tutor in 15-minute tutorial class sessions. This thesis describes the problems which the research is trying to solve, as well as its methods, results and recommendations in full. The significance of this research can be summarized into the following points:

1. This research is important as it investigates the extent a robot tutor may affect university undergraduate students' learning experience, and to explore if the usage of humanoid robots as an assistive tool for higher education may be worthwhile to improve this experience compared to human tutors alone. After the study was conducted, the results presented in Chapter 4 suggests that there are benefits in some areas of the learning experience which validates the pedagogical role of robot tutors in higher education, but not without its drawbacks. Nevertheless, its strengths in increasing student concentration and sense of reward makes it a worthwhile endeavor to further explore and improve upon its applications in classrooms.

2. This case study presents an implementation of a robot tutor in real university tutorial class environments which obtains feedback and suggestions from actual students over the course of 8 weeks of classes to give us an idea of how to design lessons and the interactions of a robot tutor in the higher education context. With this, we can explore ways to enhance learning in higher education using robots. In Chapter 5, the results of the study is thoroughly discussed and many guidelines such as speed of the robot's speech, accessibility in human-robot interaction, and implementation details were suggested for a more effective integration of robots in the university learning environment.

1.4 Outline of Chapters

Following this introductory Chapter 1 which describes the current problems of the application of robotics in higher education and the significance of this research project, Chapter 2 describes the sources and step-by-step procedures undertaken during the full literature review. It presents the number of articles, its statistics, as well as a summary of methodologies and results of the current literature which are discussed and elaborated even further in Section 2.1, Section 2.2 and Section 2.3, allowing a better understanding of the problem to be solved in this research.

The approaches used to tackle these problems are then presented in Chapter 3. The developmental phases used in this project, as well as the design of the case study is critically discussed here, pointing out its strengths and weaknesses. The philosophical approaches used in this research such as ontological and epistemological assumptions are also discussed in Section 3.1 and Section 3.2 respectively. Furthermore, the underlying theory which defines the methods to be used to assess student learning experience is presented in Section 3.3, followed by the data collection and analysis procedures to carry out the study in Section 3.4. As all research has its own biases and constraints, the threats to validity when applying these methods are examined in Section 3.5, as well as its limitations in Section 3.6. Lastly, the ethical concerns arising from the application of these methods are outlined in Section 3.7.

Chapter 4 combines data from statistical tests and qualitative sources to explore the extent in which robot tutors affect student learning experience compared to human tutors. Significant differences and sudden changes in trends are identified and discussed.

Finally, Chapter 5 further interprets the results to propose guidelines for the integration of robot tutors in higher education, and to discuss the possible advancements to be made moving forward. Following this, Chapter 6 presents an overall conclusion based on the results of the research.

Chapter 2

Literature Review

A systematic review of 62 articles on the technological and pedagogical integration of robotics in education was carried out. Search results are from Elsevier's Scopus and Thomas Reuters' Web of Science databases due to their popularity and quality as they track journal impact factors. Due to the wide topic coverage of this systematic review, any search results which contain lack of details on data collection and unreliable validity checks are automatically excluded. The selection criteria for these articles are strictly on uses of robotics in schools, universities, classrooms and other similar academic environments. It covers three main categories of articles which are related to: the usage of robots in educational research, learning theories applied in educational robotics, and learning analytics. The high number of search results based on the search criteria as shown in Table 2.1 and Table 2.2 are further refined for each topic, excluding literature which matches any of the exclusion rules defined below:

1. Robotics in Education:
   - not related to educational/social robotics
   - used for training or for disabilities/sickness
   - used as an apparatus rather than to support teaching
2. Educational Theory with Robots:
   - not related to learning theory
   - machine learning (not related to human learning)
   - not used in classrooms or by educators
3. Learning Analytics:
   - not related to students or teachers
   - not experimented in a learning environment

Scopus			
Topic	Robotics in Education	Educational Theory with Robots	Learning Analytics
Criteria[1]	(TITLE-ABS-KEY (((((“humanoid robot” OR robotics OR robots) AND education) OR “robot tutor”) AND (school OR kindergarten OR university OR classroom) AND (effect OR impact OR implication OR acceptance OR outcome))) AND NOT TITLE-ABSKEY ((disabled OR disability OR autism OR training OR mechanical OR medical))) AND PUBYEAR > 2011	(TITLE-ABS-KEY ((vygotsky OR papert OR constructivist OR constructivism OR constructionist OR constructionism OR cognitivist OR cognitivism OR behaviorist OR behaviorism OR connectivist OR connectivism OR “learning theory” OR pedagogy OR “instructional strategy”)) AND TITLE-ABSKEY ((humanoid OR robot OR robotic OR robotics)) AND NOT TITLE-ABS-KEY ((disabled OR disability OR autism OR training)))	TITLE-ABS-KEY ((“learning analytics” AND education) OR ((“data mining” OR “data collection”) AND (“student outcome” OR “academic success” OR “learning experience”)))
Results	121	416	990
Final	14	16	17

Table 2.1: Literature selection criteria – Scopus

After narrowing down the results, all of the remaining literature from each database were combined ($14 + 16 + 17 + 11 + 14 + 10 = 82$) and 31 duplicates were removed giving a total of 51 articles. 11 relevant secondary references found among these 51 articles were added to the table as well, giving a final number of 62 articles.

Based on these shortlisted results, there is an increasing trend for the research of robotics in education which is grounded on educational theory or related to learning analytics in the past 8 years (see Figure 2.1). Selected articles are more recent and are mostly from year 2015; hence, it should provide high relevance for discussing a rapid-changing technology such as educational robotics.

It is observed that most of the research employ quantitative or mixed methods of gathering data for their analysis with mostly well-grounded statistical tests for significance and correlation (see Figure 2.2). There are also a variety of articles ranging from first-quartile ranking (Q1) to unranked journals based on Thomson Reuters' Journal Citation Reports in the year 2014 (see Figure 2.3).

[1] Criteria is based on TITLE-ABS-KEY in Scopus, and TS in Web of Science; both of which covers Title, Abstract and Keywords.

Web of Science			
Topic	Robotics in Education	Educational Theory with Robots	Learning Analytics
Criteria[1]	TS=((((“humanoid robot” OR robotics OR robots) AND education) OR “robot tutor”) AND (school OR kindergarten OR university OR classroom) AND (effect OR impact OR implication OR acceptance OR outcome) NOT (disabled OR disability OR autism OR training OR mechanical OR medical))	TS=((Vygotsky OR Papert OR constructivist OR constructivism OR constructionist OR constructionism OR cognitivist OR cognitivism OR behaviorist OR behaviorism OR connectivist OR connectivism OR “learning theory” OR pedagogy OR “instructional strategy”) AND (humanoid OR robot OR robotic OR robotics) NOT (disabled OR disability OR autism OR training))	TS=((“learning analytics AND education) OR ((“data mining” OR “data collection”) AND (“student outcome” OR “academic success” OR “learning experience”)))
Results	77	197	352
Final	11	14	10

Table 2.2: Literature selection criteria – Web of Science

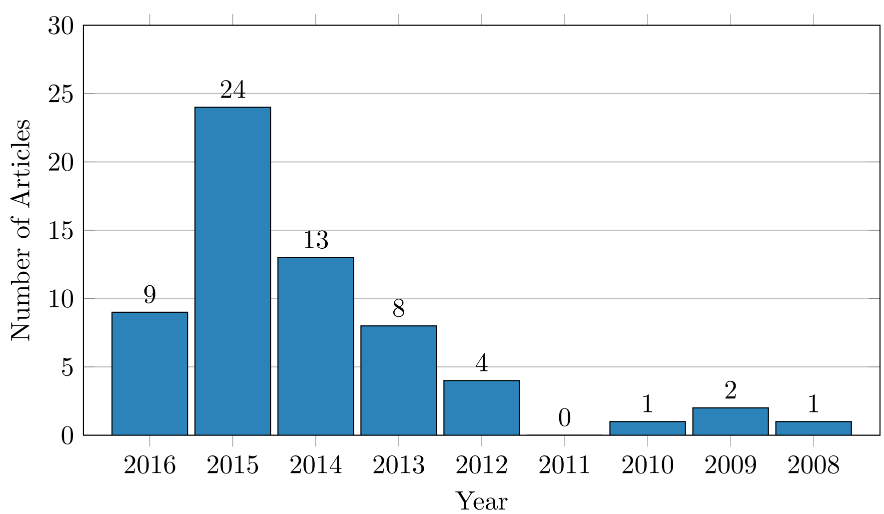


Figure 2.1: Number of reviewed literature by year.

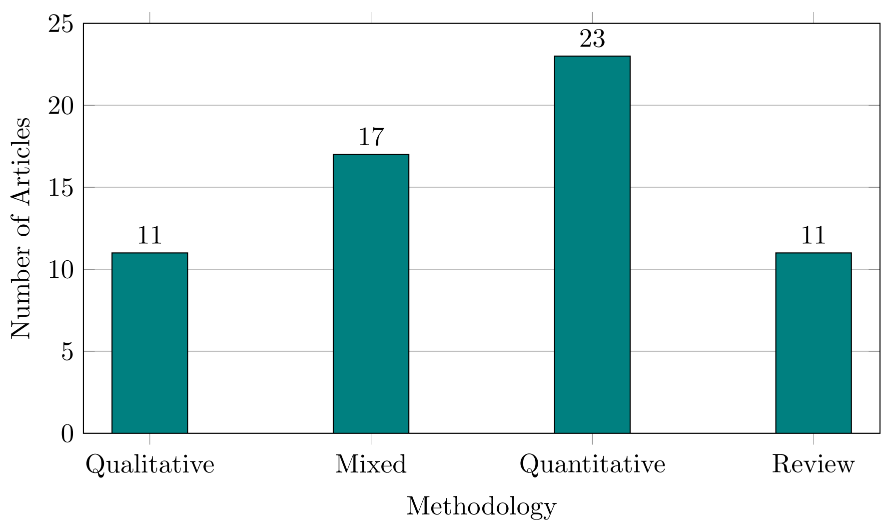


Figure 2.2: Number of reviewed literature by methodology.

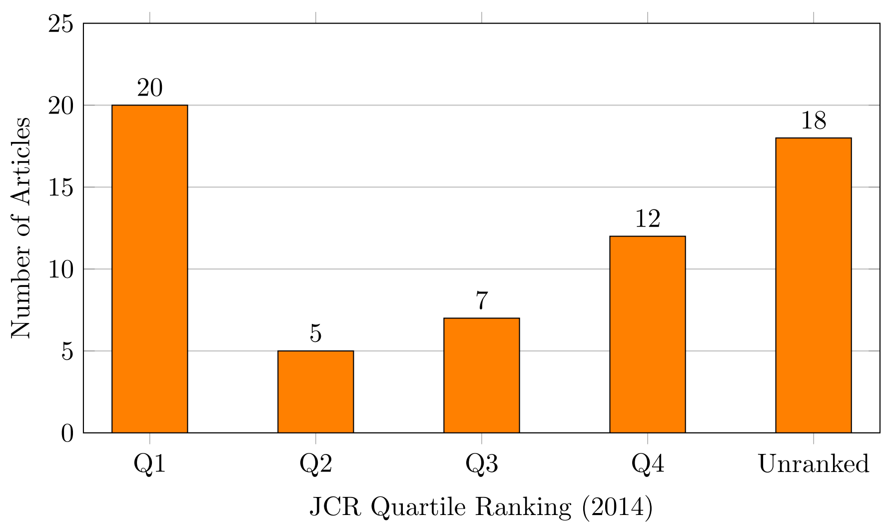


Figure 2.3: Number of reviewed literature by Journal Citation Reports (2014) ranking.

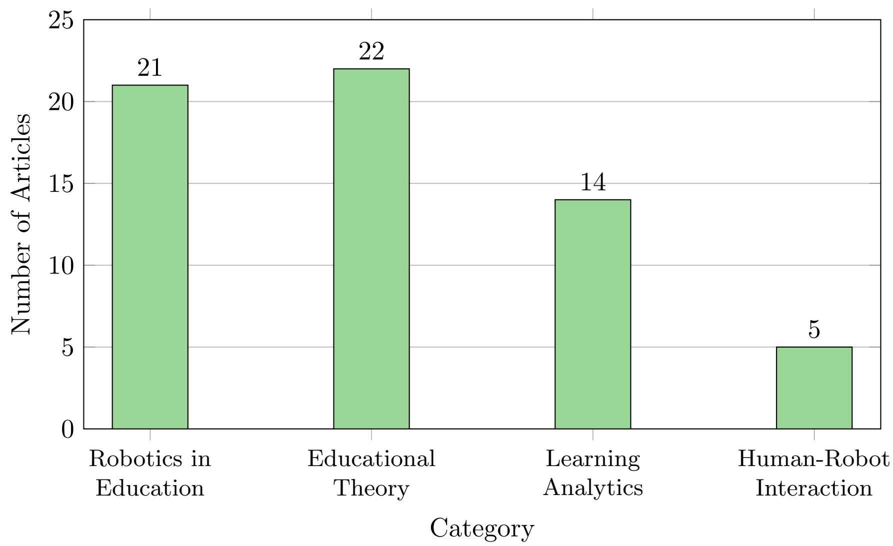


Figure 2.4: Number of reviewed literature by category.

A review summary of these articles on their methodology (keywords which demonstrate quality of the research methods are underlined in bold under the "Experiment" section), participants (including location of experiment if stated) and results were tabulated in Appendix A. Each article is marked with a shade of color to show which database it is from. These articles are then categorized into several topics (see Figure 2.4) which are then discussed in Section 2.1, Section 2.2 and Section 2.3 where relevant.

2.1 Robotics in Education

Based on the review, research suggests that robotic tutors increase motivation, performance and satisfaction of kindergarten to middle school children although there is a lack of research on robots used to assist teaching in higher education. The use of robots in education to support teaching in classrooms is limited to kindergarten (Keren and Fridin, 2014; Lee, Sullivan and Bers, 2013), elementary (Park et al., 2015; Wu et al., 2015; Zaga et al., 2015; Chin et al., 2014; Saerbeck et al., 2010) and middle schools (Alemi et al., 2015; Alves-Oliveira et al., 2015; Alemi et al., 2014; Ardito et al., 2014) showing positive results in student experience, motivation and performance. When it comes to higher education research, most research (Michieletto et al., 2016; Yi et al., 2016; Miranda et al., 2012) use robots in robotics, engineering or computer science units as an apparatus for study, operation or experimentation.

Few research using assistive robots on college students showed positive results in student engagement (Brown and Howard, 2014*b*; McGill, 2012). Using Keller's (1987) Instructional Material Motivation Survey (IMMS), McGill's (2012) research on undergraduate students only found significance in attention but not in other motivational factors such as relevance, confidence or satisfaction; this differs from the results of Chin et al.'s (2014) research on elementary school students where satisfaction and relevance were rated the highest. The lack of robotics research in the higher education context suggests that motivational and performance factors may differ based on the level of education, further emphasizing the need for assistive robotics research in university classrooms. Most of the aforementioned literature in this section employ questionnaires, pre-tests and post-tests which although presented strong evidence for correlation and statistical significance using well-established methods such as Student's t-test, Pearson's product-moment correlation and regression analysis in their results, lack analysis on the actual human-robot interaction. Student engagement is usually not monitored explicitly, but based on assumptions from responses to questions and tasks (Brown and Howard, 2014*a*).

Given the limited usage of robots in the higher education context, and the lack of analysis in the interactions during the learning process, it is important to place focus on conversations that mediate this process. The exchange of ideas through language is the fundamental instrument between teaching and learning in academic institutions which deserves a more in-depth study. For this study, the focus is on the 15-minute teaching and learning interaction with a Q&A session between the (robot) tutor and students in the university classroom which addresses this problem by attempting to investigate learning experience using flow theory, student comments and observations to understand the extent of the robot's effects compared to human tutors.

2.2 Educational Theory with Robots

Mubin et al. (2013) found that common robots used in education are electronic robot kits such as Arduino and Boebot; mechanical robot kits such as LEGO Mindstorms and Thymio; and humanoid robots such as NAO and Robovie where these robots serve as tools, peers and tutors with Papert's constructionism being the most adopted theory followed by Vygotsky's social constructivist theories. Constructionist and constructivist theories are used due to the need of a structure in the curriculum to introduce the robots into classrooms but the effects of integrating these pedagogical practices are not clearly understood. Altin and Pedaste (2013) found that in the 8 articles reviewed, the following approaches have been used for educational robotics in STEM units: discovery learning, collaborative learning, problem solving, project-based learning, competition-based learning, and compulsory learning; but lack evidence that it achieves educational goals.

For direct study of pedagogical and technological integration, there are some which have applied Vygotsky's Zone of Proximial Development (ZPD) to study its effects in classrooms (Berland et al., 2015; Silva et al., 2008) and some which propose pedagogical models by combining other technologies with robots (Chen et al., 2012; Mitnik et al., 2009). Some studies found that applying constructivist theories in educational robotics is more effective than traditional methods of teaching (Lee, Taha, Yap and Kinsheel, 2013; Plauska and Damaševičius, 2014) and that it promotes collaborative critical thinking (Bers et al., 2014; Danahy et al., 2014; Mills et al., 2013; Bilotta et al., 2009). These findings would suggest that problem-based learning and student-centered techniques such as Vygotsky's social constructivism have a more positive impact, something which this research should apply when integrating robot tutors to carry out teaching. Robot tutors can be integrated as a peer who learns with the students or as a tutor who teaches the students but this choice seems to have no differences on the amount of knowledge retrieved by the students (Blancas et al., 2015).

However, research which apply educational theories and aim to assess the effectiveness of robotics at the same time have a common limitation where they cannot easily identify factors affecting student outcomes. They would have to make assumptions in their results and will continue to do so until there are proper analysis methods to deal with this limitation. Actually, even without the use of technology, Stroet et al. (2016) could not distinguish contributing factors to students' motivation between schools which apply traditional, social constructivist and combined educational philosophies. In educational research, "it may be extremely difficult, or even impossible, to isolate and manipulate all the variables suspected of being involved in the phenomena being studied" (Berliner, 2002).

Due to this problem, the effectiveness of educational robotics may still be subjected to skepticism and its benefits for integration into classrooms is not very convincing. Although teachers generally have positive perceptions on the benefits of robots in classrooms (Khanlari, 2013; Fridin, 2014), they recognize the challenges and adaptations needed for it to work. Additionally, this barrier to adoption may be tougher to overcome as the introduction of robotics may change the lesson structure, introducing large administrative overhead and technological challenges (Khanlari, 2016; Serholt et al., 2014); as well as gravitate towards constructionist and constructivist theories which may be further complicated by cultural factors (Hăng et al., 2015; Thomas and Watters, 2015). The lack of clear evidence in relative advantage, compatibility and observability greatly hinders such technological adoptions in academic institutions (Reid, 2017).

2.3 Learning Analytics

With the problems described in Section 2.1 and Section 2.2, more refined methods for data analysis is needed to understand the effects of educational robotics in classrooms. The complex problem of isolating an absurd number of factors in educational research may require the application of big data and learning analytics to provide a better understanding. Learning analytics allows a learning design to be evaluated based on its pedagogical intent using a set of real-time, behavior-based data on learner interaction within the learning environment (Lockyer et al., 2013). Diagrams portraying social networks and interactions formed in a collaborative group can be mapped out and analyzed.

Currently, learning analytics mostly apply to virtual learning environments through Learning Management Systems (LMS) such as Moodle or Blackboard (Fidalgo-Blanco et al., 2015; Hernández-García et al., 2015; Lonn et al., 2015; van Leeuwen et al., 2015; Zacharis, 2015) where readily available data such as login duration and number of forum views can be processed and translated to student engagement or motivation, with results suggesting active participation as the major contributor of performance. However, Iglesias-Pradas et al. (2015) found no relation between LMS interactions and teamwork competency. While digital data on the actual interaction between students is easily processed, real-world analog data can be a problem to accurately obtain even with structured data collection. Eye gaze and head tracking may be useful data to assess interaction level (Fridin, 2014) but not when high-level cognitive thinking is required (Brown and Howard, 2014*a*).

With that said, there are emerging technologies to record data for classroom analysis such as The Visible Classroom Project captioning of a teacher's speech in a classroom analyzed using a rubric of pedagogical principles (Skipp and Tanner, 2015) and the low-cost distributed Multimodal Recording Device (MRD) which records video, audio, pen strokes and learning environment properties such as temperature (Domínguez et al., 2015). Combining these technologies with humanoid robotics may allow better attainment of learning analytics input from humans through social interactions (de Greeff and Belpaeme, 2015; Das et al., 2015) which can be further enhanced with emotion data (Singh et al., 2013). Perhaps one way of adapting learning analytics from virtual to real environments can be achieved in such a model where data is provided naturally through friendly human-robot interaction. The mismatch between human-like robot design and the human frame of reference may provide new ways of analyzing human creativity (Zawieska and Duffy, 2015). Analysis of such inputs from real-world classrooms may provide a more holistic insight of a student's learning in a way that may question Lundie's (2017) argument on human subjectivity in learning experience being incommensurable with information analytics.

Even if such real-time learning analytics application is successful, it may not be useful in classrooms due to high information load (van Leeuwen et al., 2015) and preprocessing or latency between collecting data and actionable decision (Madhavan and Richey, 2016). Madhavan and Richey (2016) found that methods to tackle the problem of data noise and sparseness are emerging through a new class of algorithms derived from variations of techniques known as Kalman filtering and ensemble Kalman filtering. Other efforts include Schieble et al.'s (2015) attempt to break down the complexities of student-teacher interaction using positioning theory in discourse analytics; and data mining methods proposed by Anaya et al. (2016) to solve uncertainties in collaborative interaction using Bayesian networks. Learning analytics models may also be too specific in scope as Gašević et al. (2016) found that generalized models are inaccurate in many cases; suggesting that it should instead be at the individual course level with instructional conditions factored in. Slade and Prinsloo (2013) claims that student performance and identity are highly dynamic and temporary constructs, which means that learning analytics data can have an expiry date that will become useless within an unknown period of time. In addition to the privacy and ethical concerns where collected data should be justifiably beneficial to students' learning (Rubel and Jones, 2016), learning analytics may have the problem of being useful only in an extremely narrow scope.

More research on learning analytics should be cross-disciplinary and emphasize on the impact of interactions within the learning environment. Communication and social presence of being together is important for learning (Rienties and Toetenel, 2016; Akhtar et al., 2017; Joksimović et al., 2015); therefore, it is crucial that communication among students and educators be analyzed. The implementation of real-time learning analytics is out of the scope of this research, but the aforementioned problems are recognized and will be considered when designing guidelines and frameworks for robot integration in classrooms.

Chapter 3

Research Methodology

This research involves an exploratory case study involving only Monash University students with a non-experimental design, where student participants were introduced to either a robot tutor or a human tutor in a 15-minute tutorial session once a week for 8 weeks and their experiences were recorded. Selection of participants is non-random because it was predetermined based on the cohort of students for the units which were available at the university. The case study was carried out with the following definition in mind, sharing the same post-positivist view as Yin (2013), where a case study is an empirical inquiry that investigates a phenomenon within its real-life context especially when the boundaries between this phenomenon and context is not evident; and which may rely on multiple sources of evidence. Hence, this research uses mixed-methods collecting both quantitative and qualitative data to be analyzed for a more holistic view of students' interaction with a robot in Monash University units. More elaboration on the post-positivist approach in research can be found in Section 3.1.

Student experience is evaluated through questionnaires based on a componential flow psychology model, student and tutor comments as well as observations. More details on the theoretical framework used in this research can be seen in Section 3.3. This form of methodological triangulation complements each other for a more complete data set so that their differences and similarities can be identified, compared and discussed especially for a problem such as the human learning process which may be influenced by complex neurological and societal factors yet to be fully understood.

This research also follows the single-case embedded design guidelines of Yin (2013), investigating student experiences with the robot tutor in the case of Monash University undergraduates studying in Malaysia in two different units offered by the Faculty of Information Technology. Due to the nature of this study, the topics prepared for the robot depends on the class and units at this

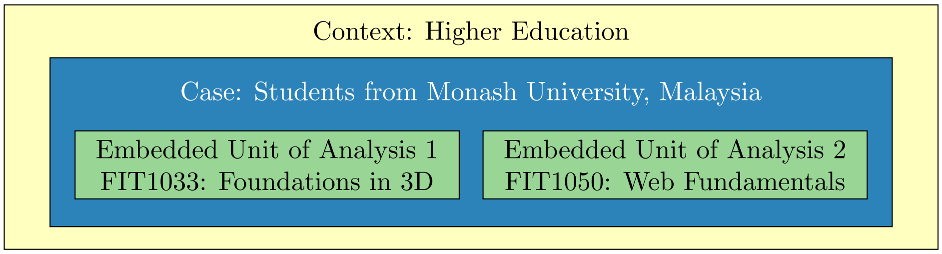


Figure 3.1: Single-case embedded design.

specific university. Other universities may have different course structures and variety of students in their classes. Looking at more than one unit provides multiple sources of evidence within a single case. This design is visualized in Figure 3.1.

Case studies have been criticized for their lack of methodological rigor, validity and generalizability. Willis (2014) provides a good overview of the advantages and disadvantages of single case study designs, acknowledging its limitations while suggesting that its weaknesses are exaggerated. One of the criticisms is that it absolves the researcher from methodological considerations due to its boundedness to context (Maoz, 2002). This allows the study to take a free-form approach which usually results in long unreadable documents with no ability to establish causal relationships. It stems from the lack of methodological rigor which jeopardizes the validity and reliability of the research. Indeed, the lack of systematic procedures is the greatest concern for case studies but this has started to change. Yin (2013) called for rigor in his book through proper development of a case study protocol to clearly define all field procedures, case study questions and context; as well as maintaining a chain of evidence that can be traced throughout the research.

In the field of international relations, Bennett (2004) discussed innovations in third-generation qualitative methods where scholars have over the last fifteen years, "revised or added to essentially every aspect of traditional case study research methods." For example, process-tracing involves the rigorous testing of alternative hypotheses against the evidence from the case similar to Bayesian inference by affirming or rejecting explanations that do not fit the evidence based on eliminative induction. Such is the approach used with the principle of falsifiability in mind; its epistemological considerations are discussed in Section 3.2. Bennett (2004) also shared the view that although single case studies are not always explicitly comparative, they are implicitly comparative. For instance, a most-likely case is one where a theory is likely to provide a sufficient explanation if it applies to any cases at all. If that theory fails to fit in a most-likely case, then it is strongly challenged. The opposite is true for least-likely cases. Therefore, single case analysis can be valuable for the testing of theoretical propositions given that the predictions are relatively precise and measurement error is low

(Levy, 2008).

Besides that, case studies are known to use qualitative methods such as observations and interviews that are highly affected by researcher subjectivity. It is also susceptible to selection bias, as a poor case selection can lead to overgeneralization or misinterpretation of observations. Concerns of validity are mitigated using methodology triangulation and pattern-matching as discussed in Section 3.5 of this thesis. Another criticism is that a single case is hardly generalizable to the population. It is important to note that unlike standard statistical methods, selected cases in a case study are not representative of a larger population. The difference between statistical and analytical generalization along with some examples are explained in Section 3.5.3. Nevertheless, all research has its limitations and they are further discussed in Section 3.6.

In summary, the research follows the process flowchart as shown in Figure 3.2. After research questions and objectives were defined at the beginning of this project, the systematic literature review was performed (see Chapter 2) to further define the problem and refine the research questions. The study was carried out using the case study design discussed above, with non-probability sampling (see Section 3.4.1) and a triangulation of questionnaire responses, comments and observations. This study requires the implementation of a humanoid robot as part of the data collection and analysis; therefore, pilot tests were crucial to ensure that it can carry out the research tasks as correctly and reliably as possible. After the robot setup has been developed to satisfaction, the field test was carried out as described in Section 3.4.2 and collected data was analyzed using both statistical and qualitative methods outlined in Section 3.4.3.

3.1 Ontological Assumptions

Ontology deals with the nature of reality. It primarily questions if social entities should be regarded as objective where they are external to social actors, or subjective where they are created from perceptions and actions of social actors. In layman terms, social actors are us humans as researchers, and the entities are the phenomenon being studied and observed.

The philosophy of this research adopts the post-positivist paradigm with the critical realist ontology whereby it is assumed that there is an objective reality separated from the subject. For this research, this makes the assumption that there are regularities in human social behavior during learning that can be generalized within a certain context. However, this reality can only be known with a probability and must be subjected to critical examination. Observations are not fixed and are open to change depending on context. Critical realism is an umbrella term coined by Bhaskar (2013) of the view that we only have access to empirical data but never the real. It recognizes the reality of the natural order as well as the events and discourses of the social world.

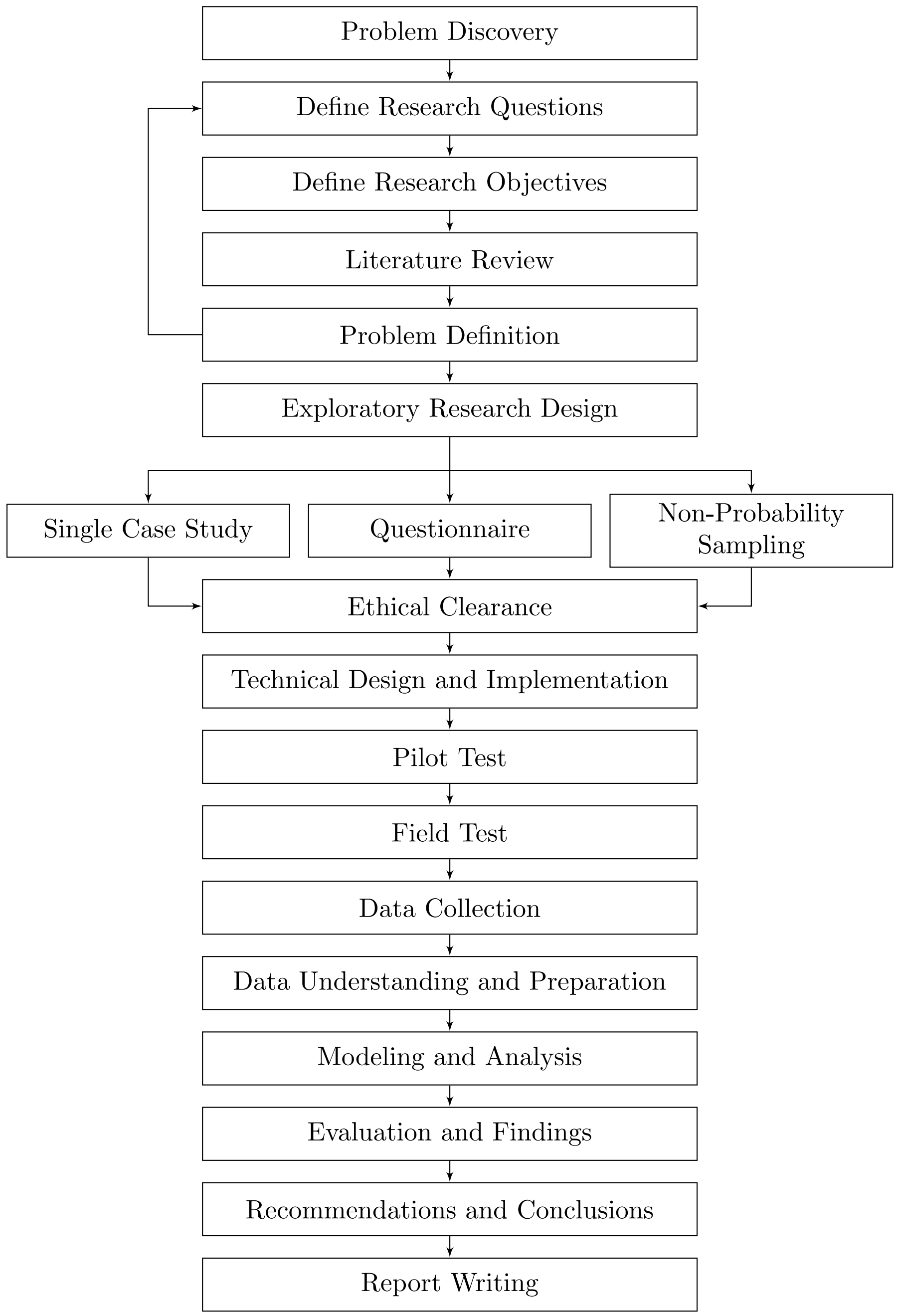


Figure 3.2: Research methodology flowchart.

We will only be able to understand and change the social world if the generative mechanisms of events and discourses are identified. A scientist's conceptualization is simply a way of knowing this reality. Therefore, there is no absolute proof to verify explanations on any of our observations. Unlike traditional positivism, this approach finds it acceptable to infer on hypothetical entities that are not directly observable in quantitative data through abductive reasoning. Models can be inferred based on their explanatory power rather than from patterns or past events. This is as what Bhaskar (2013) calls it, transcendental analysis. We transcend empiricism and question what reality is like in order for the observed phenomena to occur, placing high importance on carefully defining the conditions and context for the occurrence.

In contrast, subjectivist or constructivist ontology implies that social phenomena are not only produced through social interaction but in a constant state of revision (Bryman, 2015). The researcher always presents a specific version of social reality rather than one that can be regarded as definitive. There is no rationale for choosing either of these standpoints as these are assumptions that we have to make; yet, questions of ontology influence the way research questions are formed and carried out.

The researcher takes the critical realist standpoint and believes that there is ultimately an objective reality. Most notably, this influences the research to adopt quantitative methods in the case study design rather than just qualitative methods. The goal is to mitigate biases and emphasize on theories which can be generalized and known definitely.

3.2 Epistemological Assumptions

Epistemology is a branch of philosophy that questions what is considered knowledge in a discipline. A key question that arises in epistemology is whether or not the social world should be studied the same way as natural sciences. According to Bryman (2015), the main paradigms which are concerned about epistemology are positivism, interpretivism and realism. The positivist position is one that advocates the same methods to be used when studying both natural and social sciences. It assumes that human senses are reliable and that it can confirm observed phenomena as genuine knowledge.

On the other hand, the interpretivist position is an alternative to positivism that requires the researcher to interpret the subjective meaning of social action. Such a stance normally looks at subjective values, beliefs and experiences of participants to form a social reality rather than trying to separate them as potential biases to the understanding of reality. The interpretivist stance believes that human interpretation and understanding are constituents of scientific knowledge.

Finally, the realist position shares much of its features with positivism but branches out from there. As pointed out in the previous section on ontology, the researcher's assumptions are of the post-positivist paradigm, in-line with Bhaskarian critical realism. The key difference compared to positivism is that it does not assume that human perception is a reliable source of knowledge. The post-positivist paradigm stance of this research is of a modified objectivist epistemology. Objectivity can only be approximated because human perception is assumed to be an imperfect representation of the real world and is only one way of knowing reality. Therefore, the best model of reality which humans can know of is socially constructed. Knowledge cannot be separated from an individual and biases are inherent. In order for human observers to become more certain of the objective reality, it requires a triangulation of multiple unreliable perspectives, making subjective experiences of participants a key part of the research. Both quantitative and qualitative sources will be used to try and widen this perspective. The researcher is of the opinion that human social processes should be investigated in the same way as natural sciences while accepting that it is value-laden and complicated by a large amount factors that are highly subjective and sensitive to context.

Even so, it is still highly important for the research to maintain objectivity to the highest degree possible. This project attempts to infer a model based on pilot tests to account for the observation, relying more on discovering the mechanisms which produce effects in a specific context rather than testing a priori hypotheses. This research employs statistical tests on quantitative data to be compared against subjective qualitative data such as observations and comments, providing more insights to the objective reality that we cannot perfectly reach.

The principle behind these tests is based on falsifiability, where nothing can be proven to be true but a contradictory observation is all it takes to demonstrate the inconsistency of a theory. Popper (2002) stresses the problem of demarcation that in order for statements to be scientific, it must be capable of conflicting with possible or conceivable empirical observations. This principle was advanced as a criticism to the logical positivist idea of scientific verifiability. Post-positivism is thus a reformation of positivism to address its criticisms, and is the position which the researcher takes when carrying out this research.

3.3 Theoretical Framework

This research applies the concept of flow psychology to study student engagement and experience in the classroom when being instructed by a robot tutor. Flow, or more commonly known as "being in the zone" is the state in which a person is fully immersed in an activity and is said to be an optimal experience where people felt and performed the best (Csikszentmihalyi, 1990). It is one of the key concepts under a branch of psychology called positive psychology

(Seligman and Csikszentmihalyi, 2000), which emphasizes on the use of scientific methods to study interventions that help achieve satisfactory life focusing on positive human development. Flow does not encompass the entire student experience as it is much more complex; instead, flow is just one aspect of experience to look into and is the focus of this study. Csikszentmihalyi (1998) proposes 9 dimensions of the flow state experience which are: *challenge-skill balance (Q1), action-awareness merging (Q2), clear goals (Q3), unambiguous feedback (Q4), concentration on task at hand (Q5), sense of control (Q6), loss of self-consciousness (Q7), transformation of time (Q8),* and *autotelic experience (Q9).*

First, the experience of flow requires a balance in skill and challenge in order to feel engaging. If the task is too difficult, it can be frustrating; whereas if the task is too easy, it can be boring. Second, when experiencing flow, the person's involvement in an activity reaches a point where it merges with the self and their actions are mostly performed at the subconscious level. Third, a clear understanding of what to do is part of the flow experience as ambiguous or conflicting goals can divide a person's attention. Fourth, people who are in flow constantly require direct and clear feedback in their actions in order to respond and continue being engaged. Fifth, a person experiencing flow exerts high levels of concentration on the present activity and would not be distracted by other thoughts. Sixth, there is no sense of worry about losing control over the task at hand. Seventh, a person would be so engaged in the activity so as to not have the mental state to care about their own ego or how others think about them. Eighth, time can feel accelerated or slowed down when in the state of flow. Ninth, the activity performed is intrinsically rewarding. Carrying out the task is an end in itself and does not require external motivation.

The method used in this study to assess these experiences of flow is through a standardized scale of the componential flow model using a short version of the statistically-grounded Flow State Scale-2 (Jackson et al., 2008). For each of the aforementioned flow dimensions, there is a Likert-scale question (item) to help assess the severity in which the participant is experiencing that dimension of flow. This model is visualized in Figure 3.3.

Compared to other flow measurement methods such as flow questionnaire (FQ) and experience sampling method (ESM), this componential model using standardized scales provides a more comprehensive characterization of flow and is psychometrically more valid (Moneta, 2012). Despite these strengths, its weakness is that it imposes flow on all participants, does not measure the commonness of flow, and is poor at explaining heightened and focused attention in extreme cases. While other methods also have their own strengths and weaknesses, the short human-robot interaction time (about 15 minutes) in this research is best suited for the componential model as other methods are more suitable for sustained flow over a long period of activity.

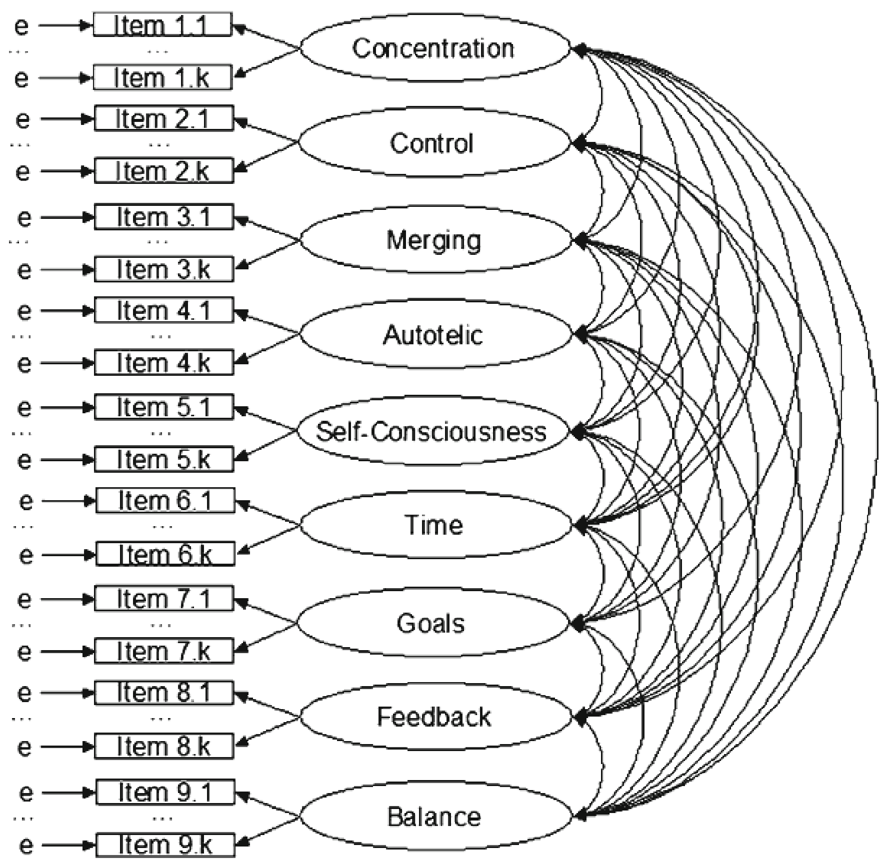


Figure 3.3: 9 dimensions of the componential flow model visualized by Moneta (2012) where e = measurement error.

3.4 Analytical Methods

This section outlines the procedures used for the field test such as data collection and analysis. The results are presented in Chapter 4.

3.4.1 Sampling

The sampling method used in this study is considered as generic purposive sampling, where a set of criteria on the contexts or cases needed to address the research questions are defined and identified (Bryman, 2015). It is a form of non-probability sampling which does not perform random selection of participants and is not representative of a wider population. Purposive sampling is done with the research goals in mind. As the research explores in the higher education context, Monash University, Malaysia is defined as the sample area. The purpose of this research tries to explore learning experience in university students from 2 different units.

This case study was carried out on a total of 63 undergraduate students aged 17 to 24 (mean $\bar{x} = 19.33$, standard deviation $\sigma^2 = 1.136$) from Monash University, Malaysia who are undertaking units offered by the Faculty of Information Technology. Two units were selected for analysis – FIT1033: Foundations in 3D and FIT1050: Web Fundamentals. For each of these units, two study groups were formed:

- First 4 sessions carried out with robot tutor, last 4 by human
  - FIT1033 Mondays 2 p.m.
  - FIT1050 Wednesdays 8 a.m.
- First 4 sessions carried out by human tutor, last 4 with robot
  - FIT1033 Wednesdays 8 a.m.
  - FIT1050 Tuesdays 3 p.m.

3.4.2 Data Collection

Every week for 8 weeks, an educational session lasting approximately 15 minutes was carried out by either a human or a robot tutor, who will give the instructions according to the teaching materials for that week. There is a single human tutor in charge of each unit during the human tutor sessions, so there are 2 tutors in total. The robot used for this study is an interactive companion robot called NAO, which is 58cm in height with 25 degrees of freedom equipped with 4 directional microphones and loudspeakers (SoftBank Robotics, 2017). Within the session, students usually have to perform a small task such as to try out a small snippet of code, perform some steps in a 3D software, or read up about a topic. For human tutors, this session is carried out as usual as in a normal tutorial class. For robots, the lesson structure is designed to be similar to human tutors and is represented in Figure 3.4, where the lesson is carried out as usual according to the teaching materials of the university unit, ending with a short Q&A voice interaction. The NAO robot is programmed using the included Choregraphe Software Development Kit (SDK) which provides a user interface to drag-and-drop nodes that trigger specific robot functions. For this study, only the "Animated Say" node was used for robot instruction and gestures, while the "Dialog" and "Tactile Head" nodes were used for the Q&A interaction.

During a session, the robot is placed on a table at the front of the classroom while a human facilitator controls the robot and the presentation slides from a short distance away using a laptop. After the lesson is complete, a student is called forward to interact with the robot at a distance of approximately 1m. Figure 3.5 shows a student interacting with the robot in front of the class.

An example of the robot setup in the learning environment is illustrated in the classroom layout shown in Figure 3.6; it is one of the 4 classrooms which the study took place in. The time taken for the robot to complete the session is 15.56 minutes on average, but there were technical issues occasionally; the minimum session time is 7 minutes, whereas the maximum is 30 minutes. At the end of the session, students are given a questionnaire form (see Appendix B) containing 3 questions about their experiences with the tutor in the session, followed by 9 questions according to the short version of the Flow State Scale-2. Along with these student responses, for every session, the session number, date, start time, end time, class, study week, topic of study, tutor comments and tutor type are recorded. This study was carried out from March 2017 to June 2017. In total, 328 student responses were collected throughout 8 weekly sessions for all 4 of the aforementioned study groups and the dataset was made available to the public (Wong, 2017).

3.4.3 Data Analysis

When it comes to demonstrating significant differences, non-parametric tests such as the Mann-Whitney-Wilcoxon test is more often than not recommended for ordinal or ranked data as it relies on the median instead of the mean and does not assume that the outcome should be normally distributed. However, the standard parametric t-test is just as effective when it comes to five-point Likert scale items, and provides enough assurance from false positives and false negatives even for sample sizes as low as 10 per group (De Winter and Dodou, 2010).

Each weekly session is compared with one another to find significant differences in each dimension of flow which is represented as a Likert-scale item (1 to 5) collected from students' responses in the short Flow State Scale-2 questionnaire (see Appendix B).

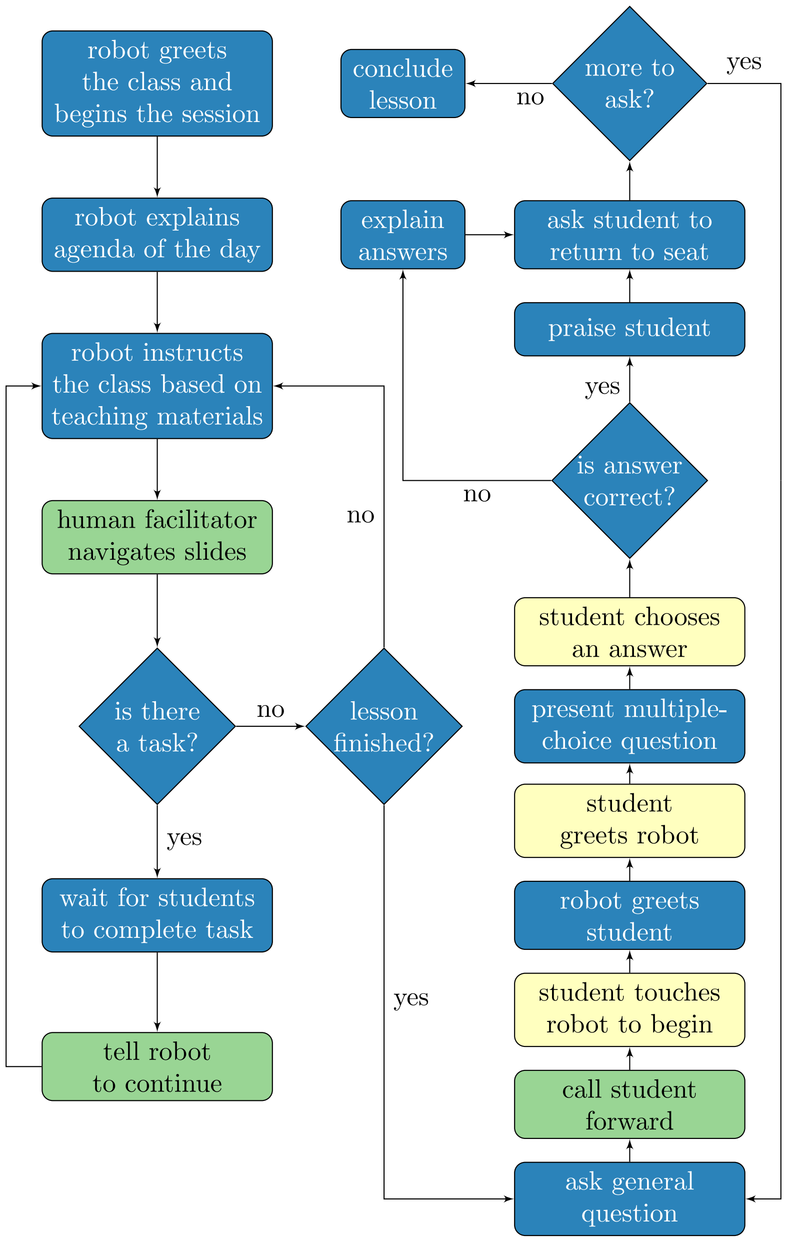


Figure 3.4: Robot lesson and interaction structure (average completion time: 15.56 minutes).

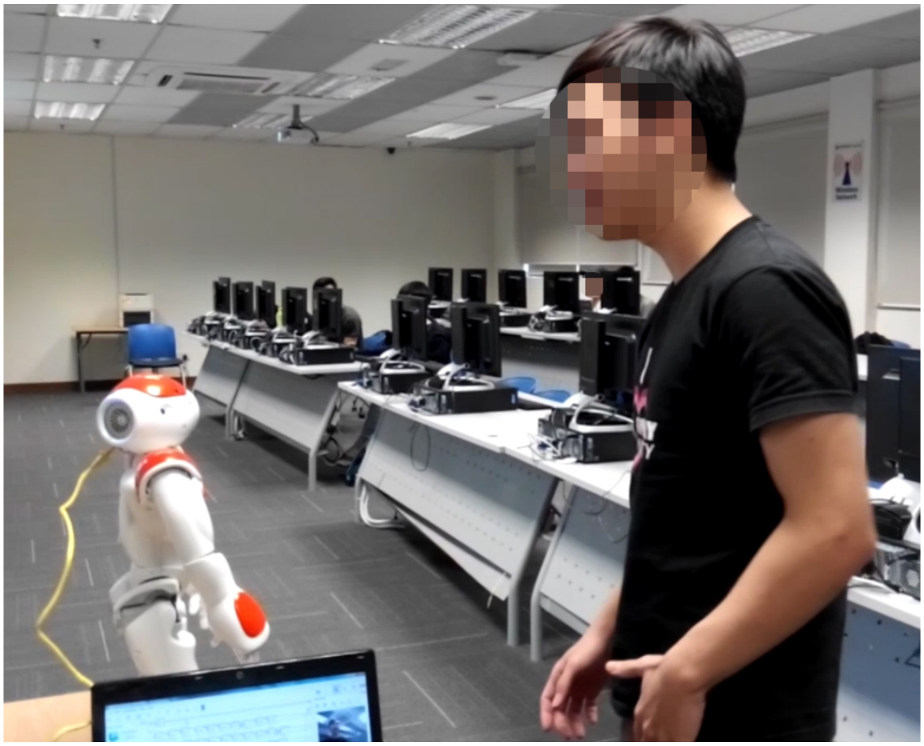


Figure 3.5: A student interacting with the robot at the front of the class in the networking computer laboratory of Monash University, Malaysia.

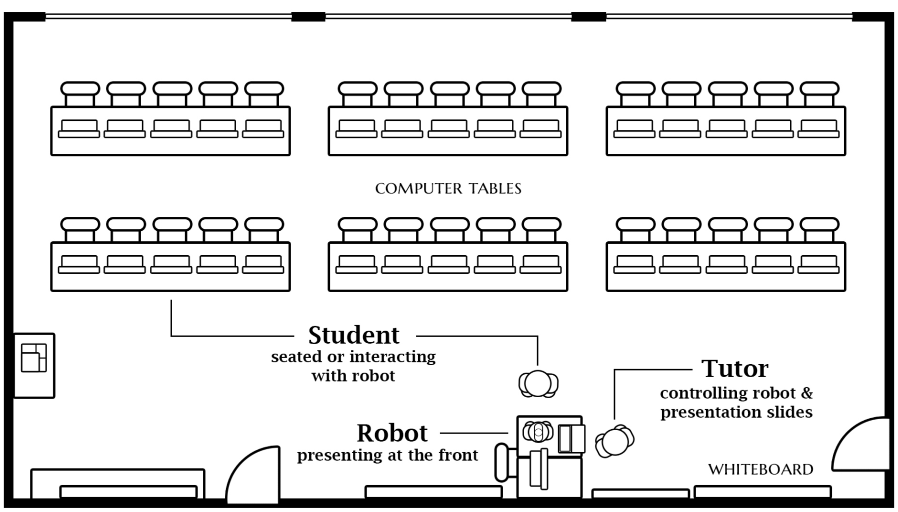


Figure 3.6: Example layout setup of the networking computer laboratory in Monash University, Malaysia.

The null hypothesis is defined as: H_0 = There is no difference in students' experience of flow dimension x between Week y and Week z. The statistical test employed here is a paired t-test (also known as Student's t-test) whereby only students who attended both weeks are included into the sample to be tested. Paired t-tests allow the identification of significant changes between variables before and after each week. For a slightly different perspective focusing on group-wise tests, unpaired t-tests (Welch's t-test) were also carried out for all week combinations. All t-tests carried out in this study are two-tailed to test significant increase and decrease in responses, assuming unequal variances due to unforeseeable and uncontrollable human factors in this study. Groups with small sample sizes of less than 6 are ignored.

Apart from week combinations, an overall test was carried out on data separated into two groups: sessions carried out with the NAO robot tutor, and those by human tutors (2 tutors in total, 1 for each unit). T-tests were carried out on these two groups for each of the flow state scale questions. In order to test for significant differences in categorical data, Pearson's chi-square test for independence is carried out between the sentiment of students' comments; and whether or not they have interacted with the tutor, expressed suggestions, or expressed boredom. Sentiment is derived from the students' remarks in all 3 of the subjective questions provided in the questionnaire. It is categorized into either: *positive, neutral, mixed* or *negative*; where "positive" is assigned for favorable remarks such as "good," "interesting," "informative," or "fun"; and "negative" for complaints such as "boring," "redundant," "annoying" or "frustrated." "Neutral" is assigned when there are neither positive or negative comments whereas "mixed" is assigned when there are both. As sentiment classification is inherently subjective and based on human judgment, it was determined that by following the aforementioned examples of rules for keywords and phrases found in student comments, that a single researcher's judgment was sufficient when the manual classification was performed.

All week combinations which are deemed significantly different ($\alpha = 0.05$) are then grouped with other combinations based on their similarities and differences. Demonstrating significant difference does not automatically suggest that the robot had any effect because none of the variables are properly isolated and there is no control group. It is worth reminding that the nature and design of this study makes it impossible to isolate the variables (Berliner, 2002) needed to pinpoint the factors that may have caused the differences in flow among human and robot tutor groups through statistical tests alone. T-tests are also carried out on a week-by-week basis because some information may be lost if only an overall test of human and robot sessions is performed. Each week can be considered a different context and can have unforeseeable social or psychological factors that may come into play to influence a student's experience of the session.

Nevertheless, by demonstrating the difference in flow states between multiple groups and university units, then combining the qualitative data such as students' comments about their experiences as a mixed-methods research through "triangulation" (Patton, 1990), it may produce insights to the bigger picture. For example, if a significant difference is found between human and robot sessions; and none found for robot-robot or human-human week combinations, then it becomes a meaningful observation. Furthermore, if multiple university units also follow the same trend, with support from the students' responses, then it would strongly suggest that the presence of a robot tutor is affecting the students' learning experience.

On an additional note, when comparing flow states across the weeks, the analysis of the two FIT1050 groups are more prominent. This is because some data are missing for 2 of the weeks conducted by a human tutor for FIT1033. The FIT1033 groups also have less statistical power when comparing trends between human and robot tutor sessions due to its smaller sample size.

3.5 Threats to Validity

Kidder and Judd (1986) summarizes four design tests which will be looked at for this research: *internal validity, construct validity, external validity* and *reliability*. Internal validity refers to the establishment of a causal relationship where certain conditions can be demonstrated to lead to other conditions. Construct validity is the most important test for this study and ensures a sufficiently operational set of measures used to collect data, questioning the subjectivity in the interpretation of such data. External validity concerns about the domain in which the results of the research can be generalized. Lastly, reliability ensures the repeat-ability of procedures used in the study with the same results given the same context.

3.5.1 Internal Validity

This research is an exploratory one which is not concerned with making causal statements. Its non-experimental design makes it clear that there are high threats to internal validity if the quantitative data in this study is used to establish cause and effect relationships. This is due to the research having no control over the assignment of participants between the sampled university units and must select from a cohort of students in a given university. The subjects of the experiment are students from different classes within the same university but they are not comparable. Such a design is most susceptible to selection bias and the results of this study may not be true if a different university or course is selected; this is expected in a case study. For example, the students who participate in the research may consist of only high-achievers. Causation cannot be proven for any effects that correlate to good student experience because the researcher has no control over extraneous variables that are not even fully known for a complex phenomenon like social interactions in the human learning process. There is no control group to be compared against, only multiple groups with the same treatment.

With that said, this does not mean that the data collected about student performance and experience is meaningless when trying to understand human learning. By comparing between these groups of students, the threat to internal validity can be mitigated through cross-verification of data. Trochim (1989) suggests that pattern-matching of empirical patterns to the predicted patterns can enhance internal and external validity, even when there are no quantitative measures. Regularities observed across multiple university units despite differences in student background, social life and personal interests are key points of discovery for generating hypotheses and a potential for more in-depth analysis on student learning.

3.5.2 Construct Validity

Patterns shown in such quantitative data may not represent actual behaviors of the subject and the results may be meaningless even when significant differences are observed. For example, if only statistical tests are relied upon, there is a risk of a false positive when determining if the robot tutor affects students' experiences, although student comments and responses are no different from human tutor sessions. This problem is mitigated through the use of methodology triangulation which relies on a comparison of data collected by the short version of the Flow State Scale-2, a subjective questionnaire, and observations on the students' experience.

3.5.3 External Validity

Usually, experimental studies attempt to generalize findings using statistical probability. It is common to hear complaints about the threat to external validity being high for case studies, yet the sampling methods used for this kind of study is clearly not meant to automatically generalize findings to a population. Yin (2013) explains that aside from the replications of results in similar cases, case studies can be generalized through theoretical abstraction by discussing how findings such as subjective experiences relate to broader issues, like things that could be done to improve those experiences. Yin (2013) gives the example of the origins of social class theory derived from a case study of Yankee City (Warner and Lunt, 1973) which significantly contributed to the understanding of social differences from "upper" to "lower" class citizens in broad situations. Another example by Yin is the derivation of urban planning theories about the role of neighborhood parks, sidewalks, small blocks and slums from Jacobs' (1992) book, *The Death and Life of Great American Cities* which is a single case study on the experiences in New York City.

Depending on the results, it may or may not be generalizable for all university students. As mentioned before, this research does not attempt to demonstrate causes and effects but instead to discover theories and mechanisms that drive the human-robot interaction within the classroom, which can help produce a model to enhance university learning with robots. It is very important to set aside the preconceived notion of trying to prove or disprove a theory. The model formed in this study is based on the data collected from 2 different units and the patterns and regularities observed between these different units of analysis can answer questions that are generalizable to university students to a certain extent.

Strictly speaking, it cannot be proven to be true for all persons, places or time due to the unknown or uncontrollable variables of the human psychology that will not be controlled or tested in this project. It can, however, serve as a foundation for researchers in future, if our technology and knowledge of human learning behaviors permit, test empirically on a larger scale. As Bennett (2004) suggests, cases are implicitly comparative and future case studies on the same topic can be compared to strengthen the external validity of the results obtained in this study.

3.5.4 Reliability

As a way of enhancing reliability, Yin (2013) proposed the case study protocol which this thesis follows, containing an overview of the project, field procedures (see Section 3.4.2 and Section 3.4.3), case study questions (see Section 1.2) and guidelines for the case study report; as well as a case study database containing organized records of data collected (Wong, 2017), documents used and reports written that are easily available and accessible. This maintains a chain of evidence, allowing external observers to trace the evidence from initial research questions to case study conclusions. This ensures that the case study procedures are repeatable for a different case, and those results can be compared with the results from this study to strengthen our understanding of student learning.

3.6 Limitations

This project only covers a specific context on Monash University students which limits the power to generalize outcomes of this research as patterns in human behavior. The qualitative data collected from Monash University students do not represent the general population of university students as they may have different values, beliefs and perspectives that are not accounted for in this research. No interviews on participants' background, past experiences and social lives will be collected in this research. The realm of social sciences is complex and the application of robot tutors in the higher education context is relatively new. A problem such as this requires a more exploratory research strategy and one that is capable of forming new theories and models such as a case study, which this research adopts. Especially for the education discipline, case studies have started gaining popularity in the past few decades (Stake, 1995).

Furthermore, this single case study looks at only two Information Technology (IT) units and therefore, the guidelines and suggestions formed lack the perspective and context from other fields of education. Moreover, the similarities and differences within the IT units alone are not representative of the whole IT field as they can be about completely different things. For a broad field such as IT, the researcher has to be careful when formulating theories or generalizations as these comparisons grant only a tiny perspective of the big picture. This study does not answer questions pertaining to the effectiveness of robot tutors; rather, it attempts to explore some of the ways in which a robot tutor may affect students experiences compared to traditional teaching by human tutors in the university setting.

Some of the limitations encountered when analyzing the data are discussed in Section 3.4.3. Triangulation is used to strengthen the validity of the results obtained in this mixed-methods research; however, biases are still inherent in the interpretation of students' responses and sentiments about the experiences. The field of educational psychology is complex and the questionnaire used in this study was not able to explain or identify certain behaviors, resulting in the need for speculation and assumptions from student and tutor comments. Moreover, all data collected in this study is self-reported by the participants, and it is assumed that they were aware and truthful of their own experiences. Lastly, although sentiment classification (see Section 3.4.3) is inherently subjective and is based on human judgment of the attitude of the respondent, an inter-rater agreement test could have been done to determine the reliability of the classification to better understand the limitations it has on this study.

3.7 Ethical Considerations

As this study deals with human subjects, the integrity of this research follows the Australian Code for the Responsible Conduct of Research by the National Health and Medical Research Committee (2007), a framework for good and ethical research practice. Upholding this code, the researcher pledges to always report truthfully on all data collected, respect the rights of all parties affected by the research, adopt appropriate methods with proper practices for safety, be accurate in citations, and conform to the policies adopted by Monash University, the institution which funds and oversees this research.

Specifically, this research involves human participants which directly requires the written approval of Monash University Human Research Ethics Committee to review the implications of this research for the following parties:

1. 63 undergraduate students from Monash University Students from the 2 selected units who agree to participate in the research will interact with the robot in a simple question and answer session on some topics. They will also fill up a questionnaire form at the end.
2. 2 tutors (1 for each of the selected units). Tutors from the 2 selected units will help provide some suitable topics and questions for the robot to ask students, as well as facilitate the robot tutor sessions and provide comments or observations of students' experiences of the session.

Students have 12 weeks of classes, and their consent was obtained in the 3rd week, after which data collection has started on the 4th week. As the tutor is also participating in the research, this student-teacher relationship may influence the consent process. In order to ensure that participation is fully voluntary, a third party who had no influence or authority over the students' grades for that unit was elected to brief the students about the research and ask for consent. The person briefing the students was simply given a script to read aloud. Students were specifically told about all details of the research procedure such as the robot's height and that there will be no physical games or activities with the robot aside from questions and answers.

In total, there are 63 consented participants for this study; all of whom are undergraduate students studying in Monash University, Malaysia. Majority, but not all participants are from the School of Information Technology. The study involves giving out a questionnaire about their experiences in a 15-minute session during the beginning of a class with or without a robot tutor, once a week for 8 weeks. The questionnaire may contain identifiable information because the participants are required to fill in a name, alias or handle which allows the researchers to track their responses throughout the 8 weeks. This information will be removed before the results are published. All details of the project and what information will be collected are made clear to the participants through an Explanatory Statement and a Consent Form already approved by the Monash University Human Research Ethics Committee (MUHREC) (see Appendix E). As an incentive, participants are given RM20 gift vouchers upon the completion of the study.

3.7.1 Principles of Integrity

Apart from the ethics review and approval from Monash University, the code requires compliance on principles of integrity, respect for persons, justice and beneficence for research participants. These principles were first defined in the historical Belmont Report (United States Department of Health and Human Services, 1979), summarizing ethical guidelines for research involving human subjects. Participants must be respected as autonomous persons receiving full disclosure of the study, procedures, risks and benefits with the extended opportunity to ask for explanations on matters regarding the research. This is ensured through an explanatory statement (see Appendix C) describing the details of this research as well as how data is collected and used. Participants are not coerced into this research just because they are the selected group for study. This project ensures voluntary participation by asking for their informed consent (see Appendix D). If students do not give their written consent, they will not be part of the study in any way.

Pertaining to justice or rights to service, this is not an issue for this research as all participants are subjected to the same conditions in the study. For medical research, there is a concern where the control group is given a placebo, making it unfair to participants as the treatment group is given real medical benefits. However, in this case, there is no control group because the aim of this experiment is not to determine effects but to explore a concept. While it is true that other Monash University students other than those within the selected units do not have an opportunity to participate in the research, there is no experimental treatment in this study that would provide a benefit to the participants over anyone else.

Beneficence is the maximization of possible benefits and reduction of possible harm. For human-robot interaction, it is possible that there is a psychological risk of discomfort with proximity despite the NAO humanoid robot's small size of 58cm in height. A distance between 1 to 2 meters is required from the robot to the student during the Q&A interaction. This distance is also suggested to be within the human social zone (Hall, 1966) which is reserved for face-to-face conversations. Participants were informed about this distance and can stop the interaction at any time if they feel uncomfortable with it. The robot will perform simple gestures but will not walk or move from its position on its own. Other than that, there are no foreseen risks on physical, social, economic or legal harms apart from minor inconveniences of filling up the questionnaire and consent forms. All university students participating in the study were given incentives for their time in the form of gift vouchers after all data collection procedures were completed. Thus, this research is considered as low-risk.

3.7.2 Data Management

All research data obtained in this study is stored and documented truthfully and accurately in the research record according to the Research Data Management: HDR Candidates Procedures (Monash University, 2010). No video or audio recordings were performed. The questionnaire responses may contain personally-identifiable information such as names or identification numbers and any of such data has been removed. Data collected from questionnaires and conversation transcripts during the human-robot interaction are non-confidential data that can be published as results. Non-confidential data will be made available and accessible for reference by other parties either through publications, web pages or both.

3.7.3 Authorship and Conflicts of Interest

Concerning authorship and data collection, the Research Outputs Data Collection Procedures (Monash University, 2013) is referred to when dealing with such issues. Monash University must be attributed as the affiliated institution in all research outputs. The researcher is aware that supervisors and members participating in the research including himself can claim authorship for any research outputs only when there is significant contribution in conception and design of the project, analysis and interpretation of data, or drafts and revisions contributing to the interpretation.

Conflicts of interest must be declared upfront and avoided so as to not compromise the integrity of the research. This interest can be any goal or value held by any persons involved in the research who personally benefits from a specific outcome of the research. There are no conflicts of interest to declare for this research.

Chapter 4

Results

Most sessions are affected to some degree by the teaching content, and some flow dimensions are more closely related to the contents of the lesson rather than the type of tutor giving the instructions. Experiences of boredom and frustration can be attributed to the lesson design and learning materials. This is observed in students' comments and comparisons between multiple sessions from different university units. Table 4.1 and Table 4.2 show the significant groups by flow dimension for paired and unpaired t-tests respectively. As seen in these tables, each t-test was carried out between two groups (weekly sessions) denoted in superscript by "A" for the first test group's week number, and "B" for the second test group's week number. For example, "Robot^{2}" indicates the second weekly session which was carried out by a robot tutor. These tests are then discussed and compared with the qualitative data to identify the extent of influence on flow which a robot tutor has in a university classroom compared to a human tutor.

4.1 Challenge-Skill Balance (Q1)

Almost no significant differences are found for challenge-skill balance (Q1) throughout the 8 weeks for all groups. There are only two cases in the paired t-test where a significant difference is observed for the Q1 section of Table 4.1 and none in unpaired t-tests; hence, there is no evidence to suggest that the introduction of a humanoid robot tutor affects this flow dimension.

#	Class	Week^{A}	Week^{B}	n	T	p
Q1	FIT1050 Tue	Human^{1}	Human^{3}	10	−2.450	0.037
	FIT1050 Wed	Robot^{1}	Human^{6}	14	−2.188	0.047
Q2	FIT1033 Mon	Robot^{1}	Human^{7}	7	−2.500	0.047
	FIT1033 Mon	Robot^{1}	Human^{8}	14	−2.511	0.026
	FIT1033 Mon	Robot^{2}	Human^{8}	14	−2.474	0.028
	FIT1050 Tue	Human^{2}	Robot^{7}	9	2.530	0.035
	FIT1050 Wed	Robot^{1}	Human^{5}	11	−2.283	0.046
	FIT1050 Wed	Robot^{1}	Human^{6}	14	−2.510	0.026
Q3	FIT1033 Mon	Robot^{1}	Robot^{4}	11	−2.887	0.016
	FIT1050 Tue	Human^{4}	Robot^{7}	10	3.000	0.015
	FIT1050 Wed	Robot^{1}	Human^{7}	9	−2.530	0.035
	FIT1050 Wed	Human^{7}	Human^{8}	9	2.530	0.035
Q4	FIT1050 Tue	Human^{1}	Robot^{6}	8	2.646	0.033
	FIT1050 Wed	Robot^{2}	Human^{8}	14	−2.589	0.022
	FIT1050 Wed	Robot^{4}	Human^{6}	12	2.345	0.039
	FIT1050 Wed	Human^{5}	Human^{8}	11	−2.283	0.046
	FIT1050 Wed	Human^{6}	Human^{8}	14	−2.876	0.013
Q5	FIT1033 Mon	Robot^{1}	Human^{7}	7	3.873	0.008
	FIT1033 Mon	Robot^{2}	Human^{7}	7	2.828	0.030
	FIT1033 Mon	Robot^{2}	Human^{8}	14	−2.511	0.026
	FIT1033 Mon	Robot^{3}	Human^{7}	7	2.500	0.046
	FIT1033 Mon	Human^{7}	Human^{8}	7	−3.873	0.008
	FIT1050 Tue	Human^{2}	Human^{4}	9	−3.411	0.009
	FIT1050 Wed	Robot^{4}	Human^{6}	12	2.803	0.017
Q6	FIT1033 Wed	Human^{1}	Human^{2}	6	−3.162	0.025
	FIT1050 Tue	Human^{1}	Robot^{6}	8	2.497	0.041
	FIT1050 Wed	Robot^{1}	Robot^{3}	11	−2.609	0.026
	FIT1050 Wed	Robot^{1}	Human^{8}	11	−3.130	0.011
	FIT1050 Wed	Robot^{2}	Human^{8}	14	−2.463	0.029
	FIT1050 Wed	Human^{6}	Human^{8}	14	−3.229	0.006
Q7	FIT1050 Tue	Human^{1}	Human^{4}	11	2.319	0.043
	FIT1050 Tue	Human^{1}	Robot^{7}	10	2.714	0.024
	FIT1050 Wed	Robot^{2}	Human^{8}	14	−2.280	0.040
	FIT1050 Wed	Robot^{3}	Human^{7}	9	−2.828	0.022
	FIT1050 Wed	Robot^{3}	Human^{8}	13	−2.919	0.013
Q8	FIT1050 Tue	Human^{1}	Robot^{5}	10	−3.279	0.009
	FIT1050 Tue	Human^{2}	Robot^{5}	10	−3.772	0.004
	FIT1050 Tue	Human^{3}	Robot^{5}	11	−3.730	0.004
	FIT1050 Tue	Human^{4}	Robot^{5}	11	−2.514	0.030
	FIT1050 Tue	Robot^{5}	Robot^{8}	9	2.828	0.022
	FIT1050 Tue	Robot^{7}	Robot^{8}	10	2.377	0.041
Q9	FIT1033 Mon	Robot^{1}	Human^{7}	7	2.500	0.047
	FIT1050 Tue	Human^{1}	Human^{4}	11	−2.390	0.038
	FIT1050 Tue	Human^{1}	Robot^{5}	10	−3.881	0.004
	FIT1050 Tue	Human^{1}	Robot^{6}	8	−2.553	0.038
	FIT1050 Tue	Human^{1}	Robot^{7}	10	−2.333	0.045
	FIT1050 Tue	Human^{1}	Robot^{8}	8	−2.646	0.033
	FIT1050 Tue	Human^{2}	Robot^{5}	10	−2.999	0.015
	FIT1050 Tue	Robot^{5}	Robot^{6}	9	2.401	0.043
	FIT1050 Tue	Robot^{5}	Robot^{7}	10	4.000	0.003
	FIT1050 Tue	Robot^{5}	Robot^{8}	9	3.592	0.007

Table 4.1: Paired t-tests with p-values < 0.05

#	Class	Week^{A}	Week^{B}	n^{A}	n^{B}	T	p
Q2	FIT1033 Mon	Robot^{2}	Human^{8}	14	15	−2.216	0.036
Q3	FIT1050 Tue	Human^{3}	Robot^{5}	12	12	2.292	0.036
	FIT1050 Tue	Human^{3}	Robot^{7}	12	13	2.275	0.033
	FIT1050 Tue	Human^{3}	Robot^{8}	12	12	2.253	0.036
Q4	FIT1050 Wed	Robot^{2}	Human^{8}	17	15	−2.204	0.035
	FIT1050 Wed	Robot^{4}	Human^{5}	13	13	2.216	0.037
	FIT1050 Wed	Human^{5}	Human^{8}	13	15	2.389	0.025
Q5	FIT1033 Mon	Robot^{2}	Human^{8}	14	15	−2.339	0.028
	FIT1033 Mon	Robot^{3}	Human^{7}	15	7	2.726	0.021
	FIT1033 Mon	Robot^{4}	Human^{8}	11	15	−2.453	0.022
	FIT1033 Mon	Human^{7}	Human^{8}	7	15	−3.726	0.002
	FIT1050 Wed	Robot^{4}	Human^{6}	13	16	2.423	0.022
	FIT1050 Wed	Robot^{4}	Human^{8}	13	15	2.469	0.020
	Overall	Robot	Human	170	148	2.027	0.044
Q6	FIT1050 Tue	Human^{1}	Robot^{6}	13	10	2.217	0.038
	FIT1050 Tue	Human^{3}	Robot^{6}	12	10	2.769	0.013
	FIT1050 Tue	Human^{3}	Robot^{8}	12	12	2.283	0.033
	FIT1050 Wed	Human^{6}	Human^{8}	16	15	−2.563	0.016
Q7	FIT1050 Tue	Human^{1}	Human^{4}	13	13	2.265	0.033
	FIT1050 Tue	Human^{1}	Robot^{6}	13	10	2.800	0.011
	FIT1050 Tue	Human^{1}	Robot^{8}	13	12	2.704	0.013
	FIT1050 Tue	Robot^{5}	Robot^{6}	12	10	2.132	0.047
	FIT1050 Wed	Robot^{1}	Human^{8}	15	15	−2.080	0.047
	FIT1050 Wed	Robot^{2}	Human^{8}	17	15	−2.213	0.035
	FIT1050 Wed	Robot^{3}	Human^{8}	15	15	−2.806	0.009
Q8	FIT1050 Tue	Human^{1}	Robot^{5}	13	12	−3.735	0.001
	FIT1050 Tue	Human^{1}	Robot^{6}	13	10	−2.483	0.022
	FIT1050 Tue	Human^{1}	Robot^{7}	13	13	−2.550	0.018
	FIT1050 Tue	Human^{2}	Robot^{5}	12	12	−4.423	0.000
	FIT1050 Tue	Human^{2}	Robot^{6}	12	10	−3.164	0.005
	FIT1050 Tue	Human^{2}	Robot^{7}	12	13	−3.179	0.004
	FIT1050 Tue	Human^{3}	Robot^{5}	12	12	−3.352	0.003
	FIT1050 Tue	Human^{3}	Robot^{6}	12	10	−2.301	0.033
	FIT1050 Tue	Human^{3}	Robot^{7}	12	13	−2.392	0.026
	FIT1050 Tue	Human^{4}	Robot^{5}	13	12	−3.034	0.006
	FIT1050 Tue	Robot^{5}	Robot^{8}	12	12	3.130	0.005
	Overall	Robot	Human	170	149	2.970	0.003
Q9	FIT1033 Mon	Robot^{3}	Robot^{4}	15	11	−2.185	0.039
	FIT1050 Tue	Human^{1}	Robot^{5}	13	12	−4.450	0.000
	FIT1050 Tue	Human^{2}	Robot^{5}	12	12	−3.664	0.001
	FIT1050 Tue	Human^{3}	Robot^{5}	12	12	−2.150	0.046
	FIT1050 Tue	Human^{4}	Robot^{5}	13	12	−2.584	0.018
	FIT1050 Tue	Robot^{5}	Robot^{6}	12	10	2.211	0.041
	FIT1050 Tue	Robot^{5}	Robot^{7}	12	13	2.594	0.016
	FIT1050 Tue	Robot^{5}	Robot^{8}	12	12	3.344	0.003
	Overall	Robot	Human	170	149	2.296	0.025

Table 4.2: Unpaired t-tests with p-values < 0.05

4.2 Merging of Action and Awareness (Q2)

For merging of action and awareness (Q2), most robot sessions score significantly lower compared to human sessions and this is clearly seen in the paired t-tests shown in the Q2 section of Table 4.1. There are also no observed cases where robot sessions score significantly higher than human sessions. These differences are only observed in weekly t-tests and no significant difference is observed if an overall test is performed. This is especially true for the first week the robot is introduced; therefore, it is possible that students require time to feel comfortable with the new tutor and lesson structure.

4.3 Clear Goals (Q3)

Similar to the previous section, it is also observed for clear goals (Q3) where in some cases, robot sessions scored significantly lower than human sessions but never the opposite. In the FIT1050 Wednesdays 8 a.m. group, the robot tutor in week 1 scored significantly lower than the human tutor in week 7 ($T = -2.530$, $p = 0.035$; see Q3 section of the paired t-tests in Table 4.1). Throughout the weeks, many students commented on the unnatural speech produced by the robot as there is a lack of intonation in the generated text-to-speech audio. There are many suggestions by the students which ask for the speed of the robot's speech to be changed. There is a consistent observation across multiple units but its occurrence is not frequent enough. It is suggested that the introduction of a robot tutor negatively affects the "clear goals" flow dimension depending on the complexity of the task in which the robot tutor is involved in. For example, in FIT1050 Tuesdays 3 p.m. group, the human tutor in week 3 which is about an assignment briefing (easy task), scored significantly higher than the robot tutors in weeks 5, 7 and 8 which is about coding and semantic web technologies (difficult tasks) as seen in the unpaired t-tests in Table 4.2. Since no significant differences were observed for simpler instructions, the robot tutor should be able to communicate the goals of low-complexity tasks to the students without problems.

4.4 Unambiguous Feedback (Q4)

Unambiguous feedback (Q4) refers to whether or not the student is aware of his/her performance during the task. In the context of this study, it is mostly a question of whether or not the tutor is capable of invoking awareness of what the student knows or doesn't know about the discussed topic. If the student asks many questions to clarify what to do, then the Q4 score is usually lower. There are significant differences between robot and human tutor for various weeks as shown in the Q4 section of Table 4.1 and Table 4.2, but not much can be said as there is not much consistency to the differences to be compared. For example in a paired t-test of 12 students in the FIT1050 Wednesdays 8 a.m. group, the robot tutor in Week 4 assignment briefing seems to have obtained a significantly higher Q4 score than the human tutor in week 6 teaching web accessibility ($T = 2.345$, $p = 0.039$). Also, week 6 is significantly lower than week 8 which is about web semantics ($T = -2.876$, $p = 0.013$); both weeks are taught by a human tutor. As the same sort of differences are also observed between human tutor sessions, it does not suggest that the robot tutor had any extra effect on this flow dimension.

Besides that, there are some human tutor sessions which score significantly higher than robot tutor sessions. In another paired t-test of 8 students in the FIT1050 Tuesdays 3 p.m. group, the human tutor session in Week 1 about web design scores significantly higher than the robot in week 6 on web accessibility ($T = 2.646$, $p = 0.033$). Week 6 on the web accessibility topic seems to have significantly lower Q4 score in both FIT1050 groups. Due to the inconsistencies in differences across multiple units, it may seem that unambiguous feedback has more to do with other factors such as the nature of the task rather than the type of tutor. Nevertheless, there is a lack of evidence suggesting any effect of the type of tutor on this flow dimension.

4.5 Concentration of Task at Hand (Q5)

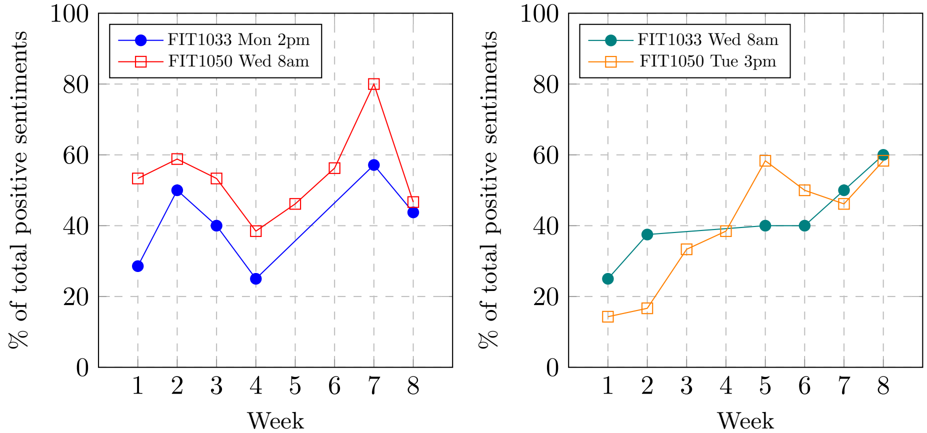
As for the concentration of task at hand (Q5), it seems to be greatly influenced by other factors. In a paired t-test of 9 students from weeks 2 and 4 of the FIT1050 Tuesdays 3 p.m. group, which are both carried out by a human tutor on assignment briefing and preparation, there seems to be a significant difference ($T = -3.411$, $p = 0.009$). This one in particular is tricky to explain as they are almost identical in many ways, even the students' responses. This could suggest that the concentration of the task is affected by the teaching materials, because in week 2, there is a student who suggested to slow the pace of instruction as well as another student who commented that the content of the tutorial class made him bored. In the FIT1033 Mondays 2 p.m. group, Week 8 scores significantly higher than weeks 2, 4 and 7 in both paired and unpaired t-tests, as it consists of a class test that requires students to focus on performing the task on their own; therefore, this is an outlier that could be omitted from the analysis. On the other hand, the week 4 robot tutor session about assignment briefing scored significantly higher than the week 6 human tutor session about web accessibility in the FIT1050 Wednesdays 8 a.m. group. In addition, an overall unpaired t-test between all robot tutor sessions (170 responses) and human tutor sessions (148 responses) regardless of unit revealed that robot tutor sessions significantly evoke more concentration on the task at hand ($T = 2.027$, $p = 0.044$). This may be the case as the presence of the robot attracts attention, as well as the unnatural speech which requires more concentration from the students in order to understand its instructions. Although concentration on the task is also dependent on the nature of the task (i.e. whether the task is of interest to the student or not), on average, there is evidence suggesting that students have higher concentration with the robot tutor compared to human tutors across multiple university units.

4.6 Sense of Control (Q6)

The results for sense of control (Q6) is similar to Section 4.2 where there are only a few cases where robot sessions score significantly lower but never vice versa. Weeks 1 and 2 of the FIT1050 Wednesdays 8 a.m. group taught by the robot tutor scored significantly lower than week 8 which was taught by a human tutor in the paired t-tests (p-values of 0.011 and 0.029 respectively). In addition, week 6 of the FIT1050 Tuesdays 3 p.m. group taught by the robot tutor also scored significantly lower than weeks 1 and 3 taught by a human tutor in the unpaired t-tests (p-values of 0.038 and 0.013 respectively). It is expected for these observations to always be influenced by other factors, yet these observed differences should not be completely disregarded especially with support from qualitative data such as the students' comments. Hence, it can be said that the introduction of a robot tutor has a weak negative influence on the sense of control due to the slightly more rigid lesson structure introduced by the robot.

4.7 Loss of Self-Consciousness (Q7)

Loss of self-conciousness (Q7) is significantly lower for robot sessions compared to human sessions, and this is consistent across two study groups of the same university unit, FIT1050. Throughout the study, it is observed that students were quite reluctant to interact with the robot. For almost all sessions, the human facilitator was required to encourage students to participate and assure them that they are free to choose any answer as it is simply to test the robot rather than their knowledge. There are 3 responses which indicate that reason for this phenomenon, is the feeling of embarrassment. The causes of this embarrassment are: (1) the students are still afraid of answering wrongly; and more interestingly, (2) the robot not being able to detect the student's response. One student wrote that, "the feeling of not getting a response from the robot is shameful" and another said, "I was scared that the robot might fail to detect my response." Instead of a feeling of disappointment, why is it that students feel embarrassed when it is the robot's fault for not being able to detect their response? The interaction with the robot requires the student to come forward, focusing the attention of the entire class on the student and the robot. This is different from the interaction with human tutors, where the student remains



(a) Weeks 1-4 carried out with robot tutor, 5-8 by human.

(b) Weeks 1-4 carried out by human tutor, 5-8 with robot.

Figure 4.1: Habituation effect and similar trend of positive sentiments over 8 weekly sessions.

seated among the crowd while the tutor asks questions to the audience in a manner similar to an open discussion. Under the limelight with the robot, and based on the students' responses, it is likely that they are simply afraid that they might not have spoken loud or clear enough; thus, reflecting poorly on their ability and being caught in an awkward situation feeling out of place.

4.8 Transformation of Time (Q8)

Transformation of time (Q8) had the most significant impact in the FIT1050 Tuesdays 3 p.m. group on Week 5, during the switch from the human tutor to the robot tutor. In the unpaired t-test (each group having a sample size of 12 students), very significant differences were detected for all human tutor sessions, Weeks 1 to 4 as compared to Week 5 (p-values < 0.01). Interestingly enough, significant differences were also consistently found for the following weeks 6 and 7. When looking at the students' comments, it is rather clear that this is due to the excitement and anticipation of finally being able to interact with the robot for the first time, but this effect gradually wears off after each week as the students get habituated to the same robot lesson structure throughout 4 weeks. Figure 4.1a shows the percentage of positive sentiments out of about 15 students (the number of students depends on the week) gradually decreasing for the first 4 weeks carried out with a robot tutor, and the trend is similar for both FIT1050 Wednesdays 8 a.m. and FIT1033 Mondays 2 p.m. groups which start the first 4 weeks with a robot tutor, followed by the next 4 weeks with

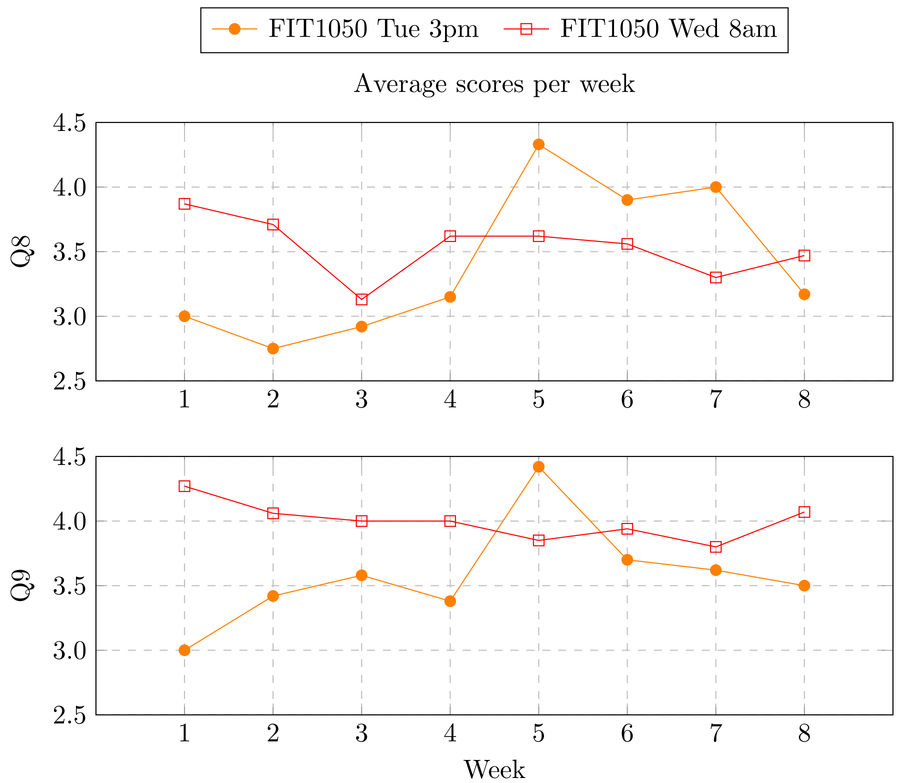


Figure 4.2: Transformation of time (Q8) and autotelic experience (Q9) average scores over 8 weekly sessions for FIT1050 groups.

a human tutor. No significant difference is observed for this flow dimension in the Wednesday group of the same unit, which highly suggests that anticipation or hype is the primary cause of the session being perceived as more engaging and rewarding. In fact, if the first 4 weeks are carried out by a human tutor, followed by the next 4 with a robot tutor, allowing anticipation to build among about 12 students (the number of students depends on the week), the positive trend gradually increases as seen in Figure 4.1b.

4.9 Autotelic Experience (Q9)

Like the previous section, the same is true for autotelic experience (Q9) albeit its effects last only 1 week rather than through the course of 3 weeks. Figure 4.2 shows the sharp change in averages during the robot tutor switch from week 4 to week 5 for the FIT1050 Tuesdays 3 p.m. group. In order to fully take advantage of the autotelic experience flow dimension, there should be variation in activity with the robot, or gradually revealing more features and capabilities of the robot. This study did not take advantage of that and instead opted for a more traditional instructional approach throughout all weeks in order to strike a better comparison with the human tutors in this study.

A summary of how robot tutor sessions affect the 9 flow dimensions are shown in Table 4.3. For each flow dimension, if there is a high consistency in both the quantitative and qualitative data that can be used as supportive evidence, it is said to have a strong influence. A plus (**+**) sign indicates that the robot tutor performs better than human tutors in that flow dimension whereas a minus (**−**) sign indicates otherwise.

4.10 Sentiment, Boredom and Student Comments

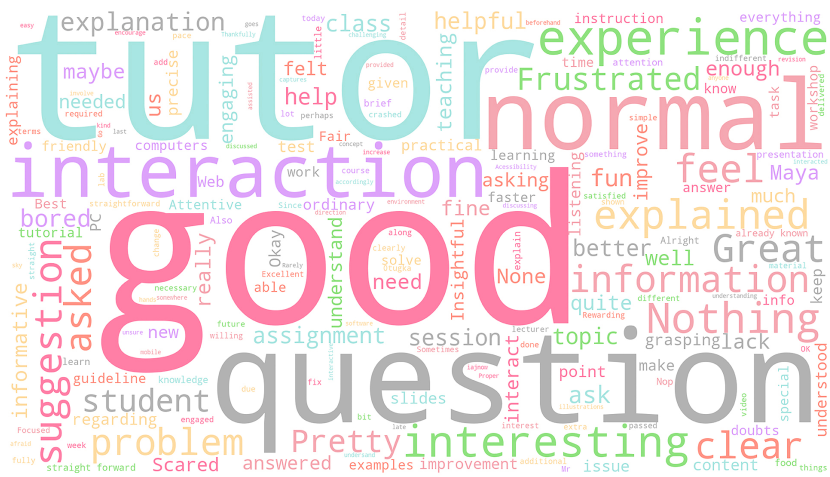
Lastly, a chi-square test between the type of tutor and sentiment yields no significant difference ($\chi^2 = 2.256, p = 0.521$). The same result is observed for experiences of boredom ($\chi^2 = 0.37, p = 0.543$) as there are a rather equal experience of excitement and boredom for both types of tutors. For robot tutor sessions, many students find it interesting (34 responses) but it also has a significantly higher number of suggestions for improvement ($\chi^2 = 28.23, p = 1.077 \times 10^{-7}$). There were 16 suggestions from the students' responses to improve the robot's speech as it is quite difficult to catch due to the unnatural tone, pace and lack of emotion in its voice; and 10 suggestions to improve its speech recognition when interacting with students. In general, students expect the robot to be more human-like in its operations. There are a few suggestions which ask to lengthen the time spent with the robot, as well as to make the robot move or walk around more. Students' comments regarding the robot tutor are visualized in Figure 4.3a.

As for human tutors, there is significantly more interaction compared to the robot tutor ($\chi^2 = 32.178, p = 1.408 \times 10^{-8}$) as the robot follows a more rigid, programmed lesson structure and cannot interact with multiple students at once. Students are able to freely ask questions to the human tutor and this is reflected quite clearly in their comments as seen in Figure 4.3b. While both types of tutors are able to present the teaching material in an informative way, the freedom of interaction is much more limited for the robot tutor; hence, proper integration and lesson design guidelines are needed.

#	Flow Dimension	Influence	Comments
Q2	Merging of Action and Awareness (Section 4.2)	Weak −	Students are slightly less spontaneous in their actions when the robot is first introduced.
Q3	Clear Goals (Section 4.3)	Weak −	Although partly due to instruction quality, students are not able to understand the task clearly due to unnatural, robotic speech.
Q5	Concentration of Task at Hand (Section 4.5)	Weak +	Students in general are more concentrated on the task when a robot tutor is involved, although it also depends on their interest on the task.
Q6	Sense of Control (Section 4.6)	Weak −	In some cases, the involvement of a robot in the task may force a more rigid lesson structure.
Q7	Loss of Self-Consciousness (Section 4.7)	Weak −	Students are more concerned with how others may think of them if/when they are interacting with the robot.
Q8	Transformation of Time (Section 4.8)	Strong +	Due to anticipation, students feel that their time spent with the robot passes more quickly. Gradually loses effect over several sessions.
Q9	Autotelic Experience (Section 4.9)	Strong +	Due to anticipation, students perceive their session with the robot as more rewarding and interesting. One-time effect.
Q1	Challenge-Skill Balance (Section 4.1)	None	Highly dependent on the task.
Q4	Unambiguous Feedback (Section 4.4)	None	Lack of consistent evidence.

Table 4.3: Summary of flow dimensions affected by the robot tutor, each corresponding to a question in the Flow State Scale-2 (see Appendix B).

(a) Robot tutor sessions.



(b) Human tutor sessions.

Figure 4.3: Word cloud of student comments from the questionnaire responses. The top 50 most frequent word counts are shown in Appendix F

Chapter 5

Discussion and Key Recommendations

Aside from the research constraints described in Section 3.6, this study revealed some technical limitations for robotic hardware and in human-robot interaction which affects students' experiences of flow. During the study, there were occurrences where the students simply could not speak loud enough for the robot to detect. Such issues inevitably extend to people with disabilities and therefore, the problem of accessibility in human-robot interaction should be considered. For example, voice interaction should not be the only method of interaction with the robot. Instead, students should be given the option to communicate with the robot using gestures or touch. There exists an accessibility-aware robot system that looks at static body language which could prove to be a useful solution to this problem (McColl et al., 2017). Furthermore, many students reported that the robot's speech is unnatural and can be difficult to listen to or understand. **If the robot is integrated with the learning environment, it is possible for the projector screen to be showing the captions or script in which the robot is reading from**. In this study, the NAO robot's speech was required to be set at 80% speed. When the speed was set to 70% or 90%, students reported that the pacing was either too slow or too fast. In order to better represent the 80% speed of the NAO robot, a sample text of 38 words in 3 sentences was completed by the robot in approximately 20 seconds; therefore, the recommended speed for robot instruction in university tutorial classes based on this case study is 114 words per minute. The sample text used is as shown below:

> Hello class, and welcome to today's lesson where we will be learning about robots. Did you know that the recommended speed of robot speech in university classes is currently being looked into? Thank you for listening, and goodbye.

Currently, the robot is very segregated from the learning environment, is never connected to the computers in the classroom and mostly requires a manual initiation. The robot needs to be part of the learning environment. It needs to be integrated with computers to be able to proceed with the slideshow presentation automatically without manual intervention, so as to not break the flow of instruction. This was suggested by some of the students, as well as the tutor of the FIT1033 unit. Aside from that, students lack the opportunity to interact with the robot and it is mostly limited to 1-to-1 interactions. Students were also generally reluctant to participate when this 1-to-1 interaction structure was imposed. This fear of getting the wrong answer may likely be caused by the competitive and collectivist Chinese culture which is also observed in Malaysia (Hodkinson and Poropat, 2014). A possible solution is to integrate web forms for students to submit their responses, while still be open to face-to-face interactions with the robot. A server, or the robot itself could process these responses and generate a response in return. This is one way of interacting with all the students, and serves as a workaround to the unnatural voice and speech recognition problem. Another suggestion would also be to reconsider the role of robots in classrooms as peers rather than tutors. In one study, 28 adults subjected to the peer condition actually graded the robot highly as a tutor compared to other conditions; and since the perceived authority of the robot is positively correlated with the clarity with which it expresses itself (Blancas et al., 2015), the limitations of the robot speech as seen in this study would suggest that the robot interactions should be redesigned to be less authoritative but more cooperative in nature.

The robot should be greatly taken into consideration during the design of the lesson structure. Currently, any existing teaching materials were just directly generated using text-to-speech with minor changes and even so, this additional amount of work on the tutor's part is one of the barriers to integration. Therefore, an automated solution or a standard for lesson planning should be established. Also, more consideration should be put into how the robot can act as an assistant in delivering the instructions to the class. In doing so, the robot should also be equipped with more ready-to-use functionality and gestures which are more suitable for educational environments, such as to grab or pick up items, and pointing at things. Simple actions like these should be at a finger's reach through a simple device such as a remote control or a smartphone application to allow the tutor to control the robot directly on demand. This way, a desired response can be triggered easily so as to not break the flow of the lesson. Any unwanted or unnatural responses from the robot when the student interacts with it can be prevented if the tutor has more direct control of the robot. With the lack of control, it was difficult to implement walking in a dynamic learning environment and the need of a robust walk controller in robots such as NAO has been expressed in the past (Shamsuddin et al., 2011). Furthermore, there is a lack of separate and parallel control between speech and physical movement of the NAO robot; hence, programming and graphical user interface (GUI) architectures in educational robotics should aim to facilitate

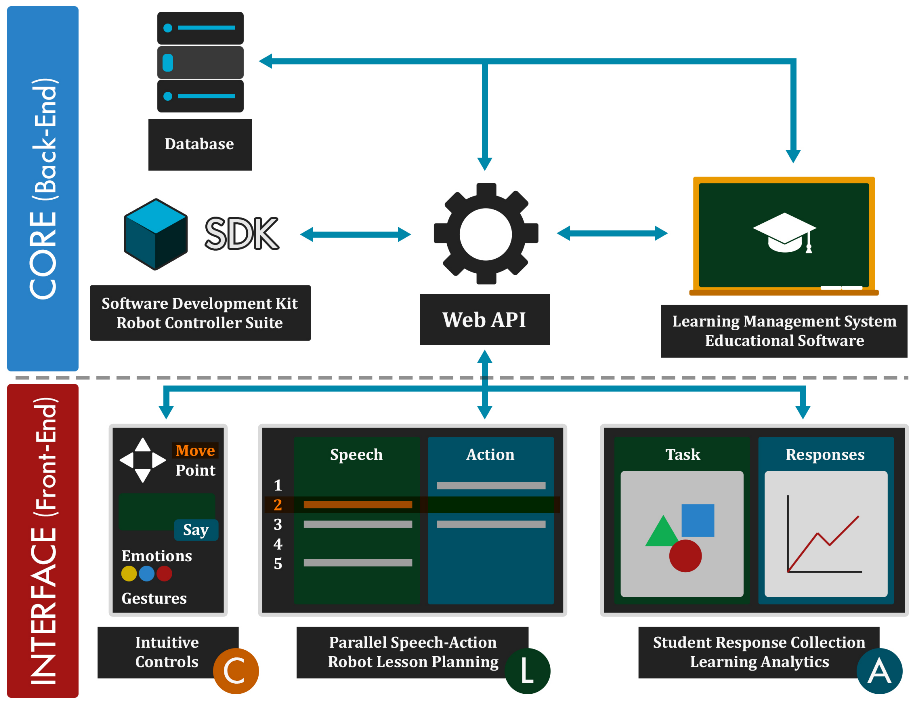


Figure 5.1: Example core API functionality and interface design for developers as part of a suggested robot integration framework for universities.

this process.

With all the suggestions combined, at least from this case study, there is a need for a web-based Application Programming Interface (API) which externalizes the functions of robot control and educational software to other devices (see Figure 5.1, with blue lines and arrows indicating data connections, such as through Local Area Network (LAN) or the Internet); promoting an Internet-of-Things (IoT) infrastructure where all devices in the classroom can be interconnected. Without such integration, the robot is mostly segregated from the lesson design process as well as the learning environment, limiting its ability as an educational robot.

The robot needs access to the university's servers and storage to contribute any data it can collect from the students or fetch lesson plans; and be controlled or interact with educators and students in a variety of ways. With this system in place, Figure 5.2 illustrates a possible framework which outlines the integration setup for robot-assisted lessons in the university learning environment. In this figure, blue lines indicate data connections; red lines indicate real-world interaction which can be through speech, touch or gestures; the yellow line indicates the light projected from the projector; and the grey line indicates the student's

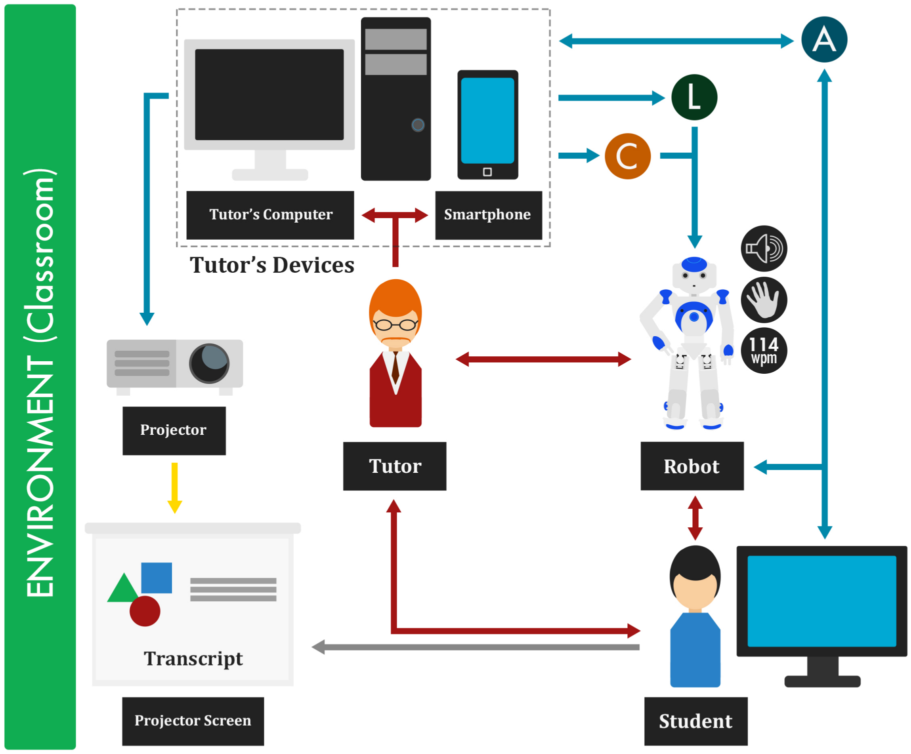


Figure 5.2: A suggested robot integration setup in a university learning environment for educators (C, L and A nodes correspond to the interface in Figure 5.1).

non-interactive observation of the projector screen.

The tutor can interact with the Web API shown in Figure 5.1 through the tutor's computer in the classroom, or with the tutor's own smartphone device (through a browser, an application, or any program that can access the web) to use the control interface (orange-colored node labeled 'C' shown in Figure 5.1 and Figure 5.2) which processes the necessary robot functions in the back-end to be translated into commands which the robot understands, and these commands are finally sent to the robot to be carried out. The tutor also has access to the "Parallel Speech-Action Robot Lesson Planning" functions (teal-colored node labeled 'L') which assigns a fixed set of actions and instructions for the robot to teach the class. "Parallel Speech-Action" is a term defined in this study to describe that actions of the robot should be executed separately from its speech, and parallel processing of these two should allow the robot to instruct the class more fluidly and effectively.

In addition, the interface for creating robot-based lessons should be designed with this concept in mind to mitigate the difficulties of robot tutor integration. Lastly, the functionality to collect student responses and perform learning analytics (blue-colored node labeled ‘A’) is needed to allow the university servers to not only receive inputs from all students at once through their computers, but also to process and analyze these responses, as well as any other inputs from the robot’s cameras and sensors, allowing for various responses from the robot, as well as assessments on student performance or experience in the learning environment.

Chapter 6

Conclusion and Future Work

In summary, compared to human tutor sessions, the introduction of a robot tutor in university tutorial classes positively affects: concentration of task at hand, transformation of time and autotelic experience; but imposes a slightly more rigid lesson and interaction structure which negatively affects: merging of action and awareness, clear goals, sense of control and loss of self-consciousness. Many students commented that the robot was fun, interesting and felt more rewarding although the introduction of the robot tutor made the lesson structure more rigid. On the other hand, many positive comments for the human tutor are about good social interaction and the ability to question freely about the lesson contents, which the robot is unable to cater for.

There are strengths and weaknesses to both types of tutors, but we can take advantage of the incorporation of robots in classrooms despite its limitations to enhance learning. A robot's role as an assistant in the classroom can be effective in providing social support, attracting students' attention and making the lesson feel more exciting. With the human tutor still being the main driving force behind the lesson, the drawbacks of robot tutors can be mitigated if more thought is put into its use in the learning environment. Therefore, an integration framework was proposed in this study to establish guidelines and recommendations in the design and implementation for robot-assisted teaching in university classes.

The strongest finding of this study is in discovering the extent in which robots affect university students' experiences of flow compared to human tutors (summarized in Table 4.3), as well as the recommended solutions to key technical limitations of robotic hardware such as unnatural speech (best set to 114 words per minute) and gesture or touch capabilities for accessibility issues during human-robot interaction.

In future, more applied research of robotics should be carried out in classrooms. By doing so, more of such case studies can strengthen or weaken the suggestions from this study which will provide us with a clearer understanding of how the introduction of robot tutors affect student learning experience. More of such studies also help expand the framework for good robot-classroom integration to understand and even enhance the role of robotics in pedagogy, as well as to take advantage of it effectively. One of such advancements is ARTIE, an integrated environment for the development of affective robot tutors with an architectural pattern to integrate emotional assessment in educational software driven by the robot's emotional pedagogical support (Cuadrado et al., 2016). As the field of educational psychology is broad, a review of these architectures with the focus shifted to the higher education context is needed, as adults may interact differently with robots compared to children, along with other social and cultural factors which may be exacerbated by age.

As more robots are integrated into classrooms, and if future artificial intelligence technology permits, the impact of human-robot interaction to educational performance and student learning experiences can be further explored with real-time data collection and learning analytics with more focus on speech recognition and facial emotion analysis (Wong et al., 2016) as emotions are key drivers of learning (Rienties and Rivers, 2014), yet have received little notice in educational research especially in the field of learning analytics (Artino, 2012). Within the 15-minute sessions of this study, students reported to have instances of boredom but yet still experience fun in the session which proves tricky to investigate through traditional data collection methods; hence, real-time learning analytics is suggested as one of the core functions in the integration framework (see Figure 5.1). As such, emotion assessment is an area which can highly benefit the field of pedagogy, and its application to educational robotics will move us a step closer to the social integration of robots in classrooms.

Bibliography

Akhtar, S., Warburton, S. and Xu, W. (2017), 'The use of an online learning and teaching system for monitoring computer aided design student participation and predicting student success', *International Journal of Technology and Design Education* **27**(2), 251–270. doi: 10.1007/s10798-015-9346-8.

Alemi, M., Meghdari, A. and Ghazisaedy, M. (2014), 'Employing humanoid robots for teaching english language in Iranian junior high-schools', *International Journal of Humanoid Robotics* **11**(03), 1450022. doi: 10.1142/S0219843614500224.

Alemi, M., Meghdari, A. and Ghazisaedy, M. (2015), 'The impact of social robotics on L2 learners' anxiety and attitude in english vocabulary acquisition', *International Journal of Social Robotics* **7**(4), 523–535. doi: 10.1007/s12369-015-0286-y.

Altin, H. and Pedaste, M. (2013), 'Learning approaches to applying robotics in science education', *Journal of Baltic Science Education* **12**(3), 365–377.
URL: *http://www.scientiasocialis.lt/jbse/files/pdf/vol12/365-377.Altin_JBSE_-Vol.12.3.pdf* [Accessed 4 December 2017]

Alves-Oliveira, P., Ribeiro, T., Petisca, S., di Tullio, E., Melo, F. S. and Paiva, A. (2015), *An Empathic Robotic Tutor for School Classrooms: Considering Expectation and Satisfaction of Children as End-Users*, *in* A. Tapus, E. André, J.-C. Martin, F. Ferland and M. Ammi, eds, 'Social Robotics: 7th International Conference, ICSR 2015, Paris, France, October 26-30, 2015, Proceedings', Springer International Publishing, Cham, pp. 21–30. ISBN: 978-3-319-25554-5. doi: 10.1007/978-3-319-25554-5_3.

Anaya, A. R., Luque, M. and Peinado, M. (2016), 'A visual recommender tool in a collaborative learning experience', *Expert Systems with Applications* **45**, 248–259. doi: 10.1016/j.eswa.2015.01.071.

Ardito, G., Mosley, P. and Scollins, L. (2014), 'We, robot: Using robotics to promote collaborative and mathematics learning in a middle school classroom', *Middle Grades Research Journal* **9**(3), 73–88.
URL: *http://blogs.ubc.ca/roboted/files/2015/07/using-roboticcs-to-promote-collaboration-and-learning-in-middle-school.pdf* [Accessed 4 December 2017]

Artino, A. R. (2012), 'Emotions in online learning environments: Introduction to the special issue', *The Internet and Higher Education* **15**(3), 137–140. doi: 10.1016/j.iheduc.2012.04.001.

Benitti, F. B. V. (2012), 'Exploring the educational potential of robotics in schools: A systematic review', *Computers & Education* **58**(3), 978–988. doi: 10.1016/j.compedu.2011.10.006.

Bennett, A. (2004), 'Case study methods: Design, use, and comparative advantages', *Models, Numbers, and Cases: Methods for Studying International Relations* pp. 19–55.
URL: *https://pdfs.semanticscholar.org/7d11/098671a75e7b289fd65adab2eb236c5cf580.pdf* [Accessed 2 December 2017]

Berland, M., Davis, D. and Smith, C. P. (2015), 'AMOEBA: Designing for collaboration in computer science classrooms through live learning analytics', *International Journal of Computer-Supported Collaborative Learning* **10**(4), 425–447. doi: 10.1007/s11412-015-9217-z.

Berliner, D. C. (2002), 'Comment: Educational research: The hardest science of all', *Educational Researcher* **31**(8), 18–20. doi: 10.3102/0013189X031008018.

Bers, M. U., Flannery, L., Kazakoff, E. R. and Sullivan, A. (2014), 'Computational thinking and tinkering: Exploration of an early childhood robotics curriculum', *Computers & Education* **72**, 145–157. doi: 10.1016/j.compedu.2013.10.020.

Bers, M. U., Ponte, I., Juelich, C., Viera, A. and Schenker, J. (2002), 'Teachers as designers: Integrating robotics in early childhood education', *Information Technology in Childhood Education Annual* **2002**(1), 123–145.
URL: *http://integratingengineering.org/stem/research/item1_earlychildhood_design-course_BersITCE.pdf* [Accessed 3 December 2017]

Bhaskar, R. (2013), *A realist theory of science*, Routledge. ISBN: 978-1844672042.

Bilotta, E., Gabriele, L., Servidio, R. and Tavernise, A. (2009), *Edutainment Robotics as Learning Tool*, *in* Z. Pan, A. D. Cheok, W. Müller and M. Chang, eds, 'Transactions on Edutainment III', Springer, Berlin, Heidelberg, pp. 25–35. ISBN: 978-3-642-11245-4. doi: 10.1007/978-3-642-11245-4_3.

Blancas, M., Vouloutsi, V., Grechuta, K. and Verschure, P. F. M. J. (2015), *Effects of the Robot's Role on Human-Robot Interaction in an Educational Scenario*, *in* S. P. Wilson, P. F. Verschure, A. Mura and T. J. Prescott, eds, 'Biomimetic and Biohybrid Systems: 4th International Conference, Living Machines 2015, Barcelona, Spain, July 28 - 31, 2015, Proceedings', Springer International Publishing, Cham, pp. 391–402. ISBN: 978-3-319-22979-9. doi: 10.1007/978-3-319-22979-9_39.

Brown, L. and Howard, A. M. (2014*a*), 'Assessment of engagement for intelligent educational agents: A pilot study with middle school students'.
URL: *http://hdl.handle.net/1853/53771* [Accessed 2 December 2017]

Brown, L. N. and Howard, A. M. (2014*b*), The positive effects of verbal encouragement in mathematics education using a social robot, *in* '2014 IEEE Integrated STEM Education Conference', pp. 1–5. doi: 10.1109/ISECon.2014.6891009.

Bryman, A. (2015), *Social research methods*, 5th edn, Oxford University Press. ISBN: 978-0199689453.

Catlin, D. (2014), *Using Peer Assessment with Educational Robots*, *in* Y. Cao, T. Väljataga, J. K. Tang, H. Leung and M. Laanpere, eds, 'New Horizons in Web Based Learning: ICWL 2014 International Workshops, SPeL, PRASAE, IWMPL, OBIE, and KMEL, FET, Tallinn, Estonia, August 14-17, 2014, Revised Selected Papers', Springer International Publishing, Cham, pp. 57–65. ISBN: 978-3-319-13296-9. doi: 10.1007/978-3-319-13296-9_6.

Chen, G.-D., Chuang, C.-K. et al. (2012), 'When a classroom is not just a classroom: Building digital playgrounds in the classroom.', *Turkish Online Journal of Educational Technology-TOJET* **11**(1), 202–211.
URL: *http://www.tojet.net/articles/v11i1/11119.pdf* [Accessed 4 December 2017]

Chin, K. Y., Hong, Z. W. and Chen, Y. L. (2014), 'Impact of using an educational robot-based learning system on students' motivation in elementary education', *IEEE Transactions on Learning Technologies* **7**(4), 333–345. doi: 10.1109/TLT.2014.2346756.

Csikszentmihalyi, M. (1990), *Flow: The Psychology of Optimal Experience*, Harper and Row, New York. ISBN: 978-0060920432.

Csikszentmihalyi, M. (1998), *Finding Flow: The Psychology of Engagement with Everyday Life*, Basic Books. ISBN: 0-465-02411-4.

Cuadrado, L.-E. I., Riesco, Á. M. and López, F. D. L. P. (2016), 'ARTIE: An integrated environment for the development of affective robot tutors', *Frontiers in Computational Neuroscience* **10**. doi: 10.3389/fncom.2016.00077.

Danahy, E., Wang, E., Brockman, J., Carberry, A., Shapiro, B. and Rogers, C. B. (2014), 'Lego-based robotics in higher education: 15 years of student creativity', *International Journal of Advanced Robotic Systems* **11**(2), 27. doi: 10.5772/58249.

Danubianu, M. (2015), A data preprocessing framework for students' outcome prediction by data mining techniques, *in* '2015 19th International Conference on System Theory, Control and Computing (ICSTCC)', pp. 836–841. doi: 10.1109/ICSTCC.2015.7321398.

Das, D., Rashed, M. G., Kobayashi, Y. and Kuno, Y. (2015), 'Supporting human-robot interaction based on the level of visual focus of attention', *IEEE Transactions on Human-Machine Systems* **45**(6), 664–675. doi: 10.1109/THMS.2015.2445856.

de Greeff, J. and Belpaeme, T. (2015), 'Why robots should be social: Enhancing machine learning through social human-robot interaction', *PLOS ONE* **10**(9), 1–26. doi: 10.1371/journal.pone.0138061.

De Winter, J. C. and Dodou, D. (2010), 'Five-point Likert items: t test versus Mann-Whitney-Wilcoxon', *Practical Assessment, Research & Evaluation* **15**(11), 1–12.
URL: *http://pareonline.net/pdf/v15n11.pdf* [Accessed 2 December 2017]

Denis, B. and Hubert, S. (2001), 'Collaborative learning in an educational robotics environment', *Computers in Human Behavior* **17**(5), 465–480. doi: 10.1016/S0747-5632(01)00018-8.

Domínguez, F., Chiluiza, K., Echeverria, V. and Ochoa, X. (2015), Multimodal selfies: Designing a multimodal recording device for students in traditional classrooms, *in* 'Proceedings of the 2015 ACM on International Conference on Multimodal Interaction', ICMI '15, ACM, Seattle, Washington, USA, pp. 567–574. doi: 10.1145/2818346.2830606.

Fagin, B. S. and Merkle, L. (2002), 'Quantitative analysis of the effects of robots on introductory computer science education', *Journal on Educational Resources in Computing* **2**(4). doi: 10.1145/949257.949259.

Fidalgo-Blanco, Á., Sein-Echaluce, M. L., García-Peñalvo, F. J. and Conde, M. Á. (2015), 'Using learning analytics to improve teamwork assessment', *Computers in Human Behavior* **47**, 149–156. doi: 10.1016/j.chb.2014.11.050.

Fridin, M. (2014), 'Storytelling by a kindergarten social assistive robot: A tool for constructive learning in preschool education', *Computers & Education* **70**, 53–64. doi: 10.1016/j.compedu.2013.07.043.

Fridin, M. and Belokopytov, M. (2014), 'Acceptance of socially assistive humanoid robot by preschool and elementary school teachers', *Computers in Human Behavior* **33**, 23–31. doi: 10.1016/j.chb.2013.12.016.

Gašević, D., Dawson, S., Rogers, T. and Gasevic, D. (2016), 'Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success', *The Internet and Higher Education* **28**, 68–84. doi: 10.1016/j.iheduc.2015.10.002.

Goggins, S., Xing, W., Chen, X., Chen, B. and Wadholm, B. (2015), 'Learning analytics at "small" scale: Exploring a complexity-grounded model for assessment automation', *Journal of Universal Computer Science* **21**(1), 66–92. doi: 10.3217/jucs-021-01-0066.

Hall, E. T. (1966), *The Hidden Dimension*, Anchor Books. ISBN: 0-385-08476-5.

Hăng, N. V. T., Meijer, M. R., Bulte, A. M. W. and Pilot, A. (2015), 'The implementation of a social constructivist approach in primary science education in Confucian heritage culture: the case of Vietnam', *Cultural Studies of Science Education* **10**(3), 665–693. doi: 10.1007/s11422-014-9634-8.

Hernández-García, A., González-González, I., Jiménez-Zarco, A. I. and Chaparro-Peláez, J. (2015), 'Applying social learning analytics to message boards in online distance learning: A case study', *Computers in Human Behavior* **47**, 68–80. doi: 10.1016/j.chb.2014.10.038.

Hodkinson, C. S. and Poropat, A. E. (2014), 'Chinese students' participation: the effect of cultural factors', *Education + Training* **56**(5), 430–446. doi: 10.1108/ET-04-2013-0057.

Iglesias-Pradas, S., de Azcárate, C. R. and Agudo-Peregrina, Á. F. (2015), 'Assessing the suitability of student interactions from Moodle data logs as predictors of cross-curricular competencies', *Computers in Human Behavior* **47**, 81–89. doi: 10.1016/j.chb.2014.09.065.

Jackson, S. A., Martin, A. J. and Eklund, R. C. (2008), 'Long and short measures of flow: The construct validity of the FSS-2, DFS-2, and new brief counterparts', *Journal of Sport and Exercise Psychology* **30**(5), 561–587. doi: 10.1123/jsep.30.5.561.

Jacobs, J. (1992), *The death and life of great American cities*, Vintage. ISBN: 978-0679741954.

Joksimović, S., Gašević, D., Kovanović, V., Riecke, B. E. and Hatala, M. (2015), 'Social presence in online discussions as a process predictor of academic performance', *Journal of Computer Assisted Learning* **31**(6), 638–654. doi: 10.1111/jcal.12107.

Kanda, T., Hirano, T., Eaton, D. and Ishiguro, H. (2004), 'Interactive robots as social partners and peer tutors for children: A field trial', *Human Computer Interaction* **19**(1), 61–84. doi: 10.1207/s15327051hci1901&2_4.

Kandlhofer, M. and Steinbauer, G. (2016), 'Evaluating the impact of educational robotics on pupils' technical- and social-skills and science related attitudes', *Robotics and Autonomous Systems* **75**(Part B), 679–685. doi: 10.1016/j.robot.2015.09.007.

Keller, J. M. (1987), 'IMMS: Instructional materials motivation survey'.

Keren, G. and Fridin, M. (2014), 'Kindergarten social assistive robot (KindSAR) for children's geometric thinking and metacognitive development in preschool education: A pilot study', *Computers in Human Behavior* **35**, 400–412. doi: 10.1016/j.chb.2014.03.009.

Khanlari, A. (2013), Effects of educational robots on learning stem and on students' attitude toward stem, *in* '2013 IEEE 5th Conference on Engineering Education (ICEED)', pp. 62–66. doi: 10.1109/ICEED.2013.6908304.

Khanlari, A. (2016), 'Teachers' perceptions of the benefits and the challenges of integrating educational robots into primary/elementary curricula', *European Journal of Engineering Education* **41**(3), 320–330. doi: 10.1080/03043797.2015.1056106.

Kidder, L. H. and Judd, C. M. (1986), *Research methods in social relations*, 5th edn, Holt, Rinehart & Winston, New York.

Lee, J. V., Taha, Z., Yap, H.-J. and Kinsheel, A. (2013), 'Constructivist game-based robotic simulator in engineering education', *International Journal of Engineering Education* **29**(4), 1024–1036.

Lee, K. T. H., Sullivan, A. and Bers, M. U. (2013), 'Collaboration by design: Using robotics to foster social interaction in kindergarten', *Computers in the Schools* **30**(3), 271–281. doi: 10.1080/07380569.2013.805676.

Levy, J. S. (2008), 'Case studies: Types, designs, and logics of inference', *Conflict Management and Peace Science* **25**(1), 1–18. doi: 10.1080/07388940701860318.

Lockyer, L., Heathcote, E. and Dawson, S. (2013), 'Informing pedagogical action: Aligning learning analytics with learning design', *American Behavioral Scientist* **57**(10), 1439–1459. doi: 10.1177/0002764213479367.

Lonn, S., Aguilar, S. J. and Teasley, S. D. (2015), 'Investigating student motivation in the context of a learning analytics intervention during a summer bridge program', *Computers in Human Behavior* **47**, 90–97. doi: 10.1016/j.chb.2014.07.013.

Lundie, D. (2017), 'The givenness of the human learning experience and its incompatibility with information analytics', *Educational Philosophy and Theory* **49**(4), 391–404. doi: 10.1080/00131857.2015.1052357.

Madhavan, K. and Richey, M. C. (2016), 'Problems in big data analytics in learning', *Journal of Engineering Education* **105**(1), 6–14. doi: 10.1002/jee.20113.

Maoz, Z. (2002), *Case study methodology in international studies: From storytelling to hypothesis testing*, *in* M. Brecher and F. P. Harvey, eds, 'Evaluating Methodology in International Studies: Millennial Reflections on International Studies', Ann Arbor: University of Michigan Press, pp. 161–186. ISBN: 0-472-08861-0.

McColl, D., Jiang, C. and Nejat, G. (2017), 'Classifying a person's degree of accessibility from natural body language during social human-robot interactions', *IEEE Transactions on Cybernetics* **47**(2), 524–538. doi: 10.1109/TCYB.2016.2520367.

McGill, M. M. (2012), 'Learning to program with personal robots: Influences on student motivation', *ACM Transactions on Computing Education* **12**(1), 4:1–4:32. doi: 10.1145/2133797.2133801.

Michieletto, S., Tosello, E., Pagello, E. and Menegatti, E. (2016), 'Teaching humanoid robotics by means of human teleoperation through RGB-D sensors', *Robotics and Autonomous Systems* **75**(Part B), 671–678. doi: 10.1016/j.robot.2015.09.023.

Mills, K. A., Chandra, V. and Park, J. Y. (2013), 'The architecture of children's use of language and tools when problem solving collaboratively with robotics', *The Australian Educational Researcher* **40**(3), 315–337. doi: 10.1007/s13384-013-0094-z.

Miranda, A., Bolea, Y., Grau, A. and Sanfeliu, A. (2012), Work in progress: A constructivist didactic methodology for a humanoid robotics workshop, *in* '2012 Frontiers in Education Conference Proceedings', pp. 1–3. doi: 10.1109/FIE.2012.6462226.

Mitnik, R., Nussbaum, M. and Recabarren, M. (2009), 'Developing cognition with collaborative robotic activities', *Educational Technology & Society* **12**(4), 317–330.
URL: *http://www.ifets.info/journals/12_4/27.pdf* [Accessed 4 December 2017]

Monash University (2010), 'Research data management: HDR candidates procedures'.
URL: *http://www.monash.edu/__data/assets/pdf_file/0005/797342/Research-Data-Management-Procedures-HDR-Candidates.pdf* [Accessed 3 December 2017]

Monash University (2013), 'Research outputs data collection procedures'.
URL: *http://www.monash.edu/__data/assets/pdf_file/0005/797360/Research-Outputs-Data-Collection-Procedures.pdf* [Accessed 3 December 2017]

Moneta, G. B. (2012), *On the Measurement and Conceptualization of Flow*, *in* S. Engeser, ed., 'Advances in Flow Research', Springer, New York, NY, pp. 23–50. ISBN: 978-1-4614-2359-1. doi: 10.1007/978-1-4614-2359-1_2.

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A. and Dong, J.-J. (2013), 'A review of the applicability of robots in education', *Technology for Education and Learning* **1**, 1–7. doi: 10.2316/Journal.209.2013.1.209-0015.

National Health and Medical Research Committee (2007), 'Australian code for the responsible conduct of research', Commonwealth of Australia, Canberra.
URL: *https://www.nhmrc.gov.au/guidelines-publications/r39* [Accessed 3 December 2017]

Park, I., Kim, D., Oh, J., Jang, Y. and Lim, K. (2015), 'Learning effects of pedagogical robots with programming in elementary school environments in Korea', *Indian Journal of Science and Technology* **8**(26), 1–5. doi: 10.17485/ijst/2015/v8i26/80723.

Patton, M. Q. (1990), *Qualitative evaluation and research methods*, 2nd edn, Sage Publications, Inc. ISBN: 978-0803937796.

Plauska, I. and Damaševičius, R. (2014), *Educational Robots for Internet-of-Things Supported Collaborative Learning*, *in* G. Dregvaite and R. Damasevicius, eds, 'Information and Software Technologies: 20th International Conference, ICIST 2014, Druskininkai, Lithuania, October 9-10, 2014. Proceedings', Springer International Publishing, Cham, pp. 346–358. ISBN: 978-3-319-11958-8. doi: 10.1007/978-3-319-11958-8_28.

Popper, K. (2002), *Conjectures and refutations: The growth of scientific knowledge*, 2nd edn, Routledge. ISBN: 978-0415285940.

Reid, P. (2017), 'Supporting instructors in overcoming self-efficacy and background barriers to adoption', *Education and Information Technologies* **22**(1), 369–382. doi: 10.1007/s10639-015-9449-6.

Rienties, B. and Rivers, B. A. (2014), 'Measuring and understanding learner emotions: Evidence and prospects', *Learning Analytics Review* **1**, 1–28.
URL: *https://pdfs.semanticscholar.org/bf00/c9780a08c132d6caf85c444145608cedf963.pdf* [Accessed 2 December 2017]

Rienties, B. and Toetenel, L. (2016), 'The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules', *Computers in Human Behavior* **60**, 333–341. doi: 10.1016/j.chb.2016.02.074.

Rubel, A. and Jones, K. M. L. (2016), 'Student privacy in learning analytics: An information ethics perspective', *The Information Society* **32**(2), 143–159. doi: 10.1080/01972243.2016.1130502.

Saerbeck, M., Schut, T., Bartneck, C. and Janse, M. (2010), Expressive robots in education - varying the degree of social supportive behavior of a robotic tutor, *in* '28th ACM Conference on Human Factors in Computing Systems (CHI2010)', ACM, Atlanta, pp. 1613–1622. doi: 10.1145/1753326.1753567.

Schieble, M., Vetter, A. and Meacham, M. (2015), 'A discourse analytic approach to video analysis of teaching: Aligning desired identities with practice', *Journal of Teacher Education* **66**(3), 245–260. doi: 10.1177/0022487115573264.

Seligman, M. E. and Csikszentmihalyi, M. (2000), 'Positive psychology: An introduction', *American Psychologist* **55**(1), 5–14. doi: 10.1037/0003-066X.55.1.5.

Serholt, S., Barendregt, W., Leite, I., Hastie, H., Jones, A., Paiva, A., Vasalou, A. and Castellano, G. (2014), Teachers' views on the use of empathic robotic tutors in the classroom, *in* 'The 23rd IEEE International Symposium on Robot and Human Interactive Communication', pp. 955–960. doi: 10.1109/ROMAN.2014.6926376.

Shamsuddin, S., Ismail, L. I., Yussof, H., Zahari, N. I., Bahari, S., Hashim, H. and Jaffar, A. (2011), Humanoid robot NAO: Review of control and motion exploration, *in* '2011 IEEE International Conference on Control System, Computing and Engineering (ICCSCE)', IEEE, pp. 511–516. doi: 10.1109/ICCSCE.2011.6190579.

Silva, A. F., Barros, R. P., Azevedo, S. O., Silva, A. and Gonçalves, L. M. G. (2008), Diagnostic robotic agent in the roboeduc environment for educational robotics, *in* '2008 IEEE Latin American Robotic Symposium', pp. 131–136. doi: 10.1109/LARS.2008.16.

Singh, A., Karanam, S. and Kumar, D. (2013), 'Constructive learning for human-robot interaction', *IEEE Potentials* **32**(4), 13–19. doi: 10.1109/MPOT.2012.2189443.

Skipp, A. and Tanner, E. (2015), 'The visible classroom evaluation report and executive summary'.
URL: *https://v1.educationendowmentfoundation.org.uk/uploads/pdf/Visible_Classroom_-(Final).pdf* [Accessed 2 December 2017]

Slade, S. and Prinsloo, P. (2013), 'Learning analytics: Ethical issues and dilemmas', *American Behavioral Scientist* **57**(10), 1510–1529. doi: 10.1177/0002764213479366.

SoftBank Robotics (2017), 'NAO robot: characteristics | SoftBank Robotics'.
URL: *https://www.ald.softbankrobotics.com/en/robots/nao/find-out-more-about-nao* [Accessed 26 November 2017]

Stake, R. E. (1995), *The art of case study research*, Thousand Oaks: Sage Publications, Inc. ISBN: 978-0803957671.

Stroet, K., Opdenakker, M.-C. and Minnaert, A. (2016), 'Fostering early adolescents motivation: a longitudinal study into the effectiveness of social constructivist, traditional and combined schools for prevocational education', *Educational Psychology* **36**(1), 1–25. doi: 10.1080/01443410.2014.893561.

Sullivan, A. and Bers, M. U. (2016), 'Robotics in the early childhood classroom: learning outcomes from an 8-week robotics curriculum in pre-kindergarten through second grade', *International Journal of Technology and Design Education* **26**(1), 3–20. doi: 10.1007/s10798-015-9304-5.

Thomas, B. and Watters, J. J. (2015), 'Perspectives on Australian, Indian and Malaysian approaches to STEM education', *International Journal of Educational Development* **45**, 42–53. doi: 10.1016/j.ijedudev.2015.08.002.

Trochim, W. M. (1989), 'Outcome pattern matching and program theory', *Evaluation and Program Planning* **12**(4), 355–366. doi: 10.1016/0149-7189(89)90052-9.

United States Department of Health and Human Services (1979), 'The Belmont report: Ethical principles and guidelines for the protection of human subjects of research'.
URL: *https://www.hhs.gov/ohrp/regulations-and-policy/belmont-report/index.html* [Accessed 3 December 2017]

van Leeuwen, A., Janssen, J., Erkens, G. and Brekelmans, M. (2015), 'Teacher regulation of multiple computer-supported collaborating groups', *Computers in Human Behavior* **52**, 233–242. doi: 10.1016/j.chb.2015.05.058.

Warner, W. L. and Lunt, P. S. (1973), *The social life of a modern community*, Greenwood Press. ISBN: 978-0837169583.

Willis, B. (2014), 'The advantages and limitations of single case study analysis', *E-International Relations Students* .
URL: *http://www.e-ir.info/2014/07/05/the-advantages-and-limitations-of-single-case-study-analysis* [Accessed 3 December 2017]

Wong, N. W. H. (2017), 'Flow State Scale-2 (short version) and subjective experience responses of undergraduate students in tutorial sessions taught either by a human or a robot tutor in Monash University, Malaysia', Mendeley Data, v1. doi: 10.17632/tz22mcg8w7.1.

Wong, N. W. H., Chew, E. and Wong, J. S.-M. (2016), The review of educational robotics research and the need for real-world interaction analysis, *in* '2016 14th International Conference on Control, Automation, Robotics and Vision (ICARCV)', IEEE, pp. 1–6. doi: 10.1109/ICARCV.2016.7838707.

Wu, W.-C. V., Wang, R.-J. and Chen, N.-S. (2015), 'Instructional design using an in-house built teaching assistant robot to enhance elementary school english-as-a-foreign-language learning', *Interactive Learning Environments* **23**(6), 696–714. doi: 10.1080/10494820.2013.792844.

Yi, H., Knabe, C., Pesek, T. and Hong, D. W. (2016), 'Experiential learning in the development of a DARwIn-HP humanoid educational robot', *Journal of Intelligent & Robotic Systems* **81**(1), 41–49. doi: 10.1007/s10846-015-0200-8.

Yin, R. K. (2013), *Case study research: Design and methods*, 5th edn, Sage Publications, Inc. ISBN: 978-1452242569.

Zacharis, N. Z. (2015), 'A multivariate approach to predicting student outcomes in web-enabled blended learning courses', *The Internet and Higher Education* **27**, 44–53. doi: 10.1016/j.iheduc.2015.05.002.

Zaga, C., Lohse, M., Truong, K. and Evers, V. (2015), *The Effect of a Robot's Social Character on Children Task Engagement: Peer Versus Tutor*, *in* A. Tapus, E. Andr, J.-C. Martin, F. Ferland and M. Ammi, eds, 'Proceedings of the 7th International Conference on Social Robotics, ICSR 2015', Lecture Notes in Computer Science, Springer Verlag, pp. 704–713. ISBN: 978-3-319-25553-8. doi: 10.1007/978-3-319-25554-5_70.

Zawieska, K. and Duffy, B. R. (2015), *The Social Construction of Creativity in Educational Robotics*, *in* R. Szewczyk, C. Zieliński and M. Kaliczyńska, eds, 'Progress in Automation, Robotics and Measuring Techniques: Volume 2 Robotics', Springer International Publishing, Cham, pp. 329–338. ISBN: 978-3-319-15847-1. doi: 10.1007/978-3-319-15847-1_32.

Appendix A: Literature Review Summary

Web of Science | Scopus | Duplicate | Others

***Quartile is based on Thomson Reuters' Journal Citation Reports 2014 ranking statistics for the journal.**

No	Literature	Journal	Quartile*	Methodology	Participants	Experiment	Results
1	Akhtar et al. (2017) [First online in 2015]	International Journal of Technology and Design Education	Q4	Quantitative	331 under-graduate students from University of Surrey, UK	Implements a Computer Supported Collaborative Learning (CSCL) environment to support lab-based CAD teaching. Student participation is monitored to identify predictors of success, **analyzed using ANOVA, Pearson correlation and linear regression**.	Attendance and average time-spent on task has a direct relation with the learning outcomes. Students who prefer to sit in groups or remain next to their fellow students tend to score better.
2	Alemi et al. (2015)	International Journal of Social Robotics	Q3	Quantitative	46 Iranian female students (aged 12)	Examines the effect of robot assisted language learning (RALL) using NAO robot on the anxiety level and attitude in English vocabulary acquisition amongst Iranian EFL junior high school students. **Two questionnaires** of anxiety and attitude were utilized to measure the students' anxiety and attitude.	t-tests indicated that there was lower anxiety and a more positive attitude towards English vocabulary acquisition in the RALL group compared with those in the non-RALL group.
3	Alemi et al. (2014)	International Journal of Humanoid Robotics	Q3	Quantitative	46 middle school Iranian students	45-item vocabulary test **(pre-test)** taken from textbook to assess vocabulary recognition and production. 30 students for NAO robot-assisted language learning (RALL) and 16 for non-RALL groups. At the end of the study, the same test was administered as an **immediate post-test** and the questions were counter balanced and administered as a **delayed post-test** two weeks after the treatment process (with Cronbach alpha of 0.89).	Mean results for pre-test: 13.53, post-test: 39.76, delayed post-test: 39.50; showing effectiveness of acquiring rudimentary linguistic skills and retention of knowledge through the socially assistive robot treatment.
4	Altin and Pedaste (2013)	Journal of Baltic Science Education	Q3	Review	8 research papers	**Systematic review** on robotics curricula for STEM subjects found that approaches used are: discovery, collaborative, problem-solving, project-based, competition-based, and compulsory learning.	Lack of quantitative evidence for applying robots in curricula to achieve educational goals. Most robotics education approaches should not be used alone.
5	Alves-Oliveira et al. (2015)	Lecture Notes in Artificial Intelligence	Q4	Quantitative	56 children, 14-16 years old, Portugal	Children are paired in groups to interact with a NAO robot tutor that guides them through a collaborative multiplayer game, EnerCities for 20 minutes. **Questionnaires were given before and after the test**.	Majority of children expected the robotic tutor to be a good game companion and revealed higher satisfaction but those who expected the robotic tutor to play best in the game showed a significant decrease in satisfaction after the interaction.

6	Anaya et al. (2016)	Expert Systems with Applications	Q1	Qualitative	23 students	A method of data mining for learning experience was formed through **vigorous literature review and analysis**, then a **survey** was carried out to obtain feedback on the developed tool.	Two approach to assess student collabration: clustering (group students according to collaboration) and metric (calculate metrics Uses influence diagrams based on Bayesian network to solve uncertainties in collaboration.
7	Ardito et al. (2014)	Middle Grades Research Journal	Unranked	Mixed	≈1600 students in Croton Harmon School, New York	Students were engaged in LEGO Mindstorms robot challenges using constructionist methods that required them to work together for 1 semester to enhance problem-solving ability. **Textual analysis** was done on student writings in a class blog, and their mathematics **subject grades were assessed**.	Exam scores are higher in concepts associated with algebra, measurement and probability. Textual analysis shows prevalent collaboration among students.
8	Benitti (2012)	Computers & Education	Q1	Review	10 articles	**Systematic review** on subjects taught with robots on the type of robot and research method used, and the sample characteristics.	Empirical evidence on effectiveness of educational robotics is limited, but has potential.
9	Berland et al. (2015)	International Journal of Computer-Supported Collaborative Learning	Q1	Quantitative	95 students (junior and high school) in Texas, USA	Students were placed in pairs on the basis of predictive CS-ZPD as indicated by AMOEBA (analytics tool) and asked to collaborate. As programming data was generated and analyzed, some students were re-paired. Students' program data were analyzed to explore how pairing with AMOEBA impacted using **IBMs SBSS statistical software package**.	Students, after having been paired on the recommendations provided by AMOEBA, evidenced more proficient program development.
10	Bers et al. (2014)	Computers & Education	Q1	Quantitative	53 kinder-garteners in Boston	TangibleK curriculum based on constructionism was designed for robotics and programming course and is carried out with CHERP programming language and LEGO Mindstorms robots. **Student performance is assessed with statistical t-tests for correlation**.	When given age-appropriate technologies, curriculum and pedagogies, young children can actively engage in learning from computer programming as applied to the field of robotics.
11	Bilotta et al. (2009)	Lecture Notes in Computer Science	Q4	Mixed	28 students in University of Calabria, Italy	An edutainment robotics program built based on constructivist theory and LEGO Mindstorms is assessed on work distribution in each student group, description of task resolution, correctness of programming strategies and number of tests completed before success.	Constructionist approach of using robotic artefacts stimulates students to collaboratively analyze processes and experiment the consequences of their behavior.

12	Blancas et al. (2015)	Lecture Notes in Artificial Intelligence	Q4	Mixed	28 adults	This study assesses whether the role a robot plays in a classroom affects knowledge retrieval, subjective experience, and the perception of the learners. The NAO robot delivers the history class in either a teacher or peer condition with differences in posture, gestures and speech (formal/informal). **Questionnaires, pre-test, post-test and video recordings** were analyzed.	There are no significant differences between the conditions in the amount of knowledge retrieved. Subjects in the peer condition graded the robot as a tutor higher than in the other conditions.
13	Brown and Howard (2014*b*)	Integrated STEM Education Conference (ISEC)	Unranked	Mixed	Trial 1: 24 college students, Trial 2: 20 high school students	Integrates a socially interactive robotic tutor (DARwIn-OP) to engage students in the classroom environment. Students are **randomly assigned** to either a control group with no robot tutor or the treatment group. **Completion time, Likert-scale survey on experience** and **freeform feedback** were analyzed.	Verbal cues are able to increase and/or maintain student engagement regardless of student age and math content level. The control group was less nervous with the robot tutor.
14	Brown and Howard (2014*a*)	Computers in Education Journal	Unranked	Mixed	13 middle school students (10-14 years old) in Atlanta, GA	15-question math test in the computer to assess total time, response accuracy and proper function execution with webcam to monitor eye gaze and pose. Data is **tested for statistical significance**. **Exit survey** and **video observations** were analyzed.	If a student is classified as being on-task, he or she is engaged regardless of speed or response. Eye gaze and head pose technique is not an effective measure of engagement when high-level cognitive thinking is required.
15	Catlin (2014)	Lecture Notes in Computer Science	Q4	Review	N/A	A review on how peer and self-assessment (PASA) is applied in educational robotics.	Black and Williams' Assessment for Learning (AfL) strategies offers a way of structuring lessons while fostering essential intellectual freedom of the student.
16	Chen et al. (2012)	Turkish Online Journal of Educational Technology	Unranked	Review	N/A	A look on the application of digital technology such as robots, projectors and computers to build a game-based learning environment for classrooms, called Digital Learning Playground (DLP) with the use of Total Scenario Response (TSR) learning design methods.	Suggests that physical things have higher potential to engage and to support authentic and possibly experiential learning.
17	Chin et al. (2014)	IEEE Transactions On Learning Technologies	Q3	Mixed	1 teacher and 52 second-grade students in Taiwan	Students are **randomly assigned** to either the proposed robot tutor learning system (using Robotis Bioloid Kit) or a PowerPoint-based learning system. The robot is used as an assistant; it performs gestures according to the instruction materials presented. A **pre-test** and **post-test questionnaire** was employed to measure attention, relevance, confidence and satisfaction.	Social interaction with humanoid robots have positive results on student motivation and performance in elementary education. Satisfaction and relevance were rated the highest.

18	Danahy et al. (2014)	International Journal of Advanced Robotic Systems	Q4	Review	4 case studies	Reflecting the role LEGO robotics has played in college engineering education over the last 15 years, starting with the introduction of the RCX in 1998 and ending with the introduction of the EV3 in 2013.	LEGO Mindstorms products have allowed students to take on complex engineering questions without experience in circuit design, artificial intelligence or programming.
19	Danubianu (2015)	19th International Conference on System Theory, Control and Computing (ICSTCC)	Unranked	Qualitative	960 university students in 130 courses of Faculty of Electrical Engineering and Computers Science in University of Suceava, Romania	A **case study** for a data preprocessing framework for students' outcome prediction using data collected by Moodle system. It shows some methods of aggregating and extracting useful data from LMS for further analysis.	Before further analysis, data may need to be preprocessed to establish the dependencies between courses and form association rules such as clustering.
20	Das et al. (2015)	IEEE Transactions on Human-Machine Systems	Q2	Quantitative	36 students from Saitama University, Japan	A human-robot interaction approach for social robots that attracts and controls the attention of a target person based on his/her current visual focus of attention. It estimates "task-related contextual cues" and "gaze pattern" to determine a suitable time to interact. Questionnaires were used to assess the performance.	Among 72 interactions, the system was able to detect 66 times the gaze point of visitors in a museum and make a successful interaction at a rate of 91.7%.
21	de Greeff and Belpaeme (2015)	PLoS One	Q1	Mixed	38 participants recruited from around a British university campus	The interaction between human participants as teachers and the robot as a learner is modelled through a language game. Participants are **randomly assigned** to social and non-social robot group. **Assessment** is done on robot learning performance, participants' choice of topic, participants' gaze, and **questionnaire** on subjective experience.	Robots might positively influence an interaction with a person through using social cues that are generally perceived as natural which can result in people offering better quality learning input to artificial systems.
22	Domínguez et al. (2015)	ACM International Conference on Multimodal Interaction, ICMI	Unranked	Review	N/A	Discusses and evaluates the design of a personal Multimodal Recording Device (MRD) to capture student actions during lectures, including its foreseeable costs, scalability, flexibility, intrusiveness and recording quality.	Low-cost devices change the paradigm from centralized to distributed recording in order to establish student behavior in class with the potential for learning analytics research.
23	Fidalgo-Blanco et al. (2015)	Computers in Human Behavior	Q1	Quantitative	110 first-year Biotechnology degree students from the Technical University of Madrid, Spain	Implements a learning analytics system using CTMTC method (Comprehensive Training Model of the Teamwork Competence) to analyze online forum data through Moodle. **Statistical tests using Pearsons correlation** were done on 5136 messages and 37,930 message views.	Active interactions have a greater relation with the individual performance in teamwork contexts than passive ones. The relationship between message views and individual final grade is inconclusive.

24	Fridin (2014)	Computers & Education	Q1	Quantitative	10 kinder-garten children (5 boys & 5 girls, aged 3-3.6) in Israel	Derives a child-robot interaction metric from eye contact and affective factor to analyze its relationship to cognitive and motor performance after NAO robot interaction based on constructivist methods. **Statistical analysis using repeated measures ANOVA and Pearson product-moment correlations**.	Children performance is positively correlated with interaction levels, and is not significantly affected by any of the between-subject factors.
25	Fridin and Belokopytov (2014)	Computers in Human Behavior	Q1	Quantitative	18 teachers	A modified Unified Theory of Acceptance and the Use of Technology model was applied using **questionnaires** following interactions with NAO robot in a workshop **(non-random sample)**. Results were analyzed using **statistical tests for correlations and linear regressions** to determine reliability of data.	Positive reactions but lack of consolidated views and there is a need for an adaptation of the model.
26	Gašević et al. (2016)	The Internet and Higher Education	Q1	Quantitative	4134 under-graduate students	The study used a **correlational design and statistical analysis** as it investigated the effects of the variables derived from the trace data and the data from the institutional student information system on the prediction of students' academic success. The data for the study were extracted from a public research-intensive university in Australia. 9 undergraduate courses were selected.	There is a need to create models for academic success prediction for individual courses, incorporating instructional conditions into the analysis model. Otherwise, several threats to the validity of the results may emerge such as overestimation or underestimation of certain predictors.
27	Goggins et al. (2015)	Journal of Universal Computer Science	Q4	Quantitative	28 groups of 3-5 students	A set of words and actions were analyzed e.g. since the task is to draw a triangle, conversations with the word "triangle" or usage of the "segment" tool indicate collaboration towards a goal. A **Tree-Augmented Naïve-Bayes assessment model** was developed, achieving the highest accuracy (95.8%) as compared to baseline models.	A web-based tool developed to visualize time-series activities and assess small group learning in real time. Many studies overlook the collaborative process and looks at final solution or grades, but this study shows that group interaction can be modeled.
28	Hernández-García et al. (2015)	Computers in Human Behavior	Q1	Mixed	656 students at Open University of Catalonia, Spain	Uses **Gephi 0.8.2 for social network visualization and analysis**. Network parameters provide a quantitative interpretation of data, while data visualization facilitates qualitative explanation.	Students who got more replies from consultant teachers tended to get higher grades. Social network analysis (SNA) parameters are related to academic performance only in some cases, not all.

29	Iglesias-Pradas et al. (2015)	Computers in Human Behavior	Q1	Quantitative	39 Masters degree students at Universidad a Distancia de Madrid, Spain	Explores the applicability of learning analytics for prediction of development of two cross-curricular competencies: teamwork and commitment. "Interactions" Moodle plugin was used for interaction data extraction and categorization. **Multiple regression analysis** on total number of interactions of each category as independent variables and the total score of each competency as dependent variable.	The results showed no relation whatsoever between interactions of any kind in the Learning Management System and the students' final level of teamwork competency. There is also no relation between any type of interaction and commitment levels.
30	Joksimović et al. (2015)	Journal of Computer Assisted Learning	Q1	Quantitative	81 students and 1747 student messages in online discussion from a public online university in Canada	Examines the relationship between indicators of social presence and academic performance. Investigation is done using minimal guidance for social interaction **(control group)** and one tailored for social and cognitive presence **(treatment group)**. **Pearson's correlation and multiple regression analysis** was performed.	Certain indicators of social presence were significant predictors of final grades. Course design that increased the level of meaningful interactions between students had a significant impact on the development of social presence and could positively affect students' academic performance.
31	Kandlhofer and Steinbauer (2016)	Robotics and Autonomous Systems	Q3	Quantitative	148 pupils (mean age 14.9 years) from 9 schools across Austria and Sweden	**Quasi-experimental two-group design** with **pre-test and post-test** using **questionnaires**. **Statistical methods with repeated MANOVA**, the gathered data were analyzed around 14 different topics ('sub-scales') arranged in three main categories.	Educational robotics has a significant positive impact on some separate sub-scales (mathematics and scientific investigation, teamwork, social skills) but not all.
32	Keren and Fridin (2014)	Computers in Human Behavior	Q1	Mixed	3 groups of children (Israeli born, 10 boys and 7 girls), 1 technician, 1 staff member	NAO robot introduced in class following a set of procedures: (1) cognitive stage - robot teaches children, (2) metacognitive stage - children teach each other how to interact with robot. **Post hoc analysis of video footage, observation checklist and interaction index** combines eye contact and emotional expressions (with Cronbach's alpha of 0.686).	Data revealed significant improvement in children's metacognitive abilities in the second stage compared to the first. Socially assistive robots can provide psychology development data in real-time for teachers to regulate.
33	Khanlari (2016) [Published online in June 2015]	European Journal of Engineering Education	Unranked	Qualitative	11 elementary teachers from Newfoundland and Labrador English Schools District, Canada	Qualitative **case study** using **online surveys** on teachers' perceptions of the effects of using robotics on students' lifelong learning skills, teachers' perceptions of the barriers of using robotics and the support they need.	Teachers perceived that robotics has positive effects on scientific inquiry skills. Challenges include lack of technical and instructional support, preparation and classroom time, knowledge about robotics and confidence.
34	Khanlari (2013)	Engineering Education (ICEED), 2013 IEEE 5th Conference	Unranked	Qualitative	6 teachers with 2-7 years of experience	**Interview** on teachers' perceptions of the effects of robotics on students' learning experiences and on their interests towards STEM subjects.	Teachers feel that learning with robotics is helpful because of its hands-on, play-and-learn nature and boosts students confidence.

35	Lee, Taha, Yap and Kinsheel (2013)	International Journal of Engineering Education	Q4	Quantitative	114 undergraduate students (aged 22-25)	Assesses the learning environment of a constructivist game-based robotics simulator compared to non game-based conventional robotics simulator on students perceptions via **Constructivist Simulation-based Learning Environment Survey (CSLES** and **Test of Robotics Related Attitudes (TORRA) questionnaires** with **statistical tests for correlation and multiple regression analysis**.	Game-based robotics simulator is more effective in terms of Negotiation, Inquiry Learning, Reflective Thinking and Challenge. There is positive but relatively weak relationship between the undergraduate students' enjoyment of robotics lessons and the game-based learning environment.
36	Lee, Sullivan and Bers (2013)	Computers in the Schools	Unranked	Mixed	19 kindergarten students in Boston, USA	Each child was **randomly assigned** in summer camp LEGO Mindstorms robotics workshop to: (a) instructional environment through pre-designed teacher-guided challenges or (b) a constructionist approach. **Assessment on reported interactions triangulated with interactions observed** across 3 videos per group.	The constructionist approach reflected higher mean numbers of interactions. A less structured learn-by-doing approach might be useful for teachers when integrating technology.
37	Lockyer et al. (2013)	American Behavioral Scientist	Q2	Review	N/A	Uses a learning design drawn from a repository established through an Australian project that identified, reviewed, and documented examples of university courses that effectively used technology to facilitate flexible learning. Suggests analytics models on a case-by-case basis comprising of individual, small group, and large group learning tasks and use of online resources and discussion forums.	Learning design consists of resources, tasks and support mechanisms. Checkpoint analytics looks at snapshot data to check if students met prerequisites. Process analytics looks at knowledge application within a tasks.
38	Lonn et al. (2015)	Computers in Human Behavior	Q1	Mixed	216 students in a Summer Bridge Program, USA	Uses **Achievement Goal Theory to measure motivation**, and **Patterns of Adaptive Learning Scales to measure achievement goal orientations** to design an early warning system (EWS) to identify students at risk. **Tests for statistical significance, multiple regression models and student self-reports** were analyzed.	Presentation of learning analytics data to students can affect motivation negatively. The next generation of learning analytics interventions must provide direction on how to tailor learning environments to learners' needs.
39	Lundie (2017) [Published online in June 2015]	Educational Philosophy and Theory	Q4	Review	N/A	Article which explores learning analytics in terms of computing philosophy and the information-theoretic account of knowledge.	The human learning subject is not reducible to informational transactions. Human subjects experience and value their own information incommensurably with the ways in which computers measure and quantify information.

40	Madhavan and Richey (2016)	Journal of Engineering Education	Q1	Review	N/A	A review of Big Data learning analytics problems in education.	Work on understanding the types of models that could be used for predictive efforts is still very much in its infancy. The latency between collecting data from students and the ability to translate these data into actionable intelligence is a significant barrier.
41	McGill (2012)	ACM Transactions on Computing Education	Unranked	Mixed	35 non-computer science undergraduate students	Usage of Institute for Personal Robots in Education (IPRE) robot to study its motivational effects on non-computer science students in a CS0 introductory programming course. Uses Keller's **Instructional Materials Motivation Survey** to measure attention, relevance, confidence, and satisfaction; then analyzed with **statistical t-tests, ANOVA and MANOVA**.	Little or no effect on relevance, confidence, or satisfaction; but significant effect on attention for non-computer science students to learn programming. A little different from results obtained using IMMS by Chin et al. (2014).
42	Michieletto et al. (2016)	Robotics and Autonomous Systems	Q3	Quantitative	About 20 postgraduate students in University of Padova, Italy	A graduate course project on humanoid robotics is presented using a project-based constructivist approach. The task combines teleoperation of NAO robot using Kinect with an integrated programming framework. **Quantitative data such as project marks, course marks and student questionnaire** are used to assess problem-solving ability.	Combining a constructivist approach with the assignation of tasks of increasing complexity (scaffolding) leads to the desired results in educational robotics.
43	Mills et al. (2013)	Australian Educational Researcher	Q4	Qualitative	24 Year 4 students (aged 8.5-9.5)	Analyzes children interactions during a series of problem solving experiments using LEGO Mindstorms and Vygotsky theory with incrementally difficult challenges through **students' speech interactions with tools, peers, and other experts, teacher interviews, and student focus group data**.	Language-mediated problem solving begins with phases of interaction for each goal or sub-goal, followed by the use of predictive questions and directive statements, and culminates in an emotive utterance of greater intensity upon realization of a likely solution.
44	Miranda et al. (2012)	Frontiers in Education Conference Proceedings	Unranked	Qualitative	20 engineering degree students in University of Catalonia, Spain	45-hour robotics workshop conducted using Robonova-I and constructivist methods by **randomly assigning** students into small groups for project-based learning. **Interviews** were conducted.	Suggests use of humanoid robots through constructivist methods enhances learning and teaching interest.
45	Mitnik et al. (2009)	Educational Technology and Society	Q2	Mixed	24 16-year-old students (10th grade)	Students work in groups of three, using a robot and wirelessly interconnected Personal Digital Assistants (PDA) based on Feuersteins Mediated Learning theory. **Statistical t-test** is made for significance and **video observations** were analyzed.	Statistically significant increase in performance. Students are highly motivated.

46	Park et al. (2015)	Indian Journal of Science and Technology	Unranked	Quantitative	27 third-grade students in Korea	Students participated in the robot learning curriculum in 3 different subjects (Korean, mathematics and music) for 12 weeks, **a paired t-test** was conducted with **pre-tests and post-tests** on creativity and class satisfaction.	Fluency and originality were significantly improved. Class satisfaction was measured by descriptive statistics with mean of 4.45 out of 5.
47	Plauska and Damaševičius (2014)	Communi-cations in Computer and Information Science	Unranked	Mixed	22 university students in Lithuania	Introduces Internet-of-Things Supported Collaborative Learning (IoTSCL) paradigm based on constructivism in a robotics course. Assessment done on student engagement using a **survey** and **four-phase interest model**. Lack of statistical tests.	Student engagement and feedback should be about physical things (in this case, robots) rather than virtual. Robots are part of the learning environment and must interact with their environment.
48	Rienties and Toetenel (2016)	Computers in Human Behavior	Q1	Quantitative	151 modules and 111.256 students (based on online data) at the Open University UK	University modules were mapped to learning design categories which are then analyzed with data from Moodle. Learner satisfaction is assessed using **Student Experience on a Module (SEaM) questionnaire**. **Correlation and multiple regression analyses** were conducted using IBM SPSS 21 statistics software package.	The primary predictor of academic retention was the relative amount of communication activities. Learner satisfaction was strongly influenced by learning design.
49	Rubel and Jones (2016)	Information Society	Q2	Review	N/A	Thoughts on information access concerns and significant student privacy problems for learning analytics in higher education institutions.	Learning analytics systems should provide controls for differential access to private student data, must be able to justify collection, full accounting of how benefits are distributed, and students should be made aware and given choices on the collection and use of their data.
50	Saerbeck et al. (2010)	Proceedings of the 28th ACM Conference on Human Factors in Computing Systems (CHI2010)	Unranked	Mixed	9 girls and 7 boys in the Primary In-ternational School of Eindhoven, Netherlands (10-11 years old)	Use of a robotic tutor (iCat from Philips Research) in 5 behavior dimensions to vary social supportiveness. Participants were **randomly assigned** between Neutral or Socially Supportive modes. **Video recordings** were analyzed, **survey** on first impressions, **language test** to grade performance, **questionnaire** to assess motivation and a **final interview**.	Participants in the social supportive condition were significantly more motivated. The neutral condition is solely on knowledge transfer while for the social supportive condition the focus is on active dialog and positive social supportive behaviors.
51	Schieble et al. (2015)	Journal of Teacher Education	Q1	Qualitative	30 preservice English teachers	The authors organized all data sources generated by individual participants and treated each individual as an intrinsic case to analyze the complexities of each teacher candidate as an individual with particular histories of participation. **Teacher reflections, discourse analysis charts, interviews, and videos** were examined.	Discourse analytic approaches using positioning theory allow candidates to focus specifically on how their linguistic and non-verbal choices impact the enactment of identities related to teaching and learning.

52	Serholt et al. (2014)	2014 RO-MAN: The 23rd IEEE International Symposium	Unranked	Qualitative	8 teachers (England, Scotland, Portugal, Sweden)	An **interview** study conducted across several European countries on teachers' views on the use of empathic robotic tutors in the classroom.	Robotic tutors should fit with existing classroom practices and social norms; and assist in recording data.
53	Silva et al. (2008)	Latin American Robotics Symposium	Unranked	Qualitative	Children up to 10 years in a public elementary school in Brazil	The interaction of each student with the Diagnostics Robotic Agent (DRA) software records the mistakes, how long the students need to perform a task and how many times requested help from a more capable partner. It then calculates ZPD metric but the accuracy and effectiveness of this method **is only observed and not validated**.	A ZPD metric may be useful to identify learning potential, allowing teachers to make better decisions to intervene and plan lessons better.
54	Singh et al. (2013)	IEEE Potentials	Unranked	Quantitative	A group of students, no specific number	Maintaining a positive learning rate of a student being taught in a classroom **using facial expression recognition and Tree Augmented Naïve-Bayes (TAN) classifier** on a biped robot platform. The TAN is a probability measurement to determine what emotion is being expressed. Measures learning rate of students through affective emotions expressed by students while the robot mimics teacher's actions.	Actions can be repeated if emotions detected are confused or frustrated. If emotions such as sad are detected, the robot tries a different set of actions instead. This type of response helped maintained a positive learning rate in this study.
55	Slade and Prinsloo (2013)	American Behavioral Scientist	Q2	Review	N/A	A socio-critical perspective on the use of learning analytics.	There is a need for consideration of the ethical dimensions and challenges of learning analytics. The proposed principles provide an ethical framework for higher education institutions to offer context-appropriate solutions and strategies to increase the quality and effectiveness of teaching and learning.
56	Sullivan and Bers (2016)	International Journal of Technology and Design Education	Q4	Quantitative	60 children from pre-kindergarten to 2nd grade	Children participate in 8-week robotics curriculum in their classrooms using the KIWI robotics kit, followed by **post-test for statistical significance** in robot and programming knowledge.	Tests for statistical significance indicates that the kindergarten, first, and second grade classes performed equally well on advanced programming in the same amount of time.
57	van Leeuwen et al. (2015)	Computers in Human Behavior	Q1	Qualitative	2 male history teachers, 51 secondary school students (mean age of 14)	Teachers make use of learning analytics information provided to them via Virtual Collaborative Research Institute (VCRI) to try and regulate the class, deciding when to intervene. Assessment done using **interviews** for teachers' reports about their strategies for diagnosing and intervening, and the associated challenges and opportunities.	Teachers must continuously choose which level (individual, group and class) to monitor and how to divide their attention. The manageability of the available information decreased due to high information load and the teachers were not always able to maintain an overview of all student activities.

58	Wu et al. (2015)	Interactive Learning Environments	Q1	Mixed	64 3rd grade students in Yunlin County, Taiwan	A teaching assistant robot named Powerful English Tutor (PET) was created to assist in teaching English as a Foreign Language. Students were **randomly assigned** to either a control group without PET and a treatment group. **Post-test** to assess performance, **survey** for motivation, **questionnaire** for perception and **observational video recordings**. **Statistical t-tests** were performed.	Treatment group had significantly better performance and motivation. Assistant robot was highly valued. Voice recognition is challenging in classroom environments which are normally noisy.
59	Yi et al. (2016)	Journal of Intelligent & Robotic Systems	Q3	Quantitative	65 under-graduates from Robotics & Mechanisms Laboratory (RoMeLa), University of California, Los Angeles	Using a Dynamic Anthropomorphic Robot with Intelligence(DARwIn)-High Performance(HP) as an educational tool in robotics undergraduate classes. The impact is experimented on number of regular, extracurricular and outreach robotics activities **before and after** participating in DARwIn-HPs development.	This study shows that undergraduates who attended the DARwIn-HP development are likely to feel strongly the necessity for studying STEM curriculum than before.
60	Zacharis (2015)	The Internet and Higher Education	Q1	Quantitative	134 university freshmen students of Computer Science and Computer Engineering courses	A **bivariate correlation** between 29 online activities with student grade resulted in 14 variables with strong impact on student final achievement, which were then used as the input in a **regression analysis**. **Exploratory univariate regression analyses** for student age, gender, previous grades, working status and ethnicity revealed that none of these variables had a significant effect on course grade.	Overall accuracy of the prediction of student risk of failing is 81.3%. Only 4 variables: reading and posting messages, content creation contribution, quiz efforts and number of files viewed - predicted 52% of the variance in the final student grade (active participation).
61	Zaga et al. (2015)	Lecture Notes in Artificial Intelligence	Q4	Mixed	20 children, 6-9 years old in Netherlands	A study on the effect of two different social characters of NAO robot (peer vs. tutor) on children engagement measured at the cognitive (attention to the task and the robot), affective (emotional response to the task), and behavioral (performance) level by frequency and duration of these attributes. **Video recording** is analyzed and a **questionnaire** is given.	In the peer character condition, children paid attention to the robot and the task for a longer period of time and solved the puzzles quicker and better than in the tutor character condition.
62	Zawieska and Duffy (2015)	Advances in Intelligent Systems and Computing	Unranked	Review	N/A	This paper argues that a new form of creativity concerns the meanings students make of anthropomorphic robots in the course of human-robot social interaction. This is based on the following assumptions: creativity is socially constructed and the main reason for students to be interested in robotics is a fascination with the illusion of life.	The combination of anthropomorphic robot design and social interaction can result in new ways to foster creativity.

Appendix B: Questionnaire Form

Subjective Experience

Based on your experience with the session, please answer the following questions honestly and as accurately as possible.

Name: ____________________________

Age: ____________________________

Questions:

1. At any point during the session, did you feel bored, scared or frustrated? If yes, what do you think caused this and why?

2. Did you interact with the tutor in any way, or not at all? Describe this interaction or the lack thereof, and how you feel about it.

3. In short, how would you describe your experience and what suggestions would you make to improve this 15-minute session?

Short Flow State Questionnaire (FSS-2)

Please answer the following questions based on your experience and interaction with the session. There are no right or wrong answers. Circle the number that best matches your experience from the options at the bottom of each statement.

Rating Scale:

Strongly disagree 1	Disagree 2	Neither agree nor disagree 3	Agree 4	Strongly agree 5

Statement:

1. I felt I was competent enough to meet the high demands of the situation.

1	2	3	4	5

2. I did things spontaneously and automatically without having to think.

1	2	3	4	5

3. I had a strong sense of what I want to do.

1	2	3	4	5

4. I had a good idea while I was performing about how well I was doing.

1	2	3	4	5

5. I was completely focused on the task at hand.

1	2	3	4	5

6. I had a feeling of total control over what I'm doing.

1	2	3	4	5

7. I was not worried about what others may have been thinking of me.

1	2	3	4	5

8. The way time passed seemed to be different from normal.

1	2	3	4	5

9. The experience was extremely rewarding.

1	2	3	4	5

Appendix C: Explanatory Statement

Explanatory Statement for Research Participants

Project: Assessing Learning Engagement using Humanoid Robots in Higher Education

You are invited to take part in a research study to analyze the role of robot tutors in the university classroom on student learning experience compared to human tutors.

This project is currently being undertaken by Nicholas Wong Wai Hong, a student of Monash University, Malaysia under the Master of Philosophy course. The research project members are:

<table>
<tr>
<td>Dr. Dharmaratne Anuja Thimali
Chief Investigator
School of Information Technology
[redacted]</td>
<td>Dr. Jojo Wong Sze-Meng
Co-Investigator
School of Information Technology
[redacted]</td>
<td>Catherine Hiew Fui Chin
Research Assistant
School of Information Technology
[redacted]</td>
</tr>
<tr>
<td>Dr. Lim Jen Nee Jones
Co-Investigator
School of Engineering
[redacted]</td>
<td>Dr. Esyin Chew
External Co-Investigator (UK)
Cardiff School of Management
[redacted]</td>
<td>Teo Bee Guan
Research Assistant
School of Information Technology
[redacted]</td>
</tr>
<tr>
<td colspan="3">Nicholas Wong Wai Hong
Co-Investigator
School of Information Technology
[redacted]</td>
</tr>
</table>

Please read this explanatory statement in full before deciding to participate in this research. This statement is to be read in conjunction with the attached consent form. If you would like further information regarding any aspect of this project, you are encouraged to contact the researchers via the phone numbers or email addresses listed above.

1. **What is the purpose of this study?**
   This research aims to compare the effects of student engagement in eight 15-minute learning sessions (once a week, for eight weeks) between a human tutor and a robot tutor. By performing a case study of Monash University students from different units, similarities and differences in student learning experience are explored to study the effects of robot tutors on engagement and flow in the higher education context.
2. **Why have I been invited to participate in this study?**
   You have been invited to participate in this research because you are a Monash University student, a potential candidate within the higher education context which is the subject of this study.
3. **What does the research involve?**
   You will be randomly assigned to a 15-minute learning session taught either by a human tutor or a robot along with other participants. If you were assigned to the human tutor group, just treat it like a normal class. Otherwise, this study involves listening and observing the NAO humanoid robot about 58cm tall placed at the front of a classroom. A human facilitator will be there to aid you and the robot. The robot will begin by introducing a topic based on what you may or may not have learned in a Monash University unit.

   After it has gone through some topics, you may need to do a class exercise or the robot may ask some questions for the topics discussed. At this time, you may or may not be called up front to interact with the robot, in which the distance would be about 1 meter away from you. You will be presented choices for each question. The session ends when all exercises or questions have been done. After this interaction, you will be given a questionnaire which collects data about your experience with the human or robot tutor. This session will be done once per week, for 8 weeks, amounting to 8 sessions in total.

Please note that your participation in this study is voluntary by signing and returning the attached consent form. You may choose to discontinue at any stage of the study without explanation. Only non-personally identifiable information collected from questionnaires are subject to publishing. Data which are no longer required will be deleted.

4. **Are there any possible benefits or risks from participation in this study?**
Your responses will help in analyzing human-robot interactions for higher education learning. A summary of the key findings will be made available to you as a participant upon request. These findings may tell us more about how a robot tutor affects your engagement and flow in the higher education context, which may help move us a step closer to the integration of robots in university classrooms. At the end of the study, you will be given a **RM 20** gift voucher as an appreciation for your time on the project.

   You may feel uncomfortable when interacting with the robot. Although the interaction is rather short, please be reminded that you may discontinue at any time.

5. **What if I have any questions about this research?**
If you would like to discuss any aspect of this study please feel free to contact Nicholas Wong Wai Hong by e-mail: nicholas.wong1@monash.edu. We would be happy to discuss any aspect of the research with you. Once we have analyzed the information we can e-mail you a summary of our findings on request. You are welcome to contact us at that time to discuss any issues relating to the research.

   Should you have any concerns or complaints about the conduct of the project, you are welcome to contact the Executive Officer, Monash University Human Research Ethics (MUHREC):

   Executive Officer
   Monash University Human Research Ethics Committee (MUHREC)
   [redacted]

   [redacted]

Thank you for taking the time to consider this study. If you wish to take part in it, please sign the attached consent form. This explanatory statement is for you to keep.

Appendix D: Consent Form

Consent Form

Project: Assessing Learning Engagement using Humanoid Robots in Higher Education

Dr. Dharmaratne Anuja Thimali Chief Investigator School of Information Technology [redacted]	**Dr. Jojo Wong Sze-Meng** Co-Investigator School of Information Technology [redacted]	**Catherine Hiew Fui Chin** Research Assistant School of Information Technology [redacted]
Dr. Lim Jen Nee Jones Co-Investigator School of Engineering [redacted]	**Dr. Esyin Chew** External Co-Investigator (UK) Cardiff School of Management [redacted]	**Teo Bee Guan** Research Assistant School of Information Technology [redacted]
Nicholas Wong Wai Hong Co-Investigator School of Information Technology [redacted]		

I agree to take part in the Monash University research project specified above. I have had the project explained to me, and I have read the Explanatory Statement, which I keep for my records. I understand that agreeing to take part means that I am willing to:

	Yes	No
Carry out eight 15-minute educational sessions, once a week for eight weeks, with either a human or robot tutor.	☐	☐
Fill up a questionnaire regarding subjective experiences with the tutor.	☐	☐
I have received the incentive of RM20 gift voucher.	☐	☐

I understand that my participation is voluntary, that I can choose not to participate in part or all of the project, and that I can withdraw at any stage of the project without being penalized or disadvantaged in any way. I understand that although the questionnaire requires an identifying characteristic such as a name, I can use my first name, an alias or any generated name so as long as the researcher can keep track of my responses throughout the 8 different sessions. I understand that any information I provide is confidential which will be kept in a secure storage accessible to the researchers listed above only, and that no information that could lead to the identification of any individual will be disclosed in any reports on the project, or to any other party. I understand that the provided information will be destroyed when no longer needed for the study. I understand that non-personally identifiable information extracted from the questionnaire may be used as case studies in the project website, presented in conference(s) or published in journal paper(s).

Participant Name	Participant Signature	Date
Investigator Name	**Investigator Signature**	**Date**

Appendix E: MUHREC Ethics Approval

MONASH University

Monash University Human Research Ethics Committee

Approval Certificate

This is to certify that the project below was considered by the Monash University Human Research Ethics Committee. The Committee was satisfied that the proposal meets the requirements of the *National Statement on Ethical Conduct in Human Research* and has granted approval.

Project Number:	7953
Project Title:	Assessing Learning Engagement using Humanoid Robots in Higher Education
Chief Investigator:	Dr a anuja
Expiry Date:	27/02/2022

Terms of approval - failure to comply with the terms below is in breach of your approval and the *Australian Code for the Responsible Conduct of Research.*

1. The Chief Investigator is responsible for ensuring that permission letters are obtained, if relevant, before any data collection can occur at the specified organisation.
2. Approval is only valid whilst your hold a position at Monash University.
3. It is responsibility of the Chief Investigator to ensure that all investigators are aware of the terms of approval and to ensure the project is conducted as approved by MUHREC.
4. You should notify MUHREC immediately of any serious or unexpected adverse effects on participants or unforeseen events affecting the ethical acceptability of the project.
5. The Explanatory Statement must be on Monash letterhead and the Monash University complaints clause must include your project number.
6. Amendments to approved projects including changes to personnel must not commence without written approval from MHUREC.
7. Annual Report - continued approval of this project is dependent on the submission of an Annual Report.
8. Final Report - should be provided at the conclusion of the project. MUHREC should be notified if the project is discontinued before the expected completion date.
9. Monitoring - project may be subject to an audit or any other form of monitoring by MUHREC at any time.
10. Retention and storage of data - The Chief Investigator is responsible for the storage and retention of the original data pertaining to the project for a minimum period of five years.

Thank you for your assistance.

Professor Nip Thomson

Chair, MUHREC

CC: Dr Jojo Wong, Dr a anuja, Nicholas Wong, Fui Hiew, Teobee Guan, Dr Esyin Chew

List of approved documents:

Document Type	File Name	Date	Version
Consent Form	consent-form-new	31/01/2017	1
Explanatory Statement	explanatory-statement-new	01/02/2017	3
Questionnaires / Surveys	questionnaire-form-new	01/02/2017	2

Appendix F: Word Count of Comments

*Only the top 50 words from the students' comments are shown.

Human Tutor		Robot Tutor	
good	39	robot	97
tutor	28	tutor	44
normal	20	interesting	34
questions	18	experience	27
interaction	16	interaction	26
experience	16	good	24
interesting	14	better	19
nothing	13	speech	18
explained	12	time	18
great	11	fun	18
clear	10	interact	18
asked	10	feel	15
information	9	bit	15
question	9	students	14
pretty	9	clear	14
frustrated	9	need	13
suggestions	9	needs	12
problem	9	question	12
fun	8	bored	12
students	8	new	12
feel	8	voice	11
bored	8	answer	10
ask	7	felt	10
better	7	able	10
help	7	recognition	10
session	7	quite	10
explanation	7	lack	10
class	7	improve	10
helpful	7	understand	10
fine	6	little	10
maybe	6	great	10
really	6	instructions	9
informative	6	however	9
well	6	session	9
us	6	student	8
enough	6	slow	8
need	6	words	8
assignment	6	fast	8
maya	6	interactive	8
quite	6	hard	7
teaching	6	talking	7
none	6	make	7
lack	6	might	7
interact	5	fine	6
engaging	5	maybe	6
scared	5	music	6
asking	5	really	6
felt	5	much	6
much	5	improved	6
answered	5	nothing	6