[Page de garde]

HEC MONTRÉAL

Impacts of Gamification on Learning and Engagement in the Context of Learning Data Analytics

par

Melissa Thériault

Sous la codirection de

Pierre-Majorique Léger, Ph. D.

et

Jean-François Plante, Ph. D.

Sciences de la gestion
(Spécialisation Expérience Utilisateur)

Mémoire présenté en vue de l'obtention
du grade de maîtrise ès sciences en gestion
(M. Sc.)

April 2021

Résumé

La venue de données massives a menée à une augmentation nette de la demande pour des scientifiques des données qualifiés, qui peuvent structurer et organiser des données pour générer des informations pertinentes, afin de guider les entreprises dans leurs prises de décisions. Dans un contexte où les entreprises sont de plus en plus axées sur les données, la recherche démontre que la demande de scientifiques des données, ainsi que les "data translators", qui permettent aux entreprises de transposer les informations clés des données en stratégie pour résoudre les problèmes de gestion, dépasse l'offre actuelle. Une solution proposée dans le but de pallier cette problématique est la ludification de l'apprentissage. La ludification, qui se définit par l'intégration de mécaniques de jeux dans différents domaines, favorise la motivation, l'engagement et une meilleure performance d'apprentissage. Cependant, il y a peu de résultats qui confirment que la ludification est également efficace pour l'apprentissage de la science des données. Or, selon la théorie de l'apprentissage ludique, la ludification n'a pas d'impact direct sur l'apprentissage. Il est toutefois probable qu'il existe plusieurs facteurs indirects pouvant influencer cette relation. Ainsi, ce mémoire évalue l'impact de certaines caractéristiques individuelles pouvant influencer l'efficacité d'une mécanique de jeu, notamment la dynamique de compétition engendrée par la présence simultanée d'autres participants dans un tableau de classement, sur l'apprentissage et l'engagement. Les effets modérateurs de l'auto-efficacité et les connaissances antérieures sur la relation susmentionnée sont également examinées.

Une étude inter-sujet de 37 participants a été conçue, où il était possible d'évaluer les effets de l'aspect compétitif d'un tableau de classement dans le contexte d'une simulation d'entreprise qui vise l'enseignement de l'analyse prédictive. Les résultats suggèrent que la nature compétitive du tableau de classement n'a pas eu d'effet direct significatif sur l'apprentissage et l'engagement. Cela étant dit, les résultats suggèrent qu'une personne avec un niveau élevé d'auto-efficacité et des connaissances antérieures qui joue avec un tableau de classement compétitif aura peu d'apprentissage additionnel sur le sujet de l'analyse prédictive, malgré le fait qu'elle aura tendance à démontrer un niveau d'engagement émotionnel supérieur.

D'un point de vue pratique, nos résultats suggèrent que les instructeurs devront développer ou adapter les formations ludiques afin de mieux prendre en compte les différences individuelles et d'améliorer l'efficacité de l'apprentissage ludique de la science des données. Ce mémoire contribue également à la littérature existante sur les effets de la ludification sur l'apprentissage, notamment en réalisant une étude basée sur la théorie.

Mots clés : Ludification, Tableaux de classement, Simulation d'entreprise, Apprentissage, Engagement, Analyses de données, Expérience contrôlée

Abstract

The rise of big data has led to an explosion of demand for well-qualified data scientists to make sense of unstructured data and generate valuable insights to help guide businesses in their decision making. As companies become more data-driven, research shows that the current demand for data scientists as well as data translators, who enable businesses to apply the insights derived from data scientists to solve problems, exceeds the current supply. A solution to train students faster to meet this demand is gamification. Gamification, which is the application of game elements in non-game contexts, has been shown to foster high learning performance, motivation, and engagement. However, there are little studies that provide evidence about the effectiveness of gamification on the learning of data science. Additionally, according to the Theory of Gamified Learning, which stipulates that gamification does not have a direct impact on learning, it is highly probable that there are factors that moderate this relationship. This suggests that certain individual characteristics of learners may influence the effectiveness of gamification on learning and engagement. Thus, this thesis evaluates whether the competitive nature of a specific game element, the leaderboard, affects learning outcomes and engagement. Moderating effects of self-efficacy and prior knowledge on the aforementioned relationship are also examined.

A between-subject experiment with 37 participants was devised, in which the effect of a competitive leaderboard was examined within the context of a business simulation that teaches predictive analytics. The results show that the competitive nature of the leaderboard did not have any significant effect on learning and engagement. However, it was found that those with high levels of self-efficacy and prior predictive modelling knowledge playing with a competitive leaderboard, showed higher levels of arousal despite having a lower probability of increasing their knowledge on predictive modelling.

From a practical standpoint, our findings suggest that instructors should develop or adapt existing gamified training to better account for individual differences to increase the effectiveness of gamification in learning data science. This thesis also contributes to the literature by providing a theory-driven study on the effects of gamification.

Keywords : Gamification, Leaderboards, Business Simulation, Learning, Engagement, Data Analytics, Controlled Experiment

Table of Contents

List of Tables and Figures

List of Tables

Chapter 2

List of Figures

Chapter 2

List of Abbreviations and Acronyms

SDT - Self-Determination Theory

SCT - Social Cognitive Theory

PBL - Problem Based Learning

EUT - End-User Training

IS - Information Systems

SSE - Software Self-Efficacy

STEM - Science, Technology, Engineering, Mathematics

Preface

This dissertation, written in the form of an article, was approved by the Administrative Direction of the Master of Science in User Experience in a Business Context program and the co-authors' consent of the article composing the dissertation was obtained.

The approval of the HEC Montréal Research Ethics Board (CER) was received for this study in May 2020.

The article has been submitted and accepted to the 2021 Americas Conference on Information Systems (AMCIS) taking place in Montreal, Canada.

The article was added to the dissertation with the signed consent of the co-authors.

Acknowledgements

I would like to start by thanking my supervisors Pierre-Majorique Léger and Jean-François Plante without whom none of this would have been possible. I would like to extend my gratitude for their guidance, patience, advice, and for empowering me to be confident in moving this research project forward. They helped me find and develop a research project that complimented my past experiences and interests and supported me until the very end. They are both very dedicated to their students and this enriching experience will follow me well into my professional career.

I would also like to thank my professor, Constantinos Coursaris, for the incredible learning experience as the first English cohort of the M.Sc. in User Experience at HEC. His expertise, teaching, and mentorship during my transition into the field of UX were invaluable.

In addition, I would like to thank Jean-François Michon, Forough Karimi-Alaghehband, and the rest of the amazing team at ERPSim Lab for their collaboration and contribution. I would like to thank them for teaching me how to use Cortex and collaborate together to implement the necessary development modifications to the game in order to bring my project to fruition.

This research project would not have come to life without the dedication and help from the research assistants and operations team at the Tech3lab. I am grateful for their patience and guidance throughout this process. I am also thankful for the other members of the laboratory that listened and advised me on my project during the past 2 years. In addition, I would like to thank Shang Lin Shen for his help and counsel with the statistical analysis.

A special thank you also goes out to Thomas Ruel for the incredible work he contributed to this project during his internship at Tech3lab. Our teamwork made the data collection period productive and fun!

Furthermore, I would like to express my gratitude to ERPSim Laboratory for funding my research.

Finally, I would like to thank my partner Alexis for his continuous support, patience, and encouragement. His belief in me gave me the force needed to preserve during this long and challenging journey.

Introduction

Context

The advent of the Internet has completely transformed the way data is collected and shared and in turn, has led to the rise of big data, providing businesses easy access to new insights about their organizations to guide important managerial decision-making choices (Cukier and Mayer-Schonbergeer 2013). Research by McAfee and Brynjolfsson published in a 2012 *Harvard Business Review* article states, “companies in the top third of their industry in the use of data-driven decision making were, on average, 5% more productive and 6% more profitable than their competitors.” As companies across various industries have become committed to data-driven transformations to stay competitive over the past years, the demand for skilled labour in data, analytics, artificial intelligence, and machine learning has exploded (Demchenko et al. 2016; Lohr 2017). Companies need trained data science professionals that can use advanced analytics and data visualisation tools to solve business problems by finding patterns and infer understandable insights out of messy, unstructured data (Mohanty et al 2013). Another important role in enabling the analytical transformation of businesses are data or analytics translators. These professionals have the organizational as well as industry expertise to leverage the insights generated by data scientists to lead practical application of these findings to the business (Henke et al. 2016). Thus, there is also a high demand for data translators who have a certain level of fluency in analytics to enable businesses to achieve real impact with their data (Henke et al. 2016; Henke, Levine, and McInerney 2018).

However, there has been a growing concern about the supply of qualified talent lagging behind the growing demand. In 2011, The McKinsey Global Institute presented a report on the future of data science/analytics job market landscape in the United States and predicted that by 2018, a talent pool of 2.8 million would emerge. They projected that by 2020, 2.72 million job openings would emerge, however, there were already 2.35 million listings open by 2015 alone (Manyika et al. 2011). This would suggest that the demand is increasing at a much faster rate than originally predicted by McKinsey. More recently, in 2016, McKinsey reported that there could be a shortfall of up to 250 000 data scientists as well as an increase in demand for business translators from 2 million to 4 million in the United States job market within the next decade (Manyika et al. 2017). Considering that the Bureau of Labor Statistics (2020) also projects a 31% growth rate for data science occupations in the United States, it would seem that the concern about imbalances in supply and demand for skilled labour continues to perpetuate the job market.

Universities and training providers have made proactive efforts to mitigate the labour shortage, offering undergraduate and/or graduate degrees in data science, certificates, online courses, or bootcamps to upskill students outside the traditional routes of statistics or computer science degrees (Kandel et al. 2012; Hong et al. 2018; National Academies of Sciences 2018). Even so, for Science, Technology, Engineering, Mathematics (STEM) programs to be successful, learners must be engaged to reduce likelihood of low attendance rates, low performance, or success rates (Campbell et al. 2004; Chen 2013; Margison et al 2013; OECD 2016).

A popular method to improve student engagement and learning in educational contexts is the application of gamification. Gamification, which is the use of video game elements in non-game contexts such as points, badges, and leaderboards (Deterding et al. 2011), has been applied in various fields including STEM-oriented education, and has shown to foster improved learning performance, increased motivation, and engagement with learning activities (Ortiz-Rojas et al. 2017; Ortiz et al. 2016; Xi and Hamari 2019; Peterson 2001; Smith and Van Doren 2004). However, few studies in the gamification literature have examined the theoretical grounds of gamification's effect on learning outcomes and overall engagement (Seaborn and Fels 2015; Santhanam et al. 2016; Nacke and Deterding 2017), which are important to understand how to engage and motivate learners in STEM to promote higher quality learning, and ultimately, create a higher supply of well-qualified professionals to close the labour gap.

Research Objective and Question

The literature suggests that gamification is an effective method to promote positive learning outcomes, motivation, attitude, engagement, and makes the overall experience of completing a task or learning more enjoyable (Landers et al 2015; Dominguez et al. 2013; de Sousa Borges et al 2014; Eickhoff et al. 2012; Halan et al 2010). While there is growing literature on gamification in STEM-oriented fields, to our knowledge, there has been little research conducted that examines gamification in the context of learning data science. The majority of gamification literature also examines multiple game elements within learning contexts, thus there are not many studies exploring the effectiveness of specific game elements that exist (Dicheva et al. 2017). Furthermore, it is also unclear if gamification has a direct effect on learning in general (Landers 2015). Thus, this thesis will seek to answer the following research questions:

RQ1. "To what extent does the competitive nature of leaderboards impact learning performance and engagement in the context of learning data analytics?"

RQ2. "To what extent does self-efficacy moderate the impact of the competitive nature of leaderboards have on learning and engagement?

RQ3. "To what extent does individual differences in expertise moderate the impact competitive nature of leaderboards have on learning and engagement?

Potential Research Contributions

Acting on the call to add more theory-driven experiments to the current gamification literature (Seaborn and Fels 2015; Santhanam et al. 2016; Nacke and Deterding 2017), this thesis will examine the effects of gamification through two leading motivational theories: self-determination theory and social cognitive theory (Deci and Ryan 1980; Bandura 1977;1986). Furthermore, this thesis isolates the effect of the competitive nature of one gamification element, the leaderboard, on learning outcomes and engagement. This will provide empirical evidence of the impact of a specific gamification element instead of a combination of multiple elements, which is rare in the gamification literature (Dicheva et al. 2017) Finally, building on Lander's (2014) Theory of Gamified Learning, which posits that gamification does not have direct effects on learning, this thesis includes two moderators, self-efficacy and prior knowledge, to provide evidence of the indirect effects of leaderboards on learning and engagement.

Regarding practical contributions, if the results are conclusive, this thesis can provide insights for professors and training and development professionals that choose to introduce business simulations that leverage leaderboards to foster better learning outcomes and engagement when teaching the complexities of data analytics.

Main Article

Article information

This article has been submitted and accepted to the *Americas Conference on Information Systems 2021* (AMCIS) (Thériault et al. 2021). It was co-authored by Thomas Ruel, Pierre-Majorique Léger, and Jean-François Plante. The article thoroughly reviewed gamification and learning's theoretical foundation and interprets the data collected from the study's experiment through theoretical lenses.

Article summary

Data scientists are amongst the most popular roles in today's job market due to the ever growing need to make sense of big data. However, there seems to be a shortage of well-qualified professionals to fill these open positions. Gamification in higher education and end-user training has been shown to be successful in teaching individuals the skills and knowledge needed to be competent in technical roles. However, since there is little research that provides evidence that gamification is an effective method to teach data science concepts and since the Theory of Gamified Learning (Landers 2015) posits that gamification does not have a direct effect on learning, we aim to investigate both direct and moderated effects of gamification on learning and engagement. We hypothesize that the competitive nature of a specific game element, the leaderboard, will lead to improved learning and engagement and that self-efficacy and prior knowledge of predictive modelling will positively moderate the aforementioned relationship. To test the latter, this article presents a between-subject experiment that investigates the effect of gamification, specifically the game element of competitive leaderboards, on learning outcomes and engagement in context of learning data analytics. A business simulation game wherein subjects needed to solve a problem using predictive analytics was used. We created a condition with a non-competitive leaderboard and one with a competitive leaderboard which included fictitious opponents to observe differences in learning and engagement. Results did not show any significant differences in learning and engagement between conditions. We did find that, for those in the competitive leaderboard condition, highly efficacious participants had a lower probability of learning predictive modelling knowledge, but also showed higher levels of arousal. We also found that prior knowledge also reduces the probability of participants' ability to improve their in-game scores when playing with the competitive leaderboard. One major implication of this work centers on highlighting the importance of individual differences of learners when using gamified learning to teach data science and providing a theory-driven study to the existing literature.

Contributions

The current study has been completed whilst working at the Tech3lab. To better understand my contribution to the present research project, the table below outlines my contributions as a percentage of work completed for each step of the project.

Step	Contribution
Identifying the research problem	Definition of the research question and the problem - 90% ● Due to my previous experience with learning and development, my supervisors proposed the possibility of examining the effect of gamification on learning while playing a business simulation created by ERPSim Lab ● My supervisors contributed to the definition of the research question and the approach to take
Literature Review	Review the literature to identify past studies, and constructs and measures to test - 100% Define and propose constructs and measures to be used in the experiment - 80% ● My supervisors offered feedback on chosen constructs and measures and validated the reliability of each.
Ethics	Complete the submission to the CER and subsequent modifications - 100%
Experimental Design	Creation of the experimental protocol and design - 90%

Recruitment and participant management	Creation of the recruitment form - 100% Solicitation and recruitment of participants - 90% ● I had help from an intern (Thomas Ruel) to recruit a few participants. Experiment schedule management - 90% ● I coordinated with participants and managed the schedule and was assisted by an intern (Thomas Ruel) when needed. Manage participant compensation - 100%
Pre-testing and data collection	Responsible for the pretests - 100% Responsible for data collection - 60% ● During the data collection, I was assisted by an intern (Thomas Ruel), but made sure the stimuli was up and running smoothly.
Data extraction and transformation	Extraction and cleaning of the data from questionnaires and video recording data - 75% ● I was assisted by an intern (Thomas Ruel) to clean half of the questionnaire data and a laboratory technician (Salima Tazi) for the extraction of the facial expression data.
Data analysis	Formatting data to be analyzed - 100% Statistical analysis - 80% ● The laboratory Statistician (Shang Lin Chen) helped me with the statistical analysis
Writing the article	Writing the scientific article - 100%

Table 1. Contributions and personal responsibilities of the research project

Thesis Structure

This thesis takes the form of an article and is structured in the following manner. First, a literature review outlining the theoretical foundations of gamification and its relationship with learning and engagement will be presented. Second, the thesis article detailing the execution and results of the experiment conducted to evaluate the effectiveness of competitive leaderboards in the context of learning data analytics will be presented. Finally, a conclusion reiterating key findings, contributions, and limitations of the thesis article and future research directions will be presented.

References

Campbell, P. B., Jolly, E. J., and Perlman, L. 2004. "Introducing the Trilogy of Success: Examining the Role of Engagement, Capacity and Continuity in Women's Stem Choices," *Women in Engineering ProActive Network*.

Chen, X. 2013. "Stem Attrition: College Students' Paths into and out of Stem Fields. Statistical Analysis Report. Nces 2014-001," *National Center for Education Statistics*.

Cukier, K., and Mayer-Schoenberger, V. 2014. *The Rise of Big Data: How It's Changing the Way We Think About the World*. Princeton University Press.

Demchenko, Y., Belloum, A., and Wiktorski, T. 2016. "Edison Data Science Framework: Part 1. Data Science Competence Framework (CF-DS) Release 1," Zenodo.

de Sousa Borges, S., Durelli, V. H., Reis, H. M., and Isotani, S. 2014. "A Systematic Mapping on Gamification Applied to Education," *Proceedings of the 29th annual ACM symposium on applied computing*, pp. 216-222.

Dichev, C., and Dicheva, D. 2017. "Gamifying Education: What Is Known, What Is Believed and What Remains Uncertain: A Critical Review," *International journal of educational technology in higher education* (14:1), pp. 1-36.

Domínguez, A., Saenz-de-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., and Martínez-Herráiz, J.-J. 2013. "Gamifying Learning Experiences: Practical Implications and Outcomes," *Computers & education* (63), pp. 380-392.

Eickhoff, C., Harris, C. G., de Vries, A. P., and Srinivasan, P. 2012. "Quality through Flow and Immersion: Gamifying Crowdsourced Relevance Assessments," *Proceedings of the 35th international ACM SIGIR conference on Research and development in information* *retrieval*, pp. 871-880.

Halan, S., Rossen, B., Cendan, J., and Lok, B. 2010. "High Score!-Motivation Strategies for User Participation in Virtual Human Development," *International conference on intelligent virtual agents*: Springer, pp. 482-488.

Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B., and Sethupathy, G. 2016. "The Age of Analytics: Competing in a Data-Driven World," *McKinsey Global Institute* (4).

Henke, N., Levine, J., and McInerney, P. 2018. "You Don't Have to Be a Data Scientist to Fill This Must-Have Analytics Role," *Harvard Business Review*, February (5).

Hong, T., Gao, D. W., Laing, T., Kruchten, D., and Calzada, J. 2018. "Training Energy Data Scientists: Universities and Industry Need to Work Together to Bridge the Talent Gap," *IEEE Power and Energy Magazine* (16:3), pp. 66-73.

Kandel, S., Paepcke, A., Hellerstein, J. M., and Heer, J.. 2012. "Enterprise Data Analysis and Visualization: An Interview Study", *IEEE Transactions on Visualization and Computer Graphics* (18:12), pp.2917–2926

Landers, R. N. 2014. "Developing a Theory of Gamified Learning: Linking Serious Games and Gamification of Learning," *Simulation & gaming* (45:6), pp. 752-768.

Landers, R. N., Bauer, K. N., Callan, R. C., and Armstrong, M. B. 2015. "Psychological Theory and the Gamification of Learning," in *Gamification in Education and Business*. Springer, pp. 165-186.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., and Hung Byers, A. 2011. *Big Data: The Next Frontier for Innovation, Competition, and Productivity*. McKinsey Global Institute.

Manyika, J., Chui, M., Lund, S., and Ramaswamy, S. 2017. "What's Now and Next in Analytics, Ai, and Automation," *McKinsey Global Institute*, pp. 1-12.

Marginson, S., Tytler, R., Freeman, B., and Roberts, K. 2013. "Stem: Country Comparisons: International Comparisons of Science, Technology, Engineering and Mathematics (Stem) Education. Final Report."

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D., and Barton, D. 2012. "Big Data: The Management Revolution," *Harvard business review* (90:10), pp. 60-68.

Mohanty, S., Jagadeesh, M., and Srivatsa, H. 2013. *Big Data Imperatives: Enterprise 'Big Data'* *warehouse ,'Bi 'implementations and Analytics*. Apress.

Nacke, L. E., and Deterding, S. 2017. "The Maturing of Gamification Research," *Computers in Human Behavior* (71), pp. 450-454.

National Academies of Sciences, E., and Medicine. 2018. *Data Science for Undergraduates: Opportunities and Options*. National Academies Press.

Lohr, S. 2017. "Where are the STEM Jobs Are (and Where They Aren't). *New York Times*, p. 10 of Education Life.

OECD. 2016. *Pisa 2015 Results (Volume I): Excellence and Equity in Education.* PISA, OECD.

Ortiz Rojas, M. E., Chiluiza, K., and Valcke, M. 2017. "Gamification and Learning Performance: A Systematic Review of the Literature," *11th European Conference on Game-Based Learning (ECGBL)*: ACAD CONFERENCES LTD, pp. 515-522.

Ortiz Rojas, M. E., Chiluiza, K., and Valcke, M. 2016. "Gamification in Higher Education and Stem: A Systematic Review of Literature," *8th International Conference on Education and New Learning Technologies (EDULEARN)*: Iated-int Assoc Technology Education A& Development, pp. 6548-6558.

Peterson. 2001. "Course Participation: An Active Learning Approach Employing Student Documentation," *Journal of Marketing Education* (23:3), pp. 187-194.

Santhanam, R., Liu, D., and Shen, W.-C. M.. 2016. "Research Note—Gamification of Technology-Mediated Training: Not All Competitions Are the Same", *IS Research* (27:2), pp. 453–465

Seaborn, K., and Fels, D. I. 2015. "Gamification in Theory and Action: A Survey," *International Journal of Human-Computer Studies* (74), pp. 14-31.

Smith, L. W., and Van Doren, D. C. 2004. "The Reality-Based Learning Method: A Simple Method for Keeping Teaching Activities Relevant and Effective," *J of Marketing Ed* (26:1), pp.66-74.

Thériault, M., Ruel, T., Léger, P.-M., and Plante, J.-F. 2020. "Can Competition Though Leaderboards Lead to Better Engagement and Learning of Data Science Concepts? An Experimental Study." *2021 AMCIS Proceedings.*

U.S. Bureau of Labor Statistics. 2020. "Fastest Growing Occupations : Occupational Outlook Handbook." Retrieved 03.03., 2021, from https://www.bls.gov/ooh/fastest-growing.htm

Xi, N., and Hamari, J.. 2019. "Does gamification satisfy needs? A study on the relationship between gamification features and intrinsic need satisfaction", *International Journal of Information Management* (46), pp. 210–22.

Chapter 1: Literature review

1.1 Introduction

As big data continues to become readily available and as more companies adopt data-centered approaches to decision-making, the need to extract useful insights to help businesses make those important decisions continues to grow in tandem (Zhang et al. 2020; Dobrin 2017). In turn, data scientists, who are often referred to as professionals that use modern machine learning techniques to make sense of and identity insights from structured and unstructured data (Kim et al. 2016; Muller et al. 2019) were predicted by Google's Chief Economist, Hal Varian, to be, "the sexy job in the next ten years," (Manyika 2009) during an interview with McKinsey in 2009. In 2012, an article published by *Harvard Business Review* coined data science as the "Sexiest Job of the 21st Century," (Davenport and Patil 2012). Several years later, the field of data science is still amongst the most sought out talent on the job market with no signs of slowing down. Indeed, the Bureau of Labor Statistics (2020) projects a 31% growth rate for data science occupations in the United States in coming years. Despite the rise in popularity, there is growing concern over the small abundance of well-qualified data science professionals. In fact, a report published by The McKinsey Global Institute in 2016 predicted a potential labour shortage of up to 250 000 data scientists in the United States in the next decade (Manyika et al. 2017). Nonetheless, this labour shortage is also suggested to be the result of challenges in terms of engaging new professionals to enter the field of data science (Song and Zhu 2017).

Based on several definitions, data scientists should be proficient in the collection, cleaning, analysis, visualization, and management of data to extract new knowledge to solve problems and communicate solutions and implications to the rest of the business (Dhar 2013; Provost and Fawcett 2013; Davenport and Patil 2012). Though universities have started introducing data science education programs (Song and Zhu 2016) to mitigate the growing labour shortage, the multidisciplinary nature of the field of data science makes incorporating the broad scope of the domain a challenge for traditional educational departments (Song and Zhu 2017). Additionally, another major barrier to increasing the number and quality of graduating students in related STEM fields lies in the poor retention rate of these programs (Tomkin et al. 2019; Seymour 2000). Studies show that this low retention in STEM programs can be attributed to the usage of traditional, lecture-centric teaching which in turn, fosters lower levels of engagement and learning when compared to active learning approaches (Freeman et al. 2014; Henderson et al. 2011).

Interestingly, the application of gamification that has been seen to be successful in improving engagement and learning of students in various educational contexts (Dichev and Dicheva 2017; Hamari et al. 2014; Seaborn and Fels 2015), may be a viable option to help engage, retain, and rapidly upskill new data science professionals. It is therefore critical to explore the relationship between gamification and learning as well as engagement. The insights derived from such an understanding may be useful in designing more engaging data science educational programs and help reduce the concerning labour shortage of well-qualified professionals. Moreover, educators and professional instructors would greatly benefit from the information helping them better adapt course work and in-class interactions to learners' needs.

Aim and scope

This literature review assesses the state of knowledge on fields relating to the research question and contains three main sections in addition to a conclusion. Self-Determination and the Social Cognitive Theory are two leading theories on human motivation and behaviour. Since the main premise of gamification posits that it is more motivating and engages learners to achieve higher learning gains, the first section will be devoted to the latter theories to understand the motivational nature of games and game elements. The second section will explore the concepts of active learning, intrinsic motivation, and engagement to understand how each dimension affects learning in a gamified context. The third section will examine approaches to end-user training and the concepts of enactive and vicarious learning as outlined by the Social Cognitive Theory. This section will also explore the concept of gamification, highlighting the leaderboard game element. The impact of gamification on learners' experience known in the literature will be revealed with a focus on studies in data science and STEM-oriented fields. Finally, the conclusion will close the circle by exploring how each psychological theory, learning approaches, and gamification drive learning and engagement.

1.2 Theories

Self-Determination Theory

Self-Determination Theory (SCT), a leading theory of human motivation, provides a succinct framework to explain human motivation and to understand the psychological concepts underlying gamer psychology (Deci and Ryan 1980; Deci 1975; Ryan et al. 2006). At its origin, the theory, as proposed by Deci and Ryan (1980), was rooted in early research of intrinsic motivation and the various factors that sustain or undermine it (Deci 1975; Deci and Ryan 1980). This new focus on intrinsic motivation diverged from the popular behaviorist theory in motivation science that proposed how external factors to an individual such as environmental reinforcements and punishments, controlled behavior (Overskeid 2018; Ryan, Bradshaw, and Deci 2019). Instead, SDT focused on the nature of *self*, not solely external sources of motivation. It proposed that in development, our primary task as humans is to assimilate, coordinate, and regulate external and internal inputs (Ryan and Deci 2019). The importance put on intrinsic motivation and the concern for actions naturally organized within humans created a shift in focus within the field of human motivation (Ryan, Di Domenico, and Deci 2019). This in turn, along with contributions from various researchers over the decades, has provided a solid theoretical framework that has helped explain underlying factors in individual differences in motivation by identifying three key basic psychological needs for intrinsic motivation: competence, autonomy, and relatedness (Deci and Ryan 1980; Gagné, Deci, and Ryan 2018; Deci, Olafsen, and Ryan, 2017). Competence refers to one's feeling of being capable and effective when presented with a challenge. However, activities must be balanced in terms of the difficulty level: challenges too easy or difficulty undermines the satisfaction of competence needs. Autonomy is defined as feelings of psychological freedom and will when fulfilling a task. Satisfaction of autonomy needs may be compromised when choices are imposed, controlled, or when individuals do not endorse actions taken. Finally, relatedness refers to feelings of meaningful connection to others and experiences of community. Cold and neglecting environments are detrimental to the satisfaction of relatedness needs (Deci and Ryan 1980; Ryan and Deci 2000).

SDT has also been successfully applied in the context of games to explain the motivations of gamers. Ryan, Rigby, and Przybylski (2006) aimed to explain the attractiveness of games by applying the theory’s framework to assess participants’ need for autonomy, competence, and relatedness. They conducted four play session studies using Super Mario 64, Zelda: Ocarina of Time, A Bug's Life, and a MMO (massive multiplayer online game). A major finding was that satisfaction of the three basic psychological needs of competence, autonomy, and relatedness predicted positive associations of immersion, persistence in play, and enjoyment, and undermining the same needs led to a negative association of the latter (Ryan, Rigby, and Przybylski 2006). Other studies have reported that autonomy and competence need satisfaction in games were positively associated with player experience outcomes such as enjoyment, value, and persistence in play (Przybylski, Rigby, and Ryan 2010). From these findings we see that, much like motivation underlying play and sport (Frederick and Ryan 1993; 1995), computer games are played by gamers because they are intrinsically satisfying and fun (Malone and Lepper 1987; Ryan, Rigby, and Przybylski 2006). Conversely, to examine extrinsic motivation for gaming and its relationship with obsessive passion, one study leveraged the SDT to explore the role of passion and gamer motivation with a sample of Singaporean high school students. They found that harmonious passion (autonomous engagement driven by pleasure) trumps obsessive passion (engagement driven by internal pressure) in terms of satisfying the three basic psychological needs (Wang et al. 2011). The results suggest that the undermining effect of extrinsic motivation when gaming is detrimental to the players’ basic needs satisfaction and overall well-being. These results are unsurprising considering that games are played because they are inherently fun and satisfy intrinsic motivational needs (Ryan and Deci 2000).

Because much of the literature in games uses the SDT framework to examine motivation of gamers, we suggest an in-depth understanding of Bandura’s Social Cognitive Theory to examine how humans learn to better understand the emergence of gamification in learning and training.

Social Cognitive Theory

Originally proposed by Bandura (1977; 1986), the Social Cognitive Theory (SCT) is a behavioral theory used to explain how humans regulate and control their behavior in social contexts. Initially coined as the Social Learning Theory (SLT) in 1971 and developed into SCT in 1986, this theory was born from the need to explore other perspectives of determinants of human behavior that contradicted the dominant unidirectional view of human behavior wherein individual dispositions or environment shape behavior (Wood and Bandura 1989). Indeed, SCT argues that the interaction between behavior, environmental, and personal factors influence how humans learn behaviors within social contexts (Bandura 1978). Environmental factors refer to physically external factors that alter an individual's behavior: these are physical or social external factors that afford new opportunities or social support (Bandura 1986; Compeau and Higgins 1995). Personal factors refer to cognitive, motivational, emotional, personality, or demographic characteristics that defines a person (Bandura 1986). Self-efficacy is one cognitive factor that is fundamental in self-regulation mechanisms in human behavior and is defined as one's belief in his/her ability to learn or achieve a behavior or goal (Bandura 1986). SCT posits that self-efficacy is augmented by four core components: (1) enactive mastery, defined as past successes influencing one's belief of future performance; (2) behavioral modeling, or known as vicarious learning, describes observing and learning from others successfully performing a task; (3) social persuasion, which is being positively influenced by the encouragement of others; (4) physical and emotional states, defined as one's mood when performing an action or task such as stress, anxiety, fatigue, etc (Bandura 1978; Carillo 2012). Thus, SCT suggests that self-efficacy influences how humans set goals and how respective behaviors are observed and evaluated. Indeed, self-efficacy can be improved by learning through observing one's own behaviors and respective consequences of given actions, and by interacting with one's environment by observing other people perform tasks and trying to replicate it (Bandura 1977; 2001; Yi and Davis 2003). This differentiation in learning approaches is defined as enactive and vicarious learning: enactive refers to learning by "doing" and vicarious refers to learning by "observing" (Bandura 1977).

SCT is one of the most widespread theories used to explore human behavior and has been empirically validated in fields such as mass media (Bandura 2001; Cantor 1994), public health (Bandura 1998; Holden 1992), and education (Dai, Moon, and Feldhusen 1998; Zimmerman 1989), and information systems (IS) (Gupta et al. 2010). Indeed, the advent of the computer in the workplace led to new applications of SCT to explore human behavior in the context of computer training (Compeau and Higgins 1995; Compeau, Higgins, and Huff 1999; Hasan and Ali 2006; Yi and Davis 2003). The personal cognitive factor, self-efficacy, was and still is, of large interest in the SCT and IS literature. Indeed, Compeau and Higgins (1995) were one of the first researchers to leverage the SCT framework to explore computer training, specifically, in the workplace. The authors pioneered and validated measures of computer self-efficacy: it measures a person's perception of his/her skills with regards to using a computer to complete tasks (Compeau and Higgins 1995). Findings from their studies suggests that higher levels of computer self-efficacy are positively related to positive emotional reactions and performance in completing tasks with computers. Another major finding was that computer self-efficacy has an impact on predicting a person's willingness to continue in a course of action in regard to computer or software use, and overall behavioral reaction to informational technology (Compeau and Higgins 1995; Compeau, Higgins, and Huff 1999).

Later research also suggests that SCT explains how gamers are motivated to maintain positive gaming behavior and outcomes as means to become more competent to achieve in-game goals (Klimmt and Hartmann 2006). One study leveraged SCT to explore leadership development in MMORPGs (massively multiplayer online games) and revealed that the gaming environment (guilds and virtual worlds), player's actions, and personal factors like self-efficacy, influenced player's confidence and assertiveness. The game afforded players with experiences that developed in-game knowledge, skills, and abilities which in turn, increased the likelihood of successful task completion and strategy execution (Ee and Cho 2012). One meta-analysis examining SCT based video game interventions created to transform health related behaviors such as diet and exercise, showed positive outcomes regarding facilitating behavioral changes (Baranowski et al. 2008). One game resulted in lower urgent care visits from players with juvenile diabetes due to gained self-efficacy about diabetes self-care after playing the game (Brown et al. 1997). These findings suggest that the game elements may foster positive behavioural change and outcomes.

In summary, gamers have the option to learn to play, alter behaviors, and successfully complete quests or tasks by interacting with their environments or simply by observing their own behaviors with trial and error. Thus, while SDT suggests that games can satisfy the three basic psychological needs, Bandura's SCT offers insight about how gamers acquire knowledge and partake in certain behaviors based on their own level of self-efficacy. This would likely imply that the effects of the game elements used in video games could also be experienced in a non-video game context such as in higher education or end-user training. However, very little research including theory-driven causal experiments on gamification has been conducted (Santhanam et al. 2016). While there is existing literature on gamification in STEM related fields, to our knowledge, few studies have examined the effects that gamification can have on learning outcomes when engaging with a business simulation aiming to teach data analytics. For this reason, in our study, we observed learning outcomes as a result of the intrinsically motivating nature of games and how games facilitate favourable behavioural changes. We also consider constructs of self-efficacy as a moderating variable in our assessment of gamification and its relationship with learning.

1.3 Active learning, intrinsic motivation, and the role of engagement

Active vs Passive Learning

Serious games and simulation game-based training as instructional approaches have become increasingly more popular in higher education and end-user training (Leger et al. 2011; Aldrich 2003; Prensky 2001). These methods contain learning and game elements that offer a more entertaining and active learning approach to teaching and end-user training (Lameras et al 2017; Maheu-Cadotte et al. 2018). Active learning is defined as any instructional method that involves learners actively and holds them accountable for their own learning. Methods such as quizzes, games, and role play are considered active learning approaches (Bonwell and Eison 1991; Ebert-May, Brewer, and Allred 1997; Sarason and Banbury 2004). Other more traditional methods of teaching such as lectures, assignments, or training presentations are called passive learning. These methods do not typically offer learners much opportunity for group discussion or to apply their knowledge (Stewart-Wingfield and Black 2005). Some studies have shown that passive learning does not foster good retention of learning material and does not hold learners' attention as well as more active teaching methods (Van Eynde and Spencer 1988; Dorestani 2005). The literature also suggests that learners are more involved, more motivated, and engage in more critical thinking skills such as analysis, synthesis, and evaluation in active learning than in passive learning (Bonwell and Eison 1991).

Games can transform passive learning materials into an interactive learning experience, or in other words, an active learning environment. Learners become players and become engrossed with the waves of challenges and strategy (Sugar and Takacs 1999). Games also support active learning as a teaching approach. Problem-based learning (PBL) is a type of active learning that is prominent in games and simulations. PBL is an instructional approach that places learners in real-world problem-solving contexts wherein they must produce a resolution for the challenges at hand. Indeed, the problems presented reflect authentic challenges and test learners with ill-structured and ambiguous situations. This process forces learners to apply their existing knowledge and be resourceful to fill in any gaps needed for the problem: this approach fosters the notion that problems are a stimulus to thinking and learners are given control over their learning. (Walker and Leary 2009; Miller 2004). Learners assume stakeholder roles and are motivated by the problem-solving process, determined to find a justifiable resolution, and are provided immediate feedback that let them know how well they are progressing (Cruickshank and Telfer 1980). New knowledge learned during the process is more easily remembered as learners have the chance to apply what is stored in memory (Barab et al. 2001). Simulation games that introduce authentic problem-solving contexts also typically foster quicker development of new skills and acquisition of knowledge needed to understand complex challenges (Randel et al. 1992; O'Neil et al. 2005; Schrage 1999). In their study using a simulation game called ERPsim that teaches operation management skills using in real time a real-life ERP system, Léger et al. (2012) found, based on a survey with instructors, this PBL approach showed significant improvements in student evaluations, motivation, competence, and engagement. Later studies have demonstrated that PBL fosters significant long-term learning gains when compared to traditional instructional approaches (Hmelo-Silver et al. 2007). PBL has also been suggested to increase motivation, improve learning, retention, and recall, and improved problem-solving skills when compared to passive learning approaches (Norman and Schmidt 2000). When examining these results through the lens of SDT and satisfaction of the need to feel competent, PBL in games and simulation activities leads to feelings of being effective and fosters chances for growth by learning new knowledge to solve novel problems (Norman and Schmidt 2000; Ryan and Deci 2000). This suggests that a simulation game experiment teaching participants new data analytic concepts and providing them with a problem to solve, which is a form of active learning, could influence the resulting learning outcomes. However, it is also important to consider the potential effects that games and their respective mechanisms may have on an individual's motivation in a demanding cognitive task such as solving a business problem.

Intrinsic Motivation

Motivation can be separated into two facets, intrinsic and extrinsic. Behaviour that is intrinsically motivated refers to when one performs a task driven by their own volition whereas extrinsically motivated behaviours are driven by external forces such as perceived consequence or reward

(Skinner and Belmont 1993; Malone 1981). The SDT and games literature suggest that players are intrinsically motivated to play games as they satisfy their basic psychological needs (Deci and Ryan 2000). Competence needs, one of the most significant factors that influence the enjoyment of games, can be easily satisfied with games that offer a balance of skill-challenge activities as well as positive performance feedback mechanism to reinforce feelings of mastery and efficacy (Rigby and Ryan 2011; Ryan, Rigby, and Przybylski 2006; Przybylski, Ryan, and Rigby. 2009; Tamborini et al. 2010). Games also satisfy needs for immediate feedback with responsive environments wherein learners know how well they are progressing (Cruickshank and Telfer 1980). Autonomy needs are satisfied with games due the voluntary engagement and freedom of choice afforded to players in terms of decision-making (Rigby and Ryan 2011). Lastly, relatedness needs are satisfied as games foster feelings of acknowledgement and support in contexts of both multiplayer and single player games: NPCs (non-player characters) in single player games induce feelings of relatedness (Rigby and Ryan 2011). Methods in games that foster the satisfaction of the three basic psychological needs and in turn, promote intrinsic motivation, have been adopted by other domains such as education and professional training. To discuss how intrinsically motivating game elements have found their way into the latter two fields, we must first consider the role of emotional engagement with respect to increased learning performance.

The Role of Engagement

When examining motivations and behaviors of learners in the context of games and simulations, it is important to consider engagement. Engagement is suggested to be a fundamental reason for the advent of serious games in educational environments (Riemer and Schrader 2016). Games create emotionally engaging experiences and have in turn, have been leveraged over the years in fields of education to augment learning experiences (Ninaus et al. 2019). Indeed, improved problem solving, and critical thinking skills, and overall learning, are suggested to be a result of increased student engagement (McCormick, Clark, and Raines 2015; Springer et al. 1999; Ruiz-Primo et al. 2011; Gasiewski et al. 2012). Conceptualizing engagement has continually evolved in the research literature. Though not an exhaustive list, engagement has been defined as participation, attachment, motivation, cognitive, emotional, and physical factors to definitions of vigor, dedication, and absorption (Kahn 1990; Schaufeli et al.2002; Fredericks et al. 2004). The early literature includes an extensive body of research on engagement but characterized it in unidimensional terms (Yonezawa et al. 2009). In later research, to review all the current and relevant measure of engagement to extract the concept's main elements, Fredericks, Blumefeld, and Paris (2004), proposed a multidimensional concept of engagement with three core facets: (1) cognitive; (2) behavioural; (3) emotional. Cognitive engagement refers to one's psychological investment and willingness in learning with respect to working hard to understand and develop skills to solve complex challenges (Fredericks et al. 2004). Behavioral engagement refers to positive conduct, involvement, and active participation in learning or social activities. Behaviors that exhibit behavioral engagement entail conforming and adhering to rules and norms, no unruly behaviors, concentration, attention, and contributing to community efforts (Fredericks et al. 2004). This type of engagement can be measured by observation or self-reported measures. Lastly, emotional engagement refers to the display of interest and emotions, such as affective reactions including boredom, happiness, sadness, and anxiety in response to learning experiences (Connell and Wellborn 1991; Skinner and Belmont, 1993). According to Russell's (1980) circumplex model of emotion, emotion has two dimensions, emotional arousal and valence. Arousal is defined by a continuum varying from calm to excited, and valence on a continuum from pleasure to displeasure (Charland et al. 2015). Interest and boredom can be mapped on the dimension of emotional arousal: interest indicates high arousal and boredom indicates low arousal. Conversely, happiness and sadness can be mapped on the dimension of emotional valence: happiness refers to positive valence and sadness refers to negative valence (Russell 1980).

Implicit and Explicit Measures of Engagement

Evaluating emotional engagement can help understand the effect games have on motivation and behavior, and in turn, learning outcomes, which is important to this current study. In line with this, methods of evaluating engagement with respect to implicit and explicit measures must be discussed. In terms of explicit measures, emotional engagement with respect to learning experiences are typically measured with self-report items or questionnaires such as the Positive and Negative Affect Schedule (PANAS) or the Self-Assessment Manikin (SAM) scales. Both mentioned scales are established, and widespread post-test questionnaires used for direct assessment of emotional reactions (Frederick and McColskey 2012; Watson, Clark, and Tellegen 1988; Bradley and Lang 1994). Implicit measures, however, are more comprehensive in terms of measuring emotional engagement in learners during learning experiences (Ninaus et al. 2019).

Automatic facial expression analysis software has been the measure of choice in the context of learning and Information Systems (IS) as it detects and recognizes emotions of learners with regards to emotional valence and arousal (Charland et al. 2015; Bosch et al. 2014; Harley et al. 2015; Whitehill et al. 2014; Al-Awni 2016; Bahreini et al. 2016). This implicit measure is completed based on Ekman's (1993) work that posits the correlation between human facial expressions and six basic emotions (happiness, sadness, surprise, fear, disgust, and anger). The method is also best for evaluations of emotional engagement in a learning context as automatic facial expression capturing is non-invasive and will not interrupt the learner's experience (Ninaus et al. 2019). In addition, since emotional engagement is not directly observable and thus can be challenging to measure, Charland et al. (2015) suggest that implicit measures such as automatic facial emotion recognition software to measure valence and electrodermal activity sensors to measure arousal provides a better understanding of the emotional dimensions of engagement. In a later study, Charland et al. (2017) examined the relationship between problem-solving performance and engagement by using both explicit and implicit measures of engagement. Results showed that explicit measures of engagement did not play a part in the prediction of problem-solving performance, whereas implicit measures did. This suggests that implicit measures are thus a non-invasive and useful methods of capturing emotional engagement and for these reasons, we chose to use them in the present study. Lackmann et al. (2021) used implicit measures for their study on the influence of lecture video design on engagement over time and learning performance. Their study consisted of a between-subject experiment with two conditions: an infographic style video with animations, images, and text, and a video recording of a regular class lecture. Video length and voice track for the instructor of the video was the same for both conditions. Participants answered a pretest and posttest assessment and had had to watch the 14-minute video as emotional and cognitive measures were being recorded with neurophysiological instruments. The results of Lackmann et al. (2021)'s experiment showed that the infographic video kept students more emotional and cognitively engaged over time and led to significant improvement of learning performance, when compared to the lecture capture video format. Furthermore, according to the researchers, higher engagement is positively correlated to learning performance, which is evidence that implicit measures are comprehensive in measuring learning and engagement in the context of learning. If, in fact, participants are more engaged from an emotional and behavioural perspective due to the nature of a gamified experience, we are likely to be able to capture the effect through our experimental manipulation. It is important to note that the construct of cognitive engagement is not applicable in the current study as appropriate neurophysiological tools needed to measure cognitive states of participants were not used.

1.4 End-user training and gamification

Enactive vs Vicarious Learning

A relevant area to discuss when studying the effects of gamification on learning and engagement is end-user training (EUT). EUT has been recognized as an important factor that drives significant organizational impact in terms of positive changes in job related performance. EUT is defined as a teaching approach wherein skills needed to effectively use computer applications are taught to end-users (Arthur et al. 2003; Gupta et al. 2010). When examining the literature through the lens of SCT, the theory's central concept, self-efficacy, has been used to understand participant learning in education and IS and gave rise to vicarious and enactive learning approaches used in end-user training (Bandura 1986; Gupta et al. 2010). Vicarious learning, also known as behavioral modelling, is explained by Bandura (1986) as how observing others perform certain behaviors and their consequent successes or failures can lead one to acquire knowledge and relevant skills. Observing attitudes, behaviors, and emotional reactions of others helps build their own mental models and in turn, guide future actions (Yi and Davis 2003). Video training involving instructors demonstrating actions, has been found to be the most common form of vicarious end-user training in the literature. Research suggests that vicarious teaching methods fosters better training outcomes than traditional methods such as lectures or studying from a manual (Gupta et al. 2010). Some empirical research also shows that vicarious learning increases self-efficacy and in turn, have has a positive impact on learning outcomes (Gist 1989, Schunk 1995; Schunk and Mullen 2012; Agarwal, Sambamurthy, and Stair 2000; Bouffard-Bouchard 1990; Moos and Azevedo 2009). Vicarious learning is also prominent in the context of video games due to the accessibility of social environments with other players which in turn, offer opportunities for social learning.

The literature states that the addition of enactive learning can enhance training outcomes when using vicarious methods (Gupta and Bostrom 2013; Gupta et al. 2010). Enactive learning refers to how humans learn from the consequences of their own actions: humans test what they think they know in environments that then provide feedback based on action. We retain and choose behaviors that result in successful consequences and alter or stop those that result in negative outcomes (Bandura 1977). SCT posits that the most optimal training method for complex tasks entails a mixture of both vicarious and enactive learning (Bandura 1986). There is also empirical evidence that suggests an optimal order that instructors should consider when combining enactive and vicarious learning methods in the context of IS training. In a study wherein subjects participated in a 30-minute training prior to performing tasks using an enterprise resources planning (ERP) software simulation game, Gaudet-Lafontaine et al. (2017) found that those who received the vicarious training first showed higher dashboard self-efficacy, when compared to those that received the enactive training first. These results suggest that providing learners with vicarious training before an enactive training may foster improved learning outcomes as increases in self-efficacy has effects on performance over time (Gaudet-Lafontaine et al. 2017).

Enactive learning in the form of feedback in response to an individual's actions is prominent in serious educational games and simulations. Improved learning outcomes, self-efficacy, and both procedural and declarative knowledge are mentioned in the literature as consequences of playing serious games (Girard, Ecalle, and Magnan 2013; Sitzmann 2011). A study conducted with 36 student trainees using an ERP software simulation game investigating the effects of neurophysiological correlates of cognitive absorption in an enactive learning context, found that perceived control seemed to have the strongest effect on training outcomes and a key factor in participants' ability to show competencies at using IS (Leger et al 2014). In their review focusing on EUT methods and effectiveness, Gupta et al. (2010) posits that feedback has a positive effect on learning outcomes, especially in the context of computer-based training, which is relevant to the technology first world we live in. These results suggest that simulated environments enable enactive learning as they provide sensory and verbal feedback about a learner's actions which in turn, creates opportunities for one to alter behaviours to attain desirable outcomes (Gupta et al 2010).

In summary, feedback and other elements related to enactive learning that are intrinsically motivating in serious games and video games (Rigby and Ryan 2011), are also used in non-gaming contexts in the form of gamification. It would, therefore, be relevant to analyze the impact of feedback in the form of a leaderboard as a gamification mechanism used in a EUT context, on learning outcomes and both behavioural and emotional engagement.

Gamification

With the motivational nature of games being so relevant to our research question, a relevant area to discuss would be gamification, since its application in higher education and EUT plays a big role in learning outcomes and engagement. Games are intrinsically motivating as they satisfy the three basic psychological needs of autonomy, competence, and relatedness, and interest in leveraging those same game features in non-game contexts has grown significantly with means to affect learning outcomes, engagement, and motivation (Rigby and Ryan 2011; Ryan, Rigby, and Pryzybylski 2006; da Rocha Seixas, Gomes, and Filho, 2016). Deterding et al. (2011) refer to this phenomenon as gamification and define it as the use of game design elements in non-game contexts. The idea of gamification is to add game elements to non-game contexts, but not to create a fully-fledged game. These game design elements are meant to transform non-game tasks to be more enjoyable, motivating, and engaging for individuals interacting with a given system, fostering positive behavioural change and outcomes (Kappen and Nacke 2013; Liu, Santhanman, and Webster 2017). Gamification leverages various game elements and mechanics such as badges, points, levels, rewards, feedback/leaderboards, competition, time constraints, limited resources, and turn based interactions (Deterding et al. 2011; Seaborn and Fels 2015). Over the past decade, these elements and mechanisms have been applied in many multidisciplinary fields such as software development (Chow and Huang 2017), commerce (Bittner and Schipper 2014), medical and health (Fleming et al. 2017), transportation (Kazhamiakin et al. 2015), finance (Altmeyer et al. 2016), and marketing (Church and Iyer 2014). Interestingly, the field of education is among the top domains of gamification research, which is relevant to our research question (Dichev and Dicheva 2017; Hamari et al. 2014; Seaborn and Fels 2015).

Furthermore, a review of 24 empirical studies of gamification by Hamari et al. (2014) found that for the majority of studies, gamification does in fact have positive effects on psychological outcomes such as motivation, attitude, and enjoyment. It is important to note that Hamari et al. (2014) also found that some studies suggested that gamification may not have long term effects and individual differences play a large role in the outcomes and benefits of a gamified experience. Regarding the role of individual differences, varying levels of prior knowledge greatly influences how one associates new knowledge with what they already know and in turn, affects performance in game-based learning environments and digital games (Spyridakis and Isakson 1991; Bulu and

Pedersen 2012). Some studies have examined students' previous understanding and knowledge of the subject matter in relation to learning outcomes, establishing that prior knowledge may positively affect one's ability to successfully complete tasks and overall experience (Blumberg et al 2008; Orvis et al 2008; Tsai et al 2012). Because of the role that prior knowledge plays in the effects of gamification on learning, we chose to consider this construct as a moderating variable in our study.

Another meta-analytic review of the effectiveness of gamification conducted by Seaborn and Fels (2015) revealed similar positive benefits in terms of increased engagement and favourable behaviours, however also showed decreases in intrinsic motivation and learning as a consequence of extrinsic rewards related to gamification. This suggests that the literature on gamification in the context learning seems to be inconclusive and is in part due to the fact that most studies lack the theoretical foundations of gamification (Harmi et al 2014; Seaborn and Fels 2015; Santhanman et al 2016; Nacke and Deterding 2017). We hope to act on the call for theory-driven causal experiments on gamification and offer empirical evidence of learning and engagement outcomes in the context of data analytics.

Gamification in Data Science and STEM

To our knowledge, few studies have examined the effects of leaderboards on learning in the field of data science and analytics. Some literature does exist on the effects of multiple game elements used to produce gamified learning systems to teach computational science skills.

For instance, there is evidence that a gamified learning experience can increase motivation to learn complex computational concepts. In this regard, Boyce and Barnes (2010) examined the effects of the implementation of a gamified layer to an existing education tool that was designed to teach computing-related mathematical concepts, on learning gains. Results showed positive self-reported measures of enjoyment and significant improvement of pretest and posttest scores. Moreover, the significant learning gains in computational thinking included learning of more complex concepts such as iteration and layering of functions (Boyce and Barnes 2010). Thus, in addition to making learning more enjoyable, the game elements motivated students to explore concepts that they would have otherwise avoided in the original educational tool. Other studies have shown that the gamification of computing courses can increase student participation, probability of passing a course, satisfaction, and voluntary time spent engaging with the course content (Iosup and Epema 2014; Lee, Jo, and Kwan 2013). Similarly, a pilot study testing the efficacy of Learn2Mine, a gamified educational system designed to teach data science, by Anderson et al. (2014) found that out of 35 students, 84% found the automatic feedback mechanism of the gamified system to be motivating. Additionally, positive feedback regarding the gamified system's useability and helpfulness was also received. A later study using Learn2Mine to teach data science by Anderson, Nash, and McCauley (2015) showed that over the course of 3 semesters, students using the gamified learning system outperformed those that did not have access to the tool in terms of assignment completion rates. These results suggest that gamification could have positive impacts on the learning of data science related skills such as computational sciences, but further experimental research will need to be carried before stronger conclusions can be drawn.

While little research has been made on gamified learning in the field of data science and analytics, there is a nascent literature on gamification and STEM learning, a related field. We suggest an in-depth understanding of applications of gamifications and consequently effects on learning of STEM-oriented concepts, as these are complementary to skills needed to be successful in data science.

Several studies have examined gamification in the context of higher education within STEM-oriented fields. Ortiz-Rojas, Chiluiza, and Valcke (2016) conducted a literature review of 30 studies on STEM fields experimenting with gamification. A major finding was that the combination of multiple game elements, namely points, badgers, and leaderboards, was used in most studies and often resulted in increases in students' class attendance, engagement, and attitudes towards the subject matter, which can have potential positive effects on learning performance. It is important to note that similarly to the general literature on gamification in education and learning (Dicheva et al. 2015), this review reported an almost equal number of positive and mixed results regarding the impact of gamification on STEM learning. This is suggested to be the result of most studies observing effects of a combination of game elements, short study periods, and small sample sizes. This implies that further experiments can be devised where game elements are studied in isolation to observe the impact of individual gamification elements.

The vast majority of the literature examining gamified STEM learning focuses on the field of computer science (CS) and software engineering (SE) (Ortiz-Rojas, Chiluiza, and Valcke 2016; Venter 2020). This may be due to the fact that many students struggle with mastering programming concepts which in turn leads to low passing rates, retention in CS or SE programs, and overall engagement. Gamification in these fields have been seen as potential methods to improve the latter challenges and foster better learning outcomes. A review on this subject reported that gamification in CS education leads to improved comprehension of the subject matter and overall engagement of students (Narasareddy, Singh, and Radermarcher 2018). Results of the review also shows that leaderboards and points as game elements were the biggest drivers of motivation for students, while badges being the lowest motivator (Narasareddy, Singh, and Radermarcher 2018). A second review by Alhammad and Moreno (2018) reviewed 21 studies related to gamification and SE and revealed a higher number of studies that indicated positive effects on learning and engagement when compared to negative or no impact. These results suggest that gamification has the potential to lower the barrier to entry for fields such as CS and SE which are complex in nature. Indeed, gamification can be used as means to retain students by making the learning of repetitive or time-consuming processes, such as code reviews, more fun and interesting (Khandelwal, Sripada, and Reddy 2017).

In summary, although there are few studies examining the impacts of gamification on learning data science, there is literature on the effects on learning in STEM, which given the multidisciplinary nature of data science, findings in STEM-oriented fields are relevant to our research question.

Gamification with Leaderboards

The most common applications of gamification in the literature are points, badges, and leaderboards (Deterding et al. 2011; Hamari et al. 2014; Seaborn and Fels 2015). Leaderboards as a gamification element is relevant to our research question and are defined as visual displays that rank a pool of players based on their achievements (Christy and Fox 2014). From a SDT perspective, feedback generated from leaderboards is a form of enactive learning that may help satisfy one's psychological needs for competence and autonomy and foster intrinsic motivated behaviors, which in turn, leads to higher quality learning (Ryan and Deci 2000). Further, according to Zichermann and Cunningham (2011) in their book *"Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps,"* leaderboards can be designed in two different types: 1) a non-disincentive leaderboard that allows players to compare their performance against others that are ranked above or below them, or 2) an infinite leaderboard which displays all players and their respective scores regardless of one's rank. For non-disincentive leaderboards, players may feel less discouraged when they rank lower since they will not see players that rank in the top 10, whereas infinite leaderboards may foster a more competitive environment as they are constantly reminded of their own rank amongst all players (Zichermann and Cunningham 2011). When competition is used as a method to motivate learners, the probability of successful behavioural change becomes dependent on the amount of effort a learner exerts to compete with others (Slavin 1980). This means that if competitors are of unequal ability, those that are less skilled may become less motivated over the course of the learning experience whereas more skilled individuals may lose motivation due to the lack of challenge (Slavin 1980). Because it is important to consider equal levels of learner ability when using a gamified approach to learning, we designed our experiment in a way that included fair competitors for our participants. Nonetheless, since we hope to determine if the competitive nature of leaderboards impacts learning outcomes and engagement, the present study observes the effects of an infinite leaderboard design.

Most of the research concerning the use of leaderboards focuses on the benefits within a learning context. A review of this literature suggests that leaderboards as a game element have high potential to foster higher motivation and in turn, learning performance increase behavioural engagement in the field of Education (Landers and Landers 2015; Ortiz-Rojar et al. 2019; Barata et al. 2013). For instance, in their experimental study with 64 university students, Landers and Landers (2015) suggests that leaderboards are an effective method to improve engagement in terms of time-on-task with course material and in turn, course performance (Landers and Landers 2015). This suggests that leaderboards have the potential to actively engage learners more than traditional methods to influence learner behaviour and motivate them to defeat the challenge they are faced with. Regarding motivational benefits of leaderboards, in a sample of 150 university business students, Cwill (2020) found that presenting academic performance scores using a leaderboard was perceived as more motivating and encouraging when compared to more traditional, unranked table-based representation of scores. This suggests that learners ranked at the lower end of a leaderboard benefit most from the game element since it motivates them to work harder to attain a higher rank. Thus, the motivational effects may be contingent on a learner's position within the leaderboard (Nebel et al. 2016) Several other studies have reported findings where leaderboards serve as motivators, increasing overall involvement and enjoyment which in turn, affects learning within gamified educational contexts (e.g., Dominguez et al. 2013; Barata et al. 2013; Cagiltay et al. 2015; Simoes et al. 2013). A recent study by Ortiz-Rojar et al. (2019) found that gamifying an engineering course using leaderboards led to significantly higher learning performance, when compared to a regular classroom environment. However, it is important to note that the aforementioned study did not observe any significant impact of the leaderboard on the intrinsic motivation of learners (Ortiz-Rojar el al. 2019). Another study by Haus and Fox (2015) has reported similar findings wherein gamifying a learning environment resulted in a decrease of intrinsic motivation, satisfaction, and empowerment among students. It is therefore possible that leaderboards may undermine the proposed intrinsically motivating benefits of gamification as they can promote extrinsic motivators and social comparison that in turn, led to feelings of stress or decreased self-efficacy that negatively influence performance (Deci et al. 2001; Lepper et al. 1973; Tang and Hall 1995; Christy and Fox2014; Cheng et al. 2009; Wu et al. 2010). Taken together, relevant literature regarding leaderboards seems to be inconsistent in terms of the effectiveness and potential detrimental influence of competition generated from this game element.

Furthermore, it is important to note the current empirical gamification literature with results related to leaderboards, often includes other game elements such as badges, stories, time constraints or any additional element, which in turn, may drive observed differences in learning or engagement outcomes. Though the literature suggests that gamification can lead to positive outcomes and benefits, increased motivation, and improved learning performance, Landers et al (2015) highlight that evidence to support the impact of individual game elements on performance is limited since most empirical studies evaluate a combination of multiple game design elements at once (Dominguez et al. 2013; de Sousa Borges et al 2014; Eickhoff et al. 2012; Halan et al 2010). In other words, isolating the effects of a specific game element in the context of learning has seldom been executed. This implies that future experiments can be devised where a particular gamification element such as a leaderboard is used as an independent variable to establish relationships between learning performance and engagement depending on leaderboard use.

1.5 Conclusion and relevance

The aim of this literature review was to provide an in-depth understanding of the relationship between learning and gamification, and more specifically, between learning and leaderboards as a form of gamification. To achieve this, we first examined psychological theories such as SDT and SCT as theoretical foundations to better understand the motivational nature of games and how individuals learn. Different learning approaches and EUT were also explored to understand how these methods of learning are related to games and gamification. Finally, we surveyed the literature on gamification and its effect on learning outcomes and engagement.

Research shows that video games are intrinsically motivating as players have opportunities to satisfy what SDT defines as the three basic psychological needs of competence, autonomy, and relatedness as (Rigby and Ryan 2011). Studies driven by SDT have established a positive relationship between the satisfaction of the previously mentioned basic need and enjoyment, immersion, and persistence in play (Ryan et al. 2006; Przybylski, Ryan, and Rigby 2009). However, the undermining effect of extrinsic motivation of games may hinder overall well-being of players (Ryan and Deci 2000). In regard to SCT, self-efficacy, which is the belief in one's ability to learn or achieve a goal or behaviour, has been suggested to heavily influence emotional reactions, task completion performance, and willingness to persist with a given task or challenge (Bandura 1986; Compeau and Higgins 1995; Compeau, Higgins, and Huff 1999). Indeed, to improve one’s self-efficacy, humans learn either by experiencing consequences of their own actions or, by observing the behaviours of others (Bandura 1977; 2001; Yi and Davis 2003). The literature also suggests that self-efficacy may influence player confidence and assertiveness which in turn, leads to higher likelihood of positive behavioural change to successfully complete tasks and execute strategies to overcome challenges in game (Ee and Cho 2012; Baranowski et al 2008; Brown et al 1997).

Regarding learning approaches, some studies have addressed the relationship between active learning environments and learning outcomes; the bulk of the evidence suggest that active learning fosters higher levels of involvement, motivation, tendencies to engage in more critical thinking skills (Bonwell and Eison 1991). Further, enactive learning, which is learning from doing, can promote better EUT outcomes when coupled with vicarious methods (Gupta and Bostrom 2013;

Gupta et al. 2010). Both active learning and enactive learning are prominent in serious educational games and simulations which studies show lead to improved learning outcomes, higher self-efficacy, knowledge mastery, engagement, and improved problem solving and critical thinking skills (Girard, Ecalle, and Magnan 2013; Sitzmann 2011; Riemer and Schrader 2016; Ninaus et al. 2019; Randel et al. 1992; O'Neil et al. 2005; Schrage 1999).

Gamification tries to replicate the positive outcomes and intrinsically motivating underpinning of games to foster higher quality learning and engagement by adding game elements to non-game contexts (Deterding et al 2011). The most common game design elements used in studies observing the effectiveness of gamification include points, badges, and leaderboards (Deterding et al. 2011; Hamari et al. 2014; Seaborn and Fels 2015). Although most studies generally show positive effects of gamification on motivation, enjoyment, engagement, and learning outcomes, most have failed to address or apply the theoretical foundations of gamification (Hamari et al 2014; Seaborn and Fels 2015). Furthermore, current literature is limited in terms of available evidence supporting the impact of individual gamification elements as most empirical studies observe more than one game element at a time (Landers et al 2015). Finally, while there is literature on gamification and STEM-oriented fields, little research has been, to our knowledge, reported on the effects of gamification on learning in the fields of data science or data analytics. Thus, this research project will add to the literature by conducting a theory-driven causal experiment studying an isolated gamficiation element, the leaderboard, and observe its effects on learning and engagement in the context of data analytics.

References

Agarwal, R., Sambamurthy, V., and Stair, R. M. 2000. "The Evolving Relationship between General and Specific Computer Self-Efficacy—an Empirical Assessment," *Information systems research* (11:4), pp. 418-430.

Al-Awni, A. 2016. "Mood Extraction Using Facial Features to Improve Learning Curves of Students in E-Learning Systems," *international journal of advanced computer science and applications* (7:11), pp. 444-453.

Aldrich, C. 2003. *Simulations and the Future of Learning: An Innovative (and Perhaps Revolutionary) Approach to E-Learning.* John Wiley & Sons.

Alhammad, M. M., and Moreno, A. M. 2018. "Gamification in Software Engineering Education: A Systematic Mapping," *Journal of Systems and Software* (141), pp. 131-150.

Altmeyer, M., Lessel, P., and Krüger, A. 2016. "Expense Control: A Gamified, Semi-Automated, Crowd-Based Approach for Receipt Capturing," *Proceedings of the 21st International Conference on Intelligent User Interfaces*, pp. 31-42.

Anderson, P. E., Turner, C., Dierksheide, J., and McCauley, R. 2014. "An Extensible Online Environment for Teaching Data Science Concepts through Gamification," *2014 IEEE frontiers in education conference (FIE) proceedings*: IEEE, pp. 1-8.

Anderson, P. E., Nash, T., and McCauley, R. 2015. "Facilitating Programming Success in Data Science Courses through Gamified Scaffolding and Learn2mine," *Proceedings of the 2015 ACM Conference on Innovation and Technology in Computer Science Education*, pp. 99-104.

Arthur Jr, W., Tubré, T., Paul, D. S., and Edens, P. S. 2003. "Teaching Effectiveness: The Relationship between Reaction and Learning Evaluation Criteria," *Educational Psychology* (23:3), pp. 275-285.

Bahreini, K., Nadolski, R., and Westera, W. 2016. "Towards Multimodal Emotion Recognition in E-Learning Environments," *Interactive Learning Environments* (24:3), pp. 590-605.

Bandura, A., Ross, D., and Ross, S. A. 1961. "Transmission of Aggression through Imitation of Aggressive Models," *The Journal of Abnormal and Social Psychology* (63:3), pp. 575-582.

Bandura, A. 1977. *Social Learning Theory*. Oxford, England: Prentice-Hall.

Bandura, A. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*.

Englewood Cliffs, NJ, US: Prentice-Hall, Inc.

Bandura, A. 2001. "Social Cognitive Theory: An Agentic Perspective," *Annual review of psychology* (52:1), pp. 1-26.

Bandura, A. 2001. "Social Cognitive Theory of Mass Communication," *Media psychology* (3:3), pp. 265-299.

Bandura, A., and Wood, R. 1989. "Effect of Perceived Controllability and Performance Standards on Self-Regulation of Complex Decision Making," *Journal of personality and social psychology* (56:5), p. 805.

Barab, S. A., Hay, K. E., Barnett, M., and Squire, K. 2001. "Constructing Virtual Worlds: Tracing the Historical Development of Learner Practices," *Cognition and instruction* (19:1), pp. 47-94.

Baranowski, T., Buday, R., Thompson, D. I., and Baranowski, J. 2008. "Playing for Real: Video Games and Stories for Health-Related Behavior Change," *American journal of preventive medicine* (34:1), pp. 74-82. e10.

Barata, G., Gama, S., Jorge, J., and Gonçalves, D. 2013. "Improving Participation and Learning with Gamification," *Proceedings of the First International Conference on gameful design, research, and applications*, pp. 10-17.

Bittner, J. V., and Shipper, J. 2014. "Motivational Effects and Age Differences of Gamification in Product Advertising," *Journal of consumer marketing*.

Blumberg, F. C., Rosenthal, S. F., and Randall, J. D. 2008. "Impasse-Driven Learning in the Context of Video Games," *Computers in Human Behavior* (24:4), pp. 1530-1541.

Bonwell, C. C., and Eison, J. A. 1991. *Active Learning: Creating Excitement in the Classroom. 1991 Ashe-Eric Higher Education Reports*. ERIC.

Bosch, N., Chen, Y., and D'Mello, S. 2014. "It's Written on Your Face: Detecting Affective States from Facial Expressions While Learning Computer Programming," *International Conference on Intelligent Tutoring Systems: Springer*, pp. 39-44.

Bouffard-Bouchard, T. 1990. "Influence of Self-Efficacy on Performance in a Cognitive Task," *The journal of social Psychology* (130:3), pp. 353-363.

Boyce, A. K., Campbell, A., Pickford, S., Culler, D., and Barnes, T. 2011. "Experimental Evaluation of Beadloom Game: How Adding Game Elements to an Educational Tool Improves Motivation and Learning," *Proceedings of the 16th annual joint conference on*

Innovation and technology in computer science education, pp. 243-247.

Boyce, A., and Barnes, T. 2010. "Beadloom Game: Using Game Elements to Increase Motivation and Learning," *Proceedings of the fifth international conference on the foundations of digital games*, pp. 25-31.

Bradley, M. M., and Lang, P. J. 1994. "Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential," *Journal of behavior therapy and experimental psychiatry* (25:1), pp. 49-59.

Brown, S. J., Lieberman, D. A., Gemeny, B., Fan, Y. C., Wilson, D., and Pasta, D. 1997. "Educational Video Game for Juvenile Diabetes: Results of a Controlled Trial," *Medical informatics* (22:1), pp. 77-89.

Bulu, S. T., and Pedersen, S. 2012. "Supporting Problem-Solving Performance in a Hypermedia Learning Environment: The Role of Students' Prior Knowledge and Metacognitive Skills," *Computers in Human Behavior* (28:4), pp. 1162-1169.

Cagiltay, N. E., Ozcelik, E., and Ozcelik, N. S. 2015. "The Effect of Competition on Learning in Games," *Computers & Education* (87), pp. 35-41.

Cantor, N. 1994. "Life Task Problem Solving: Situational Affordances and Personal Needs," *Personality and Social Psychology Bulletin* (20:3), pp. 235-243.

Carillo, K. 2012. "Understanding Is Theory: An Interpretation of Key Is Theoretical Frameworks Using Social Cognitive Theory," in *Information Systems Theory*. Springer, pp. 241-280.

Charland, P., Léger, P.-M., Sénécal, S., Courtemanche, F., Mercier, J., Skelling, Y., and Labonté-Lemoyne, E. 2015. "Assessing the Multiple Dimensions of Engagement to Characterize Learning: A Neurophysiological Perspective," *Journal of visualized experiments*: JoVE:101.

Charland, P., Léger, P.-M., Mercier, J., Skelling, Y., and Lapierre, H. G. 2016. "Measuring Implicit Cognitive and Emotional Engagement to Better Understand Learners' Performance in Problem Solving," *Zeitschrift für Psychologie* (224:4), p. 294.

Cheng, H. N., Wu, W. M., Liao, C. C., and Chan, T.-W. 2009. "Equal Opportunity Tactic: Redesigning and Applying Competition Games in Classrooms," *Computers & Education* (53:3), pp. 866-876.

Christy, K. R., and Fox, J. 2014. "Leaderboards in a Virtual Classroom: A Test of Stereotype Threat and Social Comparison Explanations for Women's Math Performance," *Computers*

& Education (78), pp. 66-77.

Chow, I., and Huang, L. 2017. "A Software Gamification Model for Cross-Cultural Software Development Teams," *Proceedings of the 2017 international conference on management engineering, software engineering and service sciences*, pp. 1-8.

Church, N. J., and Iyer, V. 2014. "The Future of Marketing: Staying Competitive in a Competitive World," *Repositorio de la Red Internacional de Investigadores en Competitividad* (8:1).

Compeau, D., and Higgins, C. 1995. "Computer Self-Efficacy: Development of a Measure and Initial Test," *MIS Q.* (19), pp. 189-211.

Compeau, D., Higgins, C. A., and Huff, S. 1999. "Social Cognitive Theory and Individual Reactions to Computing Technology: A Longitudinal Study," *MIS Quarterly* (23:2), pp. 145.

Connell, J. P., and Wellborn, J. G. 1991. "Competence, Autonomy, and Relatedness: A Motivational Analysis of Self-System Processes,".

Cruickshank, D. R., and Telfer, R. 1980. "Classroom Games and Simulations," *Theory into practice* (19:1), pp. 75-80.

Ćwil, M. 2020. "Leaderboards–a Motivational Tool in the Process of Business Education," *Joint International Conference on Serious Games*: Springer, pp. 193-203.

da Rocha Seixas, L., Gomes, A. S., and de Melo Filho, I. J. 2016. "Effectiveness of Gamification in the Engagement of Students," *Computers in Human Behavior* (58), pp. 48-63.

Dai, D. Y., Moon, S. M., and Feldhusen, J. F. 1998. "Achievement Motivation and Gifted Students: A Social Cognitive Perspective," *Educational Psychologist* (33:2-3), pp. 45-63.

Davenport, Thomas H., and D.J. Patil. 2012. "Data Scientist: The Sexiest Job of the 21st Century." 8 Harvard Business Review, October 2012. hbr.org/2012/10

Davenport, T. H., and Patil, D. 2012. "Data Scientist," *Harvard business review* (90:5), pp. 70-76.

Deci, E. L., and Ryan, R. M. 1980. "Self-Determination Theory: When Mind Mediates Behavior," *Journal of Mind and Behavior* (1:1), pp. 33-43.

Deci, E. L., and Ryan, R. 1985. "Intrinsic Motivation and Self-Determination in Human Behaviour," New York: Plenum.

Deci, E. L. 1975. "Intrinsic motivation," New York: Plenum.

Deci, E. L., Olafsen, A. H., and Ryan, R. M. 2017. "Self-Determination Theory in Work Organizations: The State of a Science," *Annual Review of Organizational Psychology and*

Organizational Behavior (4), pp. 19-43.

Deci, E. L., Koestner, R., and Ryan, R. M. 2001. "Extrinsic Rewards and Intrinsic Motivation in Education: Reconsidered Once Again," *Review of educational research* (71:1), pp. 1-27.

de Sousa Borges, S., Durelli, V. H., Reis, H. M., and Isotani, S. 2014. "A Systematic Mapping on Gamification Applied to Education," *Proceedings of the 29th annual ACM symposium on applied computing*, pp. 216-222.

Dhar, V. 2013. "Data Science and Prediction," *Communications of the ACM* (56:12), pp. 64-73.

Dicheva, D., Dichev, C., Agre, G., and Angelova, G. 2015. "Gamification in Education: A Systematic Mapping Study," *Journal of Educational Technology & Society* (18:3), pp. 75-88.

Dichev, C., and Dicheva, D. 2017. "Gamifying Education: What Is Known, What Is Believed and What Remains Uncertain: A Critical Review," *International journal of educational technology in higher education* (14:1), pp. 1-36.

Dobrin Seth and IBM Analytics. 2017. How IBM builds an effective data science team. https://venturebeat.com/2017/12/22/how-ibm-builds-an-effective-data-science-team/

Domínguez, A., Saenz-de-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., and Martínez-Herráiz, J.-J. 2013. "Gamifying Learning Experiences: Practical Implications and Outcomes," *Computers & education* (63), pp. 380-392.

Dorestani, A. 2005. "Is Interactive/Active Learning Superior to Traditional Lecturing in Economics Courses?," *Humanomics*.

Dugan, and ThomasErickson. 2019. How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation. *InProceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (CHI '19). ACM, New York, NY, USA, Forthcoming.

Ebert-May, D., Brewer, C., and Allred, S. 1997. "Innovation in Large Lectures: Teaching for Active Learning," *Bioscience* (47:9), pp. 601-607.

Ee, A., and Cho, H. 2012. "What Makes an Mmorpg Leader? A Social Cognitive Theory-Based Approach to Understanding the Formation of Leadership Capabilities in Massively Multiplayer Online Role-Playing Games," *Eludamos. Journal for Computer Game Culture* (6:1), pp. 25-37.

Ekman, P. 1993. "Facial Expression and Emotion," *American psychologist* (48:4), p. 384.

Eickhoff, C., Harris, C. G., de Vries, A. P., and Srinivasan, P. 2012. "Quality through Flow and Immersion: Gamifying Crowdsourced Relevance Assessments," *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pp. 871-880.

Fleming, T. M., Bavin, L., Stasiak, K., Hermansson-Webb, E., Merry, S. N., Cheek, C., Lucassen, M., Lau, H. M., Pollmuller, B., and Hetrick, S. 2017. "Serious Games and Gamification for Mental Health: Current Status and Promising Directions," *Frontiers in psychiatry* (7), p. 215.

Frederick, C. M., and Ryan, R. M. 1993. "Differences in Motivation for Sport and Exercise and Their Relations with Participation and Mental Health," *Journal of sport behavior* (16:3), p. 124.

Frederick, C. M., and Ryan, R. M. 1995. "Self-Determination in Sport: A Review Using Cognitive Evaluation Theory," *International journal of sport psychology*.

Fredricks, J. A., Blumenfeld, P. C., and Paris, A. H. 2004. "School Engagement: Potential of the Concept, State of the Evidence," *Review of Educational Research* (74:1), pp. 59-109.

Fredricks, J. A., and McColskey, W. 2012. "The Measurement of Student Engagement: A Comparative Analysis of Various Methods and Student Self-Report Instruments," in *Handbook of Research on Student Engagement.* Springer, pp. 763-782.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., and Wenderoth, M. P. 2014. "Active Learning Increases Student Performance in Science, Engineering, and Mathematics," *Proceedings of the National Academy of Sciences* (111:23), pp. 8410-8415.

Gagné, M., Deci, E. L., and Ryan, R. M. 2018. "Self-Determination Theory Applied to Work Motivation and Organizational Behavior," *The SAGE handbook of industrial, work & organizational psychology: Organizational psychology,* p. 97 - 121.

Gasiewski, J. A., Eagan, M. K., Garcia, G. A., Hurtado, S., and Chang, M. J. 2012. "From Gatekeeping to Engagement: A Multicontextual, Mixed Method Study of Student Academic Engagement in Introductory Stem Courses," *Research in higher education* (53:2), pp. 229-261.

Girard, C., Ecalle, J., and Magnan, A. 2013. "Serious Games as New Educational Tools: How Effective Are They? A Meta-Analysis of Recent Studies," *Journal of computer assisted* *learning* (29:3), pp. 207-219.

Gist, M. E. 1989. "The Influence of Training Method on Self-Efficacy and Idea Generation among Managers," *Personnel psychology* (42:4), pp. 787-805.

Gupta, S., and Bostrom, R. 2013. "Research Note—an Investigation of the Appropriation of Technology-Mediated Training Methods Incorporating Enactive and Collaborative Learning," *Information Systems Research* (24:2), pp. 454-469.

Gupta, S., Bostrom, R. P., and Huber, M. 2010. "End-User Training Methods: What We Know, Need to Know," *ACM SIGMIS Database: the DATABASE for Advances in Information Systems* (41:4), pp. 9-39.

Halan, S., Rossen, B., Cendan, J., and Lok, B. 2010. "High Score!-Motivation Strategies for User Participation in Virtual Human Development," *International conference on intelligent virtual agents*: Springer, pp. 482-488.

Hamari, J., Koivisto, J., and Sarsa, H. 2014. "Does Gamification Work? -- a Literature Review of Empirical Studies on Gamification," *2014 47th Hawaii International Conference on System Sciences*, pp. 3025-3034.

Harley, J. M., Bouchet, F., Hussain, M. S., Azevedo, R., and Calvo, R. 2015. "A Multi-Componential Analysis of Emotions During Complex Learning with an Intelligent Multi-Agent System," *Computers in Human Behavior* (48), pp. 615-625.

Hasan, B. and Ali, J., 2006. “The impact of general and system-specific self-efficacy on computer training learning and reactions,” *Journal of Management Information and Decision Sciences*, *9*(1), pp.17.

Henderson, C., Beach, A., and Finkelstein, N. 2011. "Facilitating Change in Undergraduate Stem Instructional Practices: An Analytic Review of the Literature," *Journal of research in science teaching* (48:8), pp. 952-984.

Hmelo-Silver, C. E., Duncan, R. G., and Chinn, C. A. 2007. "Scaffolding and Achievement in Problem-Based and Inquiry Learning: A Response to Kirschner, Sweller, And," *Educational psychologist* (42:2), pp. 99-107.

Holden, G. 1992. "The Relationship of Self-Efficacy Appraisals to Subsequent Health Related Outcomes: A Meta-Analysis," *Social work in health care* (16:1), pp. 53-93.

Iosup, A., and Epema, D. 2014. "An Experience Report on Using Gamification in Technical Higher Education," *Proceedings of the 45th ACM technical symposium on Computer s*

Science education, pp. 27-32.

Kahn, W. A. 1990. "Psychological Conditions of Personal Engagement and Disengagement at Work," *Academy of management journal* (33:4), pp. 692-724.

Kappen, D. L., and Nacke, L. E. 2013. "The Kaleidoscope of Effective Gamification: Deconstructing Gamification in Business Applications," *in: Proceedings of the First International Conference on Gameful Design, Research, and Applications.* pp. 119–122.

Kazhamiakin, R., Marconi, A., Perillo, M., Pistore, M., Valetto, G., Piras, L., Avesani, F., and Perri, N. 2015. "Using Gamification to Incentivize Sustainable Urban Mobility," *2015 IEEE first international smart cities conference (ISC2):* IEEE, pp. 1-6.

Khandelwal, S., Sripada, S. K., and Reddy, Y. R. 2017. "Impact of Gamification on Code Review Process: An Experimental Study," *Proceedings of the 10th innovations in software engineering conference*, pp. 122-126.

Klimmt, C. and Hartmann, T., 2006. “Effectance, self-efficacy, and the motivation to play video games,” *Playing video games: Motives, responses, and consequences*, pp.133-145.

Lackmann, S., Léger, P.-M., Charland, P., Aubé, C., and Talbot, J. 2021. "The Influence of Video Format on Engagement and Performance in Online Learning," *Brain Sciences* (11:2), p. 128.

Lameras, P., Arnab, S., Dunwell, I., Stewart, C., Clarke, S., and Petridis, P. 2017. "Essential Features of Serious Games Design in Higher Education: Linking Learning Attributes to Game Mechanics," *British journal of educational technology* (48:4), pp. 972-994.

Landers, R. N., and Landers, A. K. 2014. "An Empirical Test of the Theory of Gamified Learning: The Effect of Leaderboards on Time-on-Task and Academic Performance," *Simulation & Gaming* (45:6), pp. 769-785.

Landers, R. N., Bauer, K. N., and Callan, R. C. 2017. "Gamification of Task Performance with Leaderboards: A Goal Setting Experiment," *Computers in Human Behavior* (71), pp. 508-515.

Lee, M. J., Ko, A. J., and Kwan, I. 2013. "In-Game Assessments Increase Novice Programmers' Engagement and Level Completion Speed," *Proceedings of the ninth annual international ACM conference on International computing education research*, pp. 153-160.

Léger, P.-M., Charland, P., Feldstein, H. D., Robert, J., Babin, G., and Lyle, D. 2011. "Business Simulation Training in Information Technology Education: Guidelines for New Approaches in It Training," *Journal of Information Technology Education: Research*

(10:1), pp. 39-53.

Léger, P. M., Cronan, P., Charland, P., Pellerin, R., Babin, G., and Robert, J. 2012. "Authentic Om Problem Solving in an Erp Context," *International Journal of Operations & Production Management.*

Léger, P.-M., Davis, F. D., Cronan, T. P., and Perret, J. 2014. "Neurophysiological Correlates of Cognitive Absorption in an Enactive Training Context," *Computers in Human Behavior* (34), pp. 273-283.

Lepper, M. R., Greene, D., and Nisbett, R. E. 1973. "Undermining Children's Intrinsic Interest with Extrinsic Reward: A Test of the" Overjustification" Hypothesis," *Journal of Personality and social Psychology* (28:1), p. 129.

Liu, D., Santhanam, R., and Webster, J. 2017. "Toward Meaningful Engagement: A Framework for Design and Research of Gamified Information Systems," *MIS quarterly* (41:4).

Maheu-Cadotte, M.-A., Cossette, S., Dubé, V., Fontaine, G., Mailhot, T., Lavoie, P., Cournoyer, A., Balli, F., and Mathieu-Dupuis, G. 2018. "Effectiveness of Serious Games and Impact of Design Elements on Engagement and Educational Outcomes in Healthcare Professionals and Students: A Systematic Review and Meta-Analysis Protocol," *BMJ open* (8:3).

Malone, T. W. 1981. "Toward a Theory of Intrinsically Motivating Instruction," *Cognitive science* (5:4), pp. 333-369.

Malone, T. W. 1987. "Making Learning Fun: A Taxonomic Model of Intrinsic Motivations for Learning," *Conative and affective process analysis*.

Manyika, J. 2009. "Hal Varian on How the Web Challenges Managers," *McKinsey Quarterly* (2:4), pp. 1-9.

Manyika, J., Chui, M., Lund, S., and Ramaswamy, S. 2017. "What's Now and Next in Analytics, Ai, and Automation," *McKinsey Global Institute*, pp. 1-12.

McCormick, N. J., Clark, L. M., and Raines, J. M. 2015. "Engaging Students in Critical Thinking and Problem Solving: A Brief Review of the Literature," *Journal of Studies in Education* (5:4), pp. 100-113.

Miller, J. S. 2004. "Problem-Based Learning in Organizational Behavior Class: Solving Students' Real Problems," *Journal of Management Education* (28:5), pp. 578-590.

Miryung Kim, Thomas Zimmermann, Robert DeLine, and Andrew Begel. 2016. The emerging role of data scientists on software development teams. *InProceedings of the 38th*

International Conference on Software Engineering. ACM,96–107.

Muller, M., Lange, I., Wang, D., Piorkowski, D., Tsay, J., Liao, Q. V., Dugan, C., and Erickson, T. 2019. "How Data Science Workers Work with Data: Discovery, Capture, Curation, Design, Creation," *Proceedings of the 2019 CHI conference on human factors in computing systems*, pp. 1-15.

Moos, D. C., and Azevedo, R. 2009. "Learning with Computer-Based Learning Environments: A Literature Review of Computer Self-Efficacy," Review of educational research (79:2), pp. 576-600.

Nacke, L. E., and Deterding, S. 2017. "The Maturing of Gamification Research," *Computers in Human Behavior* (71), pp. 450-454.

Narasareddygari, M. R., Walia, G. S., and Radermacher, A. 2018. "Gamification in Computer Science Education: A Systematic Literature Review," *The 125th ASEE Annual Conference & Exposition.–Computers in Education Division*.

Nebel, S., Beege, M., Schneider, S., and Rey, G. D. 2016. "The Higher the Score, the Higher the Learning Outcome? Heterogeneous Impacts of Leaderboards and Choice within Educational Videogames," *Computers in Human Behavior* (65), pp. 391-401.

Ninaus, M., Greipl, S., Kiili, K., Lindstedt, A., Huber, S., Klein, E., Karnath, H.-O., and Moeller, K. 2019. "Increased Emotional Engagement in Game-Based Learning–a Machine Learning Approach on Facial Emotion Detection Data," *Computers & Education* (142), p. 103641.

Norman, G. R., and Schmidt, H. G. 2000. "Effectiveness of Problem-Based Learning Curricula: Theory, Practice and Paper Darts," *Medical education* (34:9), pp. 721-728.

O'Neil, H. F., Wainess, R., and Baker, E. L. 2005. "Classification of Learning Outcomes: Evidence from the Computer Games Literature," *The Curriculum Journal* (16:4), pp. 455-474.

Ortiz Rojas, M. E., Chiluiza, K., and Valcke, M. 2017. "Gamification and Learning Performance: A Systematic Review of the Literature," *11th European Conference on Game-Based Learning (ECGBL)*: ACAD CONFERENCES LTD, pp. 515-522.

Ortiz Rojas, M. E., Chiluiza, K., and Valcke, M. 2016. "Gamification in Higher Education and Stem: A Systematic Review of Literature," *8th International Conference on Education and New Learning Technologies (EDULEARN)*: Iated-int Assoc Technology Education A& Development, pp. 6548-6558.

Ortiz-Rojas, M., Chiluiza, K., and Valcke, M. 2019. "Gamification through Leaderboards: An Empirical Study in Engineering Education," *Computer Applications in Engineering Education* (27:4), pp. 777-788.

Orvis, K. A., Horn, D. B., and Belanich, J. 2008. "The Roles of Task Difficulty and Prior Videogame Experience on Performance and Motivation in Instructional Videogames," *Computers in Human Behavior* (24:5), pp. 2415-2433.

Overskeid, G. 2018. "Do We Need the Environment to Explain Operant Behavior?," *Frontiers in psychology* (9), p. 373.

Prensky, M. 2001. "Simulations: Are They Games," *Digital game-based learning*, pp. 210-219.

Provost, F., and Fawcett, T. 2013. "Data Science and Its Relationship to Big Data and Data-Driven Decision Making," *Big data* (1:1), pp. 51-59.

Przybylski, A. K., Rigby, C. S., and Ryan, R. M. 2010. "A Motivational Model of Video Game Engagement," *Review of general psychology* (14:2), pp. 154-166.

Przybylski, A. K., Ryan, R. M., and Rigby, C. S. 2009. "The Motivating Role of Violence in Video Games," *Personality and social psychology bulletin* (35:2), pp. 243-259.

Randel, J. M., Morris, B. A., Wetzel, C. D., and Whitehill, B. V. 1992. "The Effectiveness of Games for Educational Purposes: A Review of Recent Research," *Simulation & gaming* (23:3), pp. 261-276.

Riemer, V., and Schrader, C. 2016. "Impacts of Behavioral Engagement and Self-Monitoring on the Development of Mental Models through Serious Games: Inferences from in-Game Measures," *Computers in Human Behavior* (64), pp. 264-273.

Rigby, S., and Ryan, R. M. 2011. *Glued to Games: How Video Games Draw Us in and Hold Us Spellbound: How Video Games Draw Us in and Hold Us Spellbound*. AbC-CLIo.

Ruiz-Primo, M. A., Briggs, D., Iverson, H., Talbot, R., and Shepard, L. A. 2011. "Impact of Undergraduate Science Course Innovations on Learning," *Science* (331:6022), pp. 1269-1270.

Russell, J. A. 1980. "A Circumplex Model of Affect," *Journal of personality and social psychology* (39:6), p. 1161.

Ryan, R. M., Rigby, C. S., and Przybylski, A. 2006. "The Motivational Pull of Video Games: A Self-Determination Theory Approach," *Motivation and Emotion* (30:4), pp. 344-360.

Ryan, R. M., and Deci, E. L. 2000. "Self-Determination Theory and the Facilitation of Intrinsic

Motivation, Social Development, and Well-Being," *American psychologist* (55:1), p. 68.

Ryan, R., Bradshaw, E., and Deci, E. 2019. "A History of Human Motivation Theories in Psychology." Sternberg RJ, Pickren WE. Cambridge, UK: Cambridge University Press.

Ryan, R. M., and Deci, E. L. 2019. "Brick by Brick: The Origins, Development, and Future of Self-Determination Theory," in *Advances in Motivation Science*. Elsevier, pp. 111-156.

Ryan, R., Ryan, W., Di Domenico, S., and Deci, E. 2019. "The Nature and the Conditions of Human Flourishing: Self-Determination Theory and Basic Psychological Needs," *The Oxford handbook of human motivation.*

Sarason, Y., and Banbury, C. 2004. "Active Learning Facilitated by Using a Game-Show Format or Who Doesn't Want to Be a Millionaire?," *Journal of Management Education* (28:4), pp. 509-518.

Schaufeli, W. B., Salanova, M., González-Romá, V., and Bakker, A. B. 2002. "The Measurement of Engagement and Burnout: A Two Sample Confirmatory Factor Analytic Approach," *Journal of Happiness studies* (3:1), pp. 71-92.

Schunk, D. H. 1995. "Self-Efficacy and Education and Instruction," in *Self-Efficacy, Adaptation, and Adjustment*. Springer, pp. 281-303.

Schunk, D. H., and Mullen, C. A. 2012. "Self-Efficacy as an Engaged Learner," in *Handbook of Research on Student Engagemen*t. Springer, pp. 219-235.

Seaborn, K., and Fels, D. I. 2015. "Gamification in Theory and Action: A Survey," *International Journal of Human-Computer Studies* (74), pp. 14-31.

Seymour, E. 2002. "Tracking the Processes of Change in Us Undergraduate Education in Science, Mathematics, Engineering, and Technology," *Science Education* (86:1), pp. 79-105.

Simões, J., Redondo, R. D., and Vilas, A. F. 2013. "A Social Gamification Framework for a K-6 Learning Platform," *Computers in Human Behavior* (29:2), pp. 345-353.

Sitzmann, T. 2011. "A Meta-Analytic Examination of the Instructional Effectiveness of Computer-Based Simulation Games," *Personnel psychology* (64:2), pp. 489-528.

Skinner, E. A., and Belmont, M. J. 1993. "Motivation in the Classroom: Reciprocal Effects of Teacher Behavior and Student Engagement across the School Year," *Journal of educational psychology* (85:4), p. 571.

Song, I.-Y., and Zhu, Y. 2017. "Big Data and Data Science: Opportunities and Challenges of

Ischools," *Journal of Data and Information Science* (2:3), p. 1.

Song, I. Y., and Zhu, Y. 2016. "Big Data and Data Science: What Should We Teach?," *Expert Systems* (33:4), pp. 364-373.

Springer, L., Stanne, M. E., and Donovan, S. S. 1999. "Effects of Small-Group Learning on Undergraduates in Science, Mathematics, Engineering, and Technology: A Meta-Analysis," *Review of educational research* (69:1), pp. 21-51.

Spyridakis, J. H., and Isakson, C. S. 1991. "Hypertext: A New Tool and Its Effect on Audience Comprehension," *IPCC 91 Proceedings The Engineered Communication*, pp. 37-44 vol.31.

Sugar, S., and Takacs, G. 1999. "Games That Teach Teams: Tale of the Rat," *The Journal for Quality and Participation* (22:5), p. 54.

Schrage, M. 1999. *Serious Play: How the World's Best Companies Simulate to Innovate*. Harvard Business Press.

Tamborini, R., Bowman, N. D., Eden, A., Grizzard, M., and Organ, A. 2010. "Defining Media Enjoyment as the Satisfaction of Intrinsic Needs," *Journal of communication* (60:4), pp. 758-777.

Tang, S. H., and Hall, V. C. 1995. "The Overjustification Effect: A Meta-Analysis," *Applied Cognitive Psychology* (9:5), pp. 365-404.

Tomkin, J. H., Beilstein, S. O., Morphew, J. W., and Herman, G. L. 2019. "Evidence That Communities of Practice Are Associated with Active Learning in Large Stem Lectures," *International Journal of STEM Education* (6:1), pp. 1-15.

Tsai, F.-H., Yu, K.-C., and Hsiao, H.-S. 2012. "Exploring the Factors Influencing Learning Effectiveness in Digital Gamebased Learning," *Journal of Educational Technology & Society* (15:3), pp. 240-250.

Van Eynde, D. F., and Spencer, R. W. 1988. "Lecture Versus Experiential Learning: Their Differential Effects on Long-Term Memory," *Organizational Behavior Teaching Review* (12:4), pp. 52-58.

Venter, M. 2020. "Gamification in Stem Programming Courses: State of the Art," *2020 IEEE Global Engineering Education Conference (EDUCON)*: IEEE, pp. 859-866.

Walker, A., and Leary, H. 2009. "A Problem Based Learning Meta Analysis: Differences across Problem Types, Implementation Types, Disciplines, and Assessment Levels," *Interdisciplinary journal of problem-based learning* (3:1), p. 6.

Wang, C. K. J., Liu, W. C., Chye, S., and Chatzisarantis, N. L. 2011. "Understanding Motivation in Internet Gaming among Singaporean Youth: The Role of Passion," *Computers in Human Behavior* (27:3), pp. 1179-1184.

Watson, D., Clark, L. A., and Tellegen, A. 1988. "Development and Validation of Brief Measures of Positive and Negative Affect: The Panas Scales," *Journal of personality and social psychology* (54:6), p. 1063.

Whitehill, J., Serpell, Z., Lin, Y.-C., Foster, A., and Movellan, J. R. 2014. "The Faces of Engagement: Automatic Recognition of Student Engagement From Facial Expressions," *IEEE Transactions on Affective Computing* (5:1), pp. 86-98.

Wingfield, S. S., and Black, G. S. 2005. "Active Versus Passive Course Designs: The Impact on Student Outcomes," *Journal of Education for Business* (81:2), pp. 119-123.

Wu, M., Liao, C. C., Chen, Z.-H., and Chan, T.-W. 2010. "Designing a Competitive Game for Promoting Students' Effort-Making Behavior by Virtual Pets," *2010 Third IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning*: IEEE, pp. 234-236.

Yi, M. Y., and Davis, F. D. 2003. "Developing and Validating an Observational Learning Model of Computer Software Training and Skill Acquisition," *Information systems research* (14:2), pp. 146-169.

Yonezawa, S., Jones, M., and Joselowsky, F. 2009. "Youth Engagement in High Schools: Developing a Multidimensional, Critical Approach to Improving Engagement for All Students," *Journal of Educational Change* (10:2), pp. 191-209.

Zhang, A. X., Muller, M., and Wang, D. 2020. "How Do Data Science Workers Collaborate? Roles, Workflows, and Tools," *Proceedings of the ACM on Human-Computer Interaction* (4:CSCW1), pp. 1-23.

Zichermann, G., and Cunningham, C. 2011. *Gamification by Design: Implementing Game Mechanics in Web and Mobile Apps*. O'Reilly Media, Inc.

Zimmerman, B. J. 1989. "A Social Cognitive View of Self-Regulated Academic Learning," *Journal of educational psychology* (81:3), p. 329.

Chapter 2 : Research Article for Americas Conference on Information Systems

Can Competition Through Leaderboards Lead to Better Engagement and Learning of Data Science Concepts? An Experimental Study

Melissa Thériault[1], Thomas Ruel[2], Pierre-Majorique Leger[1], Jean-Francois Plante[1]

[1] HEC Montréal, Montréal, Canada
{melissa.theriault, pierre-majorique.leger, jean-francois.plante}@hec.ca
[2] Polytechnique Montréal, Montréal, Canada
thomas.ruel@polymtl.ca

Abstract

This paper examines the effect of gamification in engaging students to learn introductory concepts of data science, in particular, we study competitiveness through leaderboards. A between-subject experiment was conducted with 37 students and included two conditions: 1) playing against fictious opponents with a competitive leaderboard and, 2) playing alone with a leaderboard ranking only the subject's score. Our results show no effect of the competitive nature of leaderboards on learning and engagement. However, we found that highly efficacious participants with prior predictive modelling knowledge demonstrated higher levels of emotional arousal despite having a lower probability of increasing their knowledge on the subject matter. This suggests that individual differences such as self-efficacy and prior knowledge need to be accounted for when developing data science training that is augmented with competition through leaderboards as these factors may impact the learner’s ability to engage with the content.

2.1 Introduction

With the rise of data ubiquity, the desire to generate valuable insights and guide business decisions from this data is ever-growing. The demand for qualified data scientists with the necessary skills to organize and extract actionable knowledge from data has increased exponentially over the past years and a shortage of 250 000 data scientists is said to emerge in a decade (Lohr 2017; Manyika et al. 2017). s an effort to ameliorate the labour shortage, universities and online educational platforms have been offering courses in data science, relieving some of the need for extensive degrees in statistics or computer science which have historically been precursors to roles in data science (Kandel et al. 2012). However, STEM programs suffer from poor retention rates which hinders the quantity and quality of graduating students (Tomkin et al. 2019). Research suggests that the low retention is a result of perceived difficulty of the subject matter and lack of motivation (Koulouri, Lauria, and Macredie 2014). Low retention is also said to be due to the usage of traditional, lecture-centric teaching methods which ultimately lead to lower levels of engagement and learning when compared to active learning activities such as class discussions and team-projects (Freeman et al 2014; Henderson et al. 2011). In fact, studies also show that enactive learning, which is learning by doing, is important in data science education (De Veaux et al. 2017; Anderson et al. 2014)

One method that affords enactive learning is the application of gamification, which is defined by Deterding et al. as "the use of game design elements in non-game contexts." such as badges, points, levels, and more. Of the game elements studied, competition through leaderboards is the one of the most common (Hamari et al. 2014). Some studies suggest that leaderboards foster increased motivation for learners to perform better than others and in turn, leads to higher engagement as well as learning performance (Landers and Landers 2014; Ortiz-Rojar et al. 2019). Thus, the usage of competitive forms of gamification such as leaderboards to teach certain data science concepts, may be a solution to engage, retain, and quickly upskill new data scientists. However, although some research suggests that gamification can be a positive method to foster better engagement, attitudes towards subject matter, and increased class attendance in STEM-related fields (Ortiz-Rojas, Chiluiza, and Valcke 2016), others show mixed results about the effects of gamification in education (Dicheva et al. 2015). The literature suggests that studies related to gamification tend to be inconclusive since most research observe combined effects of multiple game elements, which makes it difficult to assess individual impacts of game elements (Dicheva et al. 2015).

Given this, and as a response to recent calls for theory-driven causal experiments on gamification (Seaborn and Fels 2015; Santhanam et al. 2016; Nacke and Deterding 2017), we investigate the extent to which the competitive nature of leaderboards impacts the engagement and learning outcomes of learners playing a business simulation that teaches introductory notions of predictive modelling. Specifically, the following research questions were explored in this study:

RQ1. “To what extent does the competitive nature of leaderboards impact learning performance and engagement in the context of learning data analytics?”
RQ2. “To what extent does self-efficacy moderate the impact of the competitive nature of leaderboards have on learning and engagement?
RQ3. “To what extent does individual differences in expertise moderate the impact competitive nature of leaderboards have on learning and engagement?

We hypothesize that the presence of a competitive leaderboard will lead to higher learning and engagement when compared to non-competitive leaderboard environments. This study also responds to Lander’s (2014) proposition that gamification does directly affect learning by including two moderators, self-efficacy, and prior knowledge, which we hypothesize will positively moderate the relationship between the competitiveness of leaderboards and learning as well as engagement.

The remainder of the paper is structured in the following manner. First, literature on leaderboards and the theoretical underpinnings of this research will be presented. Second, we discuss the research methodology and operationalization of the outcome variables. Third, we present the results followed by the discussion. We conclude with limitations and implications of the study.

2.2 Literature Review

Gamification with Leaderboards

Gamification, which refers to the addition of game mechanics for nongame applications, aims to use the engaging and motivational nature of games to foster learning of skills or change behaviours (Nacke and Deterding 2017). This means, the use of competition through leaderboards creates opportunities for students to engage in enactive learning and in turn, increase their knowledge. Enactive learning refers to how one learns from the consequences of their own actions. People continue behaviors that lead to successful results and refrain from behaviours that result in negative outcomes (Bandura 1977). Thus, feedback afforded by leaderboards is a form of enactive learning, which has been shown to have a positive effect on learning (Gupta et al. 2010; Leger et al. 2014).

Studies have reported findings where competition that emerges from leaderboards is effective in improving engagement with course material and in turn, course performance (Landers and Landers 2014), have been perceived as being more encouraging and motivating as well as lead to higher learning performance when compared to traditional methods of learning (Cwill 2020; Ortiz-Rojar et al. 2019). Other research suggests different results wherein learners reported decreased intrinsic motivation, satisfaction, and empowerment when exposed to gamified learning (Hanus and Fox 2015). Thus, it is possible that leaderboards have the potential to act as an extrinsic motivator and foster social comparison that may increase one's stress levels and in turn, negatively affect self-efficacy and learning performance (Cheng et al. 2009; Wu et al. 2010). Taken together, current empirical literature regarding the impact of the competitive nature of leaderboards on learning is inconsistent. Landers et al's (2015) review on the literature suggests that this inconsistency may be related to how leaderboards are often included in studies with other game elements and in turn, may cause different results in learning. Thus, to our knowledge, there are no recommendations on the causal impact of competitive leaderboards when used to teach data science concepts.

Self-Determination Theory

The self-determination theory (SDT) (Deci and Ryan 1980) is a leading theory of human motivation that has been used to understand the motivational effects of game design elements (Ryan et al. 2006; Deci and Ryan 1980). SDT divides motivation into two types: 1) intrinsic, which refers to performing an activity because of inner interest, and 2) extrinsic, which refers to motivation that results from attaining or avoiding outcomes such as rewards or punishments. When individuals feel intrinsically motivated, they perform activities for the positive feelings that emerge from the activities themselves and in turn, are more curious, determined to overcome challenges, and perform better (Deci and Ryan 1991). SDT states that the satisfaction of three basic psychological needs leads to increased intrinsic motivation: 1) competence, refers to one's mastery and effectiveness in completing a given challenge, 2) autonomy, defined as the ownership of one's behaviour while performing a task, and 3) relatedness, described as feeling connected to others (Deci and Ryan 1980). Research conducted by Ryan et al. (2006) suggests that playing video games can satisfy our physiological needs and is positively associated with immersion, persistence in play, and enjoyment. That is, playing video games leads to increased intrinsic motivation and in turn, increased performance (Ryan et al. 2006). Gamification fosters the same intrinsic motivation as in video games with the use of game elements in a non-gaming context. Similarly, competitive types of gamification may lead to greater persistence in participation, or in other words, increased engagement.

Behavioural and Emotional Engagement

Engagement is important to our research. Fredericks et al., (2004) propose a multidimensional concept of engagement which holds three core facets: behavioural, emotional, and cognitive. Behavioural engagement is described as involvement, following rules and norms, and active participating in social or learning activities (Fredericks et al. 2004). Emotional engagement refers to one's display of interest or affective emotional reactions such as boredom, happiness, sadness, and anxiety in response to participating in activity (Fredericks et al. 2004). These affective emotional reactions can be separated as arousal and valence. Arousal refers to how physiologically aroused or calm one is, whereas valence is defined as a pleasant or unpleasant emotional state (Charland et al. 2015). Cognitive engagement focuses on the psychological attention and effort in overcoming a challenge (Fredericks et al. 2004). Studies have shown positive associations between video games and student engagement (Annetta et al. 2000). The present research examines behavioural and emotional engagement types. Cognitive engagement is not applicable in this study as appropriate physiological tools needed to measure cognitive states of subjects were not used.

Moderating Effects of Self-Efficacy and Prior Experience

The current study examines the moderating effects of self-efficacy and prior experience. First, according to Social Cognitive Theory (SCT) by Bandura (1977; 1986), learning behaviours are influenced by the interactions between environmental and personal factors. Environmental factors refer to physical external factors that change one's behaviours, while personal factors are described as cognitive, emotional, of demographic characteristics that define a person (Bandura 1986). Self-efficacy, which is defined as one's belief in their ability to learn or achieve a behaviour or task, is a fundamental cognitive factor that aids in regulating human behaviour (Bandura 1986). SCT and the concept of self-efficacy has been used to explore human behaviour in various contexts such as education and information systems (Zimmerman 1989; Gupta et al. 2010). Compeau and Higgins (1995) used SCT to examine behaviours in the context of computer training and show that higher levels of computer self-efficacy are positively related to positive emotional reactions and performance in task completion. They also found that computer self-efficacy has an impact on predicting one's willingness to stay actively engaged with a computer or software.

Second, individual differences exist among learners, with prior knowledge being particularly important as it affects how one associate's new knowledge with what they currently know (Spyridakis and Isakson 1991). Studies show that prior knowledge affects one's performance in game-based learning environments and digital games, suggesting that varying levels of prior knowledge will affect how one benefits from an experience (Bulu and Pedersen 2012; Orvis et al. 2008). To this end, this study aims to examine the influences of software self-efficacy and prior knowledge on learning and engagement by considering an experience with and without a competitive leaderboard.

2.3 Research Method

Experimental Design and Sample

To test out hypotheses, we conducted a between-subject experiment. One experimental factor, the competitive leaderboard, was manipulated leading to two conditions as seen in Figure 1: Condition 1) solving a problem while playing against fictitious opponents with a competitive leaderboard that shows others' scores based on the time elapsed, and Condition 2) solving a problem alone with a leaderboard that only ranks the player's own scores.37 subjects participated (23 females, 14 males) and were randomly assigned to a condition. The average age was 25.8 years old (SD=4.4). Inclusion criteria included the completion of an introductory course to statistics. 14 participants had prior experience with predictive modeling. Of these 14 subjects with experience, 9 were in the control group and 5 in the experimental group.

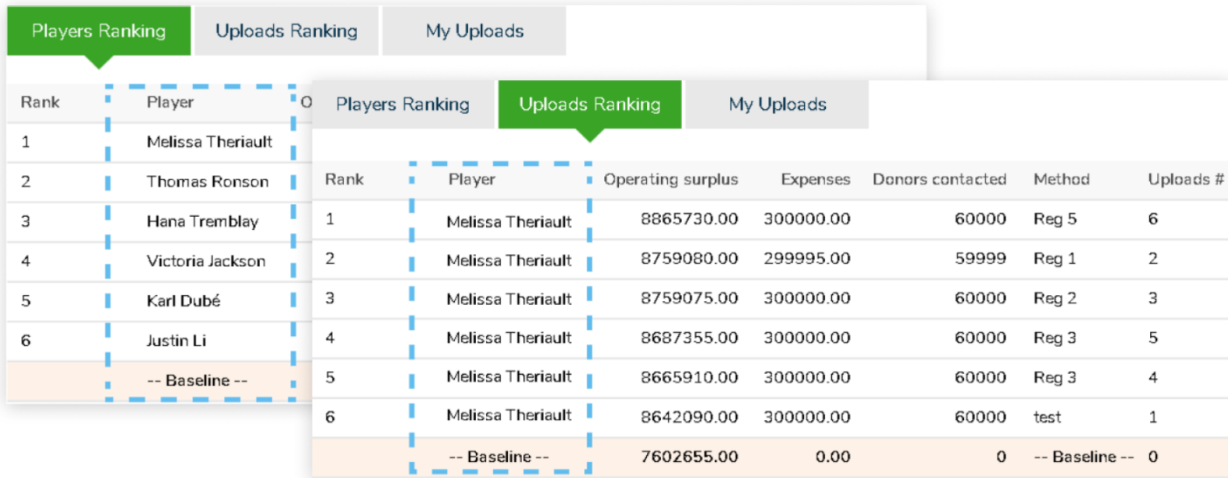

Figure 1. Left to right: Condition 1 playing against opponents with a competitive leaderboard, Condition 2 playing against oneself only.

Experimental Stimuli and Task

We employed an experimental simulation using Cortex, a business analytics simulation game by ERPSim Lab and HEC Montréal, wherein subjects used SAS Enterprise Miner® (SAS EM) to maximize funds raised for a charity given a dataset for one million potential donors and a cost structure for calling each member. For better experimental control, subjects played via an Amazon Web Services cloud instance on their PCs. For Condition 1, we built a new interface and created five fictitious opponents. To prevent motivational effects from playing against opponents that are too strong or weak, scores were created based on average scores from our pre-tests. Subjects were instructed to predict donation amounts using regression and decision tree models. Decisions were uploaded to Cortex for immediate scoring through the leaderboard.

Procedure

Participants completed a 30-minute video training to learn basic concepts of predictive modelling prior to the experiment. Once completed, we connected with them remotely using Lockback.io, a real-time moderation platform (Lookback 2020), and gave a 30-minute hands-on training to learn the basics of SAS EM. Then, participants answered a pre-test assessment. For Condition 1, subjects were told that five other players were waiting to be added to the game prior to starting. The moderator pretended to check-in with the fictitious players and after 2-minutes, declared the start of the 60-minute competition. Participants in Condition 2 were told to start the game immediately after the training. Finally, they completed a post-test assessment and filled out a questionnaire.

Subjects were recorded with a webcam via Lookback (Lookback, California, United States) during game play.

Operationalization of Research Variables

Outcome Variables

Learning outcomes were operationalized using objective and self-reported measures. Objective knowledge (Obj_kno) and certitude (Obj_cert) were assessed using the same 10 true or false questions measuring the predictive modelling knowledge of the participant with a certitude component before (pre-test) and after playing the simulation (post-test). Objective knowledge and certitude were measured using the difference between post- and pre-test results to calculate how much subjects have learned. Questions were adapted from MOOCs (massive open online courses) about predictive analytics and designed based on Bloom's revised taxonomy of educational objectives (Krathwohl 2002), targeting the lowest level of the six learning objectives - remembering. The objective with these questions was to test learners' ability to retrieve, recognize, and recall knowledge. Initially, we compiled 40 questions, but only 10 were retained to limit the participants' burden. To assess if questions targeted the correct level of learning objective and to classify them, 5 experts evaluated the level of perceived complexity of the questions using a 7-point Likert scale from very low to very high complexity, as done by Cronan et al. (2012). Questions answered incorrectly by more than 60% of experts were removed. 50% of the questions in the final assessment were of low complexity, and the other 50%, of high complexity. The self-reported predictive modelling knowledge (PM_kno) construct measured participants' agreement in terms of understanding the subject matter using 4 self-efficacy items adapted from Hollenbeck and Brief (1987) to fit the context of the predictive analytics.

Engagement was operationalized via emotional and behavioural engagement. Emotional engagement was inferred by measuring valence (Val) (positive or negative emotion) and arousal (Ars) (calm or aroused) extracted from participants' facial expressions using the automatic facial expression recognition software, FaceReader 4.0 (Noldus Information Technology Inc, Netherlands). FaceReader has been validated and used in studies similar to ours (Charland et al. 2015). Behavioural engagement was measured using two indicators that reflect the effort and persistence invested while performing in-game activities (Fredricks et al., 2004). These included in-game improvement (Imprv), which was measured using the difference between first and last upload, and the average marginal increase between each new score posted (Marginc).

Moderating Variables

The two moderator variables were software self-efficacy (SSE) and predictive modelling experience (PM_Exp). SSE measured participants' judgment of their capabilities to use SAS EM in diverse situations using 8 items adapted from Compeau and Higgins (1995) to fit the context of SAS EM. PM_Exp was measured with a yes- no question, asking subjects if they had prior experience with predictive modelling.

2.4 Data Analysis and Results

Data analysis proceeded in two stages. First, we tested the main effects of the presence of a competitive leaderboard on learning outcomes and engagement. Afterwards, we tested the moderating effects of SSE and PM_Exp. We used the Shapiro-Wilk test to assess normality and developed linear regression models to assess the impact of the competitive leaderboard on the different learning outcomes and engagement without interaction terms and then, including them. Only Obj_cert, Pred_kno, and Ars were normally distributed. When the data was not normally distributed, we treated the dependent variable as binary [median split] and conducted logistic regressions. Age, gender, and education level were included in every model as control variables.

Main Effects

We proposed that a competitive leaderboard would lead to better objective and self-reported learning, in addition to higher emotional and behavioural engagement. Results in Table 1 show that the condition the subject was in (e.g., competitive, or non-competitive leaderboard condition), did not significantly influence any of the outcome variables. Thus, our hypothesis is not supported.

Dependent Variable	Effect	df	Estimate	P-value[one-tailed]
Objective knowledge (Obj_kno)	Condition	32	-0.24	0.75
Objective certitude (Obj_cert)	Condition	32	-0.06	0.82
Self-reported predictive modelling knowledge (PM_kno)	Condition	32	0.27	0.14
Valence (Val)	Condition	32	0.59	0.43
Arousal (Ars)	Condition	32	-0.005	0.65
In-game improvement (Imprv)	Condition	32	0.47	0.54
Avg marginal increase of each upload (Marginc)	Condition	32	-1.24	0.11

Table 1. Results of if the subjects' condition impacted learning outcomes and engagement. The effect of "Condition" refers to the comparison of Condition 1 and 2 (* $p<0.10$, ** $p<0.05$).

Moderation Effects

Next, we tested moderation effects of SSE and PM_Exp on the relationship between competitive leaderboards and learning and engagement. Models used to test the moderation effects include the same variables used to test our first hypothesis, and include interaction terms involving the condition with SSE and PM_Exp.

Software Self-Efficacy

While the moderating effect of SSE on the relationship between condition and self-reported predictive modelling knowledge (PM_kno) is not significant, results in Table 2 suggest that the effect on relationship between Condition and Objective Knowledge (Obj_kno) was negative and significant ($\beta=-1.67$, $p<0.05$). Thus, there is not enough evidence to support that SSE positively moderates the relationship between condition and objective learning. For the moderating effect on the relationship between condition and emotional engagement, results suggest that the effect on arousal (Ars) was positive and stronger for participants in Condition 1 ($\beta=-0.01$, $p=0.07$) when compared to Condition 2. Results show no significant moderation effect on the relationship between condition and valence (Val). In other words, the competitive leaderboard seems to positively influence arousal levels more for participants with higher SSE but does not appear to have influenced valence. Thus, our hypothesis that self-efficacy positively moderates the relationship between the competitive nature of leaderboards and emotional engagement is generally supported. Also, SSE did not play a significant moderating role on the relationship between condition and behavioural engagement. Whether the participant was exposed to a competitive leaderboard or not does not appear to have had a different effect on their in-game improvement or average marginal increase between each upload based on their level of SSE.

Predictive Modelling Experience

The moderating effect of PM_Exp on the relationship between condition and objective and self-reported learning, and emotional engagement were not significant. However, among those with prior experience, the competitive leaderboard seemed to have had a negative and significant effect on participants' probability of attaining high in-game improvement (Imprv) (β=-4.37, p=0.07) and average marginal increase between each upload (Marginc) (β=-4.24, p=0.07) when compared to Condition 2. Thus, our hypothesis that prior experience positively moderates the relationship between competitive leaderboards and learning as well as engagement, is not supported.

Dependent Variable	Effect	df	Estimate	P-value[one-tailed]
Obj_kno	Condition * SSE	30	-1.67	0.04**
	Condition * PM_Exp	30	-2.31	0.21
Obj_cert	Condition * SSE	29	0.11	0.53
	Condition * PM_Exp	29	-0.14	0.80
PM_kno	Condition * SSE	29	0.01	0.78
	Condition * PM_Exp	29	-0.11	0.79
Val	Condition * SSE	30	0.09	0.86
	Condition * PM_Exp	30	-0.26	0.88
Ars	Condition * SSE	29	0.01	0.07*
	Condition * PM_Exp	29	-0.01	0.62
Imprv	Condition * SSE	30	0.19	0.78
	Condition * PM_Exp	30	-4.37	0.07*
Marginc	Condition * SSE	30	-0.08	0.88
	Condition * PM_Exp	30	-4.24	0.07*

Table 2. The moderating effects of SSE and prior experience on learning outcomes and engagement (* p<0.10, ** p<0.05).

2.5 Discussion

This research is an exploratory attempt to analyze how learning and engagement are impacted by a competitive leaderboard, and how this relationship is moderated by one's self-efficacy and domain experience, in the context of learning data analytics. Thus, we investigated the competitive nature of leaderboards in a new context. Three key insights emerge from our study. First, based on our results, we cannot say that the presence of a competitive leaderboard was beneficial or detrimental to learning and engagement, since no statistically significant differences were observed between groups. According to a review on gamification in education conducted by Dicheva et al. (2015), multiple studies similar to the present tend to be inconclusive. Authors found that common reasons for the high ratio of inconclusive studies in gamification research are study period length and sample size. Thus, possible explanations for our results may be a consequence of subjects only having 60 minutes to play the game and then asked to complete an assessment to evaluate their newly acquired knowledge. This might not be long enough for subjects to be comfortable with the subject matter, especially since predictive modelling is an advanced concept in data analytics.

Further, in this article, we present an integrated view of how SDT (Deci and Ryan 1980) and SCT (Bandura 1986) together predict how certain game elements influence learners' intrinsic motivation and learning behaviours. These theories imply that the use of game elements may be beneficial to learning as they aid in satisfying our 3 basic needs and shape environmental factors to alter learners' behaviours. Another possible reason for the present null results may stem from the lack of social dimension of the competitive leaderboard. That is, our experiment included fictitious opponents in Condition 1, not peers. The satisfaction of the need for relatedness as outlined by SDT, may have been absent. Similarly, according to SCT, the absence of social external factors prevented the presence of social support, and in turn, may have negatively impacted participants' beliefs and their ability to learn and perform well in the post-test assessment. Moreover, guidelines for effective gamification of business applications by Kappen and Nacke (2013) highlight the need to create possibilities for social connectedness, acceptance, and validation within a gamified system to help increase intrinsic motivation and ultimately, willingness to learn. Thus, a longer study with more subjects within the same social context may have yielded different results.

Second, our results reveal that participants with low initial self-efficacy benefited most from the competitive leaderboard as they had a higher probability of increasing their objective knowledge. Contrastingly, those with high self-efficacy had a lower probability of attaining the same level of knowledge. Without the competitive leaderboard, the effect on participants' objective learning is more stable regardless of their self-efficacy, that is, they did not reach exceptionally high levels of knowledge gained but did not regress in their learning either. Our finding might be in accordance with the well-known Dunning-Kruger effect. Kruger and Dunning (1999) argued that a person's own perception of their ability can often be overestimated and not truly reflect their actual knowledge base. The inverse is also suggested to be true; individuals that are more knowledgeable tend to underestimate their ability. This effect has been reported in studies related to game-based learning (Ocay 2019). These arguments could explain why participants that felt confident in their ability to perform well using the software at the start of the game, seem to have lower probability of attaining higher objective learning when interacting with a competitive leaderboard by the end of the simulation. This suggests that novices with low initial self-efficacy need competitive leaderboards as means of encouragement, which in turn, may increase their probability of better objective learning.

Further, we find that for highly efficacious participants, the competitive leaderboard seems to foster higher levels of arousal when compared to a non-competitive leaderboard. Under a SCT lens, self-efficacy is acquired gradually, partly by emotional arousal or the judgment of one's physiological state (Bandura et al. 1961). Stress is a form of emotional arousal that is caused in part by fear of physiological reactions to challenging situations (Bandura 1977). This may explain why self-efficacy seems to positively moderate the relationship between the presence of a competitive leaderboard and emotional arousal. Participants that had sufficient self-efficacy recognized the challenging nature of making business decisions using newly learned predictive modelling concepts in a time bound and competitive context. This seemed to have caused a higher emotional arousal response to the taxing situation in the form of stress. However, this is not necessarily a bad response. Stress or frustration, which can be related to arousal (Verona et al. 2009), during gameplay due to challenge and competition may result in more engagement (Salminen et al. 2009). Finally, from the perspective of cognitive engagement, which is the psychological investment and effort put towards a task, such negative emotions may lead to more systematic and analytical processing, thus leading to increased attention to detail (Fredericks et al.

2004). Considering the present study explores the effects of a competitive game element in the context of predictive analytics which is a highly analytical discipline, negative emotions, that is, high emotional arousal, may be beneficial to players.

Third, results suggest that experienced participants in Condition 1 have a slightly lower probability of achieving high in-game improvement and average marginal increases between uploads, when comparing to participants of the same experience level in Condition 2. Contrary to expectations, this study did not reveal higher behavioural engagement for participants that played using a leaderboard and against fictitious opponents. Keeping SDT in mind, a possible explanation for these results may be the lack of adequate satisfaction of the need to feel competent for those holding experience in developing predictive models (Deci and Ryan 1980). The combination of learning to use a new statistical software and building predictive models within a time constraint may have been unbalanced in terms of the extent of the participants' predictive modelling knowledge and activity at hand. In other words, the task during the experiment may have been too difficult, even for the more experienced participants. This can be detrimental to one's intrinsic motivation and in turn, foster lower levels of behavioural engagement in-game (Ryan et al. 2006). Thus, it would seem that before leveraging a serious game to teach predictive analytics, managers and educators should ensure that participants are given enough time to learn both the basic concepts and the statistical software used to play the game to promote optimal behavioural engagement.

2.6 Conclusion

This study is not without limitations. The major limitations are related to the study design. First, the sample size is small for a between-subject experiment with only two conditions. A larger sample size may have provided more statistically significant analyses. Second, though we aimed to investigate the competitiveness nature of leaderboards, we did not include a control group wherein subjects were not exposed to leaderboards or the simulated learning environment. It would be relevant to examine the differences in learning and engagement between groups that do not use any leaderboards or other game dynamics, and a group that uses leaderboards. Perhaps researchers who study the phenomena this way will be able to measure more complete reactions and experiences. Another limitation is the length of the study. 90 minutes to observe real change in learning is difficult, therefore future research should consider a longitudinal study such as over a semester to measure engagement and learning over time. Finally, since we simulated the opponents, the present study did not capture potential effects of social comparison that emerge from the competitiveness of leaderboards. Having subjects play against people they are familiar with such as classmates or colleagues may have yielded less inconclusive results.

Overall, though direct effects of the competitive nature of leaderboards on learning and engagement were inconclusive, we found that self-efficacy and prior knowledge have impacts on the effect of competitive leaderboards. Should managers or educators decide use leaderboard boards as means to encourage competition in an introductory training or class while teaching learners how to solve an applied business problem related to predictive analytics, perhaps, it is best to adapt the time teaching the fundamentals prior to engaging in competition as not all learners can handle the mental effort to process a new tool, concepts, and stress to be successful.

References

Anderson, P. E., Turner, C., Dierksheide, J., and McCauley, R. 2014. "An Extensible Online Environment for Teaching Data Science Concepts through Gamification," *2014 IEEE frontiers in education conference (FIE) proceedings: IEEE*, pp. 1-8.

Annetta, L. A., Minogue, J., Holmes, S. Y., and Cheng, M.-T. 2009. "Investigating the Impact of Video Games on High School Students' Engagement and Learning About Genetics," *Computers & Education* (53:1), pp. 74-85.

Bandura, A., Ross, D., and Ross, S. A. 1961. "Transmission of Aggression through Imitation of Aggressive Models," *The Journal of Abnormal and Social Psychology* (63:3), pp. 575-582.

Bandura, A. 1977. *Social Learning Theory*. Oxford, England: Prentice-Hall.

Bandura, A. 1986. *Social Foundations of Thought and Action: A Social Cognitive Theory*. Englewood Cliffs, NJ, US: Prentice-Hall, Inc.

Bulu, S. T., and Pedersen, S. 2012. "Supporting Problem-Solving Performance in a Hypermedia Learning Environment: The Role of Students' Prior Knowledge and Metacognitive Skills," *Computers in Human Behavior* (28:4), pp. 1162-1169.

Charland, P., Léger, P.-M., Sénécal, S., Courtemanche, F., Mercier, J., Skelling, Y., and Labonté-Lemoyne, E. 2015. "Assessing the Multiple Dimensions of Engagement to Characterize Learning: A Neurophysiological Perspective," *Journal of visualized experiments*: JoVE:101.

Cheng, H. N., Wu, W. M., Liao, C. C., and Chan, T.-W. 2009. "Equal Opportunity Tactic: Redesigning and Applying Competition Games in Classrooms,"*Computers & Education*(53:3), pp.866-876.

Compeau, D., and Higgins, C. 1995. "Computer Self-Efficacy: Development of a Measure and Initial Test," *MIS Q.* (19), pp. 189-211.

Cronan, T. P., Léger, P.-M., Robert, J., Babin, G., and Charland, P. 2012. "Comparing Objective Measures and Perceptions of Cognitive Learning in an Erp Simulation Game: A Research Note," *Simulation & Gaming* (43:4), pp. 461-480.

Ćwil, M. 2020. "Leaderboards–a Motivational Tool in the Process of Business Education," *Joint International Conference on Serious Games*: Springer, pp. 193-203.

De Veaux, R. D., Agarwal, M., Averett, M., Baumer, B. S., Bray, A., Bressoud, T. C., Bryant, L., Cheng, L. Z., Francis, A., and Gould, R. 2017. "Curriculum Guidelines for Undergraduate Programs in Data Science," *Annual Review of Statistics and Its Application* (4), pp. 15-30.

Deci, E. L., and Ryan, R. M. 1980. "Self-Determination Theory: When Mind Mediates Behavior," *Journal of Mind and Behavior* (1:1), pp. 33-43.

Deci, E. L., and Ryan, R. M. 1991. "A Motivational Approach to Self: Integration in Personality," *in Nebraska Symposium on Motivation*, 1990*: Perspectives on Motivation*. Lincoln, NE, US: University of Nebraska Press, pp. 237-288.

Deterding, S., Khaled, R., Nacke, L. E., and Dixon, D. 2011. "Gamification: Toward a Definition," *CHI 2011 gamification workshop proceedings:* Vancouver BC, Canada, pp. 1-79.

Dicheva, D., Dichev, C., Agre, G., and Angelova, G. 2015. "Gamification in Education: A Systematic Mapping Study," *Journal of Educational Technology & Society* (18:3), pp. 75-88.

Ekman, P. 1993. "Facial Expression and Emotion," *American psychologist* (48:4), p. 384.

Fredricks, J. A., Blumenfeld, P. C., and Paris, A. H. 2004. "School Engagement: Potential of the Concept, State of the Evidence," *Review of Educational Research* (74:1), pp. 59-109.

Hamari, J., Koivisto, J., and Sarsa, H. 2014. "Does Gamification Work? -- a Literature Review of Empirical Studies on Gamification," *2014 47th Hawaii International Conference on System Sciences*, pp. 3025-3034.

Hanus, M. D., and Fox, J. 2015."Assessing the Effects of Gamification in the Classroom:A Longitudinal Study on Intrinsic Motivation, Social Comparison, Satisfaction, Effort, and Academic Performance," *Computers & education* (80), pp. 152-161.

Hollenbeck, J. R., and Brief, A. P. 1987. "The Effects of Individual Differences and Goal Origin on Goal Setting and Performance," *Organizational Behavior and Human Decision Processes* (40:3), pp. 392-414.

Kandel, S., Paepcke, A., Hellerstein, J. M., and Heer, J. 2012. “Enterprise Data Analysis and Visualization: An Interview Study”, *IEEE Transactions on Visualization and Computer Graphics* (18:12), pp.2917–2926

Kappen, D. L., and Nacke, L. E. 2013. "The Kaleidoscope of Effective Gamification:

Deconstructing Gamification in Business Applications," *in: Proceedings of the First International Conference on Gameful Design, Research, and Applications.* pp. 119–122.

Koulouri, T., Lauria, S., and Macredie, R. D. 2014. "Teaching Introductory Programming: A Quantitative Evaluation of Different Approaches," *ACM Transactions on Computing Education* (TOCE) (14:4), pp. 1-28.

Krathwohl, D. R. 2002. "A Revision of Bloom's Taxonomy: An Overview," *Theory into practice* (41:4), pp. 212-218.

Kruger, J., and Dunning, D. 1999. "Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments," *J Pers Soc Psy* (77:6), pp.1121-1134.

Landers, R. N., Bauer, K. N., and Callan, R. C. 2017. "Gamification of Task Performance with Leaderboards: A Goal Setting Experiment," *Computers in Human Behavior* (71), pp. 508-515.

Landers, R. N., and Landers, A. K. 2014. "An Empirical Test of the Theory of Gamified Learning: The Effect of Leaderboards on Time-on-Task and Academic Performance," *Simulation & Gaming* (45:6), pp. 769-785.

Léger, P.-M., Davis, F. D., Cronan, T. P., and Perret, J. 2014. "Neurophysiological Correlates of Cognitive Absorption in an Enactive Training Context," *Computers in Human Behavior* (34), pp. 273-283.

Lohr, S. 2017."Where are the STEM Jobs Are (and Where They Aren't). *New York Times*.

Manyika, J., Chui, M., Lund, S., and Ramaswamy, S. 2017. "What's Now and Next in Analytics, Ai, and Automation," *McKinsey Global Institute,* pp. 1-12.

Nacke, L. E., and Deterding, S. 2017. "The Maturing of Gamification Research," *Computers in Human Behavior* (71), pp. 450-454.

Ocay, A. B. 2019. "Investigating the Dunning-Kruger Effect among Students within the Contexts of a Narrative-Centered Game-Based Learning Environment," *Proceedings of the 2019 2nd International Conference on Education Technology Management*, pp. 8-13.

Orvis, K. A., Horn, D. B., and Belanich, J. 2008. "The Roles of Task Difficulty and Prior Videogame Experience on Performance and Motivation in Instructional Videogames," *Computers in Human Behavior* (24:5), pp. 2415-2433.

Ortiz-Rojas, M., Chiluiza, K., and Valcke, M. 2019. "Gamification through Leaderboards: An

Empirical Study in Engineering Education," *Computer Applications in Engineering Education* (27:4), pp. 777-788.

Ortiz Rojas, M. E., Chiluiza, K., and Valcke, M. 2016. "Gamification in Higher Education and Stem: A Systematic Review of Literature," *8th International Conference on Education & New Learning Technologies: Iated-int Assoc Technology Education Development*, pp.6548-6558

Ryan, R. M., Rigby, C. S., and Przybylski, A. 2006. "The Motivational Pull of Video Games: A Self-Determination Theory Approach," *Motivation and Emotion* (30:4), pp. 344-360.

Salminen, M., Kivikangas, J. M., Ravaja, N., and Kallinen, K. 2009. "Frontal Eeg Asymmetry in the Study of Player Experiences During Competitive and Cooperative Dual Play," *Proceedings of Game and Entertainment Technologies*, pp. 44-55.

Santhanam, R., Liu, D., and Shen, W.-C. M. 2016. "Research Note—Gamification of Technology-Mediated Training: Not All Competitions Are the Same", *IS Research* (27:2), pp. 453–465

Seaborn, K., and Fels, D. I. 2015. "Gamification in Theory and Action: A Survey," *International Journal of Human-Computer Studies* (74), pp. 14-31.

Spyridakis, J. H., and Isakson, C. S. 1991. "Hypertext: A New Tool and Its Effect on Audience Comprehension," *IPCC 91 Proceedings the Engineered Communication*, pp. 37-44 vol.31.

Tomkin, J. H., Beilstein, S. O., Morphew, J. W., and Herman, G. L. 2019. "Evidence That Communities of Practice Are Associated with Active Learning in Large Stem Lectures," *International Journal of STEM Education* (6:1), pp. 1-15.

Verona, E., Sadeh, N., and Curtin, J. J. 2009. "Stress-Induced Asymmetric Frontal Brain Activity and Aggression Risk," *Journal of abnormal psychology* (118:1), p. 131.

Wu, M., Liao, C. C., Chen, Z.-H., and Chan, T.-W. 2010. "Designing a Competitive Game for Promoting Students' Effort-Making Behavior by Virtual Pets," 2010 3rd IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning: IEEE, pp. 234-236.

Zimmerman, B. J. 1989. "A Social Cognitive View of Self-Regulated Academic Learning,"*Journal of educational psychology* (81:3), p. 329.

Conclusion

Summary

The current study attempted to provide empirical evidence on the impacts of gamification on both learning outcomes and learner engagement and moderating effects of self-efficacy and prior knowledge on this relationship. To achieve this, we conducted a between-subject experimental study to observe differences between two conditions: 1) with a competitive leaderboard, and 2) with a non-competitive leaderboard, to understand the effects of a singular gamification element on learning performance and engagement. Learning outcomes were observed using objective and self-reported measures. Engagement was measured using implicit and implicit measures. Overall, our results regarding the direct relationship between leaderboards and learning outcomes and engagement were non-conclusive, however, we did find that self-efficacy and prior knowledge negatively moderates this relationship.

Main results and contributions

The results of the article in this thesis helped provide insights to the following research questions:

RQ1. “To what extent does the competitive nature of leaderboards impact learning performance and engagement in the context of learning data analytics?”

RQ2. “To what extent does self-efficacy moderate the impact of the competitive nature of leaderboards have on learning and engagement?

RQ3. “To what extent does individual differences in expertise moderate the impact competitive nature of leaderboards have on learning and engagement?

More specifically, we developed the following hypotheses for both direct and moderated effects of gamification on learning and engagement as included in the article in this thesis that has been submitted to the *Americas Conference on Information Systems* (ACMIS) in March 2021 and is currently under review.

Related Research Question	Hypothesis	Result
RQ1. To what extent does the competitive nature of leaderboards impact learning performance and engagement in the context of learning data analytics?	**H1A.** The competitive leaderboard will lead to better objective learning outcomes, when compared to a static competitive leaderboard.	Not supported
	H1B. The competitive leaderboard will lead to better self-reported learning outcomes, when compared to a static competitive leaderboard.	Not supported
	H2A. In gamified training, learners will experience high emotional engagement when a competitive leaderboard is used compared to a static competitive leaderboard.	Not supported
	H2B. In gamified training, learners will experience high behavioural engagement when a competitive leaderboard is used compared to a static competitive leaderboard.	Not supported
RQ2: To what extent does self-efficacy moderate the impact of the competitive nature of leaderboards have on learning and engagement?	**H3A.** Software self-efficacy positively moderates the relationship between the effect of competitive leaderboards on objective learning outcomes.	Not supported
	H3B. Software self-efficacy positively moderates the relationship between the effect of competitive leaderboards on self-reported learning outcomes.	Not supported
	H3C. Software self-efficacy positively moderates the relationship between the effect of competitive leaderboards on emotional engagement.	Generally supported
	H3D. Software self-efficacy positively moderates the relationship between the effect of competitive leaderboards on behavioural engagement.	Not supported
RQ3: To what extent does individual differences in expertise moderate the impact competitive nature	**H4A.** Prior predictive modelling knowledge positively moderates the relationship between the effect of competitive leaderboards on objective learning outcomes.	Not supported

of leaderboards have on learning and engagement?	**H4B.** Prior predictive modelling knowledge positively moderates the relationship between the effect of competitive leaderboards on self-reported learning outcomes.	Not supported
	H4C. Prior predictive modelling knowledge positively moderates the relationship between the effect of competitive leaderboards on emotional engagement.	Not supported
	H4D. Prior predictive modelling knowledge positively moderates the relationship between the effect of competitive leaderboards on behavioural engagement.	Not supported

Overall, this thesis' main hypothesis is not supported; we did not observe any significant differences in learning performance or engagement between our control and experimental groups, thus not providing any conclusive evidence that making a leaderboard competitive can indeed improve experiential learning in the context of learning data analytics. Inversely, we also cannot say that the integration of competitive leaderboards was detrimental to the learning or engagement of subjects. Although no main effects were observed, we did find that self-efficacy and prior knowledge about predictive modelling had negative significant moderating effects on the relationship between competitive leaderboards and objective learning and behavioural engagement. Low efficacious participants benefited most from the feedback game element as results showed they had a higher probability of increasing their objective knowledge by the end of the simulation. The inverse is also suggested to be accurate for those with high self-efficacy. Further, prior knowledge of the predictive analytics seemed to foster less observable behavioural engagement in the condition with the competitive game element than without. Finally, in accordance with the theoretical understanding of feedback game elements, the integration of the competitive leaderboard did lead to higher levels of emotional engagement for highly efficacious participants.

Although our main hypotheses were inconclusive, this thesis addressed the need for more theory-driven causal experiments in the gamification literature. Presenting the motivational underpinning of games through the perspective of SDT and human learning behaviours from SCT provided a means to interpret the results of the impact of a feedback game element on learning outcomes and engagement. In addition, this thesis contributed to literature regarding the effectiveness of gamification by limiting biased results from measuring effects of multiple game design elements, which has been flagged as an important problem in the gamification literature. This research is also, to our knowledge, amongst the first to examine the effectiveness of gamification in the context of teaching data science and analytics, thus providing a pathway for future experiments on gamification in this rapid growing field.

From a practical standpoint, this thesis brings attention to the potential learning and engagement benefits that gamification can bring as educators and trainers upskill students in fields of data science and analytics. Though direct relationships between gamification and learning were not found in this thesis, the significant moderating effects of the study highlights the importance of individual differences of learners when implementing game design elements. Educators and trainers need to be mindful of varying levels of expertise and difficulty of the content they expect to teach with the help of gamified experiences as these factors may influence the learning outcomes and engagement of the learner.

Limitations and Future Research

Considering the inconclusive results that emerged from this research project, it is necessary to discuss the limitations of the present study. Firstly, one key limitation, from a methodological standpoint, was the fact that the difference between a competitive leaderboard and a non-competitive leader may have been too small. In other words, since the leaderboard was present in both conditions and the only difference was the presence of competitors to include the social and competitive nature of leaderboards, the subtle difference may not have been enough to observe significant effects between groups.

Additionally, though our study was driven by motivational theories, we did not include a measure of intrinsic motivation such as the Intrinsic Motivation Inventory items (Ryan and Deci 2006) for participants playing the simulation game. Since the literature suggests that leaderboards are often extrinsically motivating in nature which in turn, may hinder learning (Mekler et al. 2017), a measure of motivation would offer insights to the role that intrinsic motivation plays in engagement and long-term learning. The participants in this study also played the simulation game against opponents they did not know. Perhaps if they would have been playing against individuals from their own class, subjects would have been more motivated to win and in turn, we would have observed larger differences in emotional engagement.

In the context of the long-term effect on learning, another limitation of this study was that it was short in length. A 90-minute experience does not reflect a fair environment for subjects to learn something completely new, especially something as complex as predictive modelling. Although the study only introduced basic concepts of predictive analytics, subjects were expected to learn new concepts, how to operate SAS Enterprise Miner, and solve a business problem within 60-minutes, all while doing this remotely. It is important to note that the simulation game used in the present study is typically used within a 6 weeklong university course, or within a shorter 3-day course. Therefore, having participants complete this experiment in such a short amount of time may have limited measures of learning performance and engagement. To that matter, the present study also did not measure the latter constructs over a period of time, thus, results are limited to short-term learning and engagement. This may explain why we did not observe any significant differences between conditions regarding learning and engagement.

Regarding future research, it would be interesting to replicate this between-subject experiment on the effects of leaderboards but remove the leaderboard entirely from the control condition. A future study could, for instance, measure learning improvement with a pretest and posttest assessment in addition to the emotional and behavioral engagement, between a condition that uses traditional learning methods such as a lecture or reading material and an experimental condition with a competitive leaderboard and fictitious opponents. Perhaps this type of experimental study would offer evidence of a causal relationship between the use of leaderboards and learning and engagement. Furthermore, this modification would have the potential to isolate the effect of leaderboards more accurately and offer insights to influences of competition and social comparison. The effect of relatedness as stated by SDT, could also be explored in future research by allowing participants to chat with one another while playing and view avatars and/or real-time video feed of competitors. Since there is literature suggesting that gamification may only be beneficial in the short term and since learning happens over time (Koivisto and Hamari 2014), future studies should also seek to change the study length of the present project. A longitudinal study over the course of a semester or full-day training would provide better evidence of long-term effects of leaderboards.

Future experiments could also manipulate the format of the training given to participants prior to playing the business simulation to observe effects of enactive and vicarious learning on learning gains and self-efficacy post game play. In the present study, all participants watched a video on the basic concepts of predictive modelling and brief overview of how to use SAS Enterprise Software and then participated in a 30-minute hands-on training on the statistical software package. Future studies can manipulate the order and or presence of the vicarious and enactive training prior to engaging with the simulation game to observe potential effects on learning and engagement.

Future studies may also consider adopting an experimental design that includes more than one condition with a game element to offer recommendations on which is most relevant to learning data science. Since current literature suggests that points, badges, and leaderboards are the most used game elements (Dicheva et al. 2015; Hamari et al. 2014), this type of manipulation would examine if leaderboards are in fact, the most effective method of gamification to increase learning and engagement. If future researchers have a desire to explore beyond the effects of "pontification" game elements such as points, badges, and leaderboards, it would be interesting to couple these game mechanisms with a dimension of adaptive learning to observe if learning gains would be higher as done by Khosravi, Sadiq, and Gasevic (2020). Specifically, adding an adaptive learning system that offers participants personalized feedback while trying to solve the business problem in Cortex in addition to the presence of a competitive leaderboard may offer practical insights on how to lower the barrier to learning complex data science concepts.

Finally, since learning data science can be complex and time consuming, future studies should consider measuring explicit and implicit measures of cognitive engagement using neurophysiological instruments as done in a study by Passalacqua et al. (2020) in the context of an ERP simulation game. Future studies measuring cognitive engagement may help understand which sections of learning data science need more attention and support from gamification.

Bibliography

Deci, E. L., and Ryan, R. M. 1980. "Self-Determination Theory: When Mind Mediates Behavior," Journal of Mind and Behavior (1:1), pp. 33-43.

Deci, E. L., and Ryan, R. M. 1991. "A Motivational Approach to Self: Integration in Personality," in Nebraska Symposium on Motivation, 1990: Perspectives on Motivation. Lincoln, NE, US: University of Nebraska Press, pp. 237-288.

Khosravi, H., Sadiq, S., and Gasevic, D. 2020. "Development and Adoption of an Adaptive Learning System: Reflections and Lessons Learned," *Proceedings of the 51st ACM Technical Symposium on Computer Science Education*, pp. 58-64.

Koivisto, J., and Hamari, J. 2014. "Demographic Differences in Perceived Benefits from Gamification," Computers in Human Behavior (35), pp. 179-188.

Mekler, E. D., Brühlmann, F., Tuch, A. N., and Opwis, K. 2017. "Towards Understanding the Effects of Individual Gamification Elements on Intrinsic Motivation and Performance," *Computers in Human Behavior* (71), pp. 525-534.

Landers, R. N., Bauer, K. N., and Callan, R. C. 2017. "Gamification of Task Performance with Leaderboards: A Goal Setting Experiment," Computers in Human Behavior (71), pp. 508-515.

Passalacqua, M., Léger, P.-M., Nacke, L. E., Fredette, M., Labonté-Lemoyne, É., Lin, X., Caprioli, T., and Sénécal, S. 2020. "Playing in the Backstore: Interface Gamification Increases Warehousing Workforce Engagement," *Industrial Management & Data Systems*).

Ryan, R. M., and Deci, E. L. 2006. "Intrinsic Motivation Inventory (Imi)," *Self-Determination Theory, An Approach to Human Motivation and Personality.< http://www. psych. rochester. edu/SDT/measures/IMI_description. php>.(accessed 03.29. 2021*).

Santhanam, R., Liu, D., and Shen, W.-C. M.. 2016. “Research Note—Gamification of Technology-Mediated Training: Not All Competitions Are the Same”, IS Research (27:2), pp. 453–465

Seaborn, K., and Fels, D. I. 2015. "Gamification in Theory and Action: A Survey," International Journal of Human-Computer Studies (74), pp. 14-31.

Appendix

INSTRUCTIONS INCLUDED WITH AN ANONYMOUS QUESTIONNAIRE

Evaluation of Data Analytics Training

The following pages contain three anonymous questionnaires, which we invite you to complete. These questionnaires have been developed as part of a Master's thesis at HEC.

Since your first impressions best reflect your true opinions, we would ask that you please answer the questions included in these questionnaires without any hesitation. There is no time limit for completing the questionnaires, although we have estimated that it should take about 10 minutes for each.

The information collected will be anonymous and will remain strictly confidential. It will be used solely for the advancement of knowledge and the dissemination of the overall results in academic or professional forums.

The online data collection provider agrees to refrain from disclosing any personal information (or any other information concerning participants in this study) to any other users or to any third party, unless the respondent expressly agrees to such disclosure or unless such disclosure is required by law.

You are free to refuse to participate in this project and you may decide to stop answering the questions at any time. By completing this questionnaire, you will be considered as having given your consent to participate in our research project and to the potential use of data collected from this questionnaire in future research.

If you have any questions about this research, please contact the principal investigator, Melissa Thériault, by email at the address indicated below.

HEC Montréal's Research Ethics Board has determined that the data collection related to this study meets the ethics standards for research involving humans. If you have any questions related to ethics, please contact the REB secretariat at (514) 340-6051 or by email at cer@hec.ca.

Thank you for your valuable cooperation!

Melissa Thériault
HEC Montréal
melissa.theriault@hec.ca

Prof. Pierre-Majorique Léger, Ph.D.
HEC Montréal
pml@hec.ca

Prof. Jean-François Plante, Ph.D.
HEC Montréal
jfplante@hec.ca

Pre and Post-Test Questionnaire

Questions have been adapted from MOOCs (massive open online course) found on Coursera related to Data Science, Data Analytics, and Machine Learning. Namely, courses from the University of Washington, Stanford University, and University of Colorado Boulder. All questions are true or false, and followed by questions to assess the level of confidence for a given answer.

Select True or False for the following statements:

1. Multiple linear regression should be used when we would like to predict impacts of changes in predictor variables on a target variable.
2. A Decision Tree is an approach that classifies subjects into known groups.
3. In predictive modeling, the validation set is used to evaluate the various models.
4. More complex models tend to overfit more.
5. The value of the target variable is known when using supervised learning models.
6. We generally prefer models with smaller root mean square error (RMSE).
7. Partitioning the data into three sets, training, validation, and test, can help avoid the issue of overfitting the model.
8. Predicting the most profitable customer is an example of an unsupervised learning task.
9. If the predictor variable affects the target variable in a way that is not independent of all other predictor variables, interaction terms can be added to the model to improve the fit.
10. In predictive modeling, the training set is used to fit the models.

How confident are you in your answer to the True or False statement?

(1 = Very confidence, 2 = Confident, 3 = Somewhat confident, 4 = Neutral, 5 = Somewhat unconfident, 6 = Unconfident, 7 = Very unconfident)

Software Self-Efficacy Questionnaire

The following questions asks you to indicate whether you think you could use SAS Enterprise Miner under a variety of conditions. For each of the conditions, please indicate whether you think you would be able to complete the job using a statistical software.

Then, for each condition that you answered "yes", please rate your confidence about your first judgement, by circling a number from 1 to 10, where 1 indicates "Not at all confident," 5 indicates "Moderately confident," and 10 indicates "Totally confident."

If you answered "no" to any of the conditions, do not rate your confidence level.

I could complete the job of making predictions with statistical models using SAS Enterprise Miner...

	Answer		Confidence									
	Yes	No	1	2	3	4	5	6	7	8	9	10
...if there was no one around to tell me what to do as I go.	○	○	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
...if I had only the software manuals for reference.	○	○	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
...if I had seen someone else using it before trying it myself.	○	○	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
...if I could call someone for help if I got stuck.	○	○	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
...if someone else had helped me get started.	○	○	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
...if I had a lot of time to complete the job for which the software was provided.	○	○	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
...if I had just the built-in help facility for assistance.	○	○	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
...if someone showed me how to do it first.	○	○	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
	Yes	No	1	2	3	4	5	6	7	8	9	10

Self-Reported Predictive Knowledge Questionnaire

To what extent do you agree or disagree with the following statements:

(1 = Strongly disagree, 2 = Disagree, 3 = Neither agree or disagree, 4 = Agree, 5 = Strongly agree)

1. I have mastered predictive modelling tasks
2. I do not perform predictive modelling tasks as well as I would like
3. I am certain I can perform predictive modelling tasks at the level I would like
4. I think my performance in solving predictive modelling tasks could be improved substantially