Journal of Computing Theories and Applications ISSN:3024-9104

Research Article

Learning through Simulation: A Role-Based Learning Exercise to Equip Students in Big Data Requirement Elicitation Skills and Challenges

PPG Dinesh Asanka * and Thilini V. Mahanama

Department of Industrial Management, University of Kelaniya, Sri Lanka; email: dasanka@kln.ac.lk; thilinim@kln.ac.lk
* Corresponding Author: PPG Dinesh Asanka

Abstract: Bridging the gap between theoretical knowledge and practical application remains challenging in big data education. Our paper proposes a platform where students actively participate in a role-based learning (RBL) exercise designed to simulate real-world big data projects. This RBL exercise assigns each student participant to a business analyst role tasked with eliciting requirements or a customer role that provides those big data requirements. We conducted RBL sessions for postgraduate and undergraduate students learning big data analytics/data warehousing. Through role-playing in a collaborative environment, students face challenges such as unclear objectives, data privacy concerns, and evolving requirements. We evaluated the effectiveness of RBL simulations through brainstorming sessions, which verified that the students achieved the learning outcomes through RBL exercises. We also collected students' feedback through a survey and found the RBL experience helped us understand the big data requirement elicitation skills, highlighting the significance of communication and collaboration skills. Further, we have evaluated students through a final exam question, and identified that students who participated in the RBL exercise outperformed in the big data requirement elicitation question. In summary, our research demonstrates that this RBL approach offers a valuable learning experience by enabling students to directly experience the complexities of big data requirement elicitation and identify the future requirements or challenges that encourage them to acquire the required skills.

Keywords: Big Data Analytics; Collaborative Learning; Learning and Teaching with Simulation; Requirement Elicitation; Role-Based Learning.

Received: May, 22nd 2024
Revised: July, 23rd 2024
Accepted: July, 26th 2024
Published: July, 28th 2024

1. Introduction

Eliciting requirements for big data projects presents various challenges. Firstly, the lack of tailored practices and tools for big data-related requirements hinders efficient elicitation [1]. Additionally, the dynamic nature of digital data sources such as IoT and social networks complicates traditional stakeholder-driven requirements[2]. The volume, velocity, veracity, and variety of data further intensify the complexity, making it difficult to manage effectively [3]. Lastly, the ambiguity in treating big data characteristics in software development intensifies the challenge, emphasizing the need for a clear requirement elicitation artifact model for big data applications[4].

Non-practitioners find it challenging to grasp requirement elicitation in big data projects due to its complexity and the lack of established practices. This leads to considering the unique characteristics of big data during the requirements engineering (RE) process. However, the field of RE lacks specific methods tailored for big data projects, making it difficult for students (as non-practitioners) to understand how to handle the volume, variety, velocity, veracity, and complexity of big data[3], [5]. Additionally, the lack of clarity in the literature regarding how to treat big data characteristics in the RE process further complicates understanding requirement elicitation in big data projects[6]. The absence of tailored RE practices, tools, and frameworks for big data projects contributes to the challenges students face in comprehending requirement elicitation in this context [1].

Role-playing for big data requirements elicitation allows students to simulate a practitioner's role. Therefore, assigning role-playing assessments can significantly help students understand the challenges in requirement gathering by providing a practical and immersive approach [7]–[10]. As role players in eliciting requirements, students are supposed to conduct interviews and validate ideas[11]. Role-playing techniques offer a dynamic and engaging way to prepare students for the multifaceted nature of requirement elicitation processes, such as the complexities of interacting with end-users from diverse cultural backgrounds, enhancing their ability to navigate such challenges in real-world scenarios. Moreover, this learning helps students gather software development and research project requirements. Through role-playing, students develop essential skills such as effective communication, teamwork, and problem-solving, which are crucial in requirement elicitation processes.

This research proposes utilizing role-based learning (RBL) exercises to equip students with big data requirements elicitation skills by understanding the potential challenges. To achieve this aim, this research intends to accomplish the following research objectives:

1. Assess students' understanding of challenges associated with requirements elicitation in big data projects through participation in an RBL simulation.
2. Investigate the impact of an RBL simulation on student communication and collaboration during requirements elicitation in a simulated big data project environment.
3. Evaluate the effectiveness of an RBL simulation in enhancing student learning of big data requirements elicitation skills.

This paper provides a literature review to discuss the available studies on requirement elicitation in section 2 to identify a research gap. Section 3 designs a framework for RBL exercises to accomplish our research objectives. Incorporating students' feedback on RBL sessions, the paper presents the challenges of big data requirement elicitation and recommendations to overcome them in section 4. This results and discussion section validates the effectiveness of an RBL simulation in enhancing student learning of big data requirements elicitation skills. Section 5 presents the concluding remarks on utilizing RBL sessions to teach the big data requirement elicitation process.

2. Literature Review

Effective software requirement elicitation is a crucial but challenging step in the development process [12]. While existing research explores successful elicitation practices in a global development context, big data projects present unique challenges due to the vast and dynamic nature of the data involved [13]. This section discusses the studies on requirement elicitation to justify the research gap we address in our study.

Software requirement elicitation is challenging for students to learn due to various factors. Factors include conflicting requirements, unspoken or assumed requirements, difficulty meeting relevant stakeholders, stakeholders' resistance to change, and time constraints for stakeholder meetings [12]. A paper [13] proposes an empirical study on requirement elicitation for a global software development paradigm that utilizes a two-pronged methodology to address the challenges of requirement elicitation in global software development projects. First, a systematic literature review gathers potential success factors, laying the groundwork for effective elicitation practices. Second, a survey was conducted to investigate the success factors for requirement elicitation, taking the employees at software development organizations in ten countries who worked in the requirement elicitation phase as business analysts, software engineers, software quality assurance engineers, system analysts, software developers, and business consultants. This data is analyzed considering factors such as the type and experience of experts involved, company size, and client versus vendor perspectives. This multifaceted approach helps identify a robust set of success factors and assess their importance from various viewpoints within the global software development domain. However, the study acknowledges that challenges such as geographical distances, limited face-to-face interaction, and cultural barriers inherent to global projects can make effective requirement elicitation difficult.

Software requirement elicitation poses a unique challenge in big data analysis compared to standard software development projects due to the complexities associated with handling vast amounts of dynamic data sources. The key differences between big data and operational systems are listed in Table 1 [14]. According to Table 1, there are significant differences in both systems at the requirement elicitation phase. Therefore, a special focus is required in the big data elicitation phase.

Table 1. Comparing big data and operational systems based on focus, data types, processing speed, and data accuracy

Feature	Big Data Systems	Operational Systems
Focus	Large, complex, and diverse data sets	Specific tasks and processes
Data types	Structured, semi-structured, unstructured	Primarily structured data
Processing speed	May vary depending on data size and complexity	Optimized for speed and reliability
Data accuracy	Data quality is a challenge to manage	Ensuring data integrity is critical

In big data systems development, specific tailored practices and tools for elicitation, specification, and prioritization of requirements are lacking, leading to difficulties in identifying and documenting big data-related requirements. The elicitation process becomes more complex due to the big data characteristics and technologies significantly influencing the quality requirements and system architecture [2]. This paper highlights that software requirement elicitation in big data analysis is challenging due to extracting relevant requirements from vast, dynamic, and unintended digital sources, unlike traditional stakeholder-driven approaches in standard projects. While "The Data Warehouse Toolkit" [15] doesn't explicitly list challenges as a separate section; it highlights various difficulties throughout the book by showcasing real-world case studies. Some key data warehouse requirement challenges the book addresses are described below:

- Understanding business needs: Capturing stakeholders' true needs and translating them into clear data warehouse requirements can be difficult. A study [15] emphasizes the importance of iterative communication and collaboration between business users and data warehouse developers.
- Data integration complexity: Integrating data from disparate sources with varying formats and structures can be a major challenge. Strategies for data cleansing; transformation; and extract, transform, load (ETL) processes to address these issues are discussed in [15].
- Defining the grain and dimensions: Determining the appropriate level of detail (grain) for data and identifying the relevant dimensions for analysis can be complex. Addressing these guides on selecting the grain based on business needs and designing effective dimensional models.
- Timely data availability: Ensuring data is readily available for analysis requires careful consideration of data refresh cycles and data warehousing architecture. A study [15] explores different architectures, such as data marts and data warehouses, to optimize data access.
- Managing slowly changing dimensions: Data in dimensions can change over time, so strategies are required to handle these changes. A study [15] discusses various approaches to slowly changing dimensions to maintain data integrity.

By presenting these challenges through case studies in various domains such as retail sales, inventory, procurement, order management, customer relation management, accounting, human resources management, financial services, telecommunication, transportation, education health care, electronic commerce, insurance, [15] equips readers with the knowledge and techniques to overcome them during the data warehouse requirement gathering process.

"DW 2.0: The Architecture for the Next Generation of Data Warehousing" by Inmon [16] discusses and addresses several key challenges regarding data warehouse requirements:

- Data volume, velocity, and variety: Traditional data warehouse architectures might be overwhelmed by the influx of data. The difficulty of handling the exponential growth of data volume, the ever-increasing speed of data arrival (velocity), and the inclusion of unstructured and semi-structured data varieties are discussed in [16].

- Agile business needs: A study [16] acknowledges the challenge of adapting data warehouses to rapidly changing business requirements. Traditional top-down approaches to data warehousing might not be flexible enough to accommodate these evolving needs.
- Incorporating new data sources: DW 2.0 highlights the challenge of integrating data from new sources such as social media, sensor networks, and machine-generated data. Traditional data warehouses might not be equipped to handle these diverse data types.
- Information governance and security: With the inclusion of potentially sensitive data and the vast amount of information stored, DW 2.0 emphasizes the challenge of ensuring robust information governance and security practices. Traditional approaches might need to be significantly strengthened to address these concerns.
- User adoption and skillsets: Traditional data warehouses might require additional training or user support for widespread adoption within the organization. DW 2.0 recognizes the challenge of enabling effective data analysis through user-friendly interfaces and skillsets. By outlining these challenges, [16] argues for the need for a new data warehouse architecture, DW 2.0, which aims to address these issues and create a data warehouse solution that can handle future demands.

Group discussions and hands-on practice have higher student retention rates in learning. Clearly, RBL falls into that category. Utilizing RBL for the standard lecture model, which focuses on learner-centric teaching, is discussed in [17]. Therefore, RBL is a great tool for teaching challenging concepts such as big data requirement elicitation. Role play has been widely explored as a teaching and learning strategy in higher education, focusing on specific applications and outcomes. Both [18] and [19] highlight the potential of role-play in developing practical skills and understanding complex concepts. This is further supported by [20], which discusses that role-play can enhance motivation, creativity, and collaboration. The use of technology in role-play, as demonstrated in [21], can also significantly impact engagement and learning outcomes. However, implementing role-play in higher education is not without challenges, as noted in [22] and [23]. Despite these challenges, role-play remains a valuable tool in education, particularly in promoting active and experiential learning [24].

RBL techniques can effectively be utilized in requirement elicitation by providing students with practical experience in conducting interviews and consultations. A paper [25] conducted a RBL research incorporating 43 graduate students majoring in software engineering at Kennesaw State University, GA, USA. A controlled quasi-experiment with an experimental group and a control group is used to evaluate the learning effect of RBL. The research concludes that playing the reverse role in interviews is beneficial for understanding customer perceptions and learning correct analyst behaviour. Additionally, the research emphasizes the contribution of RBL in improving interviewing skills in requirements engineering education and training.

A study [26] evaluates the use of role-playing simulations in a negotiation course, which consisted primarily of simulations involving the alternative dispute resolution processes of negotiation, facilitation, and mediation. RBL was chosen because business elicitation is integral to negotiation skills. Their research collects feedback through two sets of questionnaires from 41 graduate students before and after taking the negotiation course. This research concludes that role-playing simulations effectively help students learn negotiation skills and prepare them for real conflicts.

In our research, we examine the effectiveness of utilizing the RBL technique to help students learn (realize) the challenges in eliciting big data projects. The next section describes performing big data requirements elicitation by designing an RBL framework.

3. Designing RBL Exercise Sessions for Big Data Requirement Elicitation

This section designs RBL exercises and investigates their effectiveness in teaching big data requirements elicitation skills. On successful completion of this RBL exercise, the participants should be able to achieve the following intended learning outcomes (ILO):

1. Differentiate between a data warehouse and an operational database, articulating their purposes, structures, and uses within an organization (ILO 1: Understand the Distinctions)

2. Effectively gather and document requirements from clients and stakeholders, identifying both functional and non-functional needs related to data warehousing (ILO 2: Elicit and Document Requirements)
3. Employ effective communication strategies to bridge the gap between technical and business stakeholders, ensuring a shared understanding of project goals and requirements (ILO 3: Navigate Communication Challenges)
4. Analyze data characteristics (volume, velocity, variety, veracity, value) to determine the scope and complexity of a data warehousing project (ILO 4: Assess Data Needs)
5. Identify potential issues related to data quality and availability, and propose strategies to mitigate these risks (ILO 5: Address Data Quality Concerns)
6. Develop a basic understanding of data warehousing technologies, tools, and methodologies, including the ability to evaluate their suitability for a given project (ILO 6: Consider Technical Aspects)
7. Recognize the importance of continuous learning and knowledge sharing between clients, consultants, and other stakeholders in a data warehousing project (ILO 7: Bridge Knowledge Gaps)

Due to the potential challenges of random assignment in classroom settings, a quasi-experimental design is employed. Students are supposed to take a pre-test to assess their baseline knowledge before participating in the RBL simulation. To encounter common challenges in eliciting big data project requirements, students should role-play as either customers or business analysts. We design post-tests and surveys to assess students' learning outcomes and perspectives on the RBL experience. This design allows researchers to explore the potential of implementing RBL simulations in developing communication, collaboration, and problem-solving skills necessary for successful big data requirements elicitation.

Level three undergraduates and postgraduate students are chosen for this study as they have already developed a system for a client as a partial fulfillment of their degree program. Most of these students have served as interns and gained exposure to client projects and procedures. The students were given a basic understanding of big data fundamentals and concepts, such as big data challenges (volume, velocity, verity, veracity, and value), types of analytics (descriptive, diagnostic, predictive, prescriptive) and analytics sources (internet or things [27], internet of people, internet of events, internet of places)[16], [28], [29]. After this initial discussion, students are divided into the following two groups to act as role players in big data requirements elicitation:

1. Customer group: Represents stakeholders with a business need (for big data analysis). They are supposed to develop a business scenario, as well as the data requirements and desired project outcomes.
2. Business analyst group: Acts as the intermediary between the customer and technical team. Gathers requirements from the customer to define them as technical specifications.

Customer and business analyst groups (role players) meet to discuss project requirements. This simulation incorporates common big data requirements elicitation challenges (e.g., unclear technical needs, data limitations, evolving requirements). Once the role-playing activity ends, a test will be performed to assess students' knowledge gained on big data requirements elicitation. Participants' feedback on their experience related to their assigned role (customer or business analyst) is gathered through brainstorming sessions and as reports.

This research framework is applied in three RBL exercise sessions (Session 1, 2, and 3) carried out with different student groups. Postgraduate students who were reading a data analytics degree participated in session 1. Two human resource management professionals played customers' roles, while two professional business analysts played the business analysts' role. The entire class, including the lecturer, observed this role-playing session and provided their feedback.

Level 3 undergraduates in an information technology degree program in 2022 and 2023 participated in sessions 2 and 3. These students have some level of understanding of the industry as they have interned for six months. The students were grouped into 5-6 member teams, including at least one student per team whose ambition is to become a business analyst in the future. Every team was asked to choose a project where they had some level of experience. After the session, teams submitted a report describing their experience in role-playing, including the challenges.

Other than evaluating the students' performance on the in-class RBL exercise, we further assessed their understanding by assigning a case-based question for their end of the semester exam. In this open book exam, analyzing the students' performance for this question helps us to compare the performance of the RBL participants and non-participants' understanding. We design the following case-based big data requirement elicitation problem to address the ILOs specified below:

PharmaData, a multinational pharmaceutical company, is facing challenges in integrating and analyzing data from its various research, clinical trials, and manufacturing systems. They have decided to invest in a data warehouse to consolidate this information and gain valuable insights for decision-making. You have been brought in as a consultant to assist with the initial phases of this project.

- Stakeholders: The project involves a wide range of stakeholders, including scientists, researchers, clinical trial managers, IT professionals, and senior executives.
- Data Sources: Data comes from various sources, including laboratory information management systems (LIMS), electronic data capture (EDC) systems for clinical trials, manufacturing execution systems (MES), and enterprise resource planning (ERP) systems.
- Regulatory Requirements: The data warehouse must comply with strict regulatory requirements, such as FDA regulations for data integrity and patient privacy.

Demonstrate your ability to address the following aspects of the PharmaData data warehouse project. Address each of the below points in detail, demonstrating your understanding of the learning outcomes (in bold letters) and your ability to apply them to a real-world scenario. Use specific examples and evidence from the case study to support your responses.

1. **Same as Traditional Requirement Elicitation:** Explain why the above case can be considered as similar to requirement gathering in traditional software development system.
2. **Understanding the Distinctions:** Explain to the project stakeholders why a data warehouse is a better solution for their needs than simply enhancing their existing operational databases.
3. **Eliciting and Documenting Requirements:** Develop a plan for gathering and documenting requirements from the diverse group of stakeholders, ensuring that both technical and business needs are captured.
4. **Navigating Communication Challenges:** Propose strategies for effective communication and collaboration among the different stakeholder groups, addressing potential barriers like technical jargon and differing priorities.
5. **Assessing Data Needs:** Analyze the variety, volume, velocity, and veracity of PharmaData's data sources to determine the appropriate architecture and technology for the data warehouse.
6. **Addressing Data Quality Concerns:** Identify potential data quality issues that may arise during data integration and propose solutions to ensure the accuracy and reliability of the data in the warehouse.
7. **Considering Technical Aspects:** Recommend suitable data warehousing technologies and tools, considering factors like performance, and regulatory compliance.
8. **Bridging Knowledge Gaps:** Develop a plan to facilitate knowledge sharing between technical and non-technical stakeholders throughout the project, ensuring everyone understands the project goals and their role in its success.

Questions 1 and 2 are designed based on ILO 1. Questions 3-8 address ILOs 2-7 respectively. We designed the rubric available in Table 2 to evaluate the students' performance in achieving the ILOs.

Among the students registered for the Big Data and Data Warehouse final exam, 29 exam candidates answered this big data requirement elicitation question. Among these 29 candidates, only 14 students participated in the in-class RBL exercise while 15 others were absent. We compare the performances of these two groups under the results and discussion section.

Table 2. Designing a rubric to evaluate students' performance in achieving the ILOs

ILO	Performance Level		
	Exemplary (7-10)	**Developing (4-6)**	**Needs Improvement (1-3)**
ILO 1	Clearly articulates the differences between data warehouses and operational databases, using relevant examples and explaining their impact on project requirements.	Shows a basic understanding of the differences but may struggle to apply them to the project context.	Has difficulty distinguishing between the two or understanding their relevance to the project.
ILO 2	Systematically gathers and documents requirements using a well-structured approach, ensuring clarity, completeness, and traceability.	Gathers some requirements but may miss key details or struggle with clear documentation.	Has difficulty gathering and documenting requirements systematically.
ILO 3	Proactively identifies and addresses communication barriers, effectively translating technical concepts into business terms and ensuring a shared understanding among stakeholders.	Attempts to communicate with stakeholders but may struggle to bridge gaps in understanding or terminology.	Has difficulty communicating effectively with stakeholders, leading to misunderstandings or misalignment.
ILO 4	Thoroughly analyzes data characteristics (5 Vs) to determine the scope and complexity of the project, providing clear justifications for design decisions.	Analyzes some data characteristics but may miss key aspects or struggle to connect them to project design.	Has difficulty analyzing data characteristics or understanding their impact on project design.
ILO 5	Proactively identifies potential data quality issues and proposes detailed strategies for mitigation, including data cleansing, validation, and enrichment techniques.	Recognizes the importance of data quality but struggles to identify specific issues or propose effective solutions.	Does not demonstrate an understanding of data quality concerns or mitigation strategies.
ILO 6	Demonstrates a solid understanding of relevant data warehousing technologies, tools, and methodologies, and can justify their selection based on project requirements.	Shows limited knowledge of data warehousing technologies and methodologies, relying on others for guidance.	Lacks knowledge of data warehousing technologies and methodologies, unable to participate in technical discussions or decisions.
ILO 7	Actively facilitates knowledge transfer between technical and business stakeholders, ensuring a shared understanding of key concepts and requirements.	Recognizes the importance of knowledge sharing but does not actively participate in facilitating it.	Does not demonstrate an understanding of the importance of knowledge sharing or bridging knowledge gaps between stakeholders.

4. Results and Discussion

This section evaluates the effectiveness of an RBL simulation in enhancing student learning of big data requirements elicitation skills by assessing students' understanding of challenges associated with requirements elicitation in big data projects through RBL simulation. Then, the skills required to overcome key challenges in big data requirement elicitation are provided. The feedback received after RBL sessions 2 and 3 are categorized into multiple categories, as shown in Fig 1.

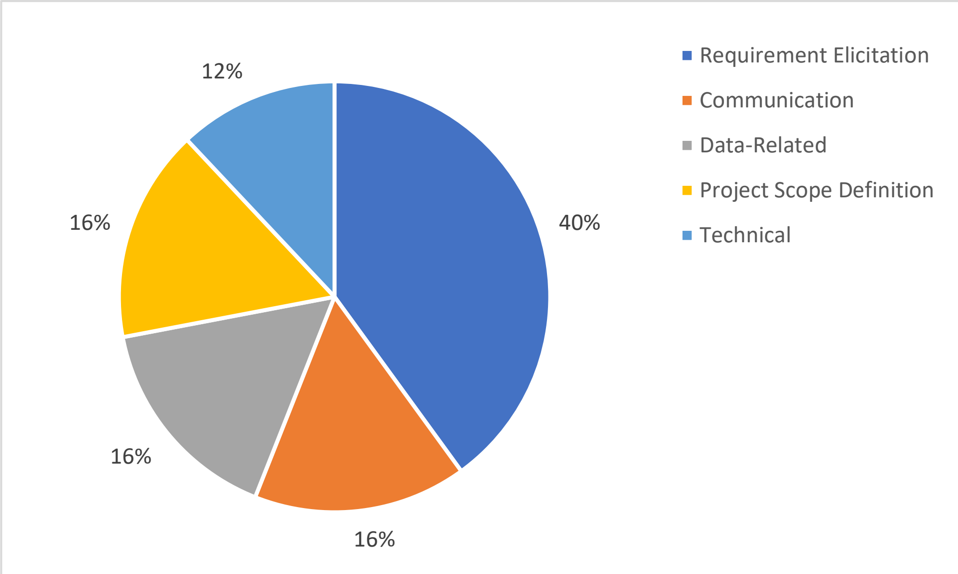

Figure 1. Distribution of Challenges in Big Data Elicitation Across Categories

According to Fig. 1, 40% of the issues fall into the requirement elicitation category as per the feedback of the session participants. This indicates that students have identified the requirements elicitation phase challenges in a big data project.

4.1 Challenges on Big Data Requirements Elicitation Identified through RBL Sessions

During the first role-playing exercise (session 1), it was noticed that business users do not have a clear idea about their analytical requirements and are always focused on the operational requirements that are familiar to them. Further, they were challenged by identifying different sources, as big data needs multiple internal and external sources. From the business analysts' perspectives, they were challenged by the ad-hoc nature of the analysis.

Students identified several key challenges faced during the requirement elicitation process, which are categorized into three key areas as follows**:**

4.1.1 Technical knowledge and understanding:

- Lack of knowledge about big data projects and their applicability in different scenarios.
- Difficulty in understanding stakeholder needs and priorities.
- Defining the scope of the big data project and prioritizing requirements amidst diverse perspectives and evolving needs.
- Ensuring alignment of requirements with organizational objectives for improved outcomes.
- Addressing data privacy regulations and compliance during the process.

4.1.2 Communication and collaboration:

- Difficulty in understanding customers' expectations and convincing them about the value of big data solutions.
- Identifying whether the customer truly needs big data capabilities.
- Managing privacy concerns when collecting data from customers.

4.1.3 Data and design challenges:

- Identifying future customer requirements and potential changes.
- Defining ad-hoc criteria for designing big data.
- Utilizing various data sources effectively.
- Understanding data complexity, volume, and scope of the warehouse.

- Lack of domain knowledge and environmental data.

In addition to the aforementioned challenges, RBL participants find it challenging to perform the following:

- Understanding the business process and its data needs.
- Effective communication with stakeholders.
- Managing technical and business knowledge during the process.
- Ensuring clarity and completeness of requirements for each aspect.

In summary, students realized the need for better communication, technical knowledge, and understanding of customer needs to ensure successful requirement elicitation in big data projects. The following section discusses how to overcome the challenges presented in section 4.1.

4.2 Recommending Helpful Skills to Overcome the Challenges of Big Data Requirements Elicitation

This section addresses common challenges (listed below) identified in RBL sessions:

- Customer's limited understanding of big data projects' benefits and potential.
- Lack of clarity in customer requirements and feasibility.
- Difficulty in distinguishing big data needs from operational database needs.
- Communication gaps and differing perspectives among stakeholders.
- Difficulty in identifying the specific data required.

Our research suggests some recommendations to overcome these challenges under the following subsections:

4.2.1 Understanding customers' requirements through facilitating technical knowledge:

Experienced business analysts who can understand customers' requirements can translate them into technical requirements and identify potential challenges early on. Customers' understanding of how big data projects will benefit their project requirements is limited. This could be addressed by filling their knowledge gap through workshops or training sessions on big data concepts, capabilities, and limitations. This will help customers identify their project requirements and effectively communicate them to business analysts to accomplish their specific project goals.

4.2.2 Defining project scope and prioritize tasks:

Defining the scope of the big data project and prioritizing requirements amidst diverse perspectives and evolving needs. A realization plan helps to clearly articulate the expected benefits of the big data project in terms of improved decision-making, operational efficiency, and cost savings. Further, implementing a phased approach and following suitable prioritizing techniques will help address these issues.

- Phased approach: Implement the big data project in phases, starting with core requirements and gradually incorporating additional functionalities based on stakeholder feedback and changing priorities.
- Prioritization techniques: Utilize prioritization frameworks such as MuSCoW (Must-have, Should-have, Could-have, Won't- have) [30] to categorize requirements based on criticality and feasibility.

It is essential to continuously track whether the project requirements align with organizational objectives to deliver better outcomes. Securing senior management's support and involvement will help ensure alignment between the big data project and the organization's strategic goals.

4.2.3 Enhancing the communication and collaboration:

Communication and collaboration between customers and business analysts can be improved by practicing active listening, documentation, and regular meeting scheduling.

- Active listening: Encourage active listening skills among all stakeholders to ensure a clear understanding of needs and expectations.

- Documentation: Maintain detailed documentation of requirements discussions, decisions, and any changes made throughout the process.
- Regular meetings: Schedule regular meetings with stakeholders to discuss progress, address concerns, and ensure alignment.

4.2.4 Ensuring data privacy and regulatory compliance:

Effective communication with stakeholders helps to understand the business process and its data needs. This will help to collect only the data necessary for the intended purpose of big data analytics, minimizing the risk of privacy violations.

The data requirement elicitation process utilizes various data sources. Consulting legal and compliance professionals helps to ensure adherence to relevant data privacy regulations throughout the requirements elicitation and development process.

4.2.5 Additional recommendations:

Big data projects hold immense potential, but their complexity can lead to challenges. The following recommendations will help to ensure the accomplishment of the project goals:

- Prototyping: Develop prototypes or mockups of the big data projects to provide stakeholders with a tangible representation of the system and gather feedback on requirements.
- Data quality assessment: Conduct a thorough data quality assessment to identify potential issues that might impact the success of the big data project.
- Domain expertise: Involve domain experts in the requirements elicitation process to provide insights into specific data needs and challenges within the organization's context.

Addressing these challenges and implementing the recommended strategies help to improve the requirements elicitation process for big data projects, leading to a more successful and impactful data-driven solution.

4.3 Evaluating the Effectiveness of Using an RBL Exercise to Teach Big Data Requirements Elicitation

An analysis of exam scores for the requirement elicitation questions, in conjunction with their participation in the RBL exercise, provides several notable findings. The average score for the question was 10.14, with a mode of 11 achieved by eight students. While the distribution of scores was somewhat dispersed, a majority clustered within the 10-13 range. Notably, students who participated in the RBL exercise achieved a higher average score of 13.29, compared to 8.13 for those who were absent, see Fig. 2. Furthermore, all students who participated in the RBL exercise attained a minimum score of 10, which aligns with the overall average.

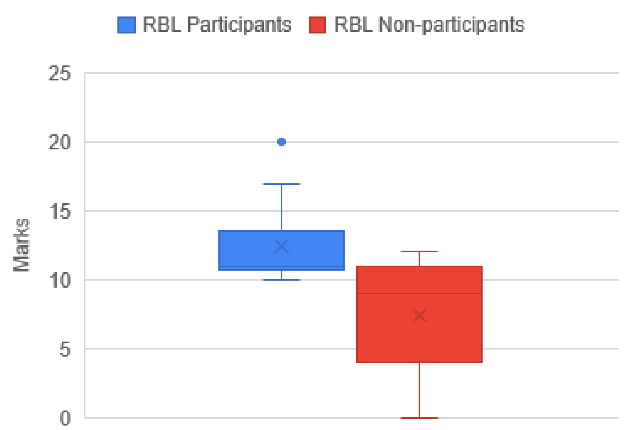

Figure 2. Comparing the performance for the big data requirement elicitation question of RBL participants and non-participants

According to Fig. 2, the RBL participants seem to have outperformed the big data requirement elicitation question at the final examination. We test the following hypothesis to evaluate further the effectiveness of using an RBL exercise to teach big data requirements elicitation which compares the students' performance for the question:

- H_0: The performance of the big data requirement elicitation problem is not higher in RBL exercise participants than in non-participants.
- H_1: The performance of the big data requirement elicitation problem is higher in RBL exercise participants than in non-participants.

We performed the Mann-Whitney U Test to test the hypothesis since the marks data do not follow a Gaussian distribution. Since the test rejects the null hypothesis (with p-value = 0.0009), we conclude that the performance for the big data requirement elicitation problem is higher in RBL exercise participants compared to non-participants at a 5% significance level. This highlights the significance of the active learning the RBL exercise participants experienced during the in-class RBL sessions.

5. Conclusion

Rather than teaching big data requirement elicitation challenges, letting students experience them enhances their understanding. Our research proposed utilizing an RBL platform that helps students realize challenges in eliciting big data requirements. Actively participating in RBL exercises helps students understand the skills required for the big data requirement elicitation.

We assigned RBL exercises on big data requirement elicitation for postgraduate students and level three undergraduates after facilitating them with a basic understanding of big data fundamentals and concepts. Then, the students were supposed to play the roles of business analysts to collect big data requirements from the students who played as customers. These RBL sessions provided a rewarding learning platform through a simulated big data project environment.

By role-playing as customers, students directly experience difficulty clearly communicating data requirements and project goals. The need to actively listen, negotiate, and translate needs into technical specifications during the RBL exercise directly addresses the communication and collaboration challenges often encountered in requirements elicitation. Therefore, RBL exercises can foster empathy and improve students' ability to communicate and collaborate effectively with stakeholders in real-world scenarios.

RBL simulation allows students to engage with diverse perspectives on what data is essential, simulating the challenges of prioritizing requirements in a collaborative environment. Students face common challenges due to unclear objectives, data privacy concerns, and continuously changing requirements within the simulation, providing valuable practice in navigating these complexities. Since role-playing is an active learning strategy that engages students in learning, students will retain the information and develop the necessary skills.

By implementing the RBL simulation and brainstorming participants' feedback, our research further evaluated its effectiveness and success in improving student learning outcomes related to requirements elicitation for big data analytical projects.

The post-simulation survey allowed students to reflect on their experiences and identify areas for improvement, further solidifying their learning. The RBL participants strongly suggested that the RBL technique can be successful in addressing the following goals in big data requirements:

- Providing a deeper understanding of the challenges in this process.
- Developing communication and collaboration skills essential for effective requirements elicitation.
- Equipping students with the necessary skills to navigate real-world big data project requirements.

We validated the effectiveness of utilizing RBL simulations for teaching big data requirement elicitation by evaluating students' performance for a follow-up final examination. The performance of the big data requirement elicitation problem in the final exam was higher in RBL exercise participants than in non-participants at a 5% significance level. This study will be extended to different groups of students to identify the effectiveness of using RBL simulations for different contexts. In particular, we will test this for some postgraduate students with more professional experience to compare their performance with that of undergraduates.

In summary, RBL participants' feedback and performance validated the effectiveness of an RBL simulation in enhancing their learning of big data requirements elicitation skills. Since RBL simulation reflects real-world requirements elicitation scenarios, it provides students with practical experience that can be directly applied to future projects. This framework could be utilized for scenario design, role allocation, and assessment methods, facilitating the wider applications of RBL in big data education.

Author Contributions: Conceptualization: Dinesh Asanka. and Thilini Mahanama; methodology, Dinesh Asanka.; validation: Dinesh Asanka, Thilini Mahanama.; formal analysis: Dinesh Asanka, Thilini Mahanama; data curation: Dinesh Asanka.; writing—original draft preparation: Dinesh Asanka.; writing—review and editing: Thilini Mahanama.; project administration: Dinesh Asanka;

Funding: This research received no external funding

Data Availability Statement: In session 1, participants' feedback was collected verbally. Session 2 and Session 3 participants' feedback was collected and shared in the following location. https://docs.google.com/document/d/1uun7OQN4TyMSX_BMj80r3ylF8rEVkRbD7r1h-krOZvU

Acknowledgements: The authors acknowledge the profound contribution of the RBL session attendees for their active participation and for providing their valuable feedback to improve the assessment.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] D. Arruda and R. Laigner, "Requirements Engineering Practices and Challenges in the Context of Big Data Software Development Projects: Early Insights from a Case Study," in *2020 IEEE International Conference on Big Data (Big Data)*, Dec. 2020, pp. 2012–2019. doi: 10.1109/BigData50022.2020.9377734.

[2] S. Lim, A. Henriksson, and J. Zdravkovic, "Data-Driven Requirements Elicitation: A Systematic Literature Review," *SN Comput. Sci.*, vol. 2, no. 1, p. 16, Feb. 2021, doi: 10.1007/s42979-020-00416-4.

[3] A. Hesham, O. E. Emam, and M. Salah, "Enhanced Framework for Big Data Requirement Elicitation," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 8, 2021, doi: 10.14569/IJACSA.2021.0120816.

[4] D. Arruda, "Requirements Engineering in the Context of Big Data Applications," *ACM SIGSOFT Softw. Eng. Notes*, vol. 43, no. 1, pp. 1–6, Mar. 2018, doi: 10.1145/3178315.3178323.

[5] D. Arruda, N. Madhavji, and I. Noorwali, "A Validation Study of a Requirements Engineering Artefact Model for Big Data Software Development Projects," in *Proceedings of the 14th International Conference on Software Technologies*, 2019, pp. 106–116. doi: 10.5220/0007927201060116.

[6] C. Djeddi, N. E. Zarour, and P. J. Charrel, "Extension of iStar for Big data projects," *CEUR Workshop Proc.*, vol. 2326, pp. 9–16, 2018.

[7] D. Kelly, J. de Boer, P. Uhlenbruck, and S. Webb, "Requirements gathering and testing by game playing," in *2015 IEEE 28th Canadian Conference on Electrical and Computer Engineering (CCECE)*, May 2015, pp. 1556–1560. doi: 10.1109/CCECE.2015.7129512.

[8] I. Soleha, E. Mu'awanah, A. R. Khamami, and N. Naqiyah, "Implementation of Role Playing Techniques on Orde Tabung in the Model Counseling of KIPAS," *EAS J. Psychol. Behav. Sci.*, vol. 4, no. 2, pp. 65–69, Mar. 2022, doi: 10.36349/easjpbs.2022.v04i02.004.

[9] S. Rueda, J. I. Panach, J. B. Cabotà, and F. Valverde, "Applying Role-Playing Game in Software Development Subjects," in *International Conference of Education, Research and Innovation*, Nov. 2016, pp. 3532–3538. doi: 10.21125/iceri.2016.1839.

[10] Z. Ul Hassan, A. Raza Sattar, M. Raheel Zafar, and W. Abbas, "Impact of Requirement Gathering Techniques on Software Development," *Inf. Knowl. Manag.*, vol. 6, no. 11, pp. 28–30, 2016, [Online]. Available: www.iiste.org

[11] B. A. Kobrinskii, "Role Playing as a Method of Learning in Knowledge Engineering," in *2022 VI International Conference on Information Technologies in Engineering Education (Inforino)*, Apr. 2022, pp. 1–5. doi: 10.1109/Inforino53888.2022.9782987.

[12] B. Attanayaka, D. Nawinna, K. Manathunga, and P. K. W. Abeygunawardhana, "Success Factors of Requirement Elicitation in the Field of Software Engineering," in *2022 4th International Conference on Advancements in Computing (ICAC)*, Dec. 2022, pp. 240–245. doi: 10.1109/ICAC57685.2022.10025166.

[13] M. Yaseen, S. Ali, A. Mustapha, and N. Mazhar, "Success factors analysis for requirement elicitation in global software development paradigm: An empirical study," *J. Softw. Evol. Process*, vol. 34, no. 7, Jul. 2022, doi: 10.1002/smr.2460.

[14] J. J. Berman, *Principles and Practice of Big Data*. Elsevier, 2018. doi: 10.1016/C2017-0-03409-2.

[15] R. Kimball and M. Ross, *The data warehouse toolkit*, 3rd ed. Nashville, TN: John Wiley & Sons, 2013. [Online]. Available: https://www.wiley.com/en-gb/The+Data+Warehouse+Toolkit%3A+The+Definitive+Guide+to+Dimensional+Modeling%2C+3rd+Edition-p-9781118530801

[16] Inmon, W. H. Inmon, D. Strauss, and G. Neushloss, *DW 2.0: The Architecture for the Next Generation of Data Warehousing*. Morgan Kaufmann, 2014.

[17] R. Abepala, "Contemporary Educational Concepts." SARA Publication, 2018.
[18] S. Mabusela and J. Adams, "Employing Role play in teaching and learning: a case of Higher Education," *South African J. High. Educ.*, vol. 27, no. 3, Jan. 2016, doi: 10.20853/27-3-263.
[19] D. Rao and I. Stupans, "Exploring the potential of role play in higher education: development of a typology and teacher guidelines," *Innov. Educ. Teach. Int.*, vol. 49, no. 4, pp. 427–436, Nov. 2012, doi: 10.1080/14703297.2012.728879.
[20] A.-J. Moreno-Guerrero, C. Rodríguez-Jiménez, G. Gómez-García, and M. Ramos Navas-Parejo, "Educational Innovation in Higher Education: Use of Role Playing and Educational Video in Future Teachers' Training," *Sustainability*, vol. 12, no. 6, p. 2558, Mar. 2020, doi: 10.3390/su12062558.
[21] M. Dracup, "Online role play stories, engagement and learning in higher education," Deakin University, 2012. doi: 10.13140/2.1.4542.7528.
[22] M. Fletcher, "Roles in support of teaching and learning," in *Beyond the comfort zone: Proceedings of the 21s ASCILITE conference*, 2004, pp. 304–313.
[23] K. Kettula and S. Berghäll, "Drama-based role-play: a tool to supplement work-based learning in higher education," *J. Work. Learn.*, vol. 25, no. 8, pp. 556–575, Oct. 2013, doi: 10.1108/JWL-04-2012-0036.
[24] A. Sivia, S. MacMath, C. Novakowski, and V. Britton, "Examining Student Engagement During a Project-Based Unit in Secondary Science," *Can. J. Sci. Math. Technol. Educ.*, vol. 19, no. 3, pp. 254–269, Sep. 2019, doi: 10.1007/s42330-019-00053-x.
[25] A. Ferrari, P. Spoletini, M. Bano, and D. Zowghi, "SaPeer and ReverseSaPeer: teaching requirements elicitation interviews with role-playing and role reversal," *Requir. Eng.*, vol. 25, no. 4, pp. 417–438, Dec. 2020, doi: 10.1007/s00766-020-00334-0.
[26] J. Andrew and J. Meligrana, "Evaluating the Use of Role Playing Simulations in Teaching Negotation Skills to University Students," *Creat. Educ.*, vol. 03, no. 06, pp. 696–707, 2012, doi: 10.4236/ce.2012.36104.
[27] I. V. Lokshina, B. J. Durkin, and C. J. M. Lanting, "Data Analysis Services Related to the IoT and Big Data," *Int. J. Interdiscip. Telecommun. Netw.*, vol. 9, no. 2, pp. 37–56, Apr. 2017, doi: 10.4018/IJITN.2017040104.
[28] S. Gathani, Z. Liu, P. J. Haas, and Ç. Demiralp, "Understanding Business Users' Data-Driven Decision-Making: Practices, Challenges, and Opportunities," *arXiv*. Dec. 27, 2022. [Online]. Available: http://arxiv.org/abs/2212.13643
[29] J. M. Carroll, "The Internet of Places," in *Social Internet of Things*, 2019, pp. 23–32. doi: 10.1007/978-3-319-94659-7_2.
[30] E. Miranda, "Moscow Rules: A Quantitative Exposé," in *Lecture Notes in Business Information Processing*, 2022, pp. 19–34. doi: 10.1007/978-3-031-08169-9_2.