UNIVERSITÀ
DI TRENTO

UNIVERSITY
OF TWENTE.

Co-funded by the
European Union

Masters Degree in

Interaction Technology
Curriculum Human Computer Interaction and Design

FINAL DISSERTATION

DEVELOPING A MULTI-METRIC DASHBOARD TO ENHANCE INTERPRETATION OF B2B CUSTOMER TRAINING EVALUATION DATA AT BAKER HUGHES

Supervisor UniTrento	Supervisor Baker Hughes	Critical Observer UniTwente
Maurizio Marchese	Mario Cesaro	Lorenzo Gatti

Student
Alessia Bertana

Academic year 2024/2025

Acknowledgement

First of all I want to express my sincere gratitude to Mario Cesaro and my two University supervisors, Maurizio Marchese and Lorenzo Gatti. They provided me with good and critical feedback and have supported me during my entire process. It has been a pleasure to work with them. I'd also like to thank all the stakeholders from the customer training team that were involved within the project for their participation in interviews, thereby providing me with valuable information about their pain points with current practices and user needs for a new tool. And last, I would like to thank my professors, friends, and others who have directly and indirectly helped me complete this dissertation.

Abstract

This research shows how a multi-metric evaluation tool can provide deeper insights into participant feedback beyond a standalone Net Promoter Score analysis in B2B customer training at Baker Hughes. The dashboard was developed using a Design Science Research approach and integrates the Net Promoter Score (NPS) with eight predefined complementary metrics, structured around Kirkpatrick's evaluation framework. Human-computer interaction (HCI) principles guided the design to support information retrieval, trend identification and interpretation of how participants perceived training sessions. Task-based usability testing with the dashboard showed strong results, with a task completion rate of 92% and a System Usability Score of 87.75%. The User Experience Questionnaire indicated positive user experience, although variation in pragmatic interpretation suggests areas for refinement. Secondary qualitative feedback confirmed the dashboard's value and highlighted opportunities to improve interactivity and to integrate more qualitative insights. Overall, the realized dashboard provides a deeper understanding of training participant feedback by integrating multiple evaluation metrics with NPS through a design tailored to stakeholder needs at Baker Hughes. This enables well-informed, actionable decision-making that facilitates targeted improvements across the B2B training portfolio.

Glossary

Business Intelligence	BI
Business-to-Business	B2B
Design Science Research	DSR
Extract, Transform and Load	ETL
Human Computer Interaction	HCI
Key Performance Indicators	KPI
Net Promoter Score	NPS
User Experience	UX
User Interface	UI

Contents

1 Introduction

1.1 Background

Providing efficient and high-quality training to clients is a top priority in today's fast-paced and knowledge-driven business environment. Since organizations invest a great deal of resources in training programs, the ability to evaluate their effectiveness becomes increasingly important. Evaluating training is crucial to ensure customer satisfaction and to help in strategic decision-making and resource allocation.

Many organizations currently use standardized evaluation metrics and frameworks to assess the effectiveness of their training programs. One of the most widely used metrics is the Net Promoter Score (NPS), which serves as a quantitative measure designed to evaluate qualitative data regarding customer loyalty [1, 2]. The NPS is calculated based on responses to a single question: *How likely are you to recommend this course?* [3]. The score indicates potential organizational growth [3, 4] and gained popularity in various industries due to its simplicity and scalability. A more detailed explanation of the NPS will be provided in *Section 2.2 Net Promoter Score*. In training, NPS is often used to understand the overall perception of participants of a course or learning experience.

At Baker Hughes, the NPS is used to assess the delivery and overall quality of customer training. Baker Hughes is a global energy technology company that operates in over 120 countries, offering a wide range of customer training programs through its subsidiary, Nuovo Pignone International S.r.l. These programs equip customer personnel with the skills to operate and maintain turbomachinery equipment and digital systems. Training is delivered through various modalities, which include in-person sessions, e-learning modules, and immersive virtual reality experiences. This thesis will be conducted within the digital customer service team at Nuovo Pignone International S.r.l.

1.2 Problem Statement

While the NPS provides an easy score to obtain and interpret, it only offers a surface-level view of customer sentiments. The NPS is a quantitative score that aims to interpret qualitative customer feedback [1, 2]. When used as a standalone score during training evaluation, it does not explain *why* participants rate a course positively or negatively. It lacks a level of detail to determine areas for improvement [5, 6]. This underscores a gap in this system and highlights the need for new approaches to understand the underlying drivers behind NPS.

At Baker Hughes, the quality of training programs currently mainly relies on the NPS as a course quality indicator. However, NPS as a standalone score does not provide insights into experiences of course participants. As a result, organizations often turn to participant comments for a deeper understanding. Yet, going through every single comment remains very time-consuming. These limitations are further reflected in the current visualization of

training evaluation data, which primarily consists of Excel graphs that highlight the score itself in graphs with little supporting contextual information.

Over the years, the company has also collected nine other evaluation metrics that offer context, such as satisfaction with content, documentation, and delivery. Within this research, eight of these metrics will be considered due to the low amount of data for the 9th metric. Despite their potential value, these eight metrics are currently only analyzed to a deeper extent when an NPS score is lower than a certain treshold, which leaves a gap in the organization's ability to see patterns accross all training feedback. This highlights the company need for a tool that can clearly visualize the context behind NPS scores that can inform decisions in improving the customer training portfolio.

Research in data visualization and decision support systems suggests that multi-metric interactive dashboards can significantly improve users' ability to interpret data and generate actionable insights [6, 7, 8, 9]. This thesis aims to develop a user-friendly *multi-metric tool* that combines NPS with the eight currently underutilized quantitative evaluation metrics. This dashboard will be informed by user experience (UX) and UI (User Interface) principles from both the academic field of Human-Computer Interaction (HCI) and Data Visualization, thereby contributing in a meaningful way to both academic literature and Baker Hughes. An appropiate multi-metric tool will be selected later in this dissertation.

1.3 Project Scope and Significance

This research aims to develop a user-friendly tool that improves analytical insights into customer training feedback for programs offered by the Customer Training division of Nuovo Pignone International S.r.l. The scope of this study is focused on analyzing and visualizing the organization's existing evaluation data. By utilizing nine quantitative metrics, including the Net Promoter Score (NPS), alongside eight additional metrics, this thesis holds significance on two levels:

- **Organizational Impact:** For Baker Hughes, this research presents a more practical approach to evaluating training quality. A multi-metric tool will be selected, designed, and finally assessed based on its usability. The goal is to enhance contextual insights and trend recognition, utilizing a dashboard that integrates the NPS along with eight evaluation metrics. When insights on course quality elements are effectively used, they have the potential to increase the effectiveness of the training programs that are offered by the company.

- **Academic Contribution:** From a research perspective, this study contributes to the training evaluation field by exploring how a multi-metric tool can improve the interpretation of participant feedback beyond what is captured by the NPS. By integrating supplementary metrics and by applying HCI and data visualization principles, the study includes a practical approach to designing user-friendly evaluation tools that support deeper insight generation in B2B training contexts.

1.4 Research Objectives and Questions

The main objective of this thesis is to design, develop, and evaluate a user-friendly multi-metric tool that improves the interpretation of course participant sentiments in B2B training evaluation. The main research question of this study is therefore as follows:

> **Main RQ:** How can a multi-metric evaluation tool be designed to provide deeper insight into participant evaluations beyond what is captured by the Net Promoter Score (NPS) in B2B customer training at Baker Hughes?

The main research question will be answered through five sub-questions:

1. What are the limitations of using NPS as a standalone metric in training evaluation in B2B contexts?
2. What are the limitations of the current NPS-based dashboard in supporting insight generation and training evaluation at Baker Hughes?
3. How can complementary training evaluation metrics be integrated with NPS to improve contextual interpretation in a multi-metric evaluation tool?
4. How can Human-Computer Interaction and data visualization principles be applied to design a user-friendly multi-metric tool for training evaluation?
5. Does the new multi-metric tool enhance the contextual interpretation of the Net Promoter Score?

The first sub-question aims to identify the limitations of using Net Promoter Score (NPS) as a standalone metric for evaluating training. Recognizing these limitations is important to justify the need for a multi-dimensional evaluation approach.

The second sub-question is posed to understand current practices at Baker Hughes, specifically how the NPS is currently used. The aim is to understand the pain points and needs of stakeholders regarding the current customer training evaluation approach.

The third sub-question explores how complementary training evaluation metrics can be integrated with NPS to improve contextual interpretation within a multi-metric evaluation tool. To address this, the study will analyze the eight supplementary metrics available to this study using an established training evaluation framework. This approach will assess the individual relevance of each metric, with the aim of providing more context beyond NPS.

The fourth sub-question examines how principles from Human-Computer Interaction and data visualization can be applied to design a clear and user-friendly multi-metric tool.

Finally, the fifth sub-question explores whether the final multi-metric design improves the interpretation of customer feedback. This will be evaluated through a task-based usability test and a requirement integration check with a key stakeholder.

1.5 Thesis Structure

The remainder of this dissertation is organized as follows:

- **Chapter 2: Literature Review** – This chapter reviews relevant literature on training evaluation methods, usage of the Net Promoter Score, multi-metric evaluation approaches, and the use of data visualization and human-computer interaction principles in the design of a tool to enhance user experience and decision-making.
- **Chapter 3: Methodology** – Describes the research approach based on Design Science Research, states the method to identify key stakeholders and gather their requirements, outlines the approach to handle the dataset, and explains the evaluation methodology.
- **Chapter 4: Tool Design and Development** – This chapter details the development process of the multi-metric tool, including requirement gathering through stakeholder interviews, questions to answer with the data, and the design of key features based on HCI and visualization principles.
- **Chapter 5: Results** – This chapter presents the results of the evaluation of the developed multi-metric tool. The evaluation will be focused on assessing the usability and effectiveness of the tool.
- **Chapter 6: Conclusion** – This chapter summarizes the main findings and interprets the evaluation results and reflects on the contributions of the research.
- **Chapter 7: Future Work** – This chapter outlines the limitations of the study and provides recommendations for Baker Hughes and suggests directions for future research.

2 Literature Review

This chapter provides the theoretical foundation for this study and offers insights into the field of Business-to-business (B2B) training evaluation, the limitations of the Net Promoter Score (NPS) in training analytics, and the need for multi-metric data visualization tools. Furthermore, this study introduces research on multi-metric data visualization techniques and human-machine interaction (HCI) principles for dashboard design, aiming to enhance the extraction of actionable insights during training quality assessments.

2.1 Training Evaluation

The value in most Business-to-Business (B2B) markets has shifted [8]. The success of a company is highly dependent on the acquisition of loyal customers who are satisfied with the products or services purchased [4]. Customers currently no longer assess value solely based on the product itself [8, 10]. Instead, they evaluate the entire service surrounding it, including onboarding support, technical assistance, and, most notably, **customer training**.

Customer training refers to a planned set of activities designed to enhance employee performance by equipping individuals with the skills and knowledge necessary to achieve predefined organizational goals [11, 12]. It plays an important role in maintaining a company's competitiveness by continuously developing workforce skills and aligning learners' capabilities with specific job demands [11, 12, 13, 14]. Without customer training, even the most innovative technologies risk failing to be implemented or not being used [15].

As training becomes increasingly central to organizational success, stakeholders are turning their attention to its effectiveness and impact. Stakeholders increasingly demand **insights** into its perceived **impact** and **quality** to ensure that training delivers real value to both the client and the learners [16]. In this study, the term *client* refers to the organization purchasing the training from the training provider, while *learners* denote the employees that the client assigns to attend the training. The primary objective of B2B training is for learners, sent by the training purchasing client, to apply their newly acquired skills in their daily roles. This concept, known as **knowledge transfer**, refers to how well training outcomes can be applied to different tasks and environments [11]. To achieve this, training should match the expected learning outcomes while engaging learners through training delivery.

Evaluation of the quality and impact is an essential component of training to understand if the training is actually providing **added value** to the learners, and as a result of **transfer**, to the client. Training evaluation most commonly refers to the systematic collection of information to judge how well a program meets its intended goals [17]. These training evaluation insights not only validate program effectiveness but also guide future training design, decide whether to continue or end a program, and justify training investments [18]. In addition, evaluation can reveal the **knowledge gap** between what trainers intend to teach and what learners actually absorb [16, 17].

Often, assumptions about training effectiveness are intuitive and may lack empirical grounding. The training might seem successful on paper, but could fail to generate meaningful performance outcomes [13]. Other common pitfalls in training evaluation include factors such as lack of objectivity, flawed evaluation methods, misinterpretation of results, poor utilization of findings, and budget limitations [17]. It is important that these pitfalls are addressed in the design of any evaluation tool.

This study utilizes a predefined dataset provided by Baker Hughes to analyze the gathered information and enhance evaluation. To achieve effective training evaluation with the available dataset, it is first necessary to know **what concept** the training provider wants to measure, and for **what purpose** it needs to be measured [17].

Furthermore, the reliability of evaluation outcomes depends not only on providing the appropriate content but also on the individual conducting the evaluation. The identity and expertise of the evaluator, who is responsible for the evaluation of the training portfolio, can significantly shape the scope, interpretation, and perceived credibility of the findings [16]. In this study, the *evaluator* is defined as the person within the training company who is responsible for assessing the quality of training. Establishing evaluation objectives in collaboration with the evaluator is a crucial starting point for any training assessment.

To interpret the existing dataset in a structured way, existing training frameworks could offer a valuable lens for analysis. There are various existing frameworks for training evaluation, but the most well-known and praised method is Kirkpatrick's four-level training evaluation [14]. Kirkpatrick's model allows for the measurement of potential effects of training at four levels: (a) participants' reaction to the training, (b) participants' learning as a result of the training, (c) participants' change in behavior as a result of the training, and (d) the subsequent impact on the organization as a result of participants' behavior change [18, 19]. Due to training evaluation complexity, high costs, and lack of direct methods, the fourth level is often not evaluated [14]. Figure 1 presents the four levels of Kirkpatrick's training evaluation framework, along with the key questions addressed at each level.

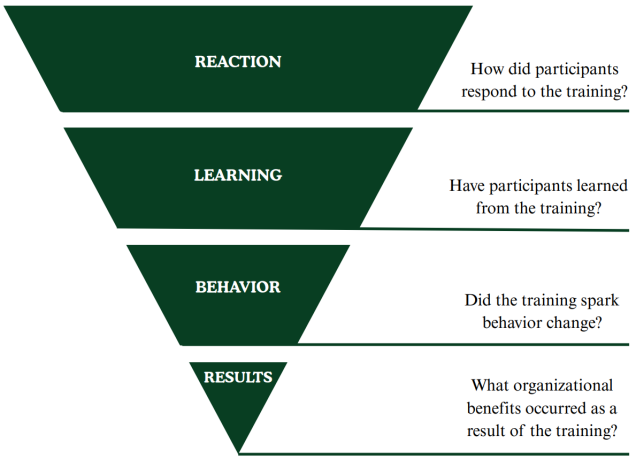


Figure 1: Kirkpatrick's four-level training evaluation framework

Following Kirkpatrick's model, practitioners can determine the extent to which participants are satisfied with a training program, whether participants learned from the program, whether participants were able to apply the learning on the job, and/or the impact on organizational outcomes. Since the evaluation data for this study has already been defined, the aim is not to select new metrics but to interpret and combine the existing ones to gain a deeper understanding of learner sentiments towards the training. Therefore, this literature study aims to examine ***'How to create a user-friendly multi-metric tool that provides insights into participant feedback beyond what is captured by the NPS from the existing dataset available to this study'***.

The dataset available to this study includes a three-year evaluation dataset collected through a summative approach, focused on assessment by learners after training. Responses were provided solely by course participants. It contains the Net Promoter Score (NPS) along with eight quantitative metrics covering aspects such as logistics, course duration, and instructor communication. This dataset will be explored in greater detail in *Section 4* to evaluate its usefulness in assessing training quality. In this exploration, Kirkpatrick's framework can be utilized to analyze the existing data by aligning each Key Performance Indicator (KPI), a quantifiable measure of performance over time, with a specific objective and its corresponding evaluation level. This mapping offers a structured lens to assess whether all levels are represented and to identify potential gaps within the current evaluation.

In summary, effective training evaluation in B2B contexts goes beyond measuring immediate satisfaction. It should assess whether the training facilitates knowledge transfer to the workplace by promoting changes in participant behavior through an engaging course experience. Building on this, it is essential to understand **what** training aspects are evaluated, for **what reasons** they are evaluated, and whether the **evaluator** recognizes the relevance of this evaluation approach. While a training evaluation method like Kirkpatrick's four-level framework offers a useful guideline on what levels to evaluate, evaluation efforts in general often still run into challenges, such as overreliance on assumptions, using metrics that don't reflect the actual learning experience, or struggles to translate qualitative experiences like engagement and skill transfer into numbers. With these considerations in mind, the following section looks at the Net Promoter Score (NPS), currently used by Baker Hughes as the main way to assess training quality.

2.2 Net Promoter Score

The Net Promoter Score (NPS) is a commonly used metric to evaluate the likelihood of a customer recommending an organization. In recent years, it has become particularly popular for measuring customer loyalty [4]. It has become an alternative to traditional customer satisfaction surveys. The NPS score is adopted in prominent companies such as General Electric, American Express, Microsoft, and Baker Hughes [6]. For some company managers, achieving a high NPS is nearly as crucial as obtaining strong economic results [2]. But what is the foundation of this score, and how can it be interpreted?

2.2.1 Definition of Net Promoter Score (NPS)

The NPS, introduced by Satmetrix and Fred Reichheld of Bain and Company, was designed as a concise, actionable metric for **customer loyalty** [3]. NPS is grounded in the idea that a customer's willingness to recommend a product or service is a strong indicator of their loyalty. Reichheld proposed that NPS is "the one number you need to grow" in his Harvard Business Review article, showing empirical correlation between NPS performance and revenue growth [2, 3, 4, 8].

Customer loyalty has become an important asset for companies today. It shows how willing people are to interact with a brand, share their opinions, and stay loyal over time, even when other companies offer better short-term deals. [4]. This loyalty can lead to steady growth and help a business stand out from its competitors [8]. Nevertheless, building it in B2B markets can be complex due to the complicated channel structures, concentrated buyer communities, or large accounts with many people influencing the relationship [8].

At its core, loyalty stems from **customer satisfaction**, a multidimensional concept that remains hard to quantify. Service quality assessment is related to customer expectations and is often connected to customer satisfaction. Customers typically evaluate their experience at two levels: their overall satisfaction with the service and perceived quality of that service [10]. Research shows that when consumers perceive their level of satisfaction as high, they are more likely to remain a loyal customer [4]. Maintaining this customer loyalty, however, requires tailored solutions and systematic tracking of customer-client relationship health [4, 8].

NPS helps to gain insight into loyalty by obtaining information on whether the customer would recommend a service or product of a company to others [4]. In the context of this study, which focuses on B2B training services, the purchasing decision lies with the client, while the effectiveness of the service depends on learner transfer. Given this dual dynamic, it becomes essential that clients perceive sustained value and remain loyal to their training provider. Most B2B clients evaluate training providers on a combination of several criteria: *Do they create economic and strategic value for our business? Do they simplify our daily operations (and my own work life)? Can I trust them? Do I like working with them?* [8] When these criteria are met, the provider-client partnership can evoke strong loyalty.

The method for calculating the NPS index is straightforward and easy to interpret. The NPS score is calculated based on responses to a single question [2]: ***"How likely are you to recommend [Company X] to a colleague?"*** Participants respond on a scale from 0 to 10. This scale is depicted in Figure 2.

Figure 2: NPS Scale

Consequently, respondents are categorized into three categories based on their likelihood to recommend a product or service:

- **Promoters:** loyal customers (scores of 9–10) who are likely to act as brand advocates and contribute to repeat business.
- **Passives:** satisfied but unenthusiastic customers (scores of 7–8) who are unlikely to promote or criticize the brand.
- **Detractors:** unhappy customers (score of 0–6) who may damage the brand reputation through negative word-of-mouth.

The foundation of the NPS metric is the belief that customers are either promoters, extremely satisfied users who will serve as enthusiastic brand ambassadors as well as sources of repeated sales, or detractors, extremely dissatisfied users who will undermine brand growth by spreading unflattering testimony of a product experience [1]. Promoters tend to stay longer with the company, purchase more products, usually cost less to serve, and are more likely to refer the supplier to colleagues and friends [8].

The final score is computed by subtracting the percentage of Detractors from the percentage of Promoters, resulting in a value between -100% and +100%. Passives are not considered as this group represents customers who had a positive experience with the company, but are either not willing to recommend or are indifferent [2]. The formula to calculate the NPS is provided below.

$$NPS = \frac{\text{Promoters} - \text{Detractors}}{\text{Total Responses}} \times 100$$

A positive score indicates a predominance of promoters, suggesting brand growth, while a negative score suggests a higher proportion of detractors, highlighting potential issues in customer satisfaction or product quality. Figure 3 describes the relations between customer attitudes and business performance according to the concept of NPS.

2.2.2 Limitations of NPS as a sole score

The NPS score rests on the assumption that an individual's likelihood to recommend correlates with their overall satisfaction and loyalty, as depicted in Figure 3. Reichheld claims that there is no need for extensive surveys and argues that NPS is a reliable standalone method to determine customer loyalty [2, 6]. Yet, despite its simplicity, ease of reporting, and widespread use, various academics have raised doubts about the efficiency of a single question score [2, 6]. In the light of these concerns, evaluation of NPS used as a sole score should be made from at least two perspectives, on empirical and conceptual grounds [6].

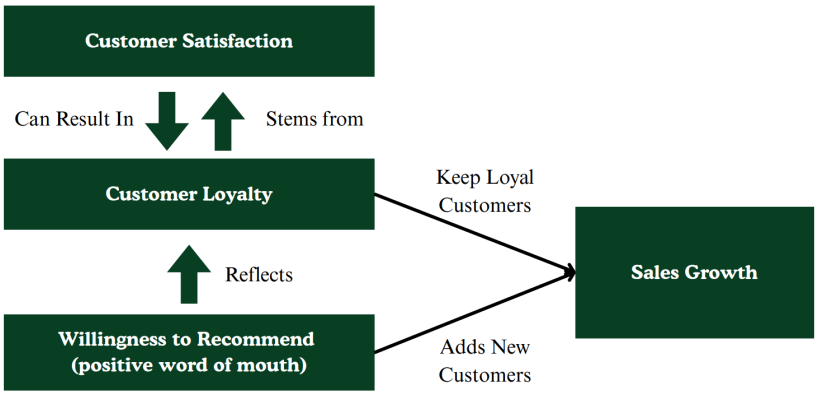


Figure 3: Relationship between customer attitudes and business performance

A) Conceptual Limitations of the NPS as a sole score

The assumption that NPS can predict sales growth seems to be intuitive: If customers spread positive word-of-mouth among their friends, some of their friends will try the brand and become customers, which will ultimately increase the brand's sales [20]. While critics state various flaws related to the conceptual framework of NPS, this thesis focuses exclusively on the benefits and downsides of using the NPS score as a standalone metric. Therefore, only the conceptual limitations that are directly related to this application will be examined.

B) Ambiguity in interpretation

NPS scores can vary between industries, business models, and even different customer segments. Without proper context, a high NPS might not necessarily indicate strong customer loyalty, and a low NPS might not always be a sign of deep dissatisfaction. For instance, in a B2B training environment, learners often have limited influence over purchasing decisions made by their organization. As a result, their likelihood to recommend a training may not directly impact the client's decision to repurchase. Learners' satisfaction alone, while important, may be outweighed by other factors, such as the client's perception of behavioral change or measurable outcomes resulting from the training. Therefore, understanding the **context** and understanding 'why' behind the NPS is essential for drawing actionable insights.

C) Oversimplification of learner experience

While the single-question NPS approach is scalable and easy to interpret, it oversimplifies the multidimensional nature of learner experience, particularly in training contexts. When translating qualitative experiences into a numeric score, it strips away context. Training environments involve complex interactions among content, instructional approach, practical applicability, relevance, and learner motivation. Yet, NPS represents all of these interactions at once without explaining why participants feel as they do or how those perceptions relate to behavioral outcomes such as engagement, application of learning, or ongoing development.

D) Lack of diagnostic value

A core conceptual limitation of NPS is its lack of diagnostic value. While it may indicate whether learners are broadly satisfied or dissatisfied, it fails to reveal which specific aspects of a training contribute to those sentiments. It offers no insight into what drives loyalty or disloyalty [21], nor does it provide insights on how to improve the score [8]. NPS does not provide statistically reliable data on what customers truly value, information that is essential for sales growth. As a standalone metric, it answers only the surface-level question: *"How are we doing with our customers?"* [21] Organizations that rely solely on NPS may struggle to address deeper, more actionable questions, such as: *"What aspects of the training need improvement?"* or *"What drives positive learner outcomes?"* Reichheld's model offers no help in answering these questions [1, 17, 21], and the path to improving the score remains unclear without additional context [6].

In summary, the theoretical foundation of NPS lacks the level of detail necessary to support informed decision-making in training evaluation. It fails to reveal which aspects of a training learners find satisfactory or dissatisfactory and it does not provide actionable insights for improvement. Moreover, without a proper context, a high NPS may not necessarily reflect strong customer loyalty. Its qualitative nature, while presented by a quantitative score, can introduce cognitive biases that further complicate interpretation. Therefore, when the NPS is used as a standalone measure, it risks providing meaning without delivering insight.

B) Empirical Limitations of the NPS

Reichheld showed a strong correlation between the NPS and a company's sales growth rate, suggesting its potential as a predictive tool. This positive relationship was confirmed by Aguinis and Burgi-Tian, who found that NPS correlated with past, present, and future revenue in the B2B market for primary products in New Zealand [4]. While several studies have linked NPS to customer intentions and behavior, empirical research that specifically examines its impact on sales growth remains limited and contains mixed results [20].

Beyond Reichheld's original study, only four academic investigations have directly assessed the relationship between NPS and sales growth. None confirmed NPS as a superior predictor. Morgan and Rego (2006) and Keiningham et al. (2007) found no significant impact of NPS on sales growth, with Morgan and Rego further concluding that NPS lacked a statistically significant relationship with future business performance [20, 22]. Meanwhile, Pingitore et al. (2007) and van Doorn et al. (2013) acknowledged NPS as a predictor of sales growth, but not one that outperforms other customer attitude metrics [20].

Additional evidence from Denmark further questions the robustness of NPS. Kristensen and Eskildsen concluded that NPS is a weak predictor of customer loyalty and satisfaction, warning that relying solely on NPS in business management could lead to misinterpretations [2]. Nevertheless, some research still supports the practical utility of NPS, suggesting it can offer a useful foundation for corporate decision-making [23].

Taken together, these findings underscore the mixed empirical support for NPS as a reliable and standalone metric. Therefore, it may be advisable for organizations to incorporate additional or alternative evaluation metrics to gain a more comprehensive and accurate understanding of customer loyalty and business performance. The following section explores how a multi-metric framework can enhance decision-making by providing richer insights.

2.3 Multi-Metric Data-Driven Decision Making

Data-driven decision making is defined as "the practice of basing decisions on the analysis of the data rather than purely on intuition" [24]. According to ISO25.012 standards, **data quality** depends on data accuracy, completeness, consistency, credibility, and currentness, as well as system-dependent aspects such as accessibility, compliance, confidentiality, efficiency, and precision [24]. When designing a data tool that should inform decisions, these aspects should be taken into account. Moreover, poor data quality or flawed measurement processes can affect decision-making, which is why it is crucial to ensure data quality and evaluation of the right metrics [24].

Raw data alone does not provide insights until it is processed and interpreted [7]. Through manipulation, data, often composed of one or more metrics, can become information, and ideally, when internalized, this information transforms into knowledge. The value of information lies in its ability to generate knowledge and support better decision-making [7, 25, 26]. Data becomes meaningful only when it informs and improves decisions [7]. Because data is built from metrics, selecting the right metrics should be guided by their potential to drive actionable decision-making [7]. To use metrics effectively, their interrelationships and limitations must be considered within the appropriate context, while ensuring they are **consistent, comparable, and repeatable** to support **actionable decision-making** [7, 24].

Viewing market dynamics from multiple angles can enable practitioners to cross-validate insights, maximize knowledge accuracy, and identify inconsistencies [7]. In the context of this study, to move beyond the interpretation of NPS as a sole score and to understand the feelings of learners to a deeper extent, a **multi-metric** approach should be adopted that pushes past the simple averages of NPS and the limitations of a single-question metric [2, 8]. This process is also known as **data triangulation**, which refers to the use of multiple data sources, or in this case metrics, to gain a comprehensive understanding of phenomena [27].

Assessment of various metrics may give a less precise picture and might add to the complexity of research. However, by considering multiple measures, it is less likely to miss a major problem [7, 9]. Triangulating NPS results with complementary quantitative metrics and stakeholder feedback can therefore provide **actionable insights** into training effectiveness [7, 6]. Nevertheless, having more metrics does not always help make better decisions [7, 25]. Therefore, the format in which the data is presented needs to be considered as well when aiming to enhance decision-making. Hence, the next section will focus on selecting the format in which the data will be given.

2.4 Data Visualization Dashboard

Information overload is a recognised phenomenon related to the continuous increase of data and the need to process that information [25]. Business Intelligence (BI) has attempted to manage overload using dashboards. **Dashboards** are defined as visual displays of the most important information needed to achieve one or more objectives. They enable concepts to be merged into a single screen [25], providing valuable information to assist stakeholders and employees in improving performance and making the most effective decisions.

Due to the limitations of relying solely on single metrics like Net Promoter Score (NPS), experts suggest using a dashboard approach [7]. Dashboards facilitate the cross-validation of multiple metrics, making it easier to identify patterns that may otherwise go unnoticed [24]. By presenting data in visual formats, dashboards can improve the interpretation of various metrics, enabling users to derive insights more effectively [7]. When effectively implemented, data visualization can help stakeholders enhance performance and make more informed decisions, saving both costs and time to analyze phenomena [25]. For these reasons, this study will utilize a dashboard as the primary tool for developing the multi-metric framework.

In today's fast-paced environment, **automation** is essential for monitoring and responding to changes [24]. Within automated systems, **organizational memory**, which refers to the accessible history of company data, supports evaluative learning and decision-making [24]. Users prefer comprehensive, integrated dashboards that support self-service analytics, which allows for independence from manual IT operations. This approach enables users across organizations to be more creative in their analysis and expression through visual reporting. To enhance the quality of the dashboard, it should preferably be automated. However, since the goal of the study is to design and develop a user-friendly multi-metric tool to enhance NPS feedback interpretation, dashboard automation is not a priority.

There are many different data visualization platforms through which the dashboard can be created. The two data visualization platforms that are available to this study, which is conducted in Baker Hughes, are Tableau and Power BI. For this study, **Power BI** has been selected due to its interactivity, scalability, and visualization features for dashboards. Power BI is a Business Intelligence and data visualization platform developed by Microsoft. It allows users to connect to and analyze data from various sources and to access it through an interactive interface. These capabilities make it suitable for the multifaceted nature of training evaluation, allowing users to explore dynamic data and derive actionable insights.

The effectiveness of any dashboard does not only depend on the selected platform but also on how the information is visually structured and presented. It is not only important to achieve specific goals but also the **user's satisfaction and experience** of the system are significant [28]. To ensure that the data supports meaningful interpretation, it is essential to apply thoughtful design principles grounded in Human-Computer Interaction (HCI), usability, and visual cognition. The following section explores these principles and their role in developing dashboards that are both functional and user-centered.

2.5 Design Principles for Dashboard Visualisations

This section explores design principles for dashboard visualizations. The focus within this section is both on user experience, which refers on how effectively and intuitively users engage with the dashboard, and user interface design, refering to the visual layout and interactive elements that shape that engagement. Together, these perspectives ensure dashboards are both functional and user-friendly.

2.5.1 User Experience Design

According to ISO 924-11, usability can be defined as "the extent to which specified users can use a product to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use." This refers to the user experience (UX) of human-computer interactions [28]. For dashboards, similar to other information systems, usability can be defined as the extent to which users use a system to achieve specific goals with high efficacy, efficiency, and satisfaction" [28]. User dashboards need to be understood by their users. Therefore, information chaos, such as information overload, erroneous information, scattered information, and incompatibility of information with job requirements, should be prevented as much as possible [7].

However, it is important to note that user experience depends on each user. There is no 'one size fits all' in this domain [29]. Although such a dashboard would be ideal, it is utopic as not every user is driven by the same goals, has the same interests, visualization literacy, and so on. These aspects include not only personal preferences, but also social factors, like biases, beliefs, or past experiences [29]. Therefore, users should be provided with tailored dashboards that fulfill their requirements and improve insight delivery to enhance the outcomes of the decisions made. Involving the end-user in the design processes supports the development of better systems, which are helpful for them and match their needs. Given these facts, it is essential to **involve the evaluator** not only in understanding the goal of the dashboard, but also to take this person into account during the development of the dashboard. For this reason, an **iterative design approach** should be applied in which the *evaluator* is consulted throughout the process. [29].

2.5.2 User Interface Design

Information presented on the dashboards is essentially visual, which can affect efficacy of dashboards [25, 26]. For example, while colours can enhance the visualisations of dashboards, overusing or misusing certain colour palettes can have a negative impact on decision-making. The aim of dashboard design is to contribute to the clarification of data to be efficiently read and understood by the user [26]. The designer plays a key role in this process of creating and unifying graphics and interaction. Ultimately, their goal is to translate information into visual and interactive systems that are easy and fast to understand for users [26].

In the context of dashboards designed for management within the business sector, it is widely recommended to prioritize functionality over aesthetics [26]. Ensuring usability and relevance requires that designers offer features and visualization options that are tailored to the specific needs of users. Information holds the greatest value when presented in user-friendly formats that not only support strategic decision-making but also facilitate the creation of actionable knowledge [7]. By incorporating data visualizations such as tables, graphs, pictograms, and numerical displays that align with the nature of the data, dashboards can enable users to interpret information from multiple angles. This multidimensional view fosters deeper analysis and ultimately leads to more informed decision-making [26].

However, graphics are not always the best solution for translating information. For example, when there is only one or two numeric data to show, the most effective way is to put the number in evidence, and not to use graphs. The use of simple text supported by a brief caption makes it possible to read the data more directly and quickly. However, when the volume of data is larger, the use of a table or graph is recommended [26]. Tables, on the other hand, usually work when comparing values. Colour maps, on the other hand, help reduce the mental process, as they use colour saturation to highlight points of interest [26].

Dashboards have the potential to reduce the probability of errors in their design [26]. Design elements such as the meaning of colour, visual coherence, information hierarchy or noise suppression are central concerns in the designer's activity. To achieve optimal results, is important to ensure construction of dashboards with **good readability** for the user and a **consistent layout** with the graphic standards of the respective company. There are various guidelines for dashboard interface design, under which **page organisation, colour, and typography** [26].

A) Page organization

When building a dashboard, the use of a grid system is an important support in the organization and coherence of the various graphic elements. As well as the concern for the existence of blank spaces in the layout in order to allow visual breaks and contribute to the focus on relevant information [26]. It is important to define the most appropriate location for the graphics and the amount of information that can be placed on the dashboard. Trying to put as much information as possible on a dashboard increases the risk of making the maximum information compression error, which makes it hard to read for users [26].

According to the Gutenberg Diagram (Figure 4), which demonstrates the reading behaviour of the western user (from left to right and from top to bottom), the main information should be inserted in the top left area of the layout, as it is in this area that the user will focus first [26]. The lower right area is the one that the user will pay least attention to [26].

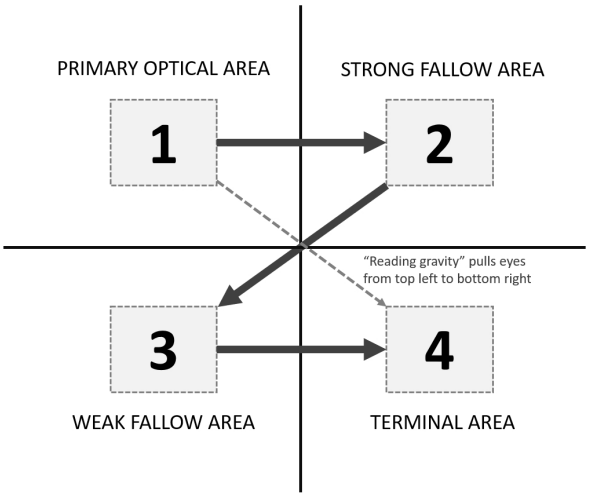


Figure 4: Gutenberg Diagram [30]

The creation of blank spaces is central to a comfortable interaction with the dashboard. Whitespace allow users to define hierarchies in order to distinguish information. A high density of information, without white spaces at the outer edges of the layout or between the elements, can make the reading of the information and the user's concentration harder [26].

B) Colour

Most dashboards make inexplicable and impulsive use of colour that doesn't add significance. Colour should be used in moderation. Dashboards should initially be designed in shades of grey, and only then should colour be added gradually so that they can convey useful information. Concerns such as increasing colour saturation to create highlights are precautions that should be taken into account when designing a dashboard [26].

It should also be noted that colour influences usability issues by its influence in capturing the attention of the user. Colour is a means that contributes to the prioritisation of information, helping to improve user interaction with the dashboard [26]. For complex data, a dashboard should consist of minimalist graphical elements with soft colours for easy reading. [26].

C) Typography

Readability of text should be the top priority. For example, headings that name the main sections of the dashboard should be highlighted in the typography through different scales or weights and the option for the adoption of a serif or sans serif typeface should be taken at the beginning of the construction of the dashboard, always preserving the coherence and readability of the layout [26].

Typography should be used to draw attention to the most important data. For example, if a specific number or trend is the key insight, the text presenting it should stand out through size, weight, or style. In other words, the way the typography is styled should always reflect what the designer wants users to notice first [26].

2.6 Literature Review Takeaways

This literature review started by exploring the concept of B2B customer training evaluation. Although the overarching goal of training is to enhance employee performance and promote skill transfer across contexts, its evaluation remains challenging due to the qualitative nature of learning and user experience. When these elements are assessed through numerical scales, cultural interpretation and individual perception can introduce bias, complicating the reliability of quantitative outcomes [11, 13]. In today's B2B landscape, effective training evaluation is essential to ensure that learning outcomes translate into real-world performance improvements, particularly through knowledge transfer.

This study, using a three-year dataset from Baker Hughes, aims to interpret existing metrics, including NPS and eight quantitative indicators, through Kirkpatrick's four-level framework to better understand learner sentiment and training impact. Rather than introducing new metrics, the goal is to create a user-friendly, multi-metric tool that reveals deeper insights into training participant feedback. With such a tool, academics emphasize the importance of clearly defining what is being measured, why it is being measured, and who will interpret the results.

While the Net Promoter Score (NPS) has gained widespread popularity as a simple, actionable measure of customer loyalty, it faces conceptual and empirical limitations when used as a standalone metric. Conceptually, NPS oversimplifies the complex, multidimensional nature of customer and learner experiences, lacks diagnostic value, and can be misleading without contextual interpretation. Empirically, although some studies show a correlation between NPS and sales growth, broader academic research reveals inconsistent findings and questions its predictive power compared to other customer attitude metrics. These limitations suggest that relying solely on NPS may obscure critical insights needed for strategic improvement. These findings underscoring the need for a multi-metric, data-driven framework to better capture learner sentiments and to inform business decisions.

While raw data holds little value without interpretation, combining multiple metrics allows for deeper insight into learner experiences and training effectiveness. This method, known as data triangulation, helps cross-validate findings, reduce bias, and uncover patterns that may otherwise be missed. However, the usefulness of metrics depends on their quality, relevance, and consistency. Selecting the right metrics and understanding their interrelationships is essential for generating actionable insights. Although more data can add complexity, it also increases the likelihood of identifying critical issues. For these reasons, a multi-metric framework was chosen as the foundation for this study, with the next step focusing on how best to visualize the data to support clear and informed decisions.

Dashboards have proven to be the most suitable approach for this study, offering a practical solution to information overload by gathering key metrics into a single visual interface that supports informed decision-making. Unlike single indicators such as Net Promoter Score (NPS), dashboards allow for the integration and cross-validation of multiple data points, making it easier to uncover patterns and derive meaningful insights. While automation and

organizational memory can enhance responsiveness and evaluative learning, the focus of this study is designing a user-friendly tool to interpret NPS feedback, making automation a secondary concern. Power BI was selected as the platform due to its interactivity, scalability, and ability to support dynamic data exploration, aligning well with the multifaceted nature of training evaluation.

Ultimately, the effectiveness of a dashboard depends not only on the technology used but also on how the information is structured and presented, with user experience and user interface design playing a significant role. Dashboards should be developed through an iterative design process that includes the evaluator at every stage. This ensures the final product is tailored to their specific needs. The design must be guided by the structure and meaning of the data. Graphic elements should be organized to support clarity and usability. The Gutenberg diagram is recommended to guide visual flow, and white space should be used to enhance readability. A grayscale layout is preferred in the initial stages to establish hierarchy. Once the structure is finalized, color can be introduced to highlight key components and attract attention. For complex data, visuals should remain minimal, and soft colors are advised to maintain focus. Typography may be adjusted in size and weight to emphasize important insights and direct attention effectively.

3 Methodology

This chapter outlines the methodology used in this research, which follows the Design Science Research (DSR) paradigm. It also highlights the methodologies that will be used to identify stakeholders, to gather stakeholder requirements, and to highlight the method that will be used to develop the tool. Finally, it details the evaluation methodology.

3.1 Design Science Research

This thesis will be executed according to the Design Science Research (DSR) framework. DSR is an iterative problem-solving approach that aims to enhance human knowledge by creating innovative artifacts [31]. The focus of DSR is on creating useful, effective solutions that can be applied in practice, rather than solely theorizing or studying existing phenomena [31][32]. The Design Science Research (DSR) frameworks includes three cycles: the relevance cycle, Rigor cycle, and the design cycle. These cycles are illustrated in figure Figure 5. Through DSR, researchers actively engage with stakeholders, gather requirements, and design tools that can be implemented and tested.

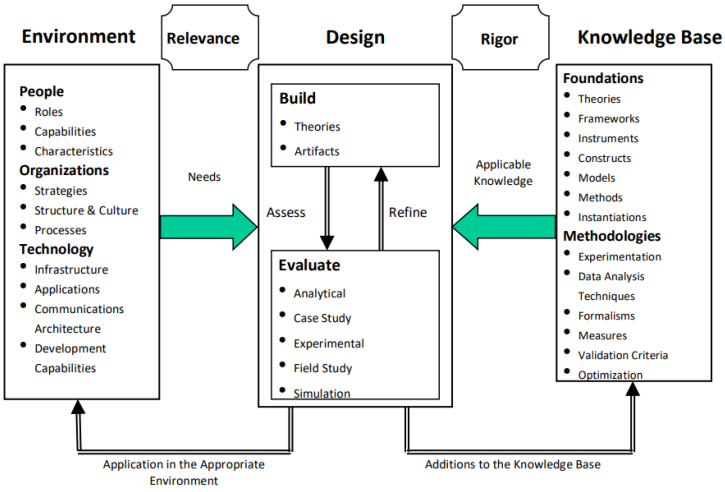


Figure 5: Design Science Research Framework [31]

The DSR process starts by gaining a clear understanding of the problem within the contextual environment of the research project. During this initial phase, known as the Relevance Cycle, requirements are gathered and objectives are established, considering what is feasible. The Rigor Cycle connects the design science activities with the knowledge base that includes scientific foundations, experience, and expertise that informs the research project. The Central Design Cycle supports a loop of research activities that are focused on constructing and evaluating design artifacts and processes. This cycle utilizes the information from the contextual environment and the gathered knowledge from literature to make informed decisions on how to design the artifact.

In the context of this study, the Design Science Research process began by exploring the real-world issue present within Baker Hughes and establishing project objectives. Through internal conversations within the organization, it was identified that the problem is that the NPS does not provide sufficient context to identify specific course strengths and weaknesses. To address this issue, the objectives were establised with stakeholders responsible for assessing the course quality at Baker Hughes. This process aligns with the **Relevance Cycle**.

Simultaneously, academic research has been conducted to gain a deeper understanding about training evaluation practices, the NPS, triangulation of metrics, and HCI and data visualization principles. This is the theoretical foundation of this study, and therefore represents the **Rigor Cycle**. This cycle was presented in *Section 2: Literature Review*.

Finally, a dashboard will be developed based on knowledge drawn from the Rigor Cycle, as well as the problem definition and objectives established during the Relevance Cycle through interviews with key stakeholders. In this **Design Cycle** phase, dashboard iterations will be proposed to key stakeholders in order to understand how it can be further refined.

3.2 Stakeholder Identification Technique

The identification of stakeholders is essential in the relevance cycle to understand which stakeholder group is important to consider during evaluation. To identify stakeholders, internal employees of the customer service team will be consulted. Once key stakeholders have been identified, the stakeholders will be analysed based on their role, interest, and influence [33]. The interest in the realization of the multi-metric dashboard and the influence in eventually using it will be either low, medium, or high.

To understand the attitudes of each individual stakeholder, the following question can be asked: *'To what extent does the stakeholder wish the data visualization tool to be produced, and how active is the role of this stakeholder to eventually use this?'*

Identified Stakeholder	**Interest**	**Influence**	**Label**

Table 1: Overview of identified stakeholders, their roles, interests, and influence.

After applying this analysis, the interest-influence matrix will be presented to identify the stakeholder that have most interest in the realization of the dashboard. The stakeholder attitudes will be presented in Table 1 to record stakeholders' attitudes toward the dashboard. The stakholders are then labeled as presented in Table 2.

Interest	**Influence**	**Description**
High	High	Manage closely: Should be closely engaged and managed actively
High	Low	Keep informed: Should be consulted and involved
Low	High	Keep satisfied: Should be monitored and involved selectively
Low	Low	Minimal effort: should be monitored occasionally

Table 2: Stakeholder labels

Once the table is filled in, a graphical matrix will be created such as depicted in Figure 6.

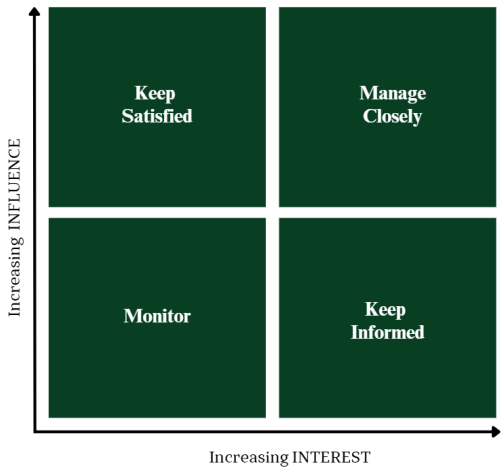


Figure 6: Stakeholder Influence Interest Matrix

3.3 Stakeholder Requirements Gathering

The aim of this study is to design, develop, and evaluate a dashboard together with the identified stakeholders. Since the design depends on their needs, a clear approach is required to gather stakeholder requirements. To understand their attitudes and pain points with the current training evaluation, feedback will be collected through interviews before starting the design of the multi-metric data visualization tool. See *Appendix II* for the interview guide.

Interview techniques include unstructured, semi-structured, and structured formats [34]. For this project, **semi-structured interviews** are the most suitable, as they allow for follow-up questions when a participant's response reveals additional insights. This approach fits well with the goal of identifying stakeholder frustrations and requirements for the new dashboard.

Gathering requirements from key stakeholders and literature is essential to develop a tool that meets its intended purpose. To ensure the most relevant aspects are prioritized, this study applies the MoSCoW prioritization method, which categorizes requirements as must-have, should-have, could-have, or would-have [35]. Must-haves are critical for the tool to be considered acceptable, while the remaining categories guide the inclusion of additional features based on their importance. The results will be presented in the format shown in Table 3.

Requirement	**MoSCoW**

Table 3: Requirement Prioritization Method

3.4 Data Handling Approach

Developing analytical tools and workflows involves handling and preparing data. A key part of this process is **Extract, Transform, and Load (ETL)**, which combines data from multiple sources into a central repository known as a data warehouse [36]. ETL applies rules to clean and organize raw data for analysis. In this study, ETL will be used to transform Excel datasheets into structured information within Power BI.

The primary data input consists of three Excel files containing survey responses from course participants. To assess data quality, an initial analysis of null values will be conducted in **Jupyter Notebook**, using Python libraries such as pandas and numpy. Based on the results, decisions will be made on whether to exclude certain metrics or handle missing values through removal or imputation.

Although Jupyter Notebook supports data cleaning, normalization, and restructuring to match the data warehouse model, it requires technical expertise for ongoing use. To ensure long-term accessibility, **Power Query** will be used for the actual data transformations. However, the initial quality assessment will still be performed in Jupyter Notebook, as Power Query lacks the analytical depth needed for this step.

Once the data from the three Excel files is cleaned and combined in Power Query, the data modelling view in Power BI will be used to structure it into a suitable warehouse schema. Finally, the visualization view in Power BI will support the creation of interactive dashboards and reports, with the platform connecting directly to the data warehouse for dynamic analysis.

While automation is not a core focus of this study, it is recommended in the literature for enabling real-time insights. If time permits, an automated data pipeline will be set up using an online **SharePoint folder**, connected to Power Query and Power BI, with scheduled refreshes managed through **Power Automate**. The automation approach is depicted in Figure 7.

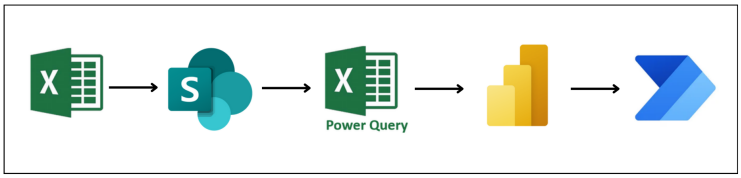


Figure 7: Automated Data Pipeline Diagram.

3.5 Evaluation Method

As previously mentioned, the dashboard will be developed using an iterative Design Science Research (DSR) approach. This method enables continuous refinement based on user feedback, to ensure the final product aligns with the evaluator's needs and supports effective data interpretation. The goal is to deliver a user-friendly, multi-metric dashboard that provides deeper insights into participant feedback.

To determine whether the dashboard meets these goals, the evaluation will address three key perspectives: **system usability**, **user experience**, and the ability of employees to **interpret participant feedback** using the tool. Additionally, a requirement check will be conducted with the evaluator to verify that all previously defined needs have been met. The evaluation methods are summarized in Table 4.

Evaluation Method	Participants	Description
Task-Based Usability Test	8–10 Baker Hughes employees, including the evaluator	Participants complete a series of tasks using the dashboard in a controlled setting, followed by a Google Forms questionnaire.
Requirement Check	Evaluator	Joint review of the dashboard to confirm alignment with stakeholder requirements.

Table 4: Evaluation Approach

3.5.1 Task-Based Usability Test

The task-based usability test focuses on evaluating the dashboard's **system usability**, **user-friendliness**, and **overall effectiveness**. In the context of training evaluation, effectiveness refers to whether participants can:

- **Granularity**, identify specific reasons behind course participant feedback
- **Correlation**, recognize patterns and relationships across metrics
- **Narrative**, construct a coherent explanation of the NPS score using dashboard insights

Each session lasts approximately 30 minutes and involves participants completing six tasks using the dashboard. Participants are selected internally from Baker Hughes to maintain data confidentiality. A sample size of 8–10 is targeted, with efforts made to ensure diversity in demographics, data literacy, and familiarity with NPS, balancing gender, age, and professional roles. The interview guide for this usability test is provided in *Appendix III*.

Sessions take place at the Baker Hughes office in Florence, in a reserved room equipped with a laptop containing the dashboard and Google Forms. If in-person attendance is not feasible, remote sessions are facilitated via Microsoft Teams. To avoid bias, participants are not incentivized.

The session begins with a Google Form covering informed consent and pre-study questions on familiarity with NPS and dashboards. A brief explanation of NPS is then provided to ensure a shared understanding. Participants are given one minute to familiarise themselves with the dashboard and ask questions before completing the six tasks. Each task is presented verbally and supported by a visual prompt to minimise misinterpretation. The tasks are designed to assess how effectively the dashboard supports data-driven decision-making.

A) Information Retrieval Tasks

- What was the overall NPS score in 2023 quartile 4?
- How many training sessions scored below the company threshold in terms of NPS in 2024, and how many participants followed these sessions?

B) Trend Analysis Tasks

- Is training performance improving or declining in terms of NPS over time? Please show how you find it.
- For Course X, what performance metric is the biggest weakness?
- Identify the key positive and negative sentiments expressed by learners in Course Y.

C) Problem Identification Tasks

- You have a budget to improve only three training sessions. Which ones would you choose and what would you improve?

These tasks directly reflect the dimensions of dashboard effectiveness outlined earlier, **granularity**, **correlation**, and **narrative**. Each question is designed to test whether users can extract meaningful insights, identify trends, and make informed decisions. To complement the task-based evaluation, participants will complete a Google Form containing both qualitative and quantitative questions. The form includes three components:

- **System Usability Scale (SUS)**, a ten-item questionnaire assessing perceived usability, with responses rated on a five-point Likert scale from "strongly disagree" to "strongly agree" (Figure 8).
- **User Experience Questionnaire – Short Version (UEQ-S)**, an eight-item scale measuring both pragmatic and hedonic aspects of user experience (Figure 9). Pragmatic aspects refer to the usability-related qualities of the dashboard, such as clarity, efficiency, and goal-oriented functionality. Hedonic aspects refer to the emotional and aesthetic appeal, including how engaging, enjoyable, and visually pleasing the dashboard feels to users.
- **Interpretation Evaluation**, custom Likert-scale questions designed to assess participants' ability to interpret dashboard data meaningfully and actionably (Figure 10).

SUS scoring involves alternating positive and negative statements. For positively worded items (1, 3, 5, 7, 9), the score is calculated by subtracting 1 from the response. For negatively worded items (2, 4, 6, 8, 10), the score is derived by subtracting the response from 5. The total is multiplied by 2.5 to yield a score from 0 to 100. Scores above 68 indicate above-average usability, while scores below may suggest usability concerns.

		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	I think that I would like to use this website frequently					
2	I found the system unnecessarily complex					
3	I thought the system was easy to use					
4	I think that I would need the support of a technical person to be able to use this system					
5	I found that the various functions in this system were well integrated					
6	I thought there was too much inconsistency in this system					
7	I would imagine that most people would learn to use this system very quickly					
8	I found the system very cumbersome to use					
9	I felt very confident using the system					
10	I needed to learn a lot of things before I could get going with this system					

Figure 8: System Usability Scale Questionnaire

obstructive	o o o o o o o	supportive
complicated	o o o o o o o	easy
inefficient	o o o o o o o	efficient
confusing	o o o o o o o	clear
boring	o o o o o o o	exciting
not interesting	o o o o o o o	interesting
conventional	o o o o o o o	inventive
usual	o o o o o o o	leading edge

Figure 9: User Experience Scale Questions - Short Version

UEQ-S responses are mapped to a scale from –3 to +3. Items 1–4 assess **Pragmatic Quality** (e.g., efficiency, clarity), while items 5–8 measure **Hedonic Quality** (e.g., enjoyment, aesthetics). Mean scores above +0.8 indicate a positive experience, while scores below –0.8 reflect negative perceptions. An automated Excel will be used to conduct the analysis [37].

		Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
1	I was able to identify reasons to explain course participant feedback					
2	I can recognize trends across data					
3	I gathered insights into why a course was good or bad					
4	I was able to identify how course participants felt about a training					
5	The goal of the dashboard is clear					
6	The dashboard helped in answering my questions					
7	The dashboard provided enough context					
8	I am confident in the conclusions I made					

Figure 10: Interpretation of learner experience

To assess dashboard effectiveness across granularity, correlation, and narrative, the **task description**, **task completion rate**, and **key findings** are presented in Table 5.

Task	Description	Completion Rate	Key Findings
T1			
T2			
T3			
T4			
T5			
T6			

Table 5: Summary of task completion rate and key findings

3.5.2 Criterias for success

The granularity, correlation, and narrative criteria evaluate how effectively the dashboard enables users to interpret participant feedback from three complementary perspectives. Granularity measures the user's ability to pinpoint specific training issues, ensuring the dashboard supports detailed diagnostic insights rather than broad generalizations. Correlation assesses whether users can identify meaningful relationships between metrics, which is essential for recognizing patterns and drawing informed conclusions. Narrative evaluates the user's capacity to contextualize Net Promoter Score (NPS) data and formulate actionable recommendations, highlighting the dashboard's role in guiding strategic decisions. For these three dimensions, a minimum success threshold of 80% is set to confirm that the dashboard fulfills its intended analytical and interpretive goals.

Metric	Target
Task Completion Rate	Target > 80%
SUS-Score	Target > 68%
Granularity	> 80% of users can retrieve information from the dashboard
Correlation	> 80% of users recognize metric relationships in trend analysis tasks
Narrative	> 80% of users can explain NPS context and provide actionable recommendations

Table 6: User Experience Evaluation Metrics

3.5.3 Requirement Assessment

Lastly, requirements that were established after interviews with stakeholders, will be judged on their integration in the resulting dashboard. The requirements will be marked with one to three stars based on their integration. One star refers to almost no integration, meanwhile three stars means success in integration of the requirement in the final product. The rating of these requirements will be conducted together with the evaluator, as this person will eventually frequently interpret the data in the dashboard.

4 Tool Design and Development

In this chapter, the methods discussed in Chapter 3 are applied to identify stakeholders, gather pain points, and collect and prioritize requirements using the MoSCoW method. The available dataset is analyzed using the Kirkpatrick framework to evaluate the relevance and impact of additional metrics. Insights from earlier sections combined with the stakholder needs help to define the dashboard goal and the key quesions it should address. Concepts are then developed through an iterative process. Finally, one final concept is selected.

4.1 Stakholder Identification

The stakeholders identified in the context of this study include stakeholder A, B, and C, as well as the evaluator, trainers, and project managers. Some roles are presented with a degree of abstraction to focus on contributions rather than individual identities. Their perspectives are discussed in the following section and summarized in Table 7.

A) Stakeholder A

Stakeholder A is responsible for ensuring the quality and relevance of course content across the Baker Hughes training portfolio. This involves overseeing course design and maintaining alignment with intended learning outcomes. When trends indicate that a course is underperforming or misaligned, it falls under the responsibility of this stakeholder to address these issues. This stakeholder takes an active role in the training quality assessment process.

Given the strategic focus on continuous improvement, stakeholder A has a **high interest** in using the training evaluation dashboard to support portolio-wide quality assessments. Due to their high involvement in decision-making, this stakeholder has a **high influence**. Therefore, this stakeholder should be **closely managed** in this dissertation.

B) Stakholder B

This stakeholder is responsible for monitoring KPI's related to training delivery. A dashboard for training evaluation would greatly benefit this stakeholder by simplifying this process. Currently, when an NPS score is low, this role must wait for the evaluator to conduct a deep dive analysis, which is time-consuming. Therefore, this role has a **high interest** in an automated dashboard. Furthermore, should a course receive a low NPS due to trainer-related issues, improvements in this domain need to be initiated by this role due to their **high influence**. All in all, this role needs to be **closely managed** in this dissertation.

C) Stakeholder C

Stakeholder C manages the commercial team. This team's main responsibility is to connect clients and their learners with the appropriate training and project managers. If a course turns out to be unsuitable for the client due to misaligned expectations, stakeholder C becomes the main contact for resolving this issue. However, since these situations are not directly

related to course quality, this stakeholder is not actively involved in evaluating trainings. Therefore, stakeholder C is classified in the **minimal effort** category with a **low interest** and **low influence**.

D) Evaluator

The evaluator compiles and presents training evaluation data. Although this role does not use the information from the dashboard to inform training-related actions, their work directly supports its purpose. This stakeholder has a **high level of influence**, but **low interest** in the outcomes of the tool. They should be **kept satisfied** to make sure that training evaluation data support is provided smoothly.

E) Training Instructors

Trainers are responsible for delivering courses and to ensure that sessions are engaging and effective for participants. While they are not directly involved in the dashboard evaluation process, they are indirectly affected by its outcomes. If a low Net Promoter Score (NPS) is linked to inadequate trainer skills, trainers will be notified.

Although trainers are not directly part of the training evaluation process with the dashboard, they are present at the location where customers complete evaluation forms. For this reason, trainers can provide valuable insights into participant feedback. They can assist in interpreting responses and identifying any underlying issues. Given their teaching responsibilities, they have **high interest**, but they lack decision-making power, resulting in **low influence**. They should be **kept informed**.

F) Project Managers

Project Managers handle the logistical aspects of trainings. They are notified when logistical issues impact NPS scores, but this stakeholder will not directly use the dashboard. Therefore, given their responsibilities, they are placed in the **minimal effort** category, as they have a **low interest** and **low influence**.

Identified Stakeholder	**Interest**	**Influence**	**Label**
Stakeholder A	High	High	Manage Closely
Stakeholder B	High	High	Manage Closely
Stakeholder C	Low	Low	Minimal Effort
Evaluator	Low	High	Keep Satisfied
Trainers	High	Low	Keep Informed
Project Managers	Low	Low	Minimal Effort

Table 7: Overview of identified stakeholders, their roles, interests, and influence.

The interest-influence attitudes of these identified stakeholders are gathered and depicted in a matrix which is presented in Figure 11.

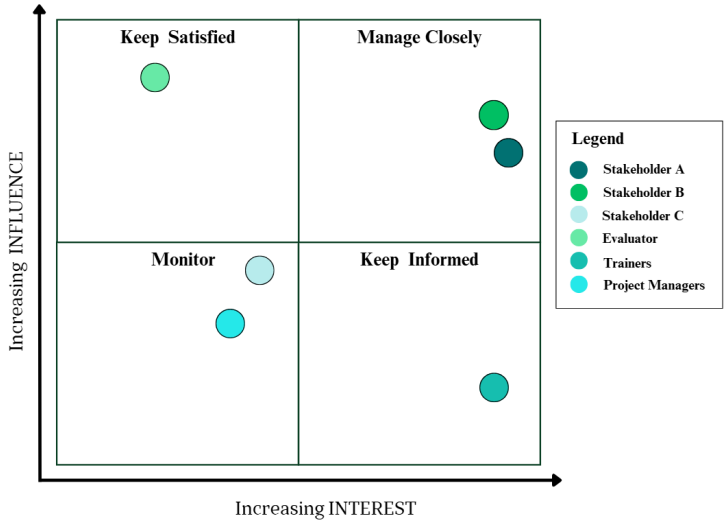


Figure 11: Stakeholder Influence Interest Analysis Matrix

4.2 Current course training evaluation method

Within the customer training team, the current training evaluation approach is primarily focused on the Net Promoter Score. The company has established several internal thresholds. When an NPS is below a confidential benchmark, a more in-depth analysis is executed by the evaluator, often in assignment of Stakeholder A and B. This deeper analysis focuses mostly on training participant comments and the supplementary metrics within the dataset.

4.3 Pain points with current course quality assessment method

Interviews with key stakeholders revealed various pain points in the current training assessment practices. The primary issue is the heavy reliance on the NPS. Other metrics are reviewed only when the NPS gets below the confidential benchmark, which can result in blind spots. Courses with average or high NPS scores can still have underlying issues that can go unnoticed simply because the overall rating seems satisfactory.

A second pain point is that NPS only provides a general overview and that it does not provide the depth needed to identify the reasons behind a trainings' success or failure. The current approach, which is mainly focused on obtaining a high NPS, is challenging because responses often lack context needed to support decision-making. This makes it difficult to determine whether a problem is a unique issue or if it forms a broader structural problem.

Data integrity issues also presents challenges. Low response rates or very small training groups can highly impact NPS results. In such cases, the feedback of a single detractor can lower the NPS significantly, even when the majority of responses are neutral or positive. This sensitivity of outliers in small group context makes the reliability of the metric even less trustworthy as a standalone metric. Therefore, triangulation of metrics is necessary, especially in a training context.

Stakeholders also raised concerns about inconsistent interpretation of the NPS scale. Cultural differences and varying explanations by instructors can influence how participants rate their experience. For example, scores of 7 or 8 may be considered positive in some cultures but are classified as passive in the NPS framework. This lack of standardization diminishes the comparability of feedback across regions, courses, and learner groups.

Another issue is the tendency for participants to base their ratings primarily on their satisfaction with the trainer rather than the course content. This misalignment can skew results and further reduce the reliability of NPS. Adressing this is important for developing a more comprehensive training evaluation dashboard.

Stakeholders also highlighted that the current approach is very time-consuming due to manual handling of the customer training evaluation data.

Finally, the absence of an aggregate view accross courses and trainings limits the ability to detect trends and broader issues. Performance indicators like instructor effectiveness and documentation quality, are viewed only on a training-by-training basis rather than across multiple courses and trainings, making it difficult to identiy recurring patterns or areas for improvement over time. The lack of filtering capabilities further complicates analysis, as feedback cannot be easily segmented by instructor, course type, or location.

Despite these limitations, it was emphasized that NPS remains a fundamental metric for understanding how training is performing across all courses over time. It is viewed it as a high-level indicator, while other metrics are needed to conduct assessment in more depth. This layered evaluation approach still suggests a need for a multi-metric dashboard that can support both broad and detailed analysis.

4.4 Requirements

Within both the literature review and interviews with the key stakeholders, requirements for the dashboard have been identified. The requirements gathered from literature are presented in Table 8. Furthermore, the requirements gathered from interviews are presented in Table 9.

Requirement	**MoSCoW**
Defining a clear dashboard goal and aligning stakeholders on its purpose	Must
Answering specific, data-driven questions based on available metrics	Must
Enabling triangulation by integrating multiple metrics for comparison	Must
Maintaining a focused and uncluttered dashboard layout	Must
Involving evaluators in the design iteration process	Must
Including interactive features such as filtering and drill-down	Must
Applying visual hierarchy to guide user attention	Should
Using consistent color schemes and readable typography	Should
Providing contextual information and comparative views for deeper insights	Should

Table 8: Gathered requirements from literature

Requirement	MoSCoW
Analyzing all courses, not just those with low NPS	Must
Detecting trends across time, courses, instructors, and locations	Must
Maintaining brand consistency with Baker Hughes guidelines	Must
Cross-validating NPS with other evaluation metrics	Must
Identifying recurring issues across training programs	Must
Providing an aggregate view of quality indicators across all courses	Must
Automating data reporting processes	Should

Table 9: Gathered stakeholder requirements from interviews

4.5 Dataset Analysis

To establish a meaningful dashboard, it is necessary to understand the structure and scope of the available dataset in this study. The Baker Hughes training evaluation dataset contains 14 general parameters, the NPS, and eight evaluation-specific metrics. Table 10 presents the general parameters that should be explored to understand the queries that can be formulated within the dashboard. For confidentiality reasons, only parameter names will be shown.

Parameter	Definition
Training Job	Training purchasing order number of a client
Customer	Company that purchased the training
End User	Company of whom the employees attend the training
Course Type	Classification under mechanical, control, or electrical domain
Region	Where the client is from
Course	The course that was conducted
Start Date	What day did the training start
End Date	What day did the training end
Training Mode	Classification under theoretical, on-job, virtual, hands-on mode
Quarter	In what quarter of the year was the course being held
Instructor	Who was the instructor throughout the course
Trainer's Company	What company is the trainer from
Resources	Is it an internal or external instructor
Venue	In what venue was the training held

Table 10: Definitions of general parameters within the dataset

In addition to the general parameters, the dataset includes the NPS and eight evaluation-specific metrics that serve as inputs for assessing the quality of training. Table 11 presents the eight metrics and maps them to the Kirkpatrick framework introduced in *Section 2.1*. This theoretical grounding is important because it addresses a limitation earlier identified in literature, which is that there is often a tendency to focus solely on participant reactions rather than learning outcomes.

Evaluation Metric	Kirkpatrick Level	Explanation
Course Introduction	Level 1: Reaction	Initial engagement and orientation effectiveness
Course Duration	Level 1: Reaction	Appropriateness of time allocation
Training Logistics	Level 1: Reaction	Venue, scheduling, and operational satisfaction
Instructor Communication	Level 1: Reaction	Delivery style and clarity perception
Course Interaction	Level 1: Reaction Level 2: Learning	Engagement and possibly learning facilitation
Instructor Competence	Level 1: Reaction Level 2: Learning	Expert knowledge transfer capability
Course Consistency	Level 2: Learning	Structured learning environment supporting comprehension
Training Documentation	Level 1: Reaction Level 2: Learning	Quality of learning materials supporting knowledge retention

Table 11: Mapping of Baker Hughes Evaluation Metrics to Kirkpatrick's Four Levels

This mapping reveals important insights regarding the depth of the analysis that it can reach. Four metrics align purely with Level 1 (Reaction), while four metrics also extend into Level 2 (Learning). This means that the dashboard will be able to provide indications of participant engagement and potential indications of knowledge acquisition. Although the dataset does not contain metrics that relate to behavior change (Level 3) or organizational results (Level 4), the presence of Level 2 metrics provides a bridge between immediate reactions and potential learning impacts.

The Net Promoter Score (NPS) is perceived as a high-level indicator of participant loyalty and their likelihood to recommend a course (*Section 2.2: NPS*). In a B2B context, even if participants are not the final decision-makers over purchasing decisions, they can still influence outcomes by sharing positive feedback with the management in their organization. Their positive feedback can influence internal discussions. Therefore, even in a B2B setting, participant loyalty plays an important role in training revenue growth.

As documented in literature as well as stakeholder interviews, the NPS as a standalone metric lacks diagnostic depth in explaining *why* a participant may rate a session positively or negatively. By triangulating NPS with the eight supplementary evaluation metrics, the dashboard provides a deeper interpretation of training evaluation feedback through a multi-dimensional analysis. Examples of diagnostic capabilities are as follows:

- Low NPS + poor Course Interaction + low Instructor Competence = delivery effectiveness issues
- High NPS + strong Training Documentation + high Course Consistency = effective learning environment
- Moderate NPS + low Course Duration + high Course Introduction = timing optimization needed

This multi-metric approach enhances the training evaluation process by enabling both aggregated views for strategic oversight and filtered perspectives for targeted improvement initiatives. Each metric is visualized both as an averaged total score and independently in radar chart format, allowing users to explore correlations, identify patterns, and develop strategies that help with course improvement.

4.6 Dashboard goal

Setting a clear goal for the dashboard and thinking of questions the dashboard should answer is both important. These aspects represent the foundation of any effective dashboard design, as emphasized in *Section 2.1. Training Evaluation*. The evaluator and key stakeholders were closely involved to ensure the dashboard goal aligns with training objectives.

At first, the dashboard goal was broadly defined as “understanding course quality.” However, through deeper stakeholder consultations, it became clear that the main objective of training is to support the knowledge transfer. Therefore, training evaluation should not only assess the satisfaction of participants, but also whether the training experience prepared participants for job-relevant skill application. It is therefore necessary to align course content with job needs. In addition, trainings should also be delivered in a way that meets learner expectations.

> "To help users in identifying trends and patterns that reveal the effectiveness of training delivery in supporting participant job readiness and knowledge transfer, moving beyond satisfaction metrics to actionable insights for systematic training improvement."

Based on stakeholder needs and theory, the dashboard will address the questions presented in Table 12. The table prioritizes questions that directly address stakeholder pain points. Data transformations and data modelling both have been applied to the dataset in accordance with the goal and the questions that the dashboard should answer.

Question	Method	Priority
What is our overall training effectiveness across all programs?	NPS + Overall satisfaction	High
Which training programs are failing and need immediate attention?	Exception reporting table	High
Is our training performance improving or declining over time?	Courses requiring attention table	High
Are our training programs preparing participants for job readiness?	Level 2 metrics analysis	High
What are the main weaknesses in our training delivery?	Radar chart analysiss	High
What specific issues are causing course failures?	Key issues in table	Medium
How many participants are affected by poor training?	Participants in low-performing courses	Medium
Which regions are underperforming and need support?	Regional performance comparison	Medium
Are certain instructors consistently delivering poor results?	Filter by instructor + performance	Medium
Do specific training modes work better than others?	Mode comparison analysis	Low
Which course types are most/least effective?	Type-based filtering	Low
Are there seasonal patterns in training effectiveness?	Quarterly trends	Low
Which customers consistently provide lower evaluation scores?	Customer filtering	Low
Are newer courses performing better than established ones?	Course name filtering	Low

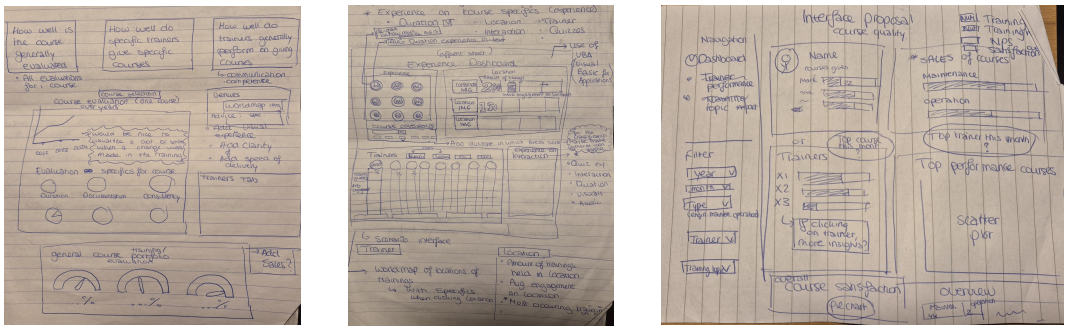
Table 12: Training Evaluation Questions by Method and Priority

4.7 Data Visualization Iterations

The dashboard design followed an iterative design approach. Each prototype was evaluated by key stakeholders to ensure the dashboard design is aligned with both theory and stakeholder needs. This section documents the iterations from initial concept to final dashboard design.

Prototype 1

The first prototypes were created as low-fidelity sketches, exploring all possible visualizations using the available dataset (Figure 12). These prototypes were meant as initial brainstorm and were therefore not shown to stakeholders.

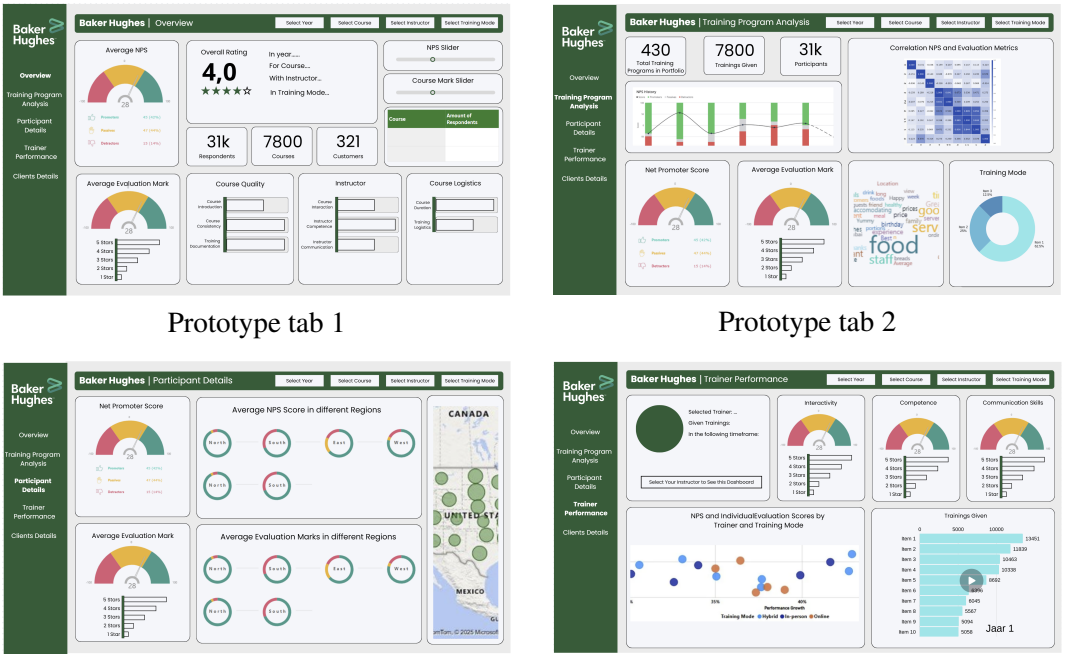


First Prototype | Second sketch iteration | Third sketch iteration

Figure 12: Overview of early prototype iterations

Prototype 2

After the three sketch iterations, a Low Fidelity prototype was created within Canva. This prototype attempted to separate and organize tabs for a better navigation (Figure 13).



Prototype tab 1 | Prototype tab 2

Prototype tab 3 | Prototype tab 4

Figure 13: Overview of Prototype tabs

Stakeholders appreciated the analytical depth but had several critical concerns. First of all, it was mentioned that there is an information overload accross multiple pages. In addition, there is a lack of clear prioritization among metrics and views. This hinders decision-making.

This feedback is aligned with literature as effective dashboards should minimize cognitive load while maximizing insight generation. Therefore, a dashboard with a more focused approach, could help to create focused simplicity.

4.8 Final Dashboard

1. Single-Page Dashboard Approach

Effective dashboards minimize cognitive burden, which is why a single-page can reduce navigation complexity. It also helps with quick comprehension and decision-making. For this reason, the final dashboard will contain a single-page design was chosen.

2. Data Visualization Choices

To faciliate filtering and aggregate views, a filter system will be applied on the left side of the screen. The filter slicers that will be used are: Course, Year, Quarter, Customer, Region, Instructor, Training Mode. These filters are selected to facilitate a broad analysis. As it is also important to know what the goal of the dashboard is, the final design will include a bar in the upper-screen which contains the dashboard title "Customer Training Evaluation", a subtitle containing the goal of the dashboard, and a text that shows when the dashboard was last updated. Knowing when the last update was, is essential for users to know if they are viewing an up-to-date version.

A horizontal arrangement of five key performance indicators, which are the NPS, an overall satisfaction score, the amount of course participants, the amount of training sessions, and the amount of customers. These KPI's reflect important numbers that should be easy to observe at a glance for users.

Additionally, since a goal within the research is to triangulate the NPS with the eight evaluation metrics, a radar chart will be added within the final dashboard design. A radar chart is suitable for simultaneously visualizing and comparing all the eight evaluation metrics. This visualization choice directly addresses the limitation of a standalone NPS by providing simultaneous visibility into all evaluation dimensions. In addition, such a radar chart illustrates areas of strenght and areas for improvement.

Obtaining an overview of trends was also identified as a stakeholder need. Therefore, the final dashboard should contain graphs that are focused on visualizing trends. A temporal visualization of NPS and Overall Satisfaction addresses the stakeholder-identified pain point regarding the absence of aggregate views across courses and time periods. The integration of NPS and satisfaction trends in a single chart supports comparative analysis while maintaining visual clarity. This supports pattern recognition and strategic planning.

According to stakeholders, people from different backgrounds perceive the evaluation scales differently. Therefore, a graph that illustrates the NPS and Overall Satisfaction per region can provide knowledge about how courses are perceived in various areas globally.

A "Training Sessions Requiring Attention" table represents a critical design decision that transforms reactive evaluation into proactive quality management. It implements the principle of management by exception, automatically identifying courses below a company threshold. This addresses the stakeholder concern about reactive rather than proactive quality management. This approach implements systematic threshold monitoring, provides actionable detail for improvement initiatives, and supports the transition from individual course analysis to program-level oversight.

3. Gutenberg Diagram Application

Information should be structured based on the urgency level. Therefore, critical KPIs will be shown first. These include the NPS and an overall satisfaction score. The overall satisfaction will be a score that averages all evaluation metrics. Graphs related to trend analysis comes right after. A table reporting the details should be presented last. Based on visual hierarchy and the Gutenberg reading patterns *(Chapter 2.5)*, the order of the elements within the dashboard should be presented as follows.

1. **Header** - With Dashboard title and context
2. **Filters** - For primary interaction controls
3. **KPI Cards** - Most critical metrics at a glance
4. **NPS and Satisfaction Trend** - Primary performance indicator over time
5. **Course Performance Radar** - Supplementary eight training effectiveness metrics
6. **Performance by Region** - Supporting information about performance by Region
7. **Courses Requiring Attention Table** - Supporting information presenting detailed actionable data

The layout follows natural reading patterns with critical information positioned in the primary optical area (top-left), supporting stakeholder requirements for immediate NPS and satisfaction visibility.

4. Brand Consistency

The Baker Hughes font and color scheme will be used. The colors serve both aesthetic and functional purposes as it helps to maintain brand identity, which is a stakeholder requirement, and which creates a visual hierarchy through color contrast.

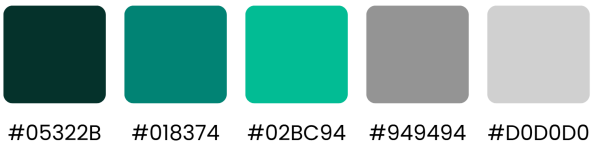


Figure 14: Brand colors

The high-fidelity final dashboard design is presented below. It is filled with a **fake dataset, which has no connection to Baker Hughes or any of its operations**. Please note that the dashboard itself is the **intellectual property** of Baker Hughes.

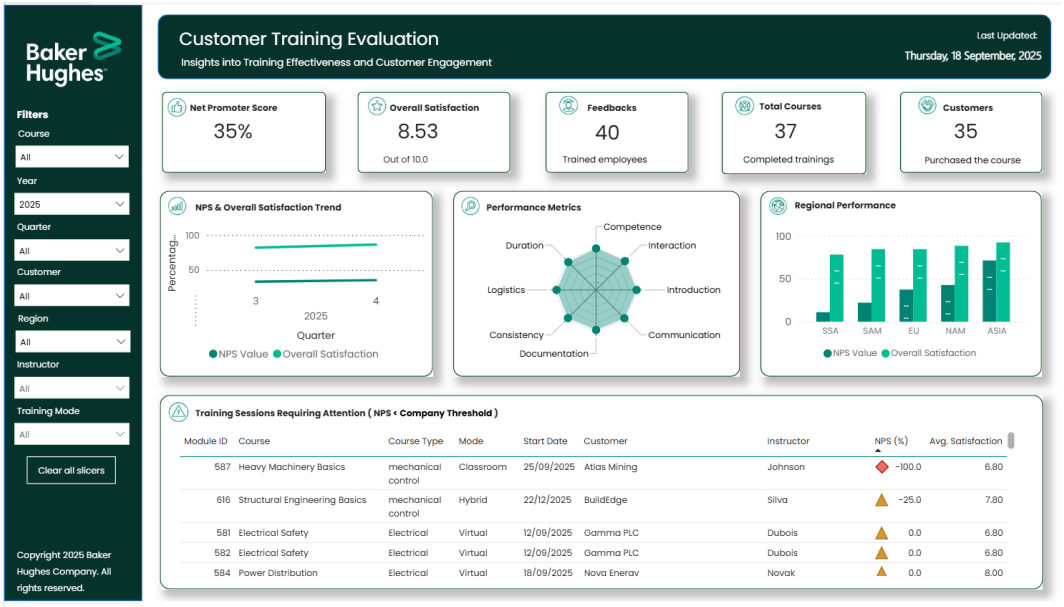


Figure 15: Final Dashboard Design

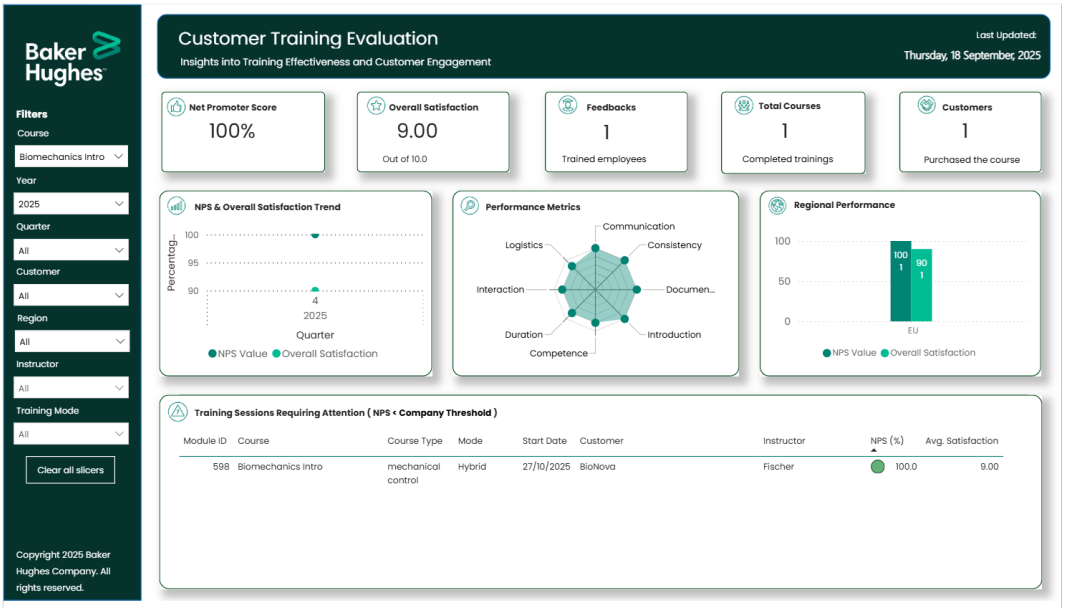


Figure 16: Final Dashboard Design - Course Selection

5 Evaluation and Results

This section presents the evaluation of the multi-metric training evaluation dashboard through task-based usability testing and stakeholder requirements validation. The evaluation addresses whether the dashboard is user-friendly and whether it improves the interpretation of training participant feedback beyond standalone NPS analysis.

5.1 Task-Based Usability Test

The dashboard was designed for a broad audience and should serve various stakeholders within the customer training team as mentioned in *Section 4.1*. To evaluate the usability of the dashboard across different user profiles, a task-based usability test was conducted with ten participants from the customer training team. The group ranged in dashboard familiarity, data literacy and NPS knowledge. This mix was intentional to assess effectiveness across experience levels, especially since the dashboard does not only supports data analysts but also other decision-makers.

Among participants, 10% had never used dashboards, 20% used them rarely, 30% monthly, 20% weekly, and 20% daily. Despite this range, 80% felt confident using data visualization tools. Regarding dashboard use for decision-making, 60% reported monthly usage, 20% weekly, and 20% rarely. Additionally, 70% were familiar with NPS, and 50% had prior experience in training evaluation. All participants found the evaluation data at least somewhat relevant to their roles, with half rating it as extremely relevant.

5.1.1 Task Completion Rate and Dashboard Effectiveness

The dashboard achieved a 92% overall task completion rate, which exceeded the predefined success criterion of 80%. The six tasks in the task-based study were designed to evaluate usability in terms of **information retrieval**, **trend analysis**, and **narrative construction**. Table 13 shows a summary of the task completion rate of the six tasks and summarizes the outcomes.

Task	Description	Rate	Findings
T1	Identify NPS score in 2023 Q4	100%	All participants quickly located NPS and satisfaction KPIs
T2	Find underperforming courses and corresponding participant count	90%	Most used table well; one struggled with filtering
T3	Analyze NPS trends over time	100%	Trends were intuitive
T4	Identify worst metric for a course	100%	Metric identification was intuitive
T5	Extract participant sentiments for a course	100%	Radar chart helped with interpretation
T6	Build improvement narrative	80%	Chose low-NPS courses and used metrics to support narrative

Table 13: Summary of dashboard usability tasks and findings

Tasks 1–5 were completed successfully by almost all participants, which indicates a strong performance in information retrieval and trend analysis. Only one participant with no prior experience with Power BI initially struggled with filtering in Task 2. However, after brief exploration, this participant adapted and completed the next tasks. This could explain the slightly lower completion rate for Task 2 (90%) and suggests that usability challenges were potentially more related to user familiarity. Another reason for this lower score could be that the dashboard did not provide enough clear cues to know how to complete this task. Overall, participants were able to navigate the dashboard to **retrieve information successfully**.

All participants demonstrated the ability to recognize relationships between metrics, especially in trend analysis tasks. Tasks 2, 3, 5, and 6 showed that users could identify and use relations between multiple metrics to provide their answers. This enabled users to **interpret patterns** across evaluation dimensions.

Task 6 required constructing a **narrative** based on data-driven reasoning and was perceived by participants as the most challenging task. Even though 80% of participants completed the task successfully, several encountered difficulties. Most participants illustrated the intended analytical approach: "I would choose to improve the three courses with the lowest NPS scores. For instance, in Course A, the main issues seem to be instructor communication and course interaction, based on the radar chart." However, several participants struggled to connect the table containing low NPS scores with performance metrics due to a technical limitation: the interactive connection between the "Training Sessions Requiring Attention" table and the performance metric radar chart was non-functional. Participants were not informed of this issue to preserve natural interaction, but it disrupted the intended triangulation, and it may have increased cognitive load.

In addition to the technical constraint, some participants may have struggled to construct narratives due to limited contextual depth in the quantitative metrics themselves. This was noted by the qualitative feedback from participants in the answers of the Google Form in the study. Without adding qualitative insights such as comments, it may have been more challenging for participants to interpret feedback and propose improvements. The task completion rate of task 6 in combination with this participant feedback suggests that while the dashboard supports storytelling, enhancements in interactivity and contextual guidance, it could improve its support for advanced analytical reasoning.

5.1.2 Additional Questions

The final section of the Google Forms survey included additional questions that were specific to this research context. These questions with their corresponding score on a scale of five have been presented in Figure 17. These answers reinforced earlier findings related to information retrieval, trend analysis, and narrative construction. Participants reported that the dashboard's purpose was clear and that it effectively helped with trend recognition across datasets. Most users felt that the dashboard provided sufficient context to support answering questions and decision-making.

However, the question regarding the clarity of whether a course was considered good or bad received the lowest rating among the additional questions. This corresponds with earlier task-based usability results, as respondents had slightly more difficulty in constructing a narrative behind participant feedback. Nonetheless, the average score of **4.1 out of 5** for this item remains relatively high, which indicates that most users were still able to form a general impression of course quality. Moreover, participants expressed strong confidence in their ability to interpret course feedback and draw conclusions, with both metrics averaging **4.2 out of 5**.

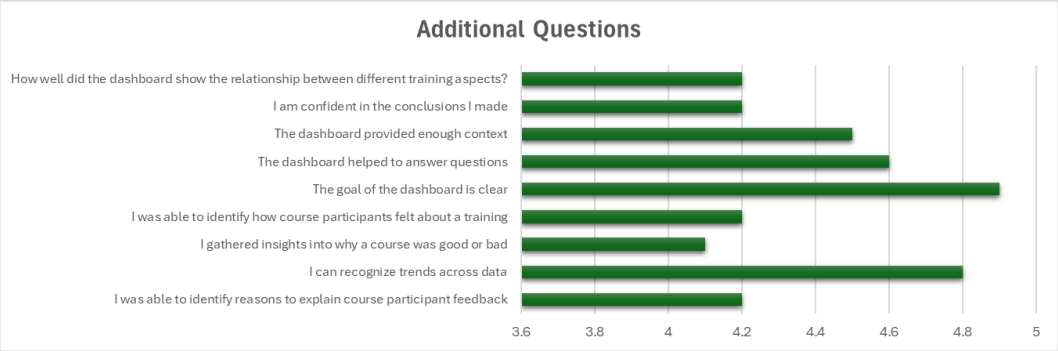


Figure 17: Additional Questions Results

These results suggest that while the dashboard performs well in guiding users through data exploration and trend recognition, there is an opportunity to improve its support for deeper evaluative insights into the course quality.

5.1.3 System Usability Scale (SUS)

The average SUS score of the study was **87.75%**. This exceeds the predefined target threshold of 68% significantly and falls within the "excellent" usability range. Individual scores varied from 77.5% to 95%, which shows that participants provided the dashoard consistently with high scores. Many reported that the dashboard was easy to learn, that its functions were well integrated, and that they felt confident using it.

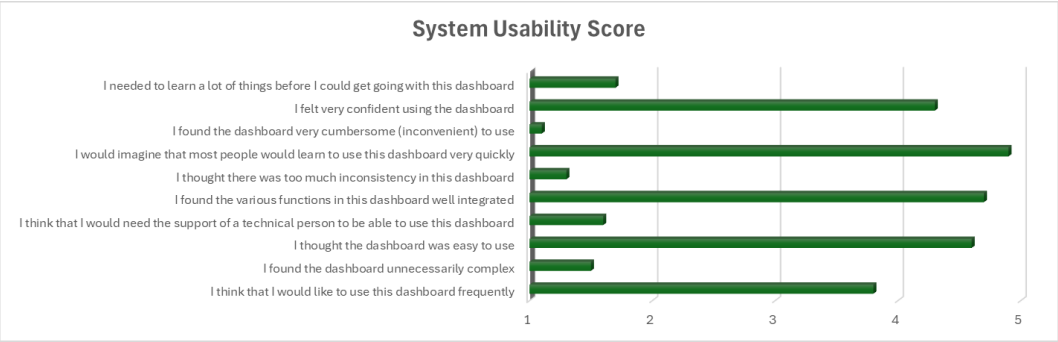


Figure 18: System Usability Score Results

Although a few participants mentioned they would not use the dashboard frequently, it's important to note that these individuals were less engaged in training evaluation, which makes their feedback understandable. Some responses showed that the dashboard might be slightly complex. Nonetheless, this concern was not alarming. Overall, the dashboard is viewed as **intuitive** and **highly usable** according to the system usability score.

5.1.4 User Experience Questionnaire (UEQ-S)

The user experience questionnaire (UEQ) is a widely used questionnaire to measure the subjective impression of users towards products [38]. It is focused on two dimensions, namely the pragmatic and hedonic quality dimension. The pragmatic dimension is focused on usability and functionality, whereas, the hedonic dimension focuses on the emotional and engagement dimension of participants. Cronbach's alpha assesses these two dimensions in terms of internal consistency. An alpha value higher than 0.70 demonstrates good internal consistency, meanwhile a score lower than 0.60 indicates poor consistency. Within this research, UEQ-S results were strongly positive across both the pragmatic and hedonic quality dimensions. The scores are presented in the table below.

Dimension	**Mean Score**	**Interpretation**	**Cronbach's Alpha**
Pragmatic Quality (efficiency, clarity)	2.425	Positive ($> +0.8$)	0.10
Hedonic Quality (stimulation, novelty)	2.500	Positive ($> +0.8$)	0.64

Table 14: UEQ-S results and internal consistency by dimension

Building on these results, it is important to note that the reliability of the scales should also be considered. The **Cronbach's alpha for the pragmatic quality dimension was 0.10**, which indicates very low internal consistency. This shows that the tasks in the study that were designed to assess dashboard usability may not have been understood well by all participants. Possible reasons for this include differences in users' experiences with dashboards, varying expectations, or unclear language. Therefore, even though the average score for pragmatic quality is high, it should be interpreted carefully.

In contrast, the **hedonic quality dimension achieved a Cronbach's alpha of 0.64**. This score reflects a moderate internal consistency. This alpha score implies that participants generally agreed on the emotional and aesthetic aspects of the dashboard, for instance its level of stimulation and perceived novelty. Moreover, while most items scored positively, "easy vs. complicated" showed the highest variance, and "inventive" and "leading edge" revealed more mixed perceptions. These findings suggest that although the dashboard is broadly appreciated for its clarity and engagement, its perceived ease of use and innovativeness vary across users.

Taken together, the dashboard is well-received, with strong scores across both the dimensions. However, the low reliability of the pragmatic scale shows the need for refinement. Improving task and dashboard clarity and potentially providing onboarding support could help to improve the experiences of users with dashboard efficiency and clarity.

5.1.5 Additional Feedback

Participants additionally provided qualitative feedback on the dashboard both verbally and through the Google Form. Overall, participants were positive regarding the dashboard's aesthetics and information organization. On the other hand, the feedback revealed opportunities for improvement.

Open-ended responses highlighted a desire for an additional filter option to quickly observe best and worst performing metrics. One of the participants suggested it would be nice to have some additional form of filter capability on the "training sessions requiring attention" table to better identify patterns in worst-performing metrics.

Furthermore, there were a few requests for more explicit connections to performance metrics in some views. As previously mentioned, the connection between the table and performance metrics did not work directly. Four participants identified the need for better connection between the table and performance metrics, of which two participants noticed it during the study. The technical fault in the dashboard might have therefore consciously or subconsciously affected the results of the task-usability test, but also the system usability score and the user experience.

Lastly, multiple participants suggested adding qualitative feedback integration by including course-specific feedback comments. For instance, one participant specifically recommended, "Maybe the report of each participant notes at the end of the course. It can really give a big explanation on why the course was good or no standing from the trainee's point of view." These suggestions indicate user recognition of the value in combining quantitative metric analysis with qualitative insights.

5.1.6 Summary of Success Criteria Achievement

The high task completion rate, combined with strong SUS and UEQ-S scores, confirms that the dashboard meets all predefined success criteria.

Success Criterion	**Status**	**Details**
Task Completion Rate >80%	**Met**	93% completion rate
SUS Score >68	**Met**	87.75% SUS achieved
Granularity >80%	**Met**	90% granularity achieved
Correlation >80%	**Met**	100% recognition across participants
Narrative >80%	**Met**	80%, with improvement potential

Table 15: Dashboard success criteria and evaluation outcomes

Overall, the results of the task-based usability test indicate that the dashboard is **highly usable** and that it helps its users with interpreting participant feedback. However, the test also revealed opportunities for improvement in terms of interactive connection between performance metrics and the NPS and potential integration of qualitative participant feedback to support the quantitative NPS and performance metrics.

5.2 Requirements Fulfillment

The evaluator confirmed that the dashboard's visualisations aligned with organisational needs. According to the evaluator, the dashboard successfully addressed all the requirements from the literature and stakeholder interviews.

Requirement	MoSCoW	Dashboard Rating
Defining a clear dashboard goal and aligning stakeholders on its purpose	Must	⋆⋆⋆
Answering specific, data-driven questions based on available metrics	Must	⋆⋆⋆
Enabling triangulation by integrating multiple metrics for comparison	Must	⋆⋆⋆
Maintaining a focused and uncluttered dashboard layout	Must	⋆⋆⋆
Involving evaluators in the design iteration process	Must	⋆⋆⋆
Including interactive features such as filtering and drill-down	Must	⋆⋆⋆
Applying visual hierarchy to guide user attention	Should	⋆⋆⋆
Using consistent color schemes and readable typography	Should	⋆⋆⋆
Providing contextual information and comparative views for deeper insights	Should	⋆⋆⋆

Table 16: Dashboard Rating for requirements from literature

Requirement	MoSCoW	Dashboard Rating
Analyzing all courses, not just those with low NPS	Must	⋆⋆⋆
Detecting trends across time, courses, instructors, and locations	Must	⋆⋆⋆
Maintaining brand consistency with Baker Hughes guidelines	Must	⋆⋆⋆
Cross-validating NPS with other evaluation metrics	Must	⋆⋆⋆
Identifying recurring issues across training programs	Must	⋆⋆⋆
Providing an aggregate view of quality indicators across all courses	Must	⋆⋆⋆
Automating data reporting processes	Should	⋆⋆⋆

Table 17: Dashboard rating from stakeholder requirements

The ninety-two percent task completion rate shows that the analytical questions were successfully addressed within the dashboard. This suggests that users can filter and gather data. Furthermore, high scores for system usability and user experience indicate a good layout and aesthetics. Flaws regarding triangulation of metrics were especially because of technical constraints rather than design flaws. Additionally, the positive requirement evaluation demonstrates that the evaluator was actively involved in the iterative design process.

Stakeholder-specific requirements were also met. The filter system enabled users to conduct a course analysis beyond low NPS courses. The ability of Power BI to show interactions between visualizations helped in providing effective trend analysis. Brand consistency was maintained through the use of the Baker Hughes logo, font, and colors. Participants showed effective use of the NPS-metric cross-validation function, despite the technical limitation that affected integration. Training issue identification was successfully applied through the exception table implementation and aggregate views were added through the multi-select filter function.

Overall, the dashboard supported effective decision-making across diverse user profiles. Most participants demonstrated the ability to analyze within the dashboard. However, improvements such as improved visualization interactivity, comment integration, and guided storytelling features could further strengthen usability.

6 Conclusion

This research aimed to investigate how a multi-metric evaluation tool can provide deeper insights into participant evaluations beyond standalone NPS analysis for Baker Hughes. To accomplish this goal, it was important to understand the limitations of NPS, the complementary training metrics, and the HCI principles of dashboard design.

Consultations with stakeholders revealed limitations of the current NPS-based evaluation approach, including manual and time-consuming analysis, absence of trend capabilities, limited insights in customer feedback, and the inability to identify systematic patterns. This disseratation focused its theoretical foundations mainly on the Kirkpatrick's evaluation framework, multi-metric decision-making, and HCI principles for dashboard design. Multiple prototypes were developed iteratively while keeping close contact with stakeholders to ensure the final dashboard aligned with organizational needs.

The dashboard development process integrated with the eight complementary metrics already captured by Baker Hughes, through visualization choices that were informed by stakeholder requirements and literature requirements. Within the dashoard, multi-selection filter capabilites allowed stakeholder to have a drill-down and aggregate view within the dashboard. The radar chart addressed the oversimplification of NPS by providing simultaneous visibility into all evaluation dimensions. Temporal trend analysis enabled pattern recognition across time periods and performance indicators. Exception-based reporting helped to provide a quick overview of the training sessions that require attention since they underperformed. The visualizations in the dashboard were organized based on the Gutenberg diagram design principle.

The evaluation of the final dashboard with ten Baker Hughes employees confirmed its effectiveness across all assessment dimensions. Task completion reached 92%, demonstrating strong usability across diverse user profiles. The System Usability Scale score of 87.75% placed the dashboard in the excellent category which showcases that the dashboard is highly usable and intuitive.

The User Experience Questionnaire results showed positive feedback for both pragmatic and hedonic dimensions. However, the Cronbach's alpha value indicated that, despite high usability ratings, participants interpreted the pragmatic items differently. This suggests that clarity, efficiency, or functionality may have been perceived in different ways by the users. This could be because of unclear elements in the dashboard's design or study tasks. Therefore, the findings highlight the need for improvements in the dashboard interface to promote a consistent understanding among different user groups and to enhance the clarity of future task-based evaluation protocols, ensuring they better accommodate diverse user profiles.

Additional questions and qualitative feedback reinforced these insights. Participants valued the dashboard's aesthetics, trend recognition capabilities, and the information organization , but some struggled with interpretation of participant feedback in the dashboard to inform their narrative. Participants suggested to include a stronger interactive connection

between the “training sessions requiring attention” table and the radar chart, to provide additional table filtering options, and to integrate qualitative participant comments to complement the quantitative data that is present in the dashboard. These recommendations underline that while the dashboard already supports effective decision-making, its qualitative insights could be further improved.

The dashboard answers the research question through several functions. First, its diagnostic capability helps to provide performance indications that are not clear from standalone NPS scores. Second, the dashboards' temporal patterns make it possible to find systematic problems and to track improvements over time. Third, the systematic exception table highlights sessions that require attention at a glance. This is good for understanding the trainings that potentially contributed to lower NPS scores. This final design combines theory with organizational needs meanwhile it ensures high usability. By moving from standalone NPS analysis to a broader multi-metric approach, the dashboard opens new opportunities for training improvement.

In summary, the research question can be answered through systematic integration of complementary evaluation metrics with NPS using appropriate visualization design and user-centered development approaches. Content-wise, the dashboard provides diagnostic detail that connects participant reactions to specific performance factors while supporting temporal pattern recognition. Design-wise, it implements visual hierarchy principles that prioritize critical information meanwhile it maintains analytical depth through aggregate views and drill-down options. At the same time, the evaluation highlights areas for improvement, especially in strengthening pragmatic consistency, improving interactivity between views, and integrating qualitative feedback to improve narrative construction. These improvements would further enhance the dashboard's role as a comprehensive and reliable decision-support tool for training evaluation.

7 Future Work

Although the dashboard met the success criteria established for this study, there is still room to improve its organizational impact and analytical capability. First of all, the metrics used in this study were predefined and focused primarily on Kirkpatrick's Level 1 (Reaction) and Level 2 (Learning). This provided useful insights, however, since it covered only half of the Kirkpatrick levels, it may have limited use to construct more in-depth narratives about training effectiveness. Future work should better highlight the distinction between Level 1 and Level 2 metrics in the dashboard. Additionally, it would be useful to extend customer feedack to Level 3 (Behavioral Change) and Level 4 (Organizational Impact). Including these levels would help to gain a more deeper understanding of training effectiveness and its impact on business performance, helping organizations connect training outcomes to knowledge transfer, on-the-job application, and key organizational results such as retention, productivity, and client success.

Another limitation of the evaluation was the potential bias introduced by the fact that the study participants were people known to the researcher. This was partly because of the fact that the study was limited to internal company members, but it may have influenced the results. Future studies should therefore aim to use a larger participant pool with people that are less familiar to the researcher to reduce bias and improve the generalizability of findings. Such a bigger sample would also make it possible to identify more patterns across different user profiles, or to focus more on one specific stakeholder group.

The usability study was designed with a small number of tasks to prevent fatigue among participants. This method was effective for the current evaluation, however, future research should include more tasks and explore them in greater depth, especially tasks associated with narrative construction. This would help with a more in-depth evaluation of how well the dashboard supports reasoning and storytelling. In addition, it would be useful to compare participants' answers with the actual reasons why courses performed well or poorly. Such a comparison would help to determine the accuracy of participants' reasoning and would provide deeper insights into how effectively the dashboard supports diagnostic analysis.

There was one technical issue during the evaluation which most probably affected Task 6. The issue was that the interactive link between the "Training Sessions Requiring Attention" table and the performance metrics radar chart did not function as intended. This disrupted the intended triangulation of metrics and may have increased cognitive load for participants. Future prototypes should therefore ensure that all interactive connections between views function reliably, as this is critical for supporting narrative construction and to support suggestions for improvement effectively.

Several participants provided feedback on potential improvements to the dashboard's functionality and clarity. Suggestions included adding more connections between performance metrics and NPS by introducing additional filtering options to the table etrics, and integrating qualitative feedback such as participant comments. Including comments from

participants would provide valuable context to complement quantitative data and could greatly improve diagnostic capability. Additionally, improving the clarity of key terms such as “customer,” “participants,” and “trainings,” and adding text for the functionalities of graphs could help users to understand the dashboard capabilities better.

Adding new functions to the dashboard could help to enhance the dashboard. For instance, adding a new column to the “Training Sessions Requiring Attention” table to highlight performance metrics that fall below a predefined threshold could help users quickly identify performance indicators into why trainings need attention. Additionally, summarizing the number of detractors, passives, and promoters would provide immediate insight into the composition of the NPS. Insights into this number are important to assess the reliability of the score.

Future updates to the dashboard could include predictive features using machine learning. This would allow it to suggest the best combinations of instructors, content, and delivery methods for different regions, which could help with improving the training portfolio.

In summary, future work should focus on minimizing bias by using a broader or more focused study group, broadening the Kirkpatrick levels included in the dashboard, diving deeper into task-based studies, validating narrative reasoning with real course outcomes, and making necessary technical and functional improvements. Together, these steps would enhance the reliability, diagnostic power and its organizational impact as a comprehensive decision-support tool for training evaluation.

References

[1] E. B. Seufert, "Chapter 4 - freemium metrics," in *Freemium Economics*, E. B. Seufert, Ed. Boston: Morgan Kaufmann, 2014, pp. 83–113. [Online]. Available: https://www.sciencedirect.com/topics/computer-science/net-promoter-score

[2] A. Baquero, "Net promoter score (nps) and customer satisfaction: Relationship and efficient management," *Sustainability*, vol. 14, no. 4, 2022. [Online]. Available: https://www.mdpi.com/2071-1050/14/4/2011

[3] F. F. Reichheld, "The one number you need to grow," *Harvard business review*, vol. 81, no. 12, pp. 46–55, 2003. [Online]. Available: https://www.nashc.net/wp-content/uploads/2014/10/the-one-number-you-need-to-know.pdf

[4] B. Pachołek, S. Sady, and W. Woźniak, "The assessment of business customers' satisfaction in the b2b market on the example of a metallurgical company," *Journal of Management and Financial Sciences*, no. 47, pp. 21–33, 2023. [Online]. Available: https://www.researchgate.net/publication/374483644_The_assessment_of_business_customers%27_satisfaction_in_the_B2B_market_on_the_example_of_a_metallurgical_company

[5] K. Kristensen and J. Eskildsen, "Is the nps a trustworthy performance measure?" *The TQM Journal*, vol. 26, no. 2, pp. 202–214, 2014. [Online]. Available: https://www.emerald.com/tqm/article-abstract/26/2/202/381179/Is-the-NPS-a-trustworthy-performance-measure?redirectedFrom=fulltext

[6] D. B. Grisaffe, "Questions about the ultimate question: conceptual considerations in evaluating reichheld's net promoter score (nps)," *Journal of Consumer Satisfaction, Dissatisfaction and Complaining Behavior*, vol. 20, pp. 36–53, 2007.

[7] N. Bendle, P. Farris, P. Pfeifer, and D. Reibstein, *Marketing Metrics*, 4th ed., ser. Pearson Business Analytics Series. Pearson Education, 2020. [Online]. Available: https://books.google.it/books?id=CtrPEAAAQBAJ

[8] D. Michels and A. Dullweber, "Do your b2b customers promote your business?" *Zugriff am*, vol. 2, p. 2023, 2014.

[9] H. Noble and R. Heale, "Triangulation in research, with examples," pp. 67–68, 2019.

[10] V. Vlckova and L. Podskubkova, "Customer service quality in b2b market from the buyer's perspective," 05 2018.

[11] "Simulation and training in work settings," in *International Encyclopedia of the Social Behavioral Sciences*, N. J. Smelser and P. B. Baltes, Eds. Oxford: Pergamon, 2001, pp. 14 097–14 101. [Online]. Available: https://www.sciencedirect.com/science/article/pii/B0080430767014248

[12] T. Khalil, "the impact of training process evaluation on the effectiveness of the training programs and improving the job performance," vol. 49, pp. 276–317, 01 2025.

[13] E. Salas, S. Tannenbaum, K. Kraiger, and K. Smith-Jentsch, "The science of training and development in organizations: What matters in practice," *Psychological Science in the Public Interest*, vol. 13, pp. 74–101, 06 2012.

[14] H. Urbancova, P. Vrabcová, M. Hudakova, and G. Ježková Petrů, "Effective training evaluation: The role of factors influencing the evaluation of effectiveness of employee training and development," *Sustainability*, vol. 13, p. 2721, 03 2021.

[15] S. Mann, "What should training evaluations evaluate?" *Journal of European Industrial Training*, vol. 20, no. 9, pp. 14–20, 12 1996. [Online]. Available: https://doi.org/10.1108/03090599610150264

[16] M. Mulder, "Customer satisfaction with training programs," *Journalof European Industrial Training 25 (2001) 6/7. - ISSN 0309-0590*, vol. 25, 08 2001.

[17] S. Dahiya and A. Jha, "Review of training evaluation," *International Journal of Computer Science and Communication*, vol. 2, no. 1, pp. 11–16, 2011.

[18] D. Kirkpatrick and J. Kirkpatrick, *Evaluating training programs: The four levels*. Berrett-Koehler Publishers, 2006.

[19] S. Falletta, "Evaluating training programs: The four levels: Donald l. kirkpatrick, berrett-koehler publishers, san francisco, ca, 1996, 229 pp." *The American Journal of Evaluation*, vol. 19, p. 259–261, 08 1998.

[20] S. Baehre, M. O'Dwyer, L. O'Malley, and N. Lee, "The use of net promoter score (nps) to predict sales growth: insights from an empirical investigation," *Journal of the Academy of Marketing Science*, vol. 50, no. 1, pp. 67–84, 2022.

[21] N. I. Fisher and R. E. Kordupleski, "Good and bad market research: A critical review of net promoter score," *Applied Stochastic Models in Business and Industry*, vol. 35, no. 1, pp. 138–151, 2019.

[22] F. Reichheld, "The ultimate question: Driving good profits and true growth," *Boston, MA*, 2006.

[23] A. Ziegler and T. Peisl, "The value of a net promoter score in driving a company's bottom line: a single-case study from ibm training services," in *European Conference on Software Process Improvement*. Springer, 2020, pp. 151–161.

[24] M. J. Diván, "Data-driven decision making," in *2017 international conference on Infocom technologies and unmanned systems (trends and future directions)(ICTUS)*. IEEE, 2017, pp. 50–56.

[25] A. Abduldaem and A. Gravell, "Principles for the design and development of dashboards: literature review," *Proceedings of INTCESS*, pp. 1307–1316, 2019.

[26] N. Martins, S. Martins, and D. Brandão, "Design principles in the development of dashboards for business management," in *Perspectives on Design II: Research, Education and Practice*. Springer, 2021, pp. 353–365.

[27] N. Carter, "The use of triangulation in qualitative research," *Number 5/September 2014*, vol. 41, no. 5, pp. 545–547, 2014.

[28] S. Almasi, K. Bahaadinbeigy, H. Ahmadi, S. Sohrabei, and R. Rabiei, "Usability evaluation of dashboards: a systematic literature review of tools," *BioMed Research International*, vol. 2023, no. 1, p. 9990933, 2023.

[29] A. Vázquez-Ingelmo, F. J. Garcia-Penalvo, and R. Theron, "Information dashboards and tailoring capabilities-a systematic literature review," *IEEE access*, vol. 7, pp. 109 673–109 688, 2019.

[30] A. Kramarzewski and E. D. Nucci, *Practical Game Design*. Packt Publishing, 2018, accessed September 28, 2025. [Online]. Available: https://www.oreilly.com/library/view/practical-game-design/9781787121799/f77e030b-0f77-4774-9025-000c0d7732e6.xhtml

[31] J. vom Brocke, A. Hevner, and A. Maedche, *Introduction to Design Science Research*. Cham: Springer International Publishing, 2020, pp. 1–13. [Online]. Available: https://doi.org/10.1007/978-3-030-46781-4_1

[32] M. Weggeman and L. Cauffman, "Using design science research to develop and validate the application of client-oriented psychological approaches illustrated by the design of a solution-focused approach," *Methods in Psychology*, vol. 11, p. 100151, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S2590260124000171

[33] R. Brugha and Z. Varvasovszky, "Stakeholder analysis: a review," *Health Policy and Planning*, vol. 15, no. 3, pp. 239–246, 09 2000. [Online]. Available: https://doi.org/10.1093/heapol/15.3.239

[34] C. GIBSON, "Semi-structured and unstructured interviewing: a comparison of methodologies in research with patients following discharge from an acute psychiatric hospital," *Journal of Psychiatric and Mental Health Nursing*, vol. 5, no. 6, pp. 469–477. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1046/j.1365-2850.1998.560469.x

[35] T. Kravchenko, T. Bogdanova, and T. Shevgunov, "Ranking requirements using moscow methodology in practice," in *Cybernetics Perspectives in Systems*, R. Silhavy, Ed. Cham: Springer International Publishing, 2022, pp. 188–199.

[36] S. K. Bansal, “Towards a semantic extract-transform-load (etl) framework for big data integration,” in *2014 IEEE International Congress on Big Data*, 2014, pp. 522–529.

[37] M. Schrepp, A. Hinderks, and J. Thomaschewski, “User experience questionnaire (ueq),” https://www.ueq-online.org/, 2024, accessed September 28, 2025.

[38] ——, “Design and evaluation of a short version of the user experience questionnaire (ueq-s),” *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 4, p. 103, 01 2017.

Appendices

7.1 Appendix I - AI disclosure

"During the preparation of this work the author used Copilot in order to refine phrasing throughout this dissertation. It was also consulted for filtering relevant information in order to avoid chaos in the thesis. Claude was used as source of inspiration to inform the dashboard design. All final decisions regarding content, structure, and interpretation were made by the author. The author reviewed and edited the content as needed and takes full responsibility for the content of the work."

Appendix II - Stakeholder Pain Points and Needs Interview Questions

Section	Interview Questions
1. Context	1. Can you describe your role and how it relates to maintaining course quality? 2. What are your main goals when reviewing course performance?
2. Current NPS Tool	3. How well does the current NPS tool support your decision-making about course improvements? 4. What types of insights do you find most valuable from the NPS data? 5. Are there aspects of course quality that the NPS score fails to capture? 6. Can you recall a situation where NPS feedback led to a concrete course change or to a moment where it didn't? 7. How confident are you in interpreting the NPS results without additional context? 8. What kind of supplementary information would help you better understand feedback from participants in training programmes?
3. Pain Points	9. What are the biggest challenges you face when using the current NPS tool? 10. Are there any usability issues that make it harder to extract meaningful insights? 11. Do you ever feel overwhelmed or unsure about how to act on the feedback?
4. User Needs	12. What features or data would make the tool more aligned with your needs? 13. How would you prefer to explore or visualize feedback trends across courses? 14. Are there specific types of feedback (e.g., qualitative, comparative) you wish were more accessible? 15. What would an ideal feedback tool look like for your role?
5. Collaboration	16. How do you currently share or discuss NPS insights with other teams (e.g., instructors, product owners)?

Section	Interview Questions
6. Closing	17. If you could improve one thing about the current feedback process, what would it be? 18. Is there anything else you'd like to share that we haven't covered?

Questions to Trainers

- How aware are you of the NPS scores your courses receive?
- How do you interpret the NPS score in relation to your teaching performance?
- Do you find the NPS feedback helpful for improving your delivery or content?
- Have you ever changed your approach based on NPS feedback? If so, how?
- What kind of feedback would be more useful to you than the current NPS score?

Questions to Stakeholder B (Trainer Skills)

- How do trainers typically respond to NPS feedback?
- Do you observe any patterns in how trainers engage with or act on NPS results?
- Are there any challenges in communicating NPS insights to trainers?
- What support do trainers need to better understand and use NPS feedback?
- In your view, how could the feedback process be improved to better support trainers?

Questions to Project Managers (Course Logistics)
Note: Ask who is in charge of course logistics in general.

- Who is generally responsible for the logistics of training courses?
- How do you currently gather and respond to logistical feedback from participants?
- Does the NPS score reflect any logistical aspects of the course (e.g., timing, materials, platform)?
- Have you ever used NPS data to make logistical improvements? If so, how?
- What kind of feedback would help you better manage course logistics?

Questions to Stakeholder A (Course Quality)

- How do you currently assess the overall quality of the training portfolio?
- What role does the NPS score play in your evaluation of course quality?
- Are there quality dimensions that NPS does not adequately capture?
- How do you prioritize which courses to improve or revise?
- What kind of feedback or data would help you make more informed decisions about course quality?

Appendix III - Task Based Usability Test Interview Guide

First of all, thank you for your time and for participating in this usability study. Today, we will be evaluating the usability of a customer training evaluation dashboard. It involves completing a set of 10 tasks related to interpreting feedback from training participants, after which you will be provided with a Google Form containing questions about the dashboard. This study is anonymous, so answers will not be linked to you as an individual. The answers will be used solely for thesis purposes, and Google Forms records will be deleted upon completion of the thesis. Just so you know, there are no wrong answers in this study, and you can choose to stop at any time without any consequences. In such a case, this usability study session will be excluded.

- **Participant Action:** Provide participants with the link to the Forms. Please complete the consent section and the pre-question section in the Forms.
- **Researcher Action:** Provide the Google Forms. *"Please fill in the consent section and pre-question section. When you arrive at the screen with the title System Usability, please let me know. You should not fill in this section yet but please do not close your Google Form Screen."*. After this, show an NPS presentation slide to explain the following shortly:

> First of all, just to provide you with a small recap of what NPS is, after course participants end the training, they are asked, 'Would you recommend this course to your colleagues?'. They can answer on a scale from 1 to 10.
>
> - 9/10: represents promoters, who are people who will recommend the training to other people
> - 7/8: are passives, people who are fine with the course but who will not say anything negative or positive about the course
> - 1-6: are detractors, people who will talk negatively about the course.
>
> The NPS Score is the difference between promoters and detractors, indicating whether you have growing sales of training or not, based on the results of people recommending or talking negatively about the course

Now let's see the dashboard. (I will provide you with a link to the dashboard in the chat.) The dashboard was created, containing both the NPS and additional metrics. Please first familiarize yourself with the dashboard for 1 minute by interacting with the interface. If you have any questions, please don't hesitate to let me know.

- **Researcher Action:** Set timer of 1 minute and let them go through the dashboard and potentially if online, provide them with the link of the dashboard
- **Participant Action:** Scroll through the dashboard

- **Researcher Action:** Only if they ask: Unfortunately, I can't answer questions about what something does, but I can clarify specific names of things you see, should you need it. (Write down the difficulties people still experience with the dashboard.)

Now that you have familiarized yourself with the dashboard, let's continue to the central part of the research, which involves completing ten short tasks. Please answer the questions in the tasks to the best of your ability but note that absolutely no answer is wrong. Answers will not be linked to you as an individual, so please be aware that everything you say will only help me understand and, consequently, improve the usability of this dashboard.

- **Researcher Action:** Pose the tasks and questions and write down all provided answers and anomalies/questions they have. Provide a written (either printed or digital) question list.
- **Participant Action:** Use the dashboard to conduct the tasks and answer the questions to the best of their ability.

A) **Information Retrieval Tasks:**

- What was the overall NPS score in 2023 quartile 4?" (KPI understanding Filtering)
- How many training sessions scored below 70 in terms of NPS in 2024 and how many participants followed these training sessions in total?" (Tests table navigation)

B) Trend Analysis Tasks:

- "Is training performance improving or getting worse NPS over time? Please show me how you find it." (Tests trend interpretation)
- "For Course X, what performance metric is the biggest weakness?"
- Identify the key positive and negative sentiments expressed by learners in Course Y.

C) Problem Identification Tasks:

- You have a budget to improve only three training sessions. Which ones would you choose and why?" (Tests comprehensive dashboard usage + decision-making) What would you improve?

Thank you so much for completing the tasks in usability study. Now the only thing that remains is to fill in the last part of the Google Forms. This questionnaire focuses on your experience with the dashboard and includes evaluation practices that have been shown to be effective in academic research. Your feedback will contribute to making this tool more intuitive, efficient, and user-friendly. I truly appreciate your honest opinion!

- **Participant Action:** Fill in the last part of the Google Forms and send it.