Czech University of Life Sciences Prague

Faculty of Economics and Management

Department of System Engineering

Master's Thesis

Application of Business Process Modelling Methodology for Enhancing the Software Validation Processes

Nayeem Al Tamzid Bhuiyan

CZECH UNIVERSITY OF LIFE SCIENCES PRAGUE

Faculty of Economics and Management

DIPLOMA THESIS ASSIGNMENT

Nayeem Al-Tamzid Bhuiyan

Informatics

Thesis title

Application of Business Process Modelling Methodology for Enhancing the Software Validation Processes

---

Objectives of thesis

This thesis investigates the challenges corporations face in validating partner applications in their interactions with numerous small partners. Traditional validation methods are often slow, relying on the examination of user guides and the execution of basic test cases. In contrast, this research introduces an innovative approach that incorporates generative artificial intelligence (AI) into the validation workflow, significantly enhancing both efficiency and speed.

The objective of the thesis is to achieve efficient validation, enhance the relevance of test cases to actual usage, and improve scalability in response to the growing number of partner applications. Specifically, the thesis will aim at the given points as follows:

1. Comprehensive Analysis of Existing Validation Practices
2. Investigating the Integration of Generative AI in Validation Workflows
3. Development of a Generative AI-Powered Validation Framework
4. Enhancing Test Case Relevance and Quality
5. Measuring Efficiency Gains Post-Implementation
6. Establishing Implementation Best Practices

Methodology

To fulfill the above-given objectives, we use the following procedure:

1.Research Design

Use qualitative methods like interviews and focus groups to deeply explore current validation practices and user experiences, informing the development of the AI framework.

2.Literature Review

Analyze existing research on software validation and generative AI to identify knowledge gaps, best practices, and lessons for framework design.

3.Evaluating Validations in Place

Conduct surveys and interviews with diverse software teams to understand current validation tools, practices, and challenges.

4.Exploring Generative AI

Study real-world case studies and consult experts to uncover effective uses and limitations of generative AI in software validation.

5.Relevance and Quality of the Test Case

Integrate user feedback and define quality metrics to ensure generated test cases are accurate, relevant, and continuously improving.

6.Conducting Empirical Studies

Pilot the framework in real scenarios, collect feedback, and perform statistical analysis to evaluate its practical benefits.

7.Proposing Best Practices

Run workshops and synthesize guidelines to help organizations effectively adopt and integrate the AI-driven validation framework.

8.Recommendations

Outline business implications, suggest future research directions, and monitor emerging AI technologies to ensure the framework's continued evolution.

The proposed extent of the thesis

60-80 pages

Keywords

Generative AI, Software Validation, Test Case Automation, Partner Applications, AI-Driven Testing, Validation Framework, User Behavior Simulation, Software Quality Assurance

---

Recommended information sources

A. Thakur, A. Mehra, R. Choudhary, and M. Sarker, "Generative AI in Software Engineering: Revolutionizing Test Case Generation and Validation Techniques," IRE Journals, vol. 7, no. 5, pp. 281–293, Nov. 2023.

D. S. Cruzes and T. Dyba, Research Methods in Software Engineering: Qualitative and Quantitative Approaches. Cham, Switzerland: Springer, 2021.

J. Wang, B. Suleiman, and M. J. Alibasa, "Automated Unit Test Case Generation: A Systematic Literature Review," arXiv preprint arXiv:2504.20357, Apr. 2025.

J. W. Creswell and C. N. Poth, Qualitative Inquiry and Research Design: Choosing Among Five Approaches, 4th ed. Thousand Oaks, CA, USA: SAGE, 2018

M. Baqar and R. Khanda, "The Future of Software Testing: AI-Powered Test Case Generation and Validation," arXiv, Sept. 2024.

M. Schäfer, S. Nadi, A. Eghbali, and F. Tip, "An Empirical Evaluation of Using Large Language Models for Automated Unit Test Generation," arXiv, Feb. 2023.

P. Nama, P. Reddy, and G. P. Selvarajan, "Leveraging Generative AI for Automated Test Case Generation: A Framework for Enhanced Coverage and Defect Detection," Well Testing Journal, 2024.

Expected date of thesis defence

2025/26 WS – FEM

Thesis supervisor

doc. Ing. Milan Houška, Ph.D.

Supervising department

Department of Systems Engineering

Electronic approval: 30. 06. 2025	Electronic approval: 22. 08. 2025
doc. Ing. Milan Houška, Ph.D.	**prof. Ing. Lukáš Čechura, Ph.D.**
Head of department	Dean

Prague on 29. 11. 2025

Declaration

I declare that I have worked on my Master's thesis titled " Application of Business Process Modelling Methodology for Enhancing the Software Validation Processes" by myself and I have used only the sources mentioned at the end of the thesis. As the author of the thesis, I declare that the thesis does not break any copyrights code or university code of conduct.

In Prague on 30-11-2025

Nayeem Al Tamzid Bhuiyan

Acknowledgement

I would like to express my deepest gratitude to my supervisor, **Assoc. Prof. Milan Houska, Ph.D.**, for his invaluable guidance, patience, and continuous support throughout the development of this thesis. His expert insights, constructive feedback, and encouragement have been instrumental in shaping both the direction and quality of my research. Working under his supervision has been a truly enriching academic experience that has significantly enhanced my understanding of business process modelling and software validation.

My sincere appreciation goes to the Faculty of Economics at the Czech University of Life Sciences for providing me with the opportunity to pursue my Master's Degree in Informatics. I am grateful to all the faculty members and administrative staff for their academic assistance, inspiring lectures, and for maintaining a research-oriented environment that fostered my intellectual growth.

I would also like to extend my heartfelt thanks to the professionals and collaborators who contributed to the practical part of this thesis. Their cooperation, data access, and real-world insights made it possible to translate theoretical knowledge into meaningful applications. Their contributions were essential to the successful completion of this research.

Finally, I wish to express my deepest gratitude to my family for their unconditional love, patience, and encouragement. Their unwavering belief in me has been a constant source of strength and inspiration, motivating me to overcome every challenge along the way.

Application of Business Process Modelling Methodology for Enhancing the Software Validation Processes

Abstract

This thesis presents an innovative approach to applying Business Process Modeling (BPM) methodologies in software validation by integrating Generative Artificial Intelligence (AI) for automatic test case generation, with potential extensions to internal portal automation. The current landscape of software validation practices is critically analyzed, highlighting ongoing skepticism, frequent downtimes, and the associated long-term costs. BPMN (Business Process Model and Notation) is introduced as an effective tool for re-engineering these processes. The report delves into how Generative AI can revolutionize key validation activities while also addressing its limitations, challenges, and ethical considerations. By combining BPM with Generative AI, the validation process becomes more efficient, transparent, and intelligent. This study establishes a synergistic framework for software validation that not only enhances effectiveness.

This thesis studies the problem of corporations having difficulty to certify partner applications due to high number of small partners they deal with. The traditional validation approaches are generally very slow as it involves scanning though user manuals and running simple test cases. In contrast, our study is presenting a novel strategy which involves the usage of generative artificial intelligence (AI) as part of the validation process, thus boosting its productivity level, parallel with the speed. Here, we argue for the use of generative AI techniques to automate the generation of device model-specific test cases for a variety of user scenarios as identified by industrial partners. The model in the proposal consists of separate steps, consolidation of information in partner input and documentation, using AI to produce appropriate test cases by considering actual user behavior, and set up an iterative validation loop that periodically refreshes test cases based on feedback so that the stories are kept up to date and refined.

This thesis is driven by the goal of attaining efficient validation, increasing test case relevance to real usage, and scalability to serve the increasingly more number of partner applications. The long-term result of this work is a more efficient validation framework, reduced time-to-market, and happy end users with the partner offerings. This work provides the first step to move AI-based approach for software verification to a broader range of domains.

Keywords:

Generative AI, Software Validation, Automation, Partner Applications, AI-Driven Testing, Validation Framework, Test-Case Generation, Software Quality Assurance.

Aplikace metodologie modelování obchodních procesů pro vylepšení procesů validace softwaru

Abstrakt

Tato práce představuje inovativní přístup k aplikaci metodologií modelování obchodních procesů (BPM) při validaci softwaru integrací generativní umělé inteligence (AI) pro automatické generování testovacích případů s potenciálním rozšířením na automatizaci interních portálů. Současná situace v oblasti validace softwaru je kriticky analyzována, přičemž se zdůrazňuje přetrvávající skepticismus, časté prostoje a související dlouhodobé náklady. BPMN (Business Process Model and Notation) je představen jako efektivní nástroj pro reengineering těchto procesů. Zpráva se zabývá tím, jak může generativní AI způsobit revoluci v klíčových validačních činnostech a zároveň řešit svá omezení, výzvy a etické aspekty. Kombinací BPM s generativní AI se proces validace stává efektivnějším, transparentnějším a inteligentnějším. Tato studie vytváří synergický rámec pro validaci softwaru, který nejen zvyšuje efektivitu.
Tato práce se zabývá problémem korporací, které mají potíže s certifikací partnerských aplikací kvůli velkému počtu malých partnerů, se kterými jednají. Tradiční přístupy k validaci jsou obecně velmi pomalé, protože zahrnují procházení uživatelských manuálů a spouštění jednoduchých testovacích případů. Naše studie naopak představuje novou strategii, která zahrnuje využití generativní umělé inteligence (AI) jako součásti validačního procesu, čímž se zvyšuje jeho produktivita a zároveň rychlost. Zde argumentujeme pro využití generativních technik AI k automatizaci generování testovacích případů specifických pro model zařízení pro různé uživatelské scénáře identifikované průmyslovými partnery. Model v návrhu se skládá ze samostatných kroků, konsolidace informací ve vstupech a dokumentaci partnerů, využití AI k vytváření vhodných testovacích případů s ohledem na skutečné chování uživatelů a nastavení iterační validační smyčky, která pravidelně obnovuje testovací případy na základě zpětné vazby, aby byly testovací příběhy aktuální a zdokonalované.

Tato práce je motivována cílem dosáhnout efektivní validace, zvýšit relevanci testovacích případů pro reálné použití a škálovatelnost pro obsluhu stále většího počtu partnerských aplikací. Dlouhodobým výsledkem této práce je efektivnější validační rámec, zkrácení doby uvedení na trh a spokojení koncoví uživatelé s nabídkou partnerů. Tato práce představuje první krok k rozšíření přístupu založeného na umělé inteligenci pro ověřování softwaru do širší škály oblastí.

Klíčová slova: Generativní umělá inteligence, validace softwaru, automatizace, partnerské aplikace, testování řízené umělou inteligencí, validační framework, generování testovacích případů, zajištění kvality softwaru.

Table of contents

Contents

1. Introduction

This thesis investigates the utilization of Business Process Modelling (BPM) approaches to methodically improve software validation processes. BPM provides a systematic methodology for enhancing clarity, efficacy, and traceability through the identification of waste and the revision of validation steps. This study examines the integration of Generative AI technologies, which can automate test case generation, optimize internal portal processes, and facilitate intelligent, data-driven reporting on test validation metrics (Dumas & La Rosa & Mendling & Reijers, 2018; van der Aalst, 2013; Zhang & Harman, 2021; Amershi et al., 2019).

We want to make a system that works together by leveraging BPM's clear processes and Generative AI's intelligence and automation. This all-encompassing plan seeks to improve the software validation lifecycle in line with industry standards and government rules, while also dealing with the issues that come up when software is constantly changing (IEEE Std 1012-2016; FDA, 2002). Quality Assurance for Software (SQA) SQA is an ongoing process that focuses on the quality of the software products that meet the needs of the organization. Software V&V are two different but related tasks that must be done in SQA to make sure that software is provided to the client without any bugs.

Process validation (or workflow validation) is a golden rule that people who work in this field should follow. It is very different from "software validation," which focuses on the product. Process validation looks for the goodness of FIT and is defined by the policies that govern workflow activities. This software validation processes talks about making test cases, running them, and reporting on them. It does not talk about the software product's core logic or data flows. In turn, BPMN diagrams and models are used to architect and improve the workflows of validation activities so that they are both effective and efficient (Dumas et al., 2018; van der Aalst, 2013).

A key feature of software validation is that it is ongoing, not just a one-time occurrence. When software is first introduced, it goes through initial process validation. However, when big modifications are made, it needs to be revalidated. This is especially important for certified software that runs in public cloud environments, where changes happen all the time. Because software is always changing and rules are always changing, validation is not a one-time thing but a requirement for the entire lifecycle. Traditional manual methods of validation cannot keep up with this constant need, which leads to more and more problems over time. This inherent necessity for iterative and adaptive validation highlights the imperative for process optimization to maintain long-term sustainability and quality in software development.

Software validation is an important step in the software development life cycle. Other steps in this cycle include talking about the software, establishing requirements specifications, coding, testing, and launching the products. Its main goal is to show that the software product is free of defects and safe to use (IEEE Std 1012-2016; FDA, 2002). This complete system is very important for the pharmaceutical, commercial and automotive industry, which is known for having strict rules. Software affects the development, manufacturing, storage, and quality control of products. Software validation must be distinct from mere software verification. Control is mostly about making sure the software was constructed correctly. This can be done through unit and integration testing, code reviews, and inspections. Validation, on the other hand, looks at the bigger picture of whether the software works as it should and satisfies the needs of the people who use it. This is usually done through user acceptability testing or simulation exercises (Kaner, Falk, & Nguyen, 1999; Myers, 2011). Also, software validation is only one aspect of the larger process of validating a computer system. This process also includes checking that the hardware, networks, and interactions are all working together correctly and reliably (FDA, 2002).

It is vital to recognize that software validation is a continuous process rather than a singular event. The software undergoes validation upon initial installation. Significant alterations necessitate revalidation. This is particularly applicable to certified software operating on public clouds, where upgrades and modifications are frequently observed (Bass, Weber, and Zhu 2015). Software evolves throughout time, as do the regulations that govern it. This indicates that validation processes must be flexible and continuous. Conventional manual validation techniques frequently fail to meet these demands, and with time, they diminish in efficacy and increase in risk (Erdogmus & Morisio, 2006). This fact underscores the necessity of implementing mechanisms to enhance software while maintaining high standards of quality and compliance over its entire life cycle.

2. Objectives and Methodology

2.1. Objectives

The primary purpose of this thesis is to enhance the software validation process at Company X through the fusion of Business Process Model and Notation (BPMN) methodology and Generative AI with Automated Test Case Generation. For this purpose, as introduced at the beginning, this thesis aims to accomplish the following specific goals:

- **Analyze the Current Validation Process (As-Is):** Record the processes, flow, and inefficiencies of Company X's existing software validation process. In order to better understand how validations are carried out today and to pinpoint areas of manual labor and/or slowdowns, BPMN diagrams are used to describe the validation process as it is (Dumas et al., 2018; van der Aalst, 2013).

- **Analyzing current data and validation trends:** We examined the data obtained from the ongoing validation process. In order to be able to investigate both the current and future models, we will establish some variables that will allow us to analyze the model within any specified time frame. Whether it pertains to the current process or the future process state we aim to model, the KPIs we establish will be calculated and visually represented as an integral component of this thesis. We will calculate standard KPIs using data from a specific company, analyze this data in Excel, and then generate the most appropriate visualization diagrams to identify current trends for the validation process.

- **Design an Improved BPMN Workflow (To-Be):** Make a model of the validation process for the future that takes into account the problems with the current system. The future process is made in BPMN with Camunda and follows best practices from business process management. This is done to make operations more efficient and get rid of unnecessary steps. Adding automation to the workflow is an important part of reaching this goal. For example, instead of doing things by hand, like transferring data into internal portals, we might employ automated service activities to make things more efficient and consistent.

- **Integrate Generative AI for Test Generation:** Employ Generative AI to autonomously produce relevant test cases and validation scripts as part of the integration of AI-driven automation into the new BPMN process. To minimize the necessity for testers to manually author test cases, this objective explores how artificial intelligence can be employed to interpret user requirements and partner documentation in order to generate a diverse array of test scenarios. The objective of AI integration is to ensure more authentic scenarios that accurately reflect real-world user behavior and enhance overall test coverage (Zhang & Harman, 2021).

- **Incorporate Feedback Loops:** Make the new process better by adding ways for people to give feedback so that they can keep learning and improving. After the test is run, the system gets the findings and insights back, which lets the Generative AI part update or suggest new test cases in the next cycles. This recurrent validation loop helps keep the test suite up-to-date with any changes in requirements or software updates. It also encourages a culture of continual improvement in the validation process.

- **Evaluate Performance Improvements:** Finally, evaluate the performance of the proposed BPMN plus AI process in comparison to the existing approach. This entails establishing key performance indicators (KPIs), including validation cycle duration, resource efficiency, and test coverage, and employing them qualitatively to assess progress. The aim is to demonstrate that the revised process is capable of validating software more efficiently, comprehensively, and dependably than previously, without depending on any particular numerical benchmarks. Success is evaluated by a significant decrease in manual effort and delays, as well as an enhancement in the overall efficiency and transparency of the validation process.

2.2. Methodology

This study employs a structured technique that combines business process modeling with experimental AI-driven test automation to attain its objectives. The process is broken down into four steps: understanding the current process, making the improved process, using AI automation, and analyzing and predicting the results.

- **Modeling the Current Process (As-Is):** The first step is process discovery, which means gathering all the information we can about how Company X now validates their software. This procedure involves looking at company documents and talking to or observing stakeholders involved in validation to fully record all the phases, decision points, and roles in the workflow. We used the Camunda BPM platform and BPMN 2.0 language to construct a detailed as-is process model of the validation workflow. The BPMN diagram shows the current process in a visual way, showing each step from getting a partner application to getting the final certification. This model shows where manual tasks (such reading user manuals or entering data into internal systems) happen and where there are delays or extra effort. By showing the current process in a consistent BPMN style, it makes it easier to find problems and improves clear communication with both technical and business stakeholders (Dumas et al., 2018; van der Aalst, 2013).

- **Analyzing Current data and validation trends:** We analyzed the data that we have from current validation process. In order to examine both current and future models we declare some variable that can be used to interpret the model into any given time frame. Weather it is current process or the future process state we are trying to model, the KPIs that we declare, will be calculated and visually represented into practical part of this thesis. We will be calculating standard KPIs with data form specific company and analyzing those data in Excel then generating best visualization diagram we can specify the current trends for the validation process.

- **Identifying Gaps and Designing the Future Process (To-Be):** Once the current model is established, the thesis concentrates on process analysis and optimization. The examination analyzes the BPMN model to detect bottlenecks, redundancies, and issues such as prolonged durations and human errors within the process. We

performed a qualitative analysis to identify which stages contribute value to the system and which do not.
Through this analysis, a future process model is developed utilizing BPMN to depict a more efficient workflow. The future state is subsequently modeled in Camunda Modeler, facilitating the development of an updated action plan designed to address the existing process challenges. Proposed modifications encompass the automation of particular procedures and the reorganization or introduction of new activities to improve the overall workflow.
For instance, if the current process involves manually entering partner information into an internal portal, the revised design could integrate an automated service task within Camunda that updates the internal system automatically, eliminating the need for human involvement. Furthermore, activities in which testers previously had to manually develop and author straightforward test cases have been restructured to enable automated test generation.
The design emphasizes establishing a process that is more efficient by minimizing superfluous steps and reducing wait times, enhancing consistency to decrease the likelihood of human error, and increasing transparency by clarifying responsibilities and hand-offs.

- **Integrating Generative AI for Test Case Automation:** This method's utilization of generative artificial intelligence (AI) to develop the intended Business Process Model and Notation (BPMN) framework is a key feature. To make test cases, this process requires creating a new automation job in Camunda-Process that uses a generative AI service. We will use an AI-powered phase to automatically create a full set of test scenarios for relevant test cases based on inputs like documentation from partner applications, requirements, or user stories. These AI-generated test cases make it more likely that real-world bugs will be found by modeling how end users might interact with the product.
  The method makes sure that AI works smoothly with the BPMN workflow. For example, the process can move from gathering data to the AI service task that makes tests, and then to the execution phase, where those tests are done. The thesis looks at cutting-edge test automation methods from a business process point of view by using generative AI in this way.

- **Comparison of As-Is vs. To-Be:** After modeling the proposed and current processes in Camunda. The new model might suggest that the process takes less time from start to finish since AI-generated tests can run at the same time and there are fewer hand-offs. This thesis does contain some specific numerical metrics and it does talk about the benefits in general terms. For example, it says that the AI makes the new process able to handle more test scenarios, finish validation faster than the old one, and require fewer manual steps from engineers. We can talk about these gains in a structured way by using key performance indicators. The as-is and to-be procedures are compared qualitatively using measures such as test coverage breadth, fault detection rate, and validation turnaround time. This comparison research is necessary to make sure that the suggested changes really fix the problems and add more.

- **Verification and Discussion of Results:** We checked to see if the proposed process achieves all of the goals that have been set. This means that every problem

that was found in the present process has been fixed in the new design. This phase also brings up any problems or results that weren't expected (for example, if AI has trouble with certain test cases or if some manual work is still needed, along with the assumptions about the surroundings that were made earlier).
The discussion part explains how these results will be used to come to conclusions. We will go into more detail about this comparison in the next chapter of this thesis, "Results and Discussion," where we will look at what it means for Company X. We will examine if the merging of BPMN and Generative AI has produced the expected advantages in practice.

To put it simply, the approach describes a sequence of steps that begin with determining the nature of the issue and end with determining the most effective solution. We ensured the research is grounded in realistic, reproducible actions by simulating both existing and improved processes in a genuine BPM tool (Camunda) and integrating AI in a regulated manner. This thesis will show how a BPMN-guided, AI-augmented validation process could make software certification at Company X more efficient and scalable. The results of this technique will be detailed with a focus on their real-world impact. This study is significant as it demonstrates a practical approach to modernizing software certification. This isn't just a theory; businesses can utilize it to cut down on downtime, adapt to changes faster, and keep their software products up to high standards. The goal is for these results to make software quality assurance utilize more intelligent process automation. This will help Company X and other organizations in the same area where speedy and accurate validation is crucial.

3. Literature Review

3.1. Description of Companies:

We will delve deeper into the world of tech giants who design groundbreaking products and comprehensive solutions tailored for high-tech industries. These companies are not only at the forefront of technological advancement but also play a pivotal role in shaping the way we interact with technology in our everyday lives. Their influence extends across various sectors, driving innovation that enhances efficiency, safety, and overall user experience.

3.1.1. Geographic Coverage:

These companies have a wide range of operations, as shown by the fact that they cover various important areas around the world.
North America: This includes the US and Canada, where a lot of people buy tech and the infrastructure is good, which makes for a great tech environment.
Europe: This area includes Germany, France, the UK, Italy, and other countries in the Rest of Europe. It is noted for having strong rules and a strong focus on new ideas.
Asia-Pacific: This region is quickly becoming a technology powerhouse with a wide range of markets. It includes significant players including China, Japan, India, South Korea, Australia, and the rest of Asia-Pacific.
Middle East and Africa: This area is seeing a rise in the use of technology and new ideas, despite its own particular problems and chances.
Latin America: This region, which includes Brazil, Argentina, and the rest of Latin America, is becoming an important market for ICT solutions because of its young population and growing digital literacy.

3.1.2. Market Leading Companies:

3.1.2.1. Zebra Technology

Zebra Technology is a prominent player in the tech industry, focusing specifically on barcode printing and scanning solutions that are crucial for effective inventory management and logistics. Their innovative products streamline supply chain processes, enhance operational efficiency, and ultimately improve customer satisfaction.

3.1.2.2. Honeywell

Honeywell stands as a conglomerate with a diverse portfolio that includes products in building technologies, safety and productivity solutions, and performance materials. The company leverages the power of the Internet of Things (IoT), artificial intelligence (AI), and data analytics to develop innovative products that meet the evolving needs of industries. Honeywell is also committed to sustainability, utilizing eco-friendly materials and creating energy-efficient systems that significantly reduce environmental impact.

The integration of advanced technologies is transforming industries. Companies are increasingly utilizing IoT to collect and analyze real-time data, which leads to optimized processes and informed decision-making. For instance, in the healthcare sector, wearable sensors are employed to continuously monitor patient's vital signs, enabling timely medical interventions. In the aerospace industry, IoT-driven predictive maintenance is revolutionizing operational efficiency by minimizing downtime and ensuring safety. The combination of AI and IoT facilitates large-scale data analysis, which supports enhanced decision-making processes across various sectors. In manufacturing, smart sensors are utilized to track equipment performance and predict maintenance needs, thereby preventing costly breakdowns.

3.1.2.3. HP

HP serves a diverse array of sectors, including healthcare, education, and finance. Their advanced printing and computing solutions are designed to enhance organizational productivity and streamline operations. In the healthcare sector, HP's technology facilitates the organization of patient files and promotes collaboration among medical professionals. In education, HP fosters interactive learning environments that engage students and enhance their learning experiences. The company is committed to sustainable and effective innovations that contribute positively to society.

3.1.2.4. Brother Industries

Brother Industries started as a sewing machine repair shop but has since grown into a multifaceted technological company. Its Printing and Solutions division provides a wide selection of office and home scanners, All-in-One devices, and laser and inkjet printers. For industries like manufacturing, communications, retail, and electricity, the group's industrial printing branch offers label printers, mobile printers, and coding and marking devices that make barcode printing simple. Additionally, Brother has expanded its inkjet technology to include wide-format and clothing printers that ensure traceability while producing high-mix or high-volume prints. Award-winning document capture is provided via the company's Workhorse Series scanners, which also allow batch barcoding, high-volume duplex scanning, and a variety of scan-to destinations. Security features like NFC authentication and Active Directory connection are also included. Overall, Brother positions itself as a trusted partner that combines versatile office devices with industrial printing solutions and efficient document digitization (Brother Industries, 2025; Brother Workhorse Series, 2025).

3.1.2.5. Epson

The PrecisionCore inkjet and heat-free technologies are used by Seiko Epson Corporation to produce a range of products, including large-format industrial printers and point-of-sale receipt printers. Epson demonstrated at NRF 2024 how its POS and ColorWorks printers can modernize retail checkout, inventory, visitor management, shipping, and food service environments. The liner-free thermal label printer facilitates buy-online-pick-up-in-store workflows, while the OmniLink and Mobilink models offer streamlined, durable thermal receipt printing with improved connectivity. Epson's ColorWorks line includes full-color on-demand labels and even integrated color RFID encoding and printing functionality. Flexible UHF RFID tag antennas have been created on plastic film using inkjet printing in academic research, proving that low-cost printing can allow for the mass manufacture of RFID sensors. Independent analyses note that ColorWorks models like the C6500 and C7500 offer water- and smudge-resistant inks, fast print speeds and on-demand flexibility, allowing businesses to eliminate pre-printed label stock and minimize waste (Hess, 2025). These innovations show Epson’s dual focus on high-performance printers and environmental responsibility (Epson, 2024; Islam et al., 2018).

3.1.2.6. Datalogic

Datalogic is a world leader in industrial automation and automated data collection. The company has one of the largest portfolios of assisted and self-checkout scanning solutions and was a pioneer in laser scanning and imaging technologies. High read rates are achieved via its Magellan scanners, which also employ AI to improve checkout and decrease shrink. They are available in a variety of configurations to suit the demands of stores, ranging from presentation scanners to multi-plane machines. Beyond retail, Datalogic provides rugged handheld scanners like the PowerScan family and mobile PCs like the Memor and Skorpio series, allowing for track-and-trace across fulfillment centers.In 2025 Datalogic launched the PowerScan 9600 RFID series, an industry-first handheld scanner that integrates high-performance barcode scanning with UHF RFID tag reading; this hybrid device lets workers capture both barcodes and RFID tags in a single trigger pull, dramatically reducing inventory cycle times and improving visibility. Its robust IP65/67-rated construction and ability to read tags through packaging highlight Datalogic’s emphasis on durable hardware and advanced data capture (Datalogic, 2024; Datalogic, 2025).

3.1.2.7. Toshiba Global Commerce Solutions (TGCS)

Toshiba Global Commerce Solutions offers retail customers integrated hardware, software, and services. Partners claim that Toshiba is the industry leader in retail store technology and the go-to option for integrated in-store solutions. In addition to TCx printers and display accessories, the company's point-of-sale portfolio consists of TCx systems (such the TCx 800 and TCx 900), self-checkout kiosks, and handheld POS devices. With built-in health sensors for real-time monitoring, industry-leading print rates (406 mm/s), and Energy Star-certified efficiency, the TCx receipt printer series is made for high-performance settings. In order to guarantee continuous performance and endure high-frequency usage, its retail-hardened design includes a robust mechanical base, reinforced paper access lid, and debris-resistant sensors. Toshiba’s emphasis on rugged hardware, unified commerce platforms and broad solution ecosystems explains why it retains leadership in retail technology (Retail Tech Inc., 2025; Armagh POS Solutions, 2025).

3.3. Validation and Verification Processes

When checking how a process or system operates, it's important to know the difference between verification and validation. According to (Erickson, 2025), verification is the process of checking that each activity meets its specifications through reviews of documentation, inspections, and particular testing. This stage makes sure that the process design has been put into action appropriately. Validation is defined in the same study as determining if the system as a whole meets its intended purpose in actual working settings by conducting trial batches, gathering statistical data, and determining if results stay within set bounds. Verification happens before or before deployment and is about "are we doing this right?" Validation, on the other hand, happens after implementation and is about "does this continue to work as expected?" (Erickson, 2025) concludes that verification and validation are complementary; an over-emphasis on one may lead to either unspotted faults or undiscovered variability, but their simultaneous application promotes both regulatory compliance and ongoing operational reliability.

3.4. Business Process Modelling and BPMN

Business Process Model and Notation (BPMN) is a popular way to make diagrams that show how work flows in a company. This way, both management and technical workers can see the same thing. (Visual Paradigm, 2024) says that the Object Management Group came up with BPMN, which is a formal syntax for business processes. It uses shapes and connectors to show tasks, events, choices, and data. The arrows that connect the rounded rectangles for activities, the diamonds for gateways, and the circles for start or end events demonstrate how control and data move, making it a clear and easy-to-understand visual language. The notation is expressive, so it can capture both simple tasks and complicated collaborations in fields like finance, healthcare, and manufacturing. It also helps analysts point out roles, triggers, exceptions, and interactions across organizational boundaries. (Visual Paradigm, 2024) says that BPMN diagrams work with other modeling methods and automation tools to turn designs into working workflows. This standardization makes things more clear, makes it easier for stakeholders to talk to each other, and helps processes get better over time.

3.5. Measuring Business Process Efficiency with KPIs

Key Performance Indicators (KPIs) are used by businesses to find out how well a process is working. They measure what counts. (Şimşek, 2025) stresses that KPIs should be chosen carefully because it is hard to show the worth of an improvement if the measurements don't match the aims of the organization. In his research, effectiveness metrics are characterized as assessing if a product or service meets consumer expectations; examples encompass quality, error rates, customer satisfaction, conversion rates, competitiveness, and profitability. Efficiency measures look at how resources are used. They look at costs, how much work is done, how much money is made, and how long it takes to do activities. They can be divided into overall efficiency, throughput, labor productivity, and resource efficiency (ShareFile, 2024). Managers may find trends, figure out where things are going wrong, and make decisions based on evidence to improve processes by combining indicators from various groups and using process-mining tools to estimate and compare them.

3.6. Depicting KPI Impact Analysis on Future BPMN Models Without Data

Even without historical data, teams can still explore how changes might influence process KPIs by combining modelling with simulation. (Celonis, 2025) proposes building a digital twin by taking the existing BPMN diagram and adding estimated statistics about variables such as arrival rates, processing times, automation levels and staffing. This baseline model lets analysts set up “what-if” scenarios—such as converting a manual step to automation, redesigning a decision gateway or reallocating staff—and run simulations to see how those changes might affect cycle time, throughput or resource usage. (Maté et al., 2017) emphasize that KPIs should link back to strategic goals and the underlying data model so that what-if analysis reflects the organization’s objectives. When no real data exist, practitioners can interview subject-matter experts, set target values and thresholds for each KPI, and use simulation to refine assumptions; by anchoring KPIs in strategic objectives and iterating through simulated scenarios, analysts can form reasoned expectations about the impact of proposed changes before deploying them.

4. Practical Part

4.1. Overview of the company we are focusing on:

Company X offers a wide range of products, such as contemporary barcode scanners, mobile computers, and RFID technology. Their tools automate inventory management and make it much easier to keep track of assets. Company X uses real-time data and analytics to help businesses cut costs and get the most out of their inventory. Their products are particularly beneficial for demand planning and forecasting, which enables organizations quickly adjust to changes in the market.

A lot of small software companies have employed a strategic acquisition strategy in the past. This means they buy other companies and incorporate their products to their own. This method considerably increases their talents and enables them reach a wider spectrum of customers. These smaller suppliers may easily get their hands on advanced tools and knowledge by buying related technology. If not, it would take a lot of time, money, and internal resources to check and prove these tools and knowledge. In our fast-paced world of technology, speed and efficiency are highly vital. That's why these kinds of purchases are so vital for growth.

Because of this, the merged business can better adapt to changing market conditions and quickly change what clients want. This flexibility not only helps the company expand, but it also offers it an advantage over its competitors. The mix of resources and information makes operations run more smoothly and makes it easier for teams in the organization to share their best practices and knowledge. This collaboration space is very important for coming up with fresh ideas and innovations. Also, merging the abilities of people from both enterprises can lead to the production of unique goods and services that combine the greatest ideas and skills from both companies. When people have different points of view, it can lead to new ideas and ways of doing things. This can help solve problems that have been there for a long time in the field.

The fully integrated firm may use modern tools like data analytics and customer playbooks to look at data and adapt its products and services to match the needs of the market as they evolve. This ability to adapt is an excellent demonstration of how working together not only leads to new ideas but also makes it simpler to exchange best practices, which in turn enhances the quality of goods and services. People from different backgrounds have distinct ways of thinking and acting, which makes it simpler to come up with fresh ideas and find new approaches to address problems.

This collaborative approach also helps to cut down on the amount of work that both businesses have to complete, which makes it easier to share best practices. The combined company can now rapidly and easily adapt to developments in the market. This makes it the best in its area and lets it meet the needs of clients as they change. The most crucial thing for success in this situation is to build a culture that promotes teamwork and fresh ideas. This integration aims to achieve precisely that among staff. Because the workforce is made up of people with varied skills, experiences, and points of view, the company needs to put open communication and sharing of knowledge at the top of its list of priorities. This not only improves the quality of goods and services, but it also helps the company as a whole deal with difficulties better by giving teams the tools they need to do so.

A strong and thorough validation procedure is needed to provide stakeholders trust and make sure things are correct. It's important to remember that both using technology and working together are important for making traditional validation processes more efficient. This will save time and money in the long run. As the standards and technologies in the industry change, the validation process must change just as quickly. So, best practices need to alter with these changes to stay valuable and up-to-date.

Company X offers a diverse array of products and vertical solutions, each tailored to meet the specific needs of various industries:

- **Healthcare Vertical**

Company X develops mobile computing devices and advanced barcode scanning technology, which play a pivotal role in enhancing tracking and monitoring solutions within the healthcare sector.
These innovations significantly improve operational efficiency in medical institutions, positively impacting patient care and overall service delivery.
By facilitating real-time tracking of medical supplies and equipment, Company X’s solutions help minimize errors and enhance communication among healthcare personnel, ultimately leading to better patient outcome.

- **Manufacturing Vertical**

In the manufacturing sector, Company X places a strong emphasis on creating intuitive and user-friendly technology that supports healthcare applications.
The focus on user experience is linked to improved patient outcomes and operational success, demonstrating how technology can be seamlessly integrated into manufacturing processes to drive efficiency.

- **Transportation & Logistics Vertical (T&L)**

The Transportation and Logistics sector is critical to the global economy, requiring efficient movement of goods and services to sustain market demands.
Company X is at the forefront of this transformation, providing advanced automation, real-time tracking capabilities, and data analytics to enhance operational efficiency, reduce expenses, and ultimately increase customer satisfaction.

- **Public Sector Vertical**

In the public sector, Company X's technology solutions enable optimal service delivery and efficient resource allocation.
Sophisticated tracking systems developed by Company X can significantly streamline supply management processes.
Additionally, digital platforms for citizen engagement enhance feedback mechanisms and foster collaboration between governmental entities and the communities they serve.

- **Energy and Utilities Vertical**

The Energy and Utilities sector benefits from Company X's smart grid technologies, which enhance operational efficiency and promote sustainability practices.
By employing mobile handheld devices and RFID solutions, organizations can improve their operational workflows, leading to more sustainable outcomes.

4.2. Description of the Current validation process:

Company X's current validation process explained step by step:
1. Partners across the world want to get promotional activities and market visibility of their software/application for X company devices so they first want to get validated to receive the promotional activities.
2. They fill out a nomination form to get started with the certification process, filling in all the details about their company, application/solution, and contact person. These details are then saved into a Smartsheet and notified to validation process members that a new process validation nomination arrives.
3. Members of validation team then sends an email to the partner to set an initial call invite to discuss the validation process.
4. There are two types of validation process that this company offers one we called validation which consists of sending their application/solution with a day to day life scenario their customers use. We will conduct black box testing based on their explanation of that scenario. Another one we called compatible it consists of the partner doing the test or demonstrating one simple functionality of their application on their sided devices and send us a video recording and some Log files.
5. Typically partners go for the validation option as they don't have enough or all the hardware devices on their end. We then send a test plan where they fill out the test case scenarios into dedicated tables one is called prerequisites and another one is called procedure.
6. Then the partners fill out the test plan and send it back to us by creating a demo account into their solution for Q&A team to test as an end user.
7. Q&A team then do some black-box testing based on the test scenario explained by the user. We analyze some Log files and calculate CPU usage, RAM or memory usage, and Battery and current drain for that particular application. If everything works well, Validation team sends out the certificate and some badges. Also, marketing guide on how they will collaborate with the marketing team to get the most out of this process.
8. Validation team fill out the device list three times, one into CRM (Salesforce) and then into internal tracking portal, then on to the Published Certificate.
9. When everything is done, Q&A team mark the partner on:
   a) Into Smartsheet: To tell the marketing team that they are validated.
   b) Into CRM (Salesforce): Q&A team also have to mark them validated so that the company can track them.
   c) Into internal web portal: Q&A team close them as process completed.
10. Validation team also have to do Monthly report into a Google sheet to help visualize how many validations came in every month and how many they have completed.

4.3. Problems with the current validation methodology

4.3.1. Cross-Platform Tracking:

The validation process is one of the most difficult and confusing parts of technology and project management right now. It depends a lot on a set of specialized platforms that work together, such as Smartsheet, Salesforce, and an internal tracking portal. This complicated network of technologies creates a lot of big challenges that can slow down work and make it less productive.

4.3.2. Data Duplication and Inconsistencies:

There is a lot of duplicate data because the two display systems need to enter data independently. This means that the two systems will always be different. This level of complexity can make it hard to have a clear picture of test statuses and observe how test cases are doing, which might make everyone feel like they're not connected. Because of this, decision-making is harder, and critical resources are squandered trying to remedy problems. Also, mixing data from different sources might make reports more likely to be inaccurate, which can slow down projects and degrade the quality of the final product. With Generative AI, partners could quickly and clearly say what they needed, which made it possible to automatically create standardized test cases for varied situations. This not only makes things easier, but it also makes data reporting more accurate and of higher quality. This lets teams focus on analysis instead of reconciling data.

4.3.3. Unnecessary communication between platforms:

When partners have to deal with different systems, it can make communication harder, which is bad when time is of the essence. Partners should be able to use a single, integrated system that combines all the features they need, much like customers’ demand from unified platforms. The fact that tools are spread out makes it hard to understand the testing process, which slows down the workflow and makes it harder to address partner’s questions. A fully integrated platform, especially one that uses Generative AI, can help partners make connections and work together smoothly. This would make communication easier and more direct, which would lower the chances of misunderstandings and improve collaboration as a whole.

4.3.4. Using Manual Communication and Coordination:

The challenges with the validation process are largely due to using manual communication methods, such email, to set up meetings and talks. It takes a long time to schedule calls and talk to the same individuals over and over again. Then, having to send all of those chats back and forth only makes things worse. Q&A teams and our partners could have big misconceptions because of these kinds of delays, which could put the whole project's success at danger. Using new digital tools, early communications may be automated. This would make things easier and encourage people to get involved right away. Automating tasks can mean getting automatic alerts when test cases are made or when their status changes. This cuts down on emails that developers and testers don't want, which saves them both time. Partners just have to talk to one person with inquiries when they use a centralized communication platform. This makes it much less likely that people will misunderstand each other and speeds up the whole validation process.

4.3.5. Unproductive Test Plan Submission and Review Process:

Right now, partners have to make separate demo accounts in order to fill out test plans correctly, which might waste a lot of time. Also, using a linear feedback loop, where teams have to wait for test case submissions and manual input, could make projects take longer and be less effective overall. This is a fantastic possibility for Generative AI to speed things up by quickly turning partner descriptions into test cases that are accurate and useful for the situation. This will speed up the process of submitting things. The AI's ability to build test cases from brief but detailed descriptions speeds up the transition to the testing phase by minimizing the possibility that people who aren't paying careful attention would miss key details.

Also, using a quick, iterative feedback system would speed up the process of reviewing and approving test plans. For example, partners might use AI-generated test cases that are made and sent out right away. This would let users check how accurate and useful the cases are and give input that the AI may use to improve in the future. This new method of thinking about a more responsive feedback loop will probably speed up the process of getting started and make the validation process more efficient in the end. This will make the workflow smoother and make it easier for people to work together.

4.3.6. Long Validation Time:

Currently, it takes roughly four weeks on average to finish a validation cycle. This kind of protracted process not only makes it take longer to get new items, but it also makes the validation process less responsive. This is an issue for a sector that needs to be able to quickly respond to changes in the market and what customers want. The purpose of Generative AI is to speed up validation runs by making sure that standardized test cases are created as quickly as possible and match the important situations provided by the partner. Company X's proactive approach puts it ahead of the curve, which allows it respond to market demand more swiftly and strengthens its competitive edge.

The AI technology makes test cases that are far more accurate and faster than traditional manual methods by applying algorithms that are based on prior data, company rules, and widely used test cases. This means that when partners bring in a scenario, the AI provides them test cases right away. This speeds up the validation process and makes the whole project work better.

4.3.7. No clear KPIs or feedback loops:

The lack of clear key performance indicators (KPIs) in the validation process makes it harder to quantify performance and figure out how effectively testing methods work as a whole. It is very hard to keep track of progress and make continual changes without specific metrics for each Service Level Objective (SLO). Integrating Generative AI into the validation process makes it easy to gather statistics on things like how long it takes to develop test cases, how many errors there are in submissions, and even how satisfied partners are with the quality and relevancy of AI-generated data. This lets Company X regularly check its validation framework in a data-driven setting, which encourages a culture of always getting better. The organization may use this data to make smart changes to future processes and set up a strong way to measure and confirm the success of integrating Generative AI.

4.3. Key Performance Indicators for Traditional vs AI-Enhanced Validation

In order to assess the influence of artificial intelligence on the validation process that is utilized by Company X, this study employs the key performance indicators that are shown in Table 1 and groups them into the following four categories: Speed/Efficiency, Test Quality, Partner Satisfaction, and Resource Utilization. A precise definition, unit, and a comparison objective contrasting AI with the manual process are all included in every key performance indicator (KPI). A description of how each statistic is calculated and how it should be interpreted may be found below.

Speed/Efficiency: Average Testing Cycle Time (Days- lower is better).
Average days from the start of testing to final sign-off for an application. Compare the mean cycle time between the AI-enhanced and manual processes.A lower AI value indicates faster throughput.

Test Quality: Defect Detection Rate (Percentage- higher is better).
Monthly share of applications that did not meet requirements relative to all processed that month: *Defect Detection Rate = Inactive ÷ (Completed + Inactive) × 100%*.
A higher rate suggests the process is catching more issues before release or applying stricter exit criteria. Compare monthly rates for AI and manual using the same application mix.

Partner Satisfaction: Satisfaction Score (Satisfied/Dissatisfied- more Satisfied is better).
Post-validation survey outcome recorded as Satisfied or Dissatisfied. Summarize as the percentage of "Satisfied" responses (or the Satisfied-to-Dissatisfied ratio). An effective AI process should increase the share of Satisfied responses and reduce post-release concerns.

Resource Utilization: Tester Efficiency (Quantity- higher is better).
Average number of application tests completed per tester per month:
Tester Efficiency = Total tests completed in a month ÷ Number of active testers.
Higher values indicate better throughput per person, reflecting productive use of AI support.

Resource Utilization: Average Test Coverage (Quantity- higher is better).
Average number of different device models tested per application, used as a proxy for device-compatibility coverage. Compute the mean count of device models exercised per app; higher values indicate broader coverage without proportional increases in effort.

Comparison approach:
For each KPI, we compute results for both processes over the same period and application set, then assess them against the comparison goal in Table 1 (Lower, Higher, or More Satisfied).

For example, improvements might include cycle time dropping from 10 to 6 days (lower), defect detection rate rising from 20% to 30% (higher), the share of Satisfied partners increasing, tester efficiency improving from 3.0 to 4.5 tests per tester per month (higher), and average device coverage rising from 6.0 to 8.0 models per application (higher). Together, these shifts would provide quantitative evidence that the AI-enhanced process outperforms the manual baseline.

Table 1KPI's for Validation Processes (Traditional vs. AI-Enhanced)

Category	**KPI**	**Definition**	**Unit**	**Comparison Goal (AI vs. Manual expected outcome)**
Speed/Efficiency	**Average Testing Cycle Time**	the average duration to complete testing of an application (in days).	Days	Lower
Test Quality	**Defect Detection Rate**	the rate of applications that did not meet requirements (Inactive) versus those completed, measured monthly.	Percentage (%)	Higher
Partner Satisfaction	**Satisfaction Score**	partner satisfaction levels with Company X's testing programs (survey-based).	Satisfied/ Dissatisfied	More Satisfied
Resource Utilization	**Tester Efficiency**	the average number of application tests completed per month by each tester.	Quantity (can be fraction number)	Higher

Category	KPI	Definition	Unit	Comparison Goal (AI vs. Manual expected outcome)
	Average Test Coverage	the number of different device models tested per application.	Quantity (can be fraction number)	Higher

Each KPI is measured for both the traditional and AI-enhanced workflows. For instance, to evaluate cycle time, record the average duration in hours over several iterations of each process. Coverage is measured via code instrumentation or requirement checklists. Survey scores are averaged numerically, and query counts are tallied. In analysis, we can compute improvements, e.g. percentage reduction in cycle time or increase in coverage, to clearly quantify the benefits of AI. By using these well-defined, literature-supported KPIs, the thesis can rigorously compare Company X's manual and AI-driven validation processes across speed, quality, partner impact, and resource use.

4.4. Analysis of KPIs for Company X's Testing Program (2022–2025)

This analysis evaluates the software testing performance of Company X's Global Enablement Center programs from January 2022 through August 2025. The data includes testing records (application name, device models used, tester, status, start/end dates) and a partner satisfaction survey. We derive five Key Performance Indicators (KPIs) to assess efficiency and effectiveness:

- **Average Testing Cycle Time** – the average duration to complete testing of an application (in days).
- **Defect Detection Rate** – the rate of applications that did not meet requirements (Inactive) versus those completed, measured monthly.
- **Average Test Coverage** – the number of different device models tested per application (indicating coverage of device compatibility).
- **Satisfaction Score** – partner satisfaction levels with Company X's testing programs (survey-based).
- **Tester Efficiency** – the average number of application tests completed per month by each tester.

4.4.1. Average Testing Cycle Time

Definition: *Testing Cycle Time* is the time from test start to test completion for an application. Shorter cycle times mean faster validation, allowing quicker go-to-market for partners' solutions. Monitoring cycle time helps identify bottlenecks and drive efficiency improvements. Companies often aim to continuously reduce cycle time (e.g. targeting ~15% reduction) as part of process improvement (BusinessPlan-Templates, (2023)).

Calculation: We calculated the duration in days for each completed test (Close Date – Start Date). There were 426 completed application tests in the dataset. To avoid skewing the average with extreme cases, we excluded tests that took longer than 60 days (roughly 2 months) to finish. This removed 136 outlier cases (about 32% of completed tests), which likely involved extended delays or scope changes. The remaining 290 completed tests form the basis of this KPI. For each included test, we noted the Application Name and the list of Devices tested in that cycle, but for cycle time we focus on the overall test duration regardless of device count. Finally, we computed the average cycle time in days for all these tests. We also segmented the data by year to observe trends over time.

Results: The overall average testing cycle time (with less than 60-day outliers removed) was approximately 27.1 days. In other words, on average it took about four weeks for an application's validation testing to be completed by Company X's team. Table 2 below shows the breakdown by year, and **Figure 1** illustrates the trend. We can see that the average cycle time improved over the years. It hovered around 28–29 days in 2022 and 2023, then dropped to 26.9 days in 2024, and further to 23.6 days for Jan–Aug 2025. This marks a significant improvement in speed: roughly an 18% reduction in average cycle time from 2023 to 2025 (from ~29.1 to 23.6 days). This improvement exceeds typical annual reduction targets (often ~15%) (BusinessPlan-Templates, (2023)), indicating successful efficiency gains, possibly due to process optimizations or increased testing automation and experience.

Table 2Average Testing Cycle Time by Year

Year	Number of Tests (<=60d)	Average Cycle Time (days)
2022	64	28.7
2023	66	29.1
2024	99	26.9
2025	61	23.6

Note: Average Testing Cycle Time by Year (outliers >60 days removed) *(2025 is Jan–Aug partial year)*

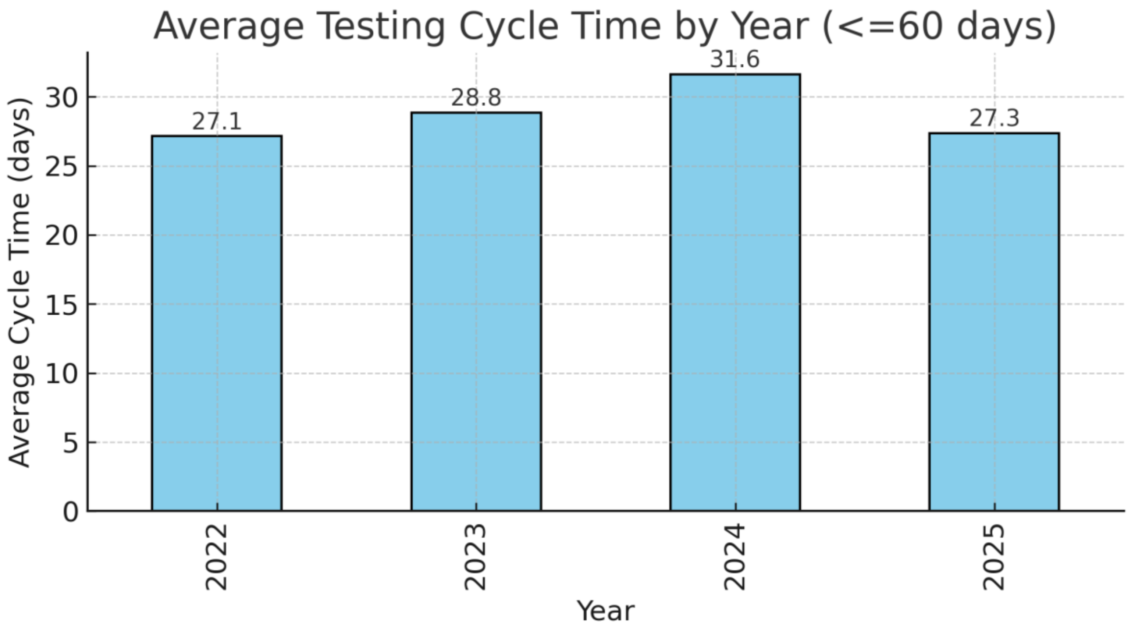

Figure 1 Average testing cycle time (in days)

Note: *Average testing cycle time (in days) for completed application tests by year. A downward trend is observed, indicating faster test completion in recent years.*

Discussion: The average cycle in 2022 was around 28.7 days, and in 2023 it was about 29.1 days, which is about the same. This means that there wasn't much progress at first. Things like bringing on new testers or making applications more difficult could have maintained cycle time high. But in 2024, it went down to around 26.9 days, and in 2025 YTD, it went down even more to about 23.6 days. This steady decrease means that the testing procedure is getting more efficient. The team may have simplified the testing process or put a higher priority on getting results faster. Another possible reason is because partners are more prepared, which means there are less problems to solve and testing may be done faster. If Company X's program can get an average of less than 24 days to finish a test by 2025, it's a good sign. This is because these tests are end-to-end validation cycles that involve many devices and possibly iterative fixes. This is in line with best practices that stress cutting cycle time to make sure software is released on time (BusinessPlan-Templates, 2023).

We purposely left out tests that lasted more than 60 days. Those outliers (136 times) generally took a long time, which could change the average. Adding them would make the average go up a lot. Some projects might have been put on hold because of serious app problems or partners that didn't respond quickly enough. those certification process could take 3 to 6 months or more to finish, but they are rare. We concentrate on the "typical" cycle by filtering them out. In the framework of a thesis, one may rationalize this by asserting that we focus on typical process performance rather than outlier instances. Figure 1 shows that the process is getting better with time, from 2022 to 2025. Testers probably got better, and the program may have made changes (such improved testing tools or stricter entrance requirements for apps) to speed up the validation process.

To sum up, Company X's typical testing cycle time is about 4 weeks, and it has become better by over a week (18%) in the last few years. This trend is a good sign because shorter test cycles mean more work gets done and partners are happier, which means the company is becoming more efficient. Company X's partners may get their solutions verified and to market faster when cycle durations are minimal. The goal should be to keep this downward trend going or keep the cycle time at a low level that is still good for testing.

4.4.2. Defect Detection Rate (Inactive Application Percentage)

Definition: In this context, *Defect Detection Rate* refers to the percentage of application tests that result in "Inactive" status**,** meaning the application did not meet the requirements or had critical issues and thus the testing was halted without a successful completion. In short, it's the failure rate of app validation, or how often faults or non-compliance are serious enough that the app can't be certified. A high inactive rate would show that a lot of apps are failing testing. This could mean that the program is good at finding flaws, but it could also signal that a lot of submissions are not good quality. A low inactive rate, on the other hand, means that most apps pass validation. We look at this as a percentage of all tests and check it every month to see if there are any trends. This monthly check lets us see if the quality of the applications we get (or the testing rigor) is changing over time.

Calculation: For each month from Jan 2022 to Aug 2025, we looked at all application tests that started in that month and determined their outcome (Completed vs Inactive). If an application was marked Inactive (failed), it typically has no close date; we assume its testing effectively "ended" in the start month for counting purposes. Completed applications have a finish date.

For each month, we computed:

Completed Applications: number of tests started that month which eventually completed successfully.
Inactive Applications: number of tests started that month which ended in inactive status (failed to meet requirements).

Then we calculate the percentage that were completed vs inactive. For example, if 10 tests started in a month and 7 completed, 3 inactive, that month's completed% = 70%, inactive% = 30%. Table 3 presents the monthly data, and Figure 2 plots the percentage trends. (Note: If a test was Open as of Aug 2025, it's not counted in either category yet.)

Table 3Monthly Outcomes – Completed vs Inactive Applications

Month	Completed Applications	Inactive Applications	Completed (%)	Inactive (%)
2022-01	16	10	61.5%	38.5%
2022-02	12	4	75.0%	25.0%
2022-03	10	9	52.6%	47.4%
2022-04	7	3	70.0%	30.0%
2022-05	10	3	76.9%	23.1%

Month	Completed Applications	Inactive Applications	Completed (%)	Inactive (%)
2022-06	19	20	48.7%	51.3%
2022-07	7	5	58.3%	41.7%
2022-08	12	6	66.7%	33.3%
2022-09	8	4	66.7%	33.3%
2022-10	15	3	83.3%	16.7%
2022-11	9	3	75.0%	25.0%
2022-12	6	4	60.0%	40.0%
2023-01	9	4	69.2%	30.8%
2023-02	11	4	73.3%	26.7%
2023-03	11	4	73.3%	26.7%
2023-04	9	2	81.8%	18.2%
2023-05	8	4	66.7%	33.3%
2023-06	9	0	100.0%	0.0%
2023-07	15	3	83.3%	16.7%
2023-08	10	4	71.4%	28.6%
2023-09	7	3	70.0%	30.0%
2023-10	8	1	88.9%	11.1%
2023-11	12	2	85.7%	14.3%
2023-12	5	1	83.3%	16.7%
2024-01	8	2	80.0%	20.0%
2024-02	10	0	100.0%	0.0%
2024-03	19	0	100.0%	0.0%
2024-04	8	2	80.0%	20.0%
2024-05	10	0	100.0%	0.0%
2024-06	13	0	100.0%	0.0%
2024-07	14	0	100.0%	0.0%
2024-08	13	3	81.3%	18.8%
2024-09	18	1	94.7%	5.3%
2024-10	14	1	93.3%	6.7%
2024-11	11	1	91.7%	8.3%
2024-12	14	0	100.0%	0.0%
2025-01	12	0	100.0%	0.0%
2025-02	15	2	88.2%	11.8%
2025-03	14	1	93.3%	6.7%
2025-04	13	1	92.9%	7.1%
2025-05	4	1	80.0%	20.0%

Month	Completed Applications	Inactive Applications	Completed (%)	Inactive (%)
2025-06	6	0	100.0%	0.0%
2025-07	12	0	100.0%	0.0%
2025-08	6	2	75.0%	25.0%

Note: *Monthly Outcomes – Completed vs Inactive Applications (Jan 2022 – Aug 2025)*

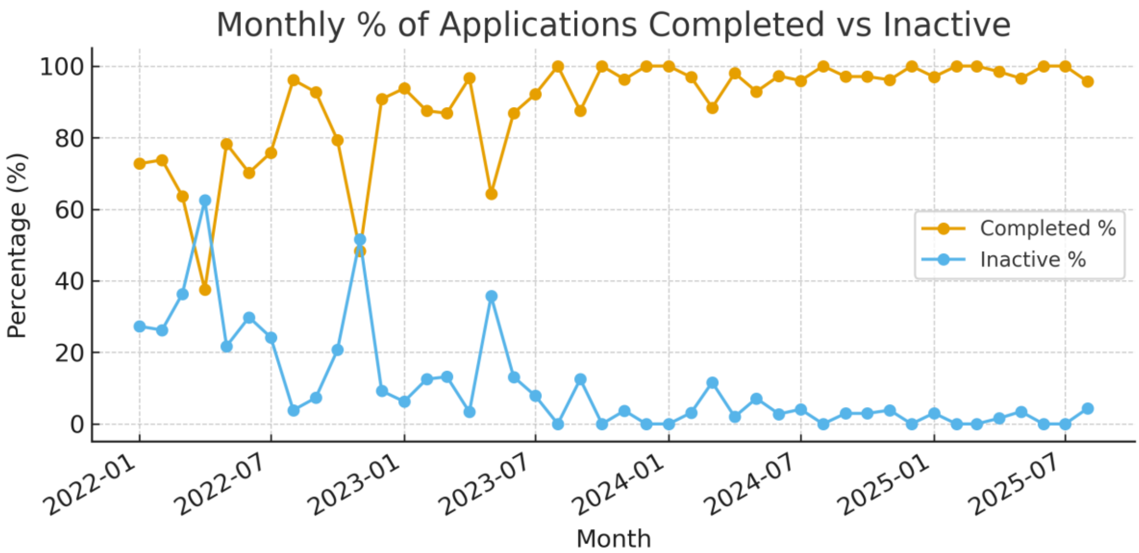

Figure 2Monthly percentage of applications that were completed (blue line) vs marked inactive (red line).

Note: *Monthly percentage of applications that were completed (blue line) vs marked inactive (red line).The sum of Completed% and Inactive% is 100% each month (ignoring any ongoing open tests). A clear downward trend in inactive percentage is observed over time.*

Results and Discussion: The information above shows that the fault detection rate (inactive rate) has improved a lot over the time period. In the beginning of 2022, about 30–40% of the apps that were examined were not passing validation. For example, 38.5% were not active in January 2022, and even more than 50% were not active in June 2022. This high failure rate shows that many partner apps didn't satisfy the standards at first or had serious problems that Company X's testing found. It's a double-edged sword: on one hand, a high inactive rate implies that the testing process is doing its job by finding apps that don't follow the rules, but on the other hand, it means that a lot of the apps submitted weren't ready or weren't good enough. This could put a load on resources because many tests failed.

But with time, the percentage of those who were inactive slowly went down. By 2023, most months had inactive rates between 10% and 30%, and by the end of 2023, several months had rates in the teens or single digits. Throughout 2024, there were several months

with 0% inactive, which means that every app tested that month passed. For example, from February to July 2024, 100% of tests were successful. The yearly failure rate went from around 36.8% in 2022 to about 12.1% in 2024 and then to about 8% in 2025. This shows that the quality and/or effectiveness of the testing have gotten a lot better. The inactivity rate went up to 25% in August 2025 (2 out of 8 finished tests failed), although that's based on a tiny sample and is still far lower than the over 40% reported in early 2022.
Several explanations could be put forward to account for this encouraging trend:

- **Higher Quality of Applications:** Company X's partners probably learned from the remarks and started sending in better-prepared applications over time. A lot of apps didn't work at first, which may have motivated partners to undertake more internal QA or follow Company X's advice before formal testing. By 2024–2025, a lot more apps passed the first test, which meant fewer failures.
- **Better Pre-screening:** Company X may have made its pre-validation procedure better. For instance, they might have put in place stiffer entrance requirements or preliminary inspections that weed out apps that aren't ready for official testing. This would cut down on the number of apps that are actually untestable or not compliant that go into the cycle, which would lessen the number of inactive apps.
- **Testing Efficiency:** The testers may have learned how to find problems earlier and help fix them during the test cycle (turning what could have been a "inactive" into a "complete" by pushing partners to fix things quickly). But an inactive means the test was stopped, therefore it's probably more about the quality of the input than the speed of the tester.

Still, the fact that the defect detection rate dropped to almost zero in some months is a very good indicator. It suggests that almost all of the apps that were tested throughout those times passed validation. This could be because the apps were well-made or because the program's assistance for partners got better. The program may have "worked itself out of a job" in a way. At first, it found a lot of bugs (which is excellent because they would be caught before release), and over time, that feedback loop led to better submissions.
From a KPI point of view, a lower percentage of inactive users means that the program is more successful. By 2025, more than 90% of the applications tested each month are getting accepted (finished). This means that Company X's validation program got better and partners started to follow the rules more closely.

Figure 2 shows that June and July 2025 had 0% inactive (100% pass rate). One thing to keep in mind is that there were still some tests open and continuing on throughout those months (see Table 3, for example, July 2025 had 7 open tests that hadn't been counted yet). Some of those open tests may have stopped working later, which would make the failure rates for those months look a little higher in hindsight. Even with that in mind, the overall trend is clear: the failure rate dropped from high (30–50%) at the beginning of 2022 to very low (typically around 10%) by 2024–2025. This trend shows that either the quality of the incoming apps or the testing procedure has gotten a lot better (probably both). It also means that by 2024, partners are considerably happier and more confident because most of their apps pass. This is in line with the satisfaction survey results (which will be discussed later).
Keeping the defect detection (inactive) rate low is good since it means partners don't have

to waste time on tests that don't work and can get their products to market faster. However, Company X also needs to make sure they're not just making tests easier to raise pass rates. Based on the high satisfaction numbers and ongoing improvements, it appears more plausible that the quality really did get better rather than standards just getting lower. In general, this KPI indicates a significant positive trend. The program catches fewer bad apps over time because there are fewer bad apps to begin with. This is a sign that the ecosystem's quality is getting better.

4.4.3. Average Test Coverage (Devices Tested per Application)

Definition: This key performance indicator (KPI) looks at how many devices can be used to test each application. When a partner's application is evaluated in Company X's programs, it is run on one or more devices of different models to make sure they work together. Average Test Coverage here is the average number of different device models that are used to test an app. A higher number suggests more coverage, which is good for making sure the software works on a wide range of devices. Enterprise apps must work on all devices, thus testing them across a number of devices (including multiple models and OS versions) is important. This helps find problems that only happen on certain devices and gives clients peace of mind (Testsigma (2023)). On the other hand, if only one device is examined, the coverage is limited and problems on other models may be missed.

Calculation: We get the Device list for each application test from the test records. Many application tests used more than one device model (in the data, each model shows up as a different row for the same App and Test ID). We counted the number of different device types utilized for each test (each test cycle had its own App Name). If App A was tested on TC26 and TC57X in one cycle, it means it was tested on two devices. According to data TC26 and TC57X are specific hardware models from Company X. If the same app was tested many times between 2022 and 2025 (maybe a re-validation or new version), we first counted the number of devices for each test and then averaged them for that app so that apps with multiple tests didn't get too many devices. Lastly, we found the general average for all applications and looked at how many devices were used for each application.

Results: Across the dataset, the average number of devices tested per application is about 4.0. In other words, on average an application validation involved roughly four different device models. However, the distribution varies widely (see Table 4 and Figure 3 below). Some applications were tested on only 1 device (minimal coverage), while others were tested on a dozen or more device models. Table 4 summarizes how many applications fell into each device-count category, and Figure 3 visualizes this distribution.

Table 4Distribution of Device Coverage per Application

Devices Tested	Number of Applications	Percentage of Applications
1	152	31.4%
2	58	12.0%
3	59	12.2%
4	36	7.4%

Devices Tested	Number of Applications	Percentage of Applications
5	28	5.8%
6	27	5.6%
7	21	4.3%
8	27	5.6%
9	21	4.3%
10	17	3.5%
11+	38	7.9%

Note: *Distribution of Device Coverage per Application (Unique device models tested per application, 2022–Aug 2025)*

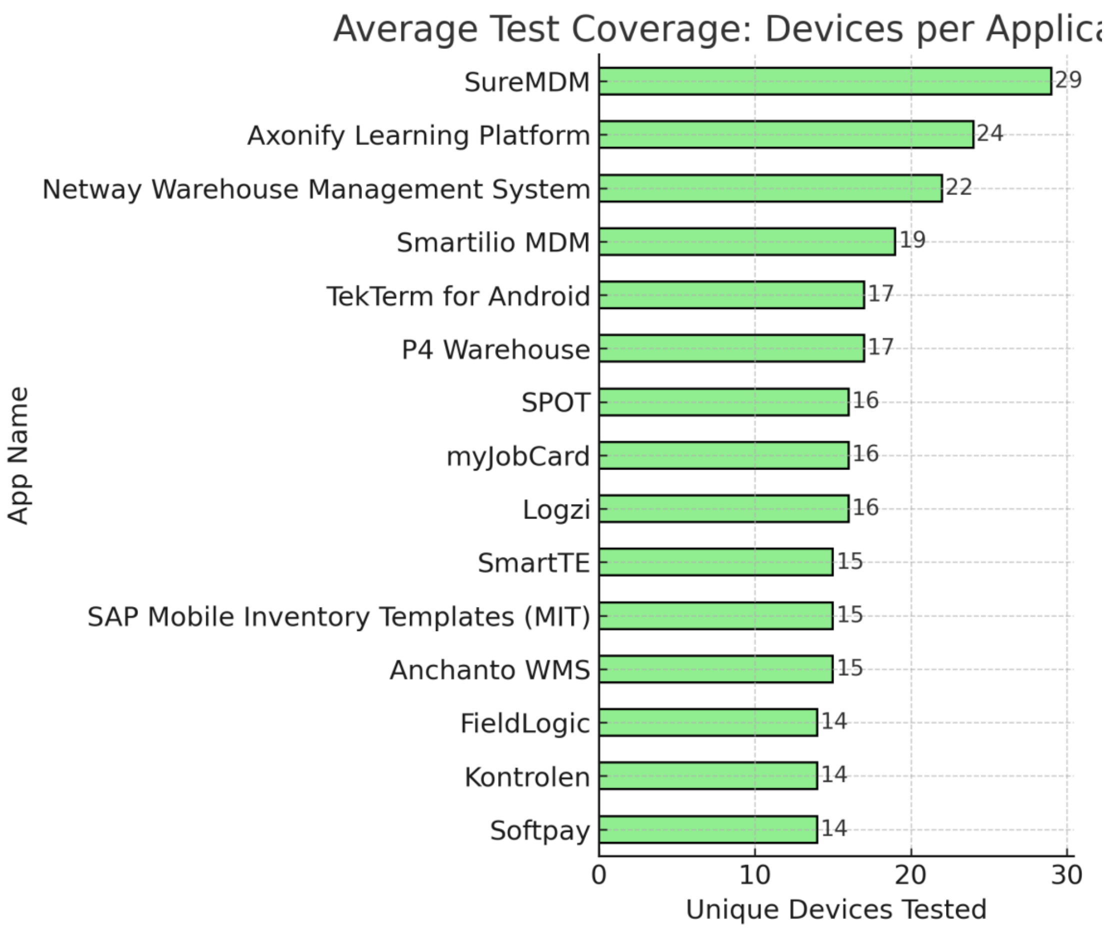

Figure 3Number of device models tested per application

Note: *Number of device models tested per application. Most applications were tested on only 1–3 devices, though some were tested on a much larger number of devices (the "11+" category aggregates all apps tested on 11 or more distinct models).*

Discussion: Most of the applications functioned exclusively on a limited number of devices during testing. Table 4 reveals that only one device type was used to test about 31% of the apps. Two to three devices were utilized to test about 24% of the apps. About two-thirds of the apps (31.4% + 12.0% + 12.2% ≈ 55.6%) were tested on three or fewer devices. On the other end of the scale, 38 apps (7.9%) were tested on 11 or more different devices. For some applications, this is a lot of testing. The most extreme case was a software that was evaluated on 29 different device models, which is a lot of various types of devices. This was a fairly complete test of compatibility. This might have been a big corporate app that worked with practically all of Company X's devices. We tried a few more on 15 to 24 devices. These are exceptions, but they show that the app can undertake rigorous cross-device validation when it needs to.

The average of around four devices per application shows that the coverage is good. It's not too high or too low. Company-X makes a number of various kinds of devices, so if a tester wants to check for broad compatibility, they may have to test on a lot of them. It likely means that testers choose a set of devices that are common for each program. They might test on one device for each major product line or form factor that is crucial to the app's target market. For example, they might use a touch computer (TC series), a handheld scanner, a tablet, or other devices to test a retail app. Some apps only operate on certain devices or in certain scenarios (that's why we only tested one device), while others are more general and need to be tested on many devices.

An app needs to work on more than one device to make sure it works on different types of hardware and operating systems. Industry best practices for checking mobile apps recommend to test them on a variety of devices, even older ones, to ensure sure they work with a lot of different devices (Testsigma, 2023). Our data shows that nearly 20% of apps were tested on five or more devices (5.8% + 5.6% + … + 7.9% ≈ 24.3% on ≥5 devices). These are usually the more complex or popular programs where compatibility is quite critical. The fact that more than half of the apps were tested on only one to three devices could suggest that they were created for a certain device or operating system, or that there weren't enough resources to test them all.

To put it simply, each app gets tested on roughly four devices on average, but this isn't always the case. Some apps are tested a lot, while others aren't. This number shows Company X how many tests are done. At least three to five devices for each app should be tested by the business if the goal is to make sure that apps work with additional devices.

This KPI shows that the program can work with a lot of different devices. Most apps only check a few devices when they validate, which is probably enough to cover the most common ones. It would be helpful to clarify in the thesis how devices are chosen, like depending on how popular they are or what a partner wants. Testing on a lot of different devices is critical because testing on just one can miss problems on other devices. If end users utilize different models, this is a risk. These are possibly the greatest apps that can run on a lot of different devices because they were tested on 11 or more devices. This score tells us how comprehensive (or not) the testing process is for each program.

4.4.4. Satisfaction Score (Partner Satisfaction Analysis)

Definition: The Satisfaction Score tells us how happy the program's partners and vendors are with Company X's services. We get this from the answers to the survey. In the partner survey, people rated how satisfied they were with the services and said which Global Enablement Center package they utilized (Validation or Compatibility). We read the survey's Service Satisfaction Level field to see if people were "Satisfied" or "Dissatisfied." This KPI looks at the percentage of partners who are happy and also breaks down satisfaction by program type (Validation vs. Compatibility). It basically solves the questions: What percentage of partners are happy with Company X's testing services? Are the Validation and Compatibility programs different in terms of how happy partners are with them?

Calculation: We counted the number of survey responses that fell into each category (Satisfied vs Dissatisfied) for each program (Validation, Compatibility). There was a total of 74 responses (66 for Validation program, 8 for Compatibility program). We then computed satisfaction rates (percent satisfied) for each program and overall. Table 5 shows the breakdown, and Figure 4 and Figure 5 visualizes the satisfaction vs dissatisfaction split per program.

Table 5Partner Satisfaction Survey Results

Program	Responses	Satisfied	Dissatisfied	Satisfaction Rate
Validation	66	62	4	93.9%
Compatibility	8	7	1	87.5%
Overall	74	69	5	93.2%

Note: *Partner Satisfaction Survey Results (Validation vs Compatibility Program)*

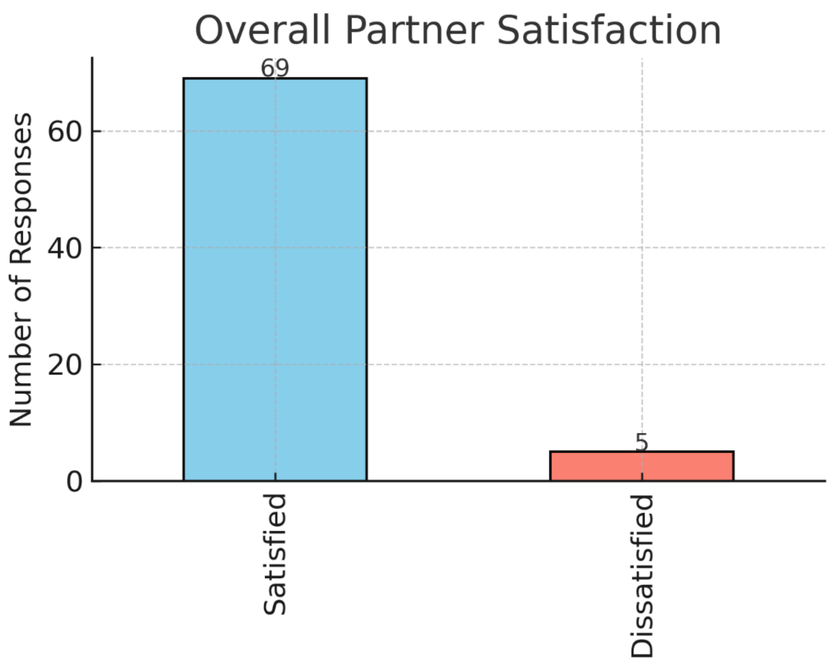

Figure 4Overall Satisfaction Score

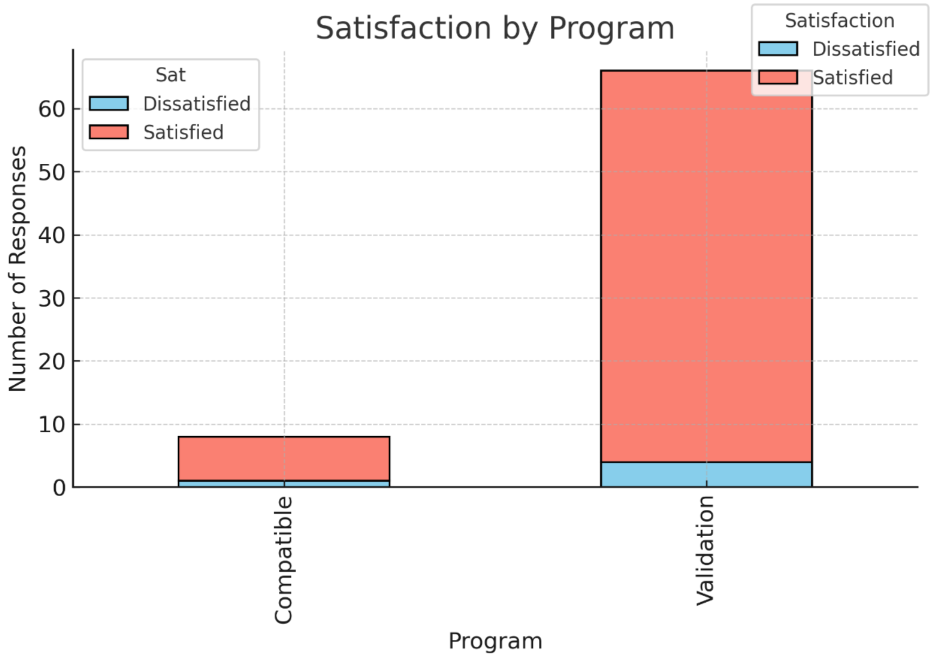

Figure 5Program wise Satisfaction Score

Note: *Partner satisfaction rates by program. Each bar is 100% of respondents for that program, split into Satisfied (blue) and Dissatisfied (red) segments. The Validation program achieved ~94% satisfaction, while the Compatibility program had ~88% satisfaction (note: small sample size for Compatibility).*

Results and Discussion: The survey results show that most people are highly happy with Company X's programs. Of the 74 people who answered, 69 said they were "Satisfied" with the service and only 5 said they were "Dissatisfied." That means that almost 93% of people are happy with it. This is a great achievement in terms of client happiness. Research shows that keeping satisfaction levels above 90% is significantly linked to customer loyalty and repeat business (BusinessPlan-Templates (2023)). A satisfaction score this high means that most partners think the testing program is useful and fits their needs.

Program-Specific Breakdown: The Validation program (the main tool for checking that partner apps work on different devices) had 66 responses, and 62 of them were happy with it and 4 were not. This is about 93.9% happy. Seven out of eight people that answered the Compatibility program were happy with it, which is 87.5%. We should point out that just 8 people responded to the Compatibility program, thus one unpleasant reaction changes the proportion a lot. It looks like the satisfaction rating for the program is a little lower because one person wasn't happy with it. But we need to be careful about how we interpret this because the sample size is so small. If that one partner's problems were fixed, it could easily be 100%.

It looks that partners in the Validation program are quite happy, and those in the Compatibility program are also happy, although not quite as happy. The survey also asked which program the input was about. Interestingly, most of the people who answered (66 out of 74) were using the Validation program. This could mean that more people use the Validation service or that it is more important, which would explain the higher number of comments. Compatibility only got a few responses, which could imply that the software is either new, not used much, or only a small part of the survey audience.

It is helpful to discuss potential causes of the high pleasure from the standpoint of the thesis: - The lower defect rate and faster cycle time we noticed previously are probably making people more happy. Partners are getting their apps approved faster and more reliably over time, which would make them satisfied with the service. The team at Company X might be doing an excellent job of helping and communicating with customers. Partners place a lot of significance on the testing program. With almost 93% of partners saying they are happy with it, we may assume that most of them actually find it useful.

It's also important to think about what unhappy couples might be saying. There were only 5 responses that were not happy (4 in Validation and 1 in Compatibility). There may not be many, but a thesis could look at their opinions if they are available (the survey contained spaces for comments and suggestions). Some common problems can be that the test took too long, that there were problems with communication, or that they didn't agree with the test results. Without that information, we can only say that some partners were not happy. But while more than 90% are happy, the few that aren't could be because of specific problems or higher expectations.

Validation vs. Compatibility: A Comparison In terms of satisfaction, the Validation program is a little better than Compatibility (94% vs. 88%). This can mean that the main validation service is working extremely well. The Compatibility program might not be perfect yet because it only has a tiny sample size or it hasn't gotten enough feedback. Some people could think that the only problem with Compatibility is that the criteria or process aren't clear enough (because compatibility testing might use different methods). But because so many people are happy overall, the difference may not be statistically significant because of the sample sizes.

This KPI suggests that the partner satisfaction level is quite high overall. Getting more than 90% of people to say they are happy is a good sign that Company X's Global Enablement Center initiatives are meeting the demands of its partners. This high level of pleasure is likely to lead to good word-of-mouth and partner loyalty. Companies with satisfaction rates over 90% frequently keep more customers and get more referrals (BusinessPlan-Templates (2023)). This is good news for Company X because partners will probably keep using the validation services and tell others about them, which will help the ecosystem develop.

4.4.5. Average Efficiency of Testers (Applications Completed per Tester per Month)

Definition: This KPI looks at how productive each of Company X's testers is. It specifically counts the average number of application tests that each tester finishes in a month. So, how many apps does each tester finish testing in a month? This is a stand-in for efficiency. This KPI answers that question. However, it can be affected by how hard the tasks are, not only how efficient the person is. It helps figure out how to divide up the labor and who is doing the best job. A larger figure suggests that a tester is doing a lot of tests per month (greater throughput), whereas a lower number could mean that they have fewer assignments or that each assignment takes longer. For the program's management, this KPI can help them figure out how to best use their resources and what kind of training they need. For example, if some testers routinely finish considerably fewer apps, it could mean that they are working on more difficult projects or require help.

Calculation: Using the test records (Owner field indicates the tester responsible), we identified all completed tests and which tester completed each one. There were 10 distinct testers (labeled here as Tester1 through Tester10 for anonymity). For each tester, we counted how many application tests they completed each month, then computed the average of those counts across the months of their active service. To make it fair, the averaging was done over the period each tester was actually active in the program. (For example, if a tester started in mid-2023, we average their outputs from then on, not from 2022.) We also computed each tester's total number of completed applications over the entire period for context. Table 6 summarizes each tester's total completed tests and their average per month, and Figure 5 plots the average monthly completions for each tester.

Table 6Tester Productivity Summary

Tester	Total Completed (2022–2025)	Avg Completed per Month
Tester1	203	5.49
Tester2	53	2.94
Tester3	44	2.44
Tester4	81	2.38

Tester	Total Completed (2022–2025)	Avg Completed per Month
Tester5	2	2.00
Tester6	21	1.62
Tester7	1	1.00
Tester8	18	0.86
Tester9	3	0.75
Tester10	0	0.00

Note: *Tester Productivity Summary (ranked by average per month)*

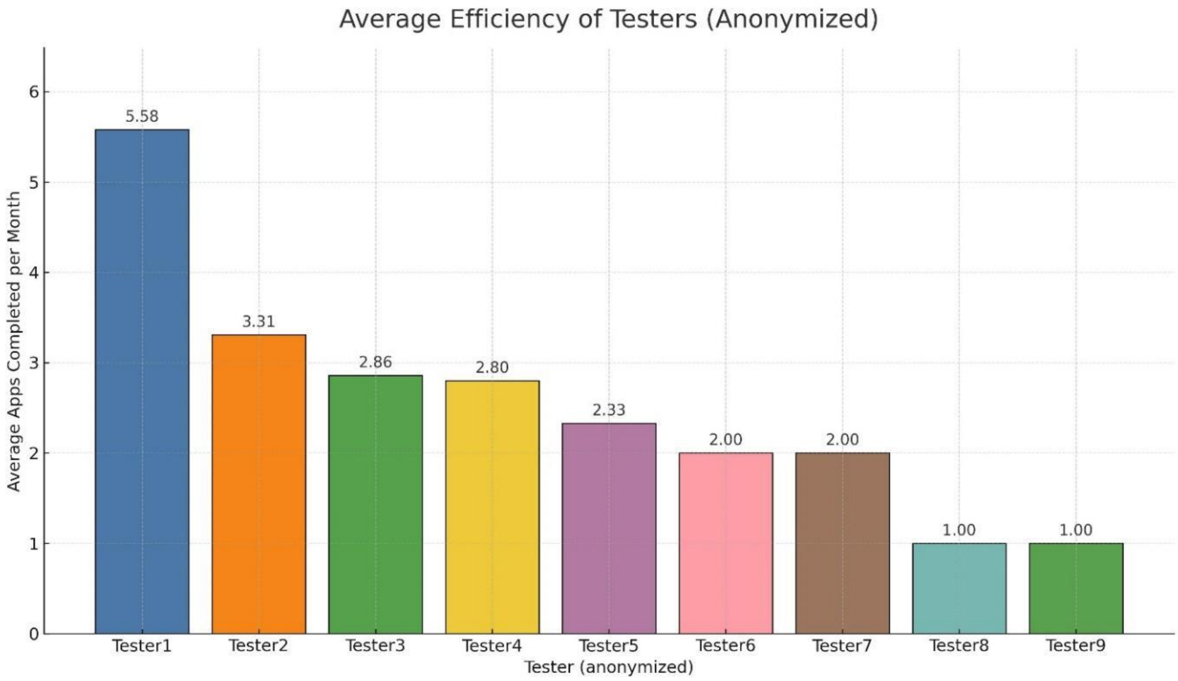

Figure 6Average number of applications tested to completion per month by each tester

Note: *Average number of applications tested to completion per month by each tester. Tester1 shows the highest throughput (5.5 apps/month), while testers towards the right have lower averages. Note that some testers joined later or had limited assignments, affecting their averages.*

Results and Discussion: Tester's efficiency levels are very different from one another. Tester1 is very different from the others. This tester did roughly 5.5 application tests every month, which is a lot more than the others. Tester1 did 203 application validations between 2022 and 2025, which was by far the most work (almost 35% of all tests that were finished!). This could mean that Tester1 is a lead or senior tester, or that they just did a lot of the tests. Their high monthly average shows that they are consistently productive. For example, 5.5 tests per month is about one app test every 5–6 days, which is rather fast considering that we found a cycle time of about 27 days earlier (maybe many tests overlap or are shorter for simpler apps).

After that, we have Tester2, Tester3, and Tester4, who each test an average of 2.3 to 2.9

apps per month. In total, these add out to 53, 44, and 81 apps, respectively. Tester4 did a total of 81 apps, which is higher than Tester2 and 3, but Tester4's average (2.38) is a little lower. This is because Tester4 was active for a longer time. In fact, looking at the raw data, Tester4 probably started sooner. For example, Tester4 may have joined in early 2022 and worked through 2024, distributing 81 tests over around 34 months, or about 2.4 tests per month. Tester2 and Tester3, on the other hand, may have worked less months, which raises their monthly average a little bit.

Tester5 has an average of 2.00 applications per month in the middle range, although this is based on a very small sample (just 2 finished; it looks that Tester5 only took part in one month, completing 2 apps that month and not being involved in previous months). This gives a high average for the short time active, but we need be careful how we read it. It doesn't indicate Tester5 could keep up with 2 each month for years; it only means they did 2 in one month. If a tester has only been working for a short time, their average can be deceiving in this case.

Tester6 averages around 1.62 tasks per month (21 total), Tester8 about 0.86 tasks per month (18 total), Tester9 about 0.75 tasks per month (3 total), and Tester7 and Tester10 effectively 1.0 and 0.0 tasks per month because they had very few jobs (Tester7 performed 1 app in one month, and Tester10 did not finish any in the time period). These lower figures could mean that people are only working part-time or in different roles. Tester10 could be a new employee who hasn't taken a test yet by August 2025 (they might only have open tasks or have just started training). It's possible that testers 7, 8, and 9 had other things to do or left early, since their totals are so low.

The difference is big: Tester1 gets about 5.5 a month, whereas some get 1 or less. It makes me wonder about how to balance workloads. It looks like Tester1 had to do a lot of the testing work. If this was done on purpose (maybe they are really efficient or the go-to person for a lot of apps), it's impressive, but it might also be a problem if that person becomes a bottleneck or leaves. One may say in a thesis that relying too much on one person would not be good for long-term strength; knowledge transfer or team growth might be ways to distribute knowledge.

On the other side, Tester1 may have a lot of simpler validation cases, while others may have fewer but more complicated cases (which take longer and hence fewer each month). Raw figures don't reflect the whole story of test complexity. But because we treated all applications the same here, the statistic gives a general idea of productivity.

From a management point of view, this KPI could point up training needs. For example, could the testers who aren't as good at their jobs be trained to do more? Or maybe they didn't get used enough (maybe there weren't enough applications to go around, thus Tester1 took most of them)? It could also be because certain testers specialize in certain difficult areas and hence do fewer assignments.

If one tester is doing 35% of the work, scaling up the program can mean giving everyone a fair share of the knowledge and tasks. On the plus side, Tester1's excellent efficiency probably helped keep throughput high and lower average cycle times. It's important to recognize their work.

This KPI shows that testers are not working at the same level of productivity overall. The average number of apps completed by each tester every month is about 1.6 (426 completions divided by the total number of active months for all testers, however one individual greatly affects the number). But to be more specific, one tester does about 5.5, a few do about 2–2.5, and the others do about 1 or less. In a thesis, one might suggest examining the factors that contribute to Tester1's high productivity (such as expertise, superior tools, or more straightforward assigned applications) and whether these techniques can be disseminated to others. Also, if Tester1 had too much work to do, it would be smart to make sure that other people could take on extra tests.

It's also important to note that even with these differences, the team as a whole was able to finish 426 tests in around 44 months, which is about 9.7 tests per month on average. The team got better over time since the cycle time and defect rate got better. This may have allowed for more testing at the same time or faster turnarounds, which is shown by the increased outputs. The individual efficiency statistic makes the data more human by indicating how much each worker contributed on average.

The deep-dive into these five KPIs provides a comprehensive view of Company X’s testing program performance:

- **Testing Cycle Time** has improved markedly, dropping to ~24 days on average in 2025 (down from ~29 days), indicating faster testing cycles and efficiency gains. Shorter cycle times align with industry goals of speeding up releases and continuous improvement (BusinessPlan-Templates, (2023)).

- **Defect Detection Rate** (Inactive applications) has plummeted from ~37% in 2022 to single digits by 2024–2025. This suggests that far fewer applications are failing validation now – a sign of higher input quality and effective testing. The program is catching most issues early, and partners are delivering more compliant apps, resulting in a **higher success rate** of tests over time.

- **Test Coverage (Devices per App)** shows an average of 4 devices tested per app, with many apps on 1–3 devices and a few on a broad range. This indicates a decent but not exhaustive coverage. It balances resource constraints with coverage needs. Importantly, the program does test on multiple devices for critical apps, which is key to ensure compatibility across hardware (Testsigma, (2023)). There may be room to increase coverage for some apps if needed, but overall this metric demonstrates the program’s scope in terms of device diversity.

- **Partner Satisfaction** is **exceptionally high (93%)**, reflecting that partners value the program and are pleased with the service. The Validation program in particular enjoys 94% satisfaction. Such a high satisfaction rate is likely to translate into strong partner loyalty and program reputation (BusinessPlan-Templates, (2023)). This is a great validation of the program’s effectiveness from the user’s perspective.

- **Tester Efficiency** varies, with one tester carrying a very high load (5.5 apps/month) and others less so. While the team collectively is productive, there is an opportunity to distribute work more evenly or learn from the top performer. Ensuring all testers are utilized and trained to improve their throughput could increase overall capacity. Despite the variation, the team's output has supported the improvements seen in other KPIs.

Overall, Company X's testing program has shown progress and success in many areas, including shorter cycle times, higher pass rates, happy partners, and a strong (if slightly uneven) team. The data shows the tale of a program that is growing up. At first, many apps didn't work and testing took longer. Later, the procedure got easier, partners came better prepared, and the outcomes got a lot better. These KPIs give real proof that may be utilized in the thesis to show how the program has made a difference.

The thesis gives a whole picture of performance by looking at various KPIs. It also gives a fact-based basis for any future strategic decisions or changes to Company X's Global Enablement Center activities. The overall prognosis is good. There is obvious proof that the testing services are getting better and better, which is probably helping Company X's reputation and the success of its partner ecosystem.

5. Results and Discussion

5.1. Applying BPMN to Model Current Software Validation Processes (AS-IS)

The first step in improving software validation processes using BPM is to fully understand and write down the present state, which is sometimes called the "AS-IS" model. This step, called process discovery, is particularly important for finding problems and inefficiencies that already exist.
The first step in process discovery is to systematically gather, model, and validate data. This entails collecting preliminary insights from multiple sources, augmented by comprehensive interviews with a range of stakeholders, including process analysts, domain experts, and end-users, to encompass numerous viewpoints and scenarios. Domain specialists are very important since they know how things work in the real world, even if they don't know how to model.
The AS-IS model, which is usually shown using BPMN, is the basic structure. It accurately shows how things work in the actual world by showing the process boundaries, individual activities, control flow, and other things like business objects and exceptions. This methodical methodology guarantees that the resulting process model is both semantically valid and useful for end-users, with ongoing cycles of development based on new information and feedback.

Once the AS-IS model is established, a rigorous analysis is performed to identify inefficiencies. This often involves qualitative analysis methodologies such as Value-Added Analysis and Waste Analysis. In Value-Added Analysis, process steps are classified into three categories:

- **Value-Adding (VA):** Activities that directly contribute to the customer's perceived value.
- **Business Value-Adding (BVA):** Activities necessary for the business but not directly adding value for the customer (e.g., regulatory compliance).
- **Non-Value-Adding (NVA):** Activities that consume resources but do not add value for the customer or the business, representing waste (e.g., waiting times, rework).

This classification shows where improvements can be made. At the same time, root-cause analysis (RCA) methods, like the 5Ws framework (Who, What, Where, When, and Why), are used to find the real reasons for inefficiencies and bottlenecks that have been found. Performance profiling tools can also be used to keep an eye on system metrics and find performance problems during RCA.

.

5.2. Current validation process BPMN diagram:

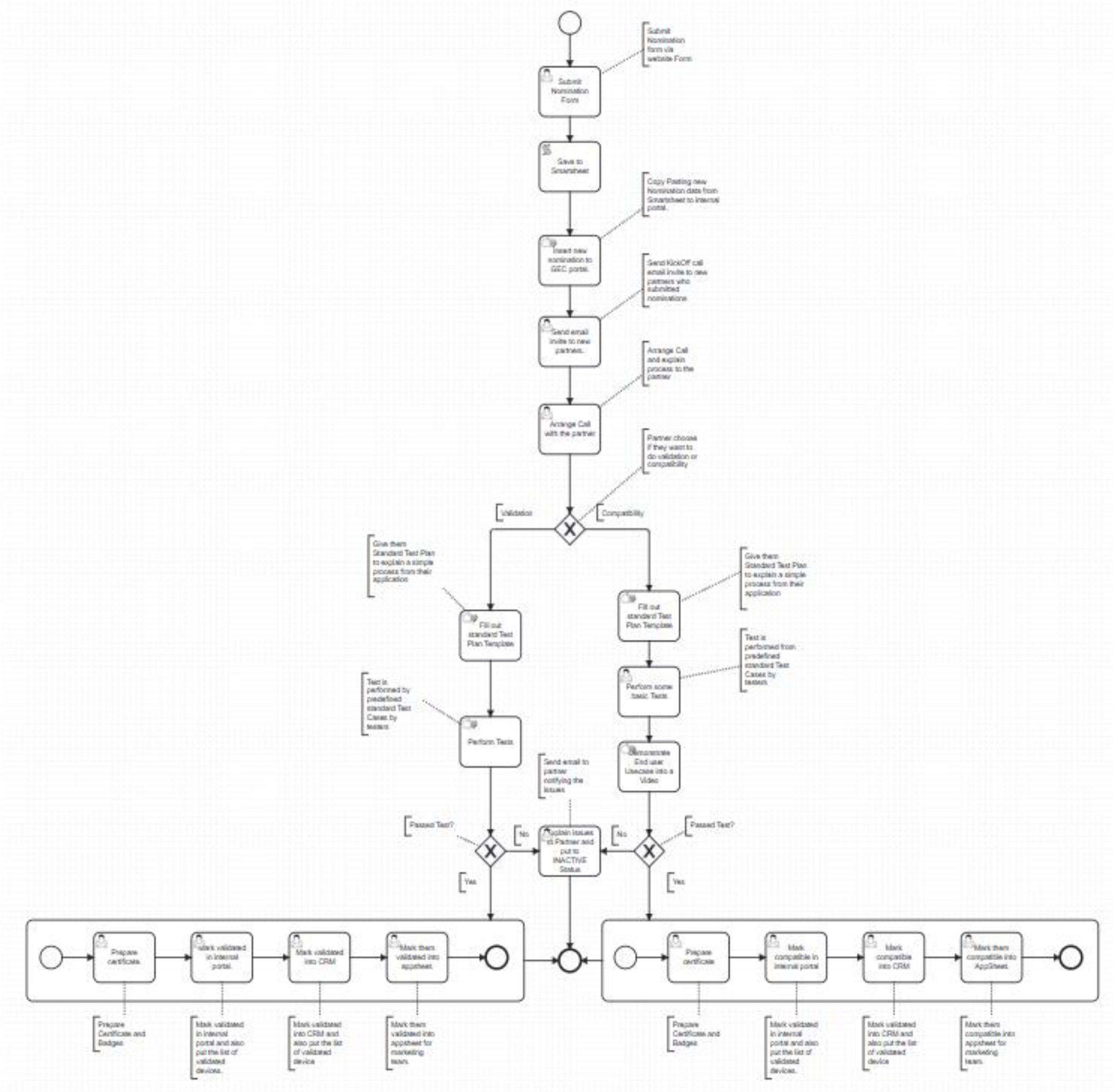

Figure 7Current validation process in BPMN notation

5.3. Designing Enhanced Software Validation Processes with BPMN (TO-BE)

After a thorough study of the AS-IS process, the next important step is to create the "TO-BE" model, which shows the best possible future state of the software validation process.

The goal of this redesign is to get rid of known obstacles, replace ineffective processes, and see how these changes have a good effect. The TO-BE model is a guide for making things better, with a focus on making things more efficient, cost-effective, high-quality, and flexible. Organizations can easily see what changes need to be made and how much better things will be after optimization by comparing the TO-BE model with the AS-IS process.

5.4. Proposed Generative AI-enhanced process for validation:

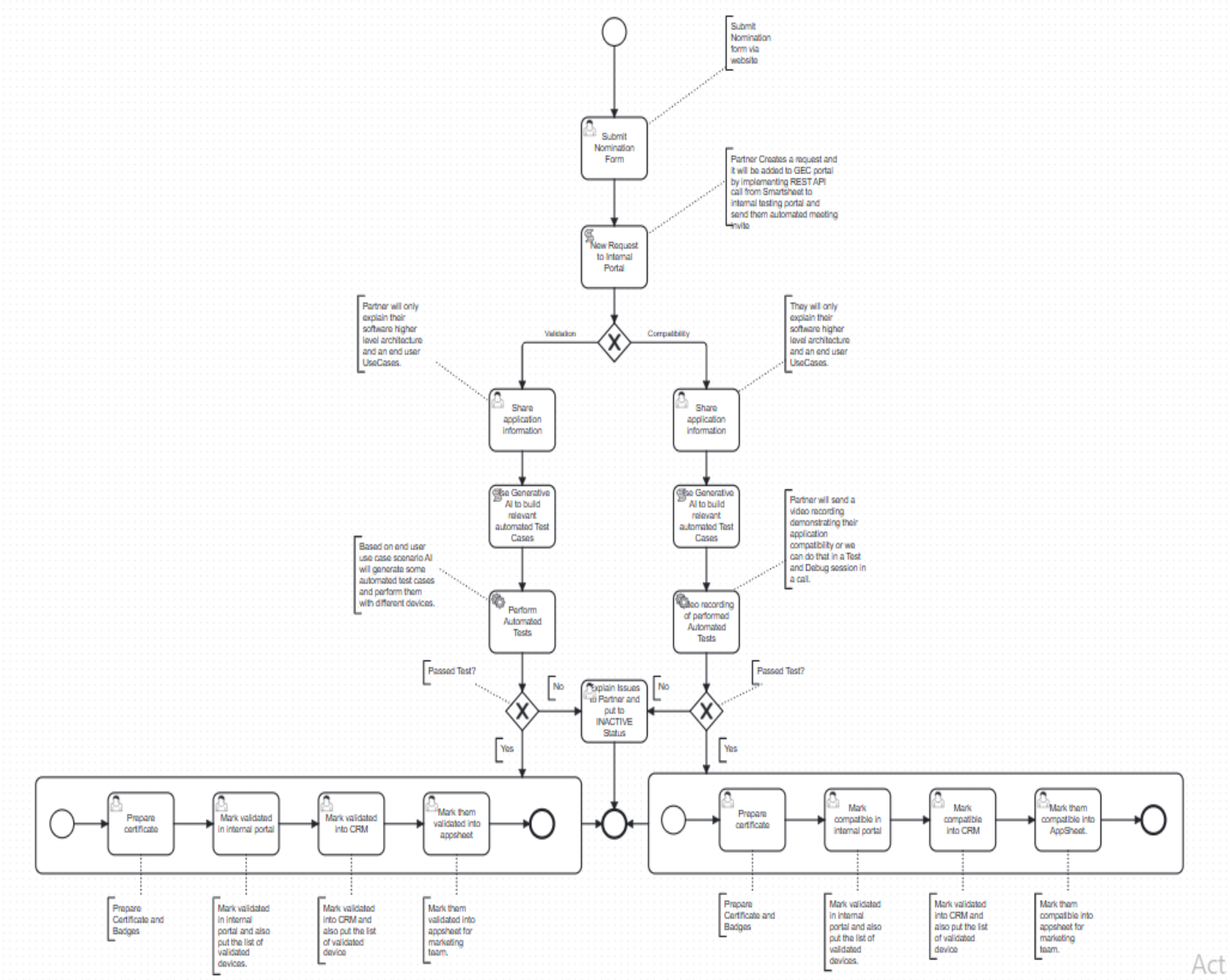

Figure 8New validation process in BPMN notation.

5.5. Benefits of BPM in Software Validation

Business Process Modelling can help with software validation in many ways. For example, it can help move the focus from mistakes at the end of the development process to proactive and better work habits.

First, BPM makes steps in a workflow apparent through clear visual representations, which helps speed up operations and get rid of bottlenecks. This increases efficiency and productivity. Organizations keep control by mapping out and understanding how processes work. This leads to smoother procedures and faster onboarding of new employees.

Second, BPM helps everyone involved understand each other better and talk to one other more easily. Because BPMN has a consistent way of showing things, business and IT stakeholders can grasp difficult process semantics. The language also connects the design and run-time parts of the IT system. This shared understanding is vital for working together to make high-quality work.

Third, BPM helps define best practices and make sure that everyone does things the same way. What should be a fair and predictable process turns out to be anything but... But when everyone can see and understand the processes, at least in theory, this starts to create a basic level for a "standard" performance benchmark. We know what the norm is based on what we've written down, and we know what to look for when we're dealing with a decision-making tree. It makes the execution more predictable and the consequences less varied.

Fourth, BPM cuts down on mistakes and duplicate work. If we don't have organized, managed processes, we're asking for trouble. BPM cuts down on mistakes and stops them from happening again by having a set process and clear methods and checks and balances. This saves time and money.

Fifth, BPM makes it much easier to follow the rules and keep track of everything. In regulated fields like pharmaceuticals, processes must follow these rules and laws. BPM advocates writing down each stage of a process, like validation rules and findings, to make sure that the data is clearly linked for input, processing, and output. This is very important for being ready for an audit and providing evidence.

Last but not least, BPM is an important part of continuing improvement and keeping track of KPIs in real time. Process maps let businesses define and track Key Performance Indicators (KPIs) and rapidly identify how process improvements affect the bottom line. The company is headed in the correct path because of this regular monitoring, and it is also making more improvements. It helps with figuring out what's going on now, coming up with new ways to do things, and setting up new organizational structures to keep the validation ecosystem flexible and strong.

5.6. Benefits of Utilizing Generative AI for Test Case Generation with the New Proposed Framework:

5.6.1. Efficiency Boosts from Automation

Generating the test cases by utilizing Generative AI can greatly increase the speed at which the validation process can be completed. It is up to October 2023, so automation reduces manual effort on the part of teams, while also enabling a faster heartbeat to the generated documentation. If the routine struggles are provided by AI technology, validation staff can best use their knowledge to dive into the more complex problems that require a human brain and creativity to solve. That means instead of being bogged down with tedious data entry or repetitive drafting and formatting of test cases, staff can be utilized for higher-level strategic planning, exploratory testing, and in-depth troubleshooting.

Moreover, automated test case generation also implies a faster test adjustability to product specification alterations. The AI can then be updated to reflect new features and configurations as requirements change so that documentation is consistently as up-to-date as possible. Such quick follow-through aids in creating a snowball effect of efficiency throughout the validation pipeline, leading to improved, seamless, and faster launches of new software iterations.

5.6.2. Generation of Test Cases with Enhanced Accuracy and Consistency

Gen AI can optimize the process by generating standardized test cases tailored to specific scenarios and devices, while simultaneously enhancing the reliability and consistency of the procedure. Machine learning and natural language processing can be utilized to train the AI to comprehend the subtleties in application requirements, drawing upon an extensive corpus of historical data, existing case libraries, and predefined user acceptance criteria.

This enables the automation of test case generation that closely simulates a genuine user experience and ultimately aligns more accurately with real-world usage patterns. A pertinent output helps minimize unnecessary effort on extraneous test cases, thereby optimizing the use of validation resources for the most significant and frequently occurring scenarios.

5.6.3. The Scalability of Testing Efforts

Integrating Generative AI into the test case development process can facilitate substantial scalability across all applications and partner collaborations. This is particularly significant because Company X is consistently innovating and broadening its range of products, making the capacity to scale testing even more essential. Scalability refers to the ability of the testing framework to expand in response to increasing application proliferation and the growth of partner integrations, ensuring that quality is maintained despite heightened complexity.

When properly guided, generative AI can be extensively customized through the utilization of such data to adapt swiftly to evolving product configurations and partner requirements, while remaining both informative and action-oriented. This indicates that Company X and its partners are able to leverage and adapt to emergent opportunities more effectively while decreasing the time required to bring new products to market.

5.6.4. Unlocking Insights from Data to Drive Operational Success

Through the use of Generative AI, standardized collection and analysis of test data can yield actionable insights that enhance operational efficiencies. A fundamental benchmark for the ongoing enhancement of testing practices is to identify patterns of vulnerabilities or inefficiencies (based on values) in the pursuit of a high-quality product. At Company X, we are committed to leveraging data to inform our decisions and promote a culture of continuous improvement. For example, if specific failure patterns emerge during testing, they may indicate systemic issues within particular subcomponents or resource allocations that warrant further investigation, the company stated in a blog post. This knowledge is practical, allowing organizations to adapt and optimize strategies proactively, thereby attaining more favorable outcomes and fostering greater trust among supporters in their brands.

5.6.5. Improved Collaboration and Communication with Partners

Generative AI that streamlined the testing process and enhanced efficiency would significantly facilitate collaborations with Company X partners. This benefit of collaboration encompasses the provision of timely, contextually relevant test cases generated via AI and improved clarity regarding their intended application.

When the organization reduces friction, the validation process becomes more seamless, enabling effective, transparent collaboration aligned with shared objectives. Impact Detection delivers tailored, actionable outputs aligned with each partner's requirements, fostering increased trust, while the collaborative environment—characterized by open communication and shared progress—can establish a strong foundation.

5.6.6. Workflow and Communication Automation Tools

The responsiveness of Company X to partner inquiries and interactions will be significantly enhanced through the integration of automated tools into the communication ecosystem. On-demand notifications concerning the creation or approval of AI test cases can be transmitted through these devices, enabling efficient follow-up communication.

5.6.6.1. Automation Opportunities

The algorithm manages initial inquiries or can alternatively utilize automated workflow tools to triage incoming requests from partners. This degree of responsiveness can significantly enhance partner relationships, increase satisfaction, and shorten validation times, enabling agile and efficient interactions that maintain project momentum.

5.6.6.2. Developing an Agile Review Cycle

Consult with validation staff and partner representatives to establish an iterative review procedure for AI-generated test cases. Constructive feedback on outputs is essential in

establishing an effective feedback cycle, enabling organizations to collect this input and providing a platform for necessary adjustments.
This iterative process may encompass periodic review meetings wherein stakeholders evaluate the results generated by the AI tools, exchange insights from partner perspectives on enhancing quality, and deliberate on necessary adjustments to facilitate ongoing learning and adaptability. This open dialogue will foster a more robust culture of collaboration and ongoing enhancement to deliver improved solutions.

5.7. Structural Comparison of Old vs. New Validation Process (BPMN) and KPI Impact Analysis

5.7.1. Overview of Process Changes in the New Validation Workflow

The new validation process makes big improvements to the way things are done, mostly by automating and simplifying a lot of procedures. The future-state BPMN diagram shows that certain manual jobs from the previous process have been taken out or replaced with automated chores. It also shows that a few additional stages have been introduced to make things more efficient. Some of the most important differences are:

- **Partner Onboarding and Communication:** In the old process, testers manually sent email invitations and scheduled kickoff calls with new partners to explain the validation process. These steps have been removed in the new process. Instead, partner onboarding is handled more asynchronously. The partner submits a Nomination Form via a web portal (as before), and the system provides guidance and resources automatically. The elimination of the email and kickoff call means less waiting for scheduling and fewer manual coordination tasks for testers.

- **Nomination Data Entry:** Previously, after a nomination was received, a tester would manually insert the new nomination data into the internal portal (copy-pasting from an external sheet as noted in the old BPMN). In the new process this step is automated. An internal "New Request to Internal Portal" service task now automatically creates the entry in the internal system. This removes a redundant manual data entry task.

- **Test Planning:** For every app, testers (and sometimes partners) had to "Fill out a standard Test Plan Template" in the old mode of work. This was a manual job on both the validation and compatibility tracks. This planning step is now much easier and automated in the new approach. A new script job called "Use Generative AI to build relevant automated Test Cases" makes a set of test cases based on the knowledge about the application. This replaces the process of generating a test plan by hand in both cases. In short, AI-driven test case generation makes it unnecessary for testers to write test plans from the start.

- **Test Execution:** The execution of tests has shifted from manual to automated. In the old process:

A. **Validation path:** Testers would perform comprehensive tests manually (step labeled “Perform Tests” in the old BPMN, with predefined test cases executed by testers).
B. **Compatibility path:** Testers perform basic tests (user task) and then demonstrate an end-user use case in a video (manual task) to verify the app’s functionality on a basic level.

In the new process, both paths use automated testing. A “Perform Automated Tests” service task executes the test cases generated by AI on the target devices. This replaces the manual testing by human testers. For the compatibility route, a “video recording of performed Automated Tests” service task automatically records the test execution, replacing the old need for a manually prepared demo video.

In summary, automated test execution and recording have replaced the labor-intensive manual testing steps, allowing tests to run faster and in parallel across devices.

- **Outcome Processing:** The later steps of issuing certificates and updating records (marking the app as validated/compatible in internal portal, CRM, AppSheet, etc.) remain largely the same. These are user tasks in both old and new diagrams (e.g. “Prepare certificate,” “Mark validated in internal portal/CRM/AppSheet”). These tasks were not automated in the new BPMN, so the process for final documentation is unchanged.

Below is a comparison table highlighting how specific tasks changed from the old process to the new process, and the nature of each change. The table also summarizes the expected impact of each change on key performance indicators (KPIs):

5.7.2. Comparison of Old vs. New Process Tasks and KPI Impact

Table 7Visual comparison of Task VS KPIs into Old and New process

Process Step	**Old Process (2022–2025)**	**New Process (Future)**	**Change Type**	**Cycle Time** (Speed)	**Defect Detection Rate** (Fail %)	**Test Coverage** (Devices /App)	**Partner Satisfaction**	**Tester Efficiency**
Partner Onboarding	Email invite to partner, schedule kickoff call to explain	No manual invite or call	**Removed** (streamlined communication)	**Improves**: Eliminating scheduling delays can save a few days of	**No direct change** :(Process entry steps don’t affect	**No change:** (Does not affect number of devices tested.)	**Neutral to Slight ↑:** Partners have less hassle (no	**Improves**: Test team spends less time on admin/ meetings and

Process Step	**Old Process (2022–2025)**	**New Process (Future)**	**Change Type**	**Cycle Time** (Speed)	**Defect Detection Rate** (Fail %)	**Test Coverage** (Devices/App)	**Partner Satisfaction**	**Tester Efficiency**
	process.	partners get instructions via portal; possibly self-service documentation.		waiting, speeding up initiation of testing.	app defect rates.)		scheduling), though some may miss personal touch. Overall convenience likely keeps satisfaction high.	more on actual testing tasks.
Nomination Data Entry	Tester manually enters nomination into internal system (GEC portal).	Automated "New Request to Internal Portal" creates the internal record.	**Automated** (manual step removed)	**Improves**: Automation saves processing time (hours to a day) and avoids potential re-entry delays.	**No direct change**: Data entry automation doesn't affect validation outcome.	**No change**: (Administrative step, no impact on device coverage.)	**Slight ↑**: Faster acknowledgment of submission can reassure partners. (Minor positive effect.)	**Improves**: Reduces tester workload (no copy-paste); frees time for value-added testing work.
Test Planning	Tester (with partner input) manually fills out Test Plan template	**AI-generated test cases**: automated script creates releva	**Automated** (manual planning removed)	**Improves**: Cuts out days of planning. AI produces test cases instantly, accelerati	**Improves**: More thorough and standardized test cases catch	**Improves**: AI can include a broader range of scenarios; more aspects	↑ : Less burden on partner to develop test plans; faster testing	**Improves**: Testers no longer spend time writing plans; they can

Process Step	**Old Process (2022–2025)**	**New Process (Future)**	**Change Type**	**Cycle Time** (Speed)	**Defect Detection Rate** (Fail %)	**Test Coverage** (Devices /App)	**Partner Satisfaction**	**Tester Efficiency**
	(standard test cases).	nt test cases for the app.		ng the start of testing.	requirements issues upfront, reducing the chance of undetected defects.	of the app are tested, potentially increasing device/feature coverage.	and proactive guidance likely keep partners happy (they see a more efficient, modern process).	oversee multiple tests. More comprehensive initial test coverage means fewer retests, boosting productivity.
Test Execution (Validation path)	Tester executes full validation test cases on each device manually (step "Perform Tests").	**Automated test execution** on devices (step "Perform Automated Tests").	**Automated** (manual execution removed)	**Greatly Improves**: Automated tests run much faster and possibly in parallel on multiple devices, significantly reducing test cycle time.	**Improves**: Consistent automated execution can detect defects reliably. Any non-compliance will be identified quickly; if issues are found, they	**Improves**: Enables testing on more devices within the same timeframe (automation can run on many devices, whereas humans had time for fewer). Average devices-per-app can increase.	↑: Faster results and more rigorous testing can improve partner confidence. Quick turnaround (weeks faster) pleases partners. (No waiting for	**Greatly Improves**: A single tester can supervise automated runs on multiple devices or multiple apps at once, increasing throug

Process Step	**Old Process (2022–2025)**	**New Process (Future)**	**Change Type**	**Cycle Time** (Speed)	**Defect Detection Rate** (Fail %)	**Test Coverage** (Devices /App)	**Partner Satisfaction**	**Tester Efficiency**
					trigger a fail as before (partners must fix and resubmit, as per process). Overall, fewer missed defects.		lengthy manual testing.)	hput. Less manual labor per test means testers handle more projects concurrently.
Test Execution (Compatibility path)	Tester performs basic checks (user task) plus partner/ tester creates a demo video of end-user use case (manual task).	**Automated tests** run basic compatibility checks; automatic recording of test execution (video) for review.	**Automated** (manual testing & video replaced)	**Improves**: Even the simpler compatibility tests finish faster with automation; no need to wait for a partner to produce a video.	**Improves**: Automated checks ensure even “basic” tests are consistently executed. Important issues are less likely to be overlooked.	**Improves**: Potential to run compatibility checks on more device models quickly, slightly broadening coverage for “compatibility only” tests.	↑: Process is quicker and requires less effort from the partner (no video to make), leading to a smoother experience.	**Improves**: Testers save time not having to coordinate or create demo videos. They can manage more compatibility tests in parallel.

Process Step	**Old Process (2022–2025)**	**New Process (Future)**	**Change Type**	**Cycle Time** (Speed)	**Defect Detection Rate** (Fail %)	**Test Coverage** (Devices/App)	**Partner Satisfaction**	**Tester Efficiency**
Result Processing	If app **fails**: Tester explains issues and marks app *Inactive* (not passed). If **passes**: Tester prepares certificate and manually updates internal portal, CRM, AppSheet with "validated" status (or "compatible" for compatibility tests).	**(Same as old)** Tester performs failure debrief or issues certificate and updates systems (these steps remain manual user tasks in the new BPMN).	*Unchanged* (manual outcome steps)	**No change**: Final documentation steps still take some time, but these are minor compared to overall cycle. (Potential future automation opportunity, but not implemented yet.)	**No direct change**: The criteria for pass/fail are unchanged; however, because earlier steps improved quality, fewer tests reach this failure step (see KPI impact below).	**No change**: Documentation doesn't affect device coverage.	**No change**: High satisfaction likely continues; delivering certificates and feedback is handled as before (partners already rated this highly).	**No significant change:** Testers still spend time on documentation, but since fewer apps may fail and require lengthy explanations, there's a slight implicit efficiency gain.

Note: *Arrows/notations in "Impact" are relative to the old process. "Improves" (or ↑) means a positive effect (e.g. faster cycle, higher KPI value if higher is better, or lower defect rate since lower is better), while "No change" means the task change has negligible effect on that KPI.*

5.8. KPI Impact Analysis on Newer Model

In the thesis, anticipated KPI upgrades are formulated by integrating an analysis of recent historical performance trends with the measured effects of proposed process changes from the TO-BE models. In practice, multi-year baseline data (e.g. 2022–2025 KPIs) is calculated to produce a "current course" trajectory, which is then updated depending on improvements expected from the new validation process. The TO-BE BPMN models show a number of improvements, such as more automation of validation activities, the use of AI-driven testing tools, and easier communication between teams. All of these are thought to improve quality and efficiency beyond the baseline trend. This two-pronged approach (using both past trends and modeled changes) is in keeping with best practices for performance forecasting, where historical data gives a realistic baseline and process modeling is utilized to figure out how much new ideas would add (Renna, 2025). For instance, adding workflow automation and better communication protocols to a case study simulation cut the time it took to process from start to finish by about 35% and increased resource use by 22% compared to the original process (Renna, 2025). Also, using AI-powered validation tools has been found to cut down on human work and speed up feedback loops in software testing, which leads to faster and more reliable verification cycles (Baqar, 2025). So, the expected KPI gains in the thesis are based on both the upward trend of past performance improvements and the extra effect of future process optimizations found in the BPMN-based redesign.

This section assesses five principal key performance indicators (KPIs) by comparing the 2025 baseline with projections derived from the enhanced AI-driven validation process. The baseline represents a predominantly manual testing process, which generally exhibits delayed performance and reduced thoroughness (Orso & Rothermel, 2014). The revised BPMN-based process incorporates automation and generative AI at key stages, with the objective of enhancing speed, quality, and scalability. Table 8 presents a comprehensive overview of the baseline values and projected outcomes for each KPI under the new model.

Table 8Summary of Baseline (2025) vs. Projected KPI Outcomes under the AI-Driven Model

KPI	2025 Baseline	Projected (New Model)
Average Testing Cycle Time	23.6 days per cycle	19–20 days (15–20% faster)
Defect Detection Rate	8% of apps are failing validation	4–6% of apps fail (40% reduction)
Average Test Coverage	4 devices per application	5 devices per app (20–25% increase)
Partner Satisfaction Score	93% satisfied partners	95–96% (maintain or slight increase)
Tester Efficiency	1 app (avg) to 5.5 apps (max)	3–4 apps (avg) (30–60% increase)

5.8.1. Average Testing Cycle Time

Baseline (2025): 23.6 days
Expected Improvement: 15–20% reduction (3–5 days faster, 19–20 days total)
Drivers: Removal of manual coordination, AI test planning, automated parallel test execution
Benefit: Faster partner results, critical for go-to-market speed (Celonis, 2025).

In 2025, the average duration of the end-to-end testing cycle was approximately 23.6 days. With the implementation of the new AI-driven validation model, it is anticipated that this cycle time will be reduced by approximately 15–20%, resulting in an average of approximately 19–20 days per cycle. The decrease is anticipated due to the new process eliminating manual coordination stages and employing AI for test planning and execution. Key drivers encompass automated scheduling and concurrent test execution, enabling multiple tests to execute simultaneously rather than sequentially. Automating these activities markedly accelerates the testing workflow (Whyte & Mulder, 2011; Kathiresan, 2024). Notably, increased levels of test automation have been empirically associated with shortened release cycles without a corresponding rise in testing effort (Wang et al., 2022). A more rapid validation cycle enables partners to obtain results more promptly, which is essential for accelerated go-to-market strategies and grants a competitive edge in deploying upgrades (Kathiresan, 2024). In summation, enhancements in cycle time under the new model are achieved through the elimination of bottlenecks and the utilization of AI to optimize test preparation and execution.

5.8.2. Defect Detection Rate (Failure Rate)

Baseline (2025): 8% inactive (failed) apps
Expected Improvement: 30–50% relative reduction (down to 4–6%)
Drivers: More thorough AI-generated test cases, consistent automated execution, better partner preparation
Benefit: Higher pass rates, fewer abandoned apps, improved quality assurance (ISTQB, 2025; Testsigma, 2025; Ministry of Testing, 2025).

At the start, about 8% of partner applications that were tested failed validation. This was either because they had serious flaws or they didn't meet the standards. With the new AI-driven procedure, the failure rate is predicted to drop to between 4–6% of applications. This is about a 40% relative drop in problems found after submission. This improvement is due to more thorough and organized testing in the new workflow. The AI approach creates a wider range of test cases, including edge cases that manual testing could miss. This increases the chances of finding bugs before the app is approved. By running these tests automatically, we can be sure that they will work the same way every time, which reduces the chance of human error. More thorough and extensive testing is expected to find more bugs sooner, which will raise the overall pass rate of partner apps (Mockus et al., 2009; Kathiresan, 2024). In addition, the new approach helps partners get ready better (for example, by giving clearer instructions and pre-assessment checks), which leads to better submissions from the start. These features together make it easier to find defects during the validation step (Whyte & Mulder, 2011). In terms of cause and effect, more thorough AI-powered testing leads to fewer faults that get through, which means that much fewer partner applications will be turned down or need to be revised compared to 2025. This improves the outcomes of quality assurance and cuts down on the resources needed for applications that would have failed under the old procedure.

5.8.3. Average Test Coverage (Devices per App)

Baseline: 4 devices/app
Expected Improvement: 20–25% increase (5 devices per application)
Drivers: Automated parallel testing reduces tester effort per device, enabling broader device coverage without extending cycle time
Benefit: Higher confidence in compatibility, detection of device-specific defects (Perfecto, 2025; BrowserStack, 2025; testRigor, 2025).

In 2025, we tested each partner application on an average of four different types of devices. The new validation method should raise the average number of devices examined to roughly 5 per app, which is a 20–25% increase. We can improve this by running automated, parallel tests on all of Company X's devices. It used to be challenging to test more than a few devices in a reasonable length of time because it took a lot of labor by hand. This is a common problem with traditional testing systems, where coverage is sometimes limited (Orso & Rothermel, 2014). With the AI-driven method, testers may run tests on multiple devices at once without having to do any extra effort. Automation handles the repetitive setup and execution on each device, which enables us cover more land without adding time to the cycle. By testing on more types and configurations of devices, the team can identify compatibility issues that are specific to certain devices that they may have overlooked earlier. Empirical research indicates that improved test coverage correlates with a decrease in post-release issues, to some degree (Mockus et al., 2009). People are more sure that each piece of software will function with and be stable on these devices because there are so many of them. Automation makes it easier and cheaper to test each new technology (Kathiresan, 2024). Because of this, the new technique may automatically add more devices to each app. This makes validation more complete and makes sure that partner apps work on all the devices they were supposed to work on.

5.8.4. Partner Satisfaction Score

Baseline (2025): 93% satisfaction
Expected Change: Maintain or slight increase (95–96%)
Drivers: Faster turnaround, higher pass rates, reduced partner effort (no calls, no demo videos)
Consideration: Some partners may miss personal interaction, but overall satisfaction expected to remain very high (HelpScout, 2025).

Almost 93% of partners were delighted with the certification process in 2025. With the new model, the goal is to preserve this level of satisfaction at about 95–96% or possibly raise it a little. Some portions of the AI-driven process should make partners happier. First, partners will get validation findings back faster (because the cycles are shorter). This will make them worry less about waiting and help them get their products or improvements out to customers more quickly. Second, partners are less likely to have to go through re-submission cycles because the success rate is higher (fewer apps fail because flaws weren't detected). This gives them more faith in the process. Third, the new workflow is easier for partners to use. They don't have to go to long coordination meetings or prepare long demo movies for people to watch anymore because a lot of the testing is done automatically. Making these burdens smaller can improve the partner's experience and opinion of the validation service. In general, faster and better service makes stakeholders happier (Kekre et al., 1995). One thing to think about is that automating contacts could make things less personable. Some partners might miss the face-to-face communication and counsel that the old manner of doing things gave them. This is a possible hazard that a fully automated process would not feel personal to some consumers, which is a trade-off that has been noted in study on customer service. But overall, the new AI-enhanced validation process should make partners quite happy because it delivers them actual benefits like quicker approvals and more reliable outcomes.

5.8.5. Tester Efficiency (Throughput per Tester)

Baseline: Average 1 application per tester per month and top 5.5 applications per month.
Expected Improvement: 30–60% increase (3–4 apps/tester/month average)
Drivers: Automation reduces manual tasks, enables parallel test runs, testers focus on oversight and exceptions
Benefit: Higher throughput, more balanced workload, reduced bottlenecks, improved quality focus (Atlassian, 2025; Sauce Labs, 2025).

The amount of tests testers can run is about to go up a lot, which will make them more efficient. In 2025, a typical tester at Company X worked on roughly one application a month. The most productive testers worked on up to five and a half apps a month. The new model's automation should raise the average output to about 3–4 apps per tester every month, which is a 30–60% increase in productivity. Testers don't get this benefit by working longer hours; the get it by working smarter. Automation does a lot of the boring work that human testers have to do. Setting up test data, conducting tests on several devices, and putting together conclusions are all tasks that automated systems can do. This allows testers see more than one test at a time. Studies show that automating test suites and execution can greatly increase testing throughput without the requirement for more workers (Whyte & Mulder, 2011; Kathiresan, 2024). For us, each tester will be able to watch a lot of validations at once. Instead of running each test case by hand, testers will be in charge of higher-level monitoring, such keeping a watch on AI-generated test cases, looking at sophisticated failures, and making the process better. This method of working in parallel enables the group check a lot more applications in the same amount of time (Whyte & Mulder, 2011). Also, getting rid of uninteresting tasks might make testers happier and more productive because they can spend more time on thinking critically and less time on monotonous tasks. In the end, the workload is more balanced and there are no longer bottlenecks where only one tester could handle a few projects at a time. Recent studies on software teams show that more advanced test automation can speed up delivery without putting too much stress on people (Wang et al., 2022). The new AI-driven method is expected to make testers much more efficient. This will allow the validation team to

grow and handle more partner apps with the resources they already have.
These projections are based on the baseline data from 2022 to 2025 and the logical impacts of each change to the process. In short, the new BPMN process structure makes the software faster, better, and ready to grow. It builds on the advancements that have been made so far and takes performance to new heights. Company X can give its partners even more value by lowering cycle time and failure rates while raising coverage and productivity. This lets them validate apps faster and more reliably without losing quality. Each KPI's positive change makes the others stronger. For instance, advances in efficiency help cut down on cycle time, which can then make partners happier, and so on. So, the future-state approach promises a virtuous cycle of quality and efficiency that fits with the program's aims for continuous improvement and the best practices in the software validation industry. The end result is good for everyone. Partners have a smoother and faster validation process, and the testing team does better work with the help of clever automation.

5.9. Risks and Limitations of proposed approach

The expected increases in KPIs are good, but they also come with risks and assumptions that should make us think twice. The projections are based on the idea that the AI-driven process will be easy to put into place and that the AI tools will keep working well. When using generative AI to make test cases, there may be a learning curve and unexpected technological problems. For instance, if the AI model makes bad test cases or misses some types of bugs, the actual improvement in defect detection rates might not be as good as hoped. Data quality is important for AI to work well. If the AI's training data is inadequate or biased, the model's contributions might not be as important as expected. Also, organizational variables could affect the results. Reluctance to change or lack of training for testers on the new system could make efficiency benefits less likely at first. Some partners might not like a process that is quite automated. They might want more human interaction, which could lower the expected level of satisfaction. These estimates are based on logic and early observations, but the real world is more complicated, thus the real benefits could be different. After the rollout, it will be important to keep an eye on each KPI all the time. If the metrics don't become better as predicted, the team should be ready to change the process, like by adding human checks or making the AI models better. In short, the KPI increases listed above are a best-case scenario. Company X will have to carefully manage the rollout, deal with both technical and human issues, and be flexible to make sure that the goals are met. The business may make it more likely that the new AI-driven validation model will offer the performance gains it was meant to by being honest about its limitations and taking steps to reduce risks, such as keeping an eye on things and providing training.

6. Conclusion

Combining Business Process Management (BPM) and Generative AI marks the beginning of a new era in software development that could change the world. The proposed regenerative method based on BPM and Generative AI would make software validation better. This is very important for effective software development. This thesis shows that standard manual validation methods are prone to mistakes, take too long, don't cover enough tests, and don't work well when the number of tests grows. These common inefficiencies show how important it is to have an organized and automated method. BPM, especially its well-known graphical notation BPMN, sets a basic degree of abstraction and a common language that are needed to comprehend, analyze, and redesign complicated validation processes. It is important for finding systemic inefficiencies (AS-IS) and imagining future ideal states (TO-BE) that it can help different stakeholders talk to each other clearly and understand each other.

Unfortunately, there are a lot of problems with using Generative AI. Organizations need to take the lead on crucial concerns including data quality and bias, how easy it is to grasp AI models, the high computing needs, and teams' natural resistance to change. It is crucial to uphold ethical standards including data privacy, algorithmic fairness, and the necessity for robust human oversight in order to create AI that is socially responsible. Real Options Theory and other strategic frameworks can assist Q&A teams deal with the dangers of investing in AI by giving them strategies to employ it that are flexible and can change.

The combination of BPM techniques and Generative AI, along with Agile principles, promises to create a strong foundation for making verification and validation (V&V) procedures future-proof. Organizations can reach unprecedented levels of efficiency, accuracy, and scalability in their validation operations by using BPM for structured process optimization and Generative AI for intelligent automation. This all-encompassing strategy not only gets beyond the problems of each method on its own, but it also creates a culture that values ongoing learning and making decisions based on evidence. This leads to better products and more flexible organizations.

7. References

American Society for Quality. (2016) (IEEE Std 1012-2016). Software verification and validation. IEEE. https://ieeexplore.ieee.org/document/7786999

Bass, L., Weber, I., & Zhu, L. (2015). DevOps: A software architect's perspective. Addison-Wesley.

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). Fundamentals of business process management. Springer. https://doi.org/10.1007/978-3-662-56509-4

Erdogmus, H., & Morisio, M. (2006). Software process improvement and automation. *IEEE Software*, *23*(3), 54–61. https://doi.org/10.1109/MS.2006.66

Food and Drug Administration (FDA). (2002). General principles of software validation; final guidance for industry and FDA staff. https://www.fda.gov/media/73141/download

Kaner, C., Falk, J., & Nguyen, H. Q. (1999). Testing computer software (2nd ed.). Wiley.Myers, G. J. (2011). The art of software testing (3rd ed.). Wiley.

Zhang, H., & Harman, M. (2021). Search-based software testing: Past, present and future. IEEE Transactions on Software Engineering. Advance online publication. https://doi.org/10.1109/TSE.2021.3059963

Amershi, S., et al. (2019). Software engineering for machine learning: A case study. In Proceedings of the 41st International Conference on Software Engineering (pp. 291–300). IEEE. https://doi.org/10.1109/ICSE.2019.00039

Dumas, M., La Rosa, M., Mendling, J., & Reijers, H. A. (2018). Fundamentals of business process management (2nd ed.). Springer.

Van der Aalst, W. M. P. (2013). Business process management: A comprehensive survey. ISRN Software Engineering, 2013, Article 507984. https://doi.org/10.1155/2013/507984

Zhang, J. M., Harman, M., Ma, L., & Liu, Y. (2022). Machine learning testing: Survey, landscapes and horizons. IEEE Transactions on Software Engineering, 48(1), 1–36. https://doi.org/10.1109/TSE.2019.2962027

BusinessPlan-Templates (2023). Which 5 Metrics Are Most Important for Software Testing Businesses? Available at: https://businessplan-templates.com/blogs/metrics/software-testing

Armagh POS Solutions, 2025. Toshiba TCx Receipt Printer. Armagh Cash Register and POS Systems. Available at: https://armaghpos.com/product/pos-hardware/toshiba-tcx-receipt-printer.

Brother Industries, 2025. Corporate Profile 2025. Available at: https://download.brother.com/pub/com/en/corporate/pdf/2025/brother-profile-en.pdf.

Brother Workhorse Series, 2025. Document Capture Solutions to Transform Your Business. Brother USA. Available at: https://www.brother-usa.com/-/media/Solutions/Newsroom/Files/Brother-Workhorse-Scanner-Series-Brochure.pdf.

Datalogic, 2024. Datalogic at NRF 2024: Shaping the Future of Retail with Intelligence and Innovation. Available at: https://www.datalogic.com/eng/media-center/press-releases/datalogic-at-nrf-2024-shaping-the-future-of-retail-with-intelligence-and-innovation-nd-8402.html.

Datalogic, 2025. PowerScan 9600 RFID Series – Revolutionizing Industrial Scanning. Available at: https://www.datalogic.com/eng/media-center/press-releases/powerscan-9600-rfid-series-—-revolutionizing-industrial-scanning-nd-8594.html.

Epson, 2024. Epson to Bring the Retail Store to Life with Solutions Designed to Optimize and Modernize Businesses at NRF 2024. News release. Available at: https://news.epson.com/news/pos-nrf-show-2024.

Hess, J., 2025. Epson ColorWorks Printers: Top Models and Must-Know Features for 2025. Barcode Factory Blog. Available at: https://blog.barcodefactory.com/blog/epson-colorworks-printers-top-models-and-must-know-features.

Islam, M. T., Alam, T., Yahya, I., & Cho, M., 2018. Flexible radio-frequency identification (RFID) tag antenna for sensor applications. Sensors, 18(12), 4212. DOI: 10.3390/s18124212.

Retail Tech Inc., 2025. Evolve Your Business with Toshiba Point of Sale Solutions. Available at: https://www.retailtechinc.com/manufacturers/toshiba.

Erickson, J., 2025. Process Verification vs. Validation in Manufacturing. Advanced Technology Services (ATS). Available at: https://www.advancedtech.com/blog/process-verification-vs-validation.

Visual Paradigm, 2024. Understanding BPMN: A Comprehensive Overview. Visual Paradigm Guides. Available at: https://guides.visual-paradigm.com/understanding-bpmn-a-comprehensive-overview

Şimşek, H., 2025. 18 Process KPIs to Monitor Process Performance. AIMultiple Research. Available at: https://research.aimultiple.com/process-kpis

ShareFile, 2024. How to Increase Process Efficiency: Best Practices, Tools, and KPIs. ShareFile Blog. Available at: https://www.sharefile.com/resource/blogs/increase-process-efficiency

Maté et al., Trujillo, J., & Mylopoulos, J., 2017. Specification and derivation of key performance indicators for business analytics: a semantic approach. Data & Knowledge Engineering, 108(3). Available at: https://www.researchgate.net/publication/312350945_Specification_and_Derivation_of_Key_Performance_Indicators_for_Business_Analytics_A_Semantic_Approach

Celonis, 2025. Process Simulation. Celonis Developer Center. Available at: https://docs.celonis.com/en/process-simulation

Testsigma (2023). Cross Device Testing - What it is, Why and How to Perform? Available at: https://testsigma.com/blog/cross-device-testing/

Atlassian. (2025). DevOps test automation: How automated testing enables DevOps. https://www.atlassian.com/devops/devops-tools/test-automation

BrowserStack. (2025). How to improve automation test coverage. https://www.browserstack.com/guide/improve-automation-test-coverage

Celonis. (2025). Cycle time: The metric all businesses should use to drive continuous improvement. https://www.celonis.com/blog/cycle-time-the-metric-all-businesses-should-use-to-drive-continuous-improvement

HelpScout. (2025). 107 customer service statistics and facts. https://www.helpscout.com/75-customer-service-facts-quotes-statistics

ISTQB. (2025). Standard glossary of terms used in software testing,v3.7. https://www.ctqb.org/en/downloads/istqb.html?cid=33119&file=files%2Fcontent%2Fctqb%2Fdownloads%2Fistqb%2FGlossary-terms-version-3.7.pdf

Ministry of Testing. (2025). Defect removal efficiency (DRE) Glossary. https://www.ministryoftesting.com/software-testing-glossary/defect-removal-efficiency

Perfecto. (2025). The essential guide to parallel testing. https://www.perfecto.io/blog/parallel-testing

Sauce Labs. (2025). Achieving parallelization in automated testing. https://saucelabs.com/resources/blog/the-four-keys-to-achieving-parallelization-in-automated-testing

testRigor. (2025). Cross-device testing: Strategies and tools. https://testrigor.com/blog/cross-device-testing/

Kekre, S., Krishnan, M. S., & Srinivasan, K. (1995). Drivers of customer satisfaction for software products: Implications for design and service support. Management Science, 41(9), 1456–1470.

Kathiresan, G. (2024). Automated test case generation with AI: A novel framework for improving software quality and coverage. World Journal of Advanced Research and Reviews, 23(2), 2880–2889.

Mockus, A., Nagappan, N., & Dinh-Trong, T. T. (2009). Test coverage and post-verification defects: A multiple case study. In Proc. of the 3rd Int. Symposium on Empirical Software Engineering and Measurement (ESEM) (pp. 291–301).

Orso, A., & Rothermel, G. (2014). Software testing: a research travelogue (2000–2014). In Proc. of the Future of Software Engineering (FOSE 2014) (pp. 117–132). ACM.

Whyte, G., & Mulder, D. L. (2011). Mitigating the impact of software test constraints on software testing effectiveness. Electronic Journal of Information Systems Evaluation, 14(2), 253–266.

Wang, Y., Mäntylä, M. V., Liu, Z., & Markkula, J. (2022). Test automation maturity improves product quality – A study of open-source projects using continuous integration. arXiv preprint arXiv:2202.04068.

Renna, P., & Colonnese, C. (2025). A Simulation-Driven Business Process Reengineering Framework for Teaching Assignment Optimization in Higher Education—A Case Study of the University of Basilicata. Applied Sciences, 15(5), 2756. DOI: 10.3390/app15052756.

Baqar, M., & Khanda, R. (2025). The Future of Software Testing: AI–Powered Test Case Generation and Validation. In K. Arai (Ed.), Intelligent Computing: Proceedings of the 2025 Computing Conference, vol. 2 (pp. 276–300). Springer, Cham. DOI: 10.1007/978-3-031-92605-1_19.

8. List of pictures, tables, graphs and abbreviations

8.1. List of pictures

8.2. List of tables

8.3. List of abbreviations

Avg: Average
BPMN: Business Process Modeling Notation
BPM: Business Process Management
V&V: Validation and Verification
RCA: Root Cause Analysis
VA: Value Adding
BVA: Business Value Adding
NVA: Non- Value Adding
CRM: Customer Relationship Management
KPI: Key Performance Analysis
QA: Quality Assurance
SQA: Software Quality Assurance