



MIDOCEAN UNIVERSITY

College of Media and Marketing

Master of Digital Marketing

A thesis submitted for the Master's degree in Digital Marketing

***The Role of A\B Digital Marketing Testing in
Boosting Online Sales in Saudi Arabia: A field
study***

***Student: HAMZAH AHMED ALI BIN
SHAAIB
251000062***

Supervisor: Dr. Mohamed Ewies

ABSTRACT

In a highly competitive digital era, improving online sales requires making data-driven decisions. A/B marketing testing has emerged as a key analytical technique that enables businesses to empirically evaluate differences in marketing components and determine which solutions best appeal to customers. This study aimed to examine the impact of three A/B testing characteristics, namely audience segmentation differentiated effectiveness, call-to-action effectiveness, and A/B variation design attractiveness, on online sales in Saudi Arabia. It also investigated the moderating role of the accuracy of A/B testing data analysis in influencing the relationship between these characteristics and online sales outcomes.

This study adopted a mixed-method approach to provide both depth and empirical validation. In the first phase, exploratory insights were gathered through interviews with six professionals, including digital marketing managers, e-commerce managers and executives, an online store manager, and a data analysis specialist. These interviews helped to better understand real-world practices and informed the development of the research model. In the second phase, the proposed model was empirically tested using survey data collected from 250 respondents. The interview data were analyzed using thematic analysis to identify key patterns and themes, while the survey data were examined through regression analysis using SmartPLS to test the study's hypotheses.

The findings reveal that variation design attractiveness has the strongest positive impact on online sales, suggesting that visually appealing and well-structured designs play a crucial role in capturing consumer attention and enhancing engagement. The effectiveness of call-to-action messaging also emerged as a key factor, highlighting the importance of clear, persuasive, and action-oriented messages in encouraging consumer responses. In addition, audience segmentation was found to significantly improve marketing effectiveness, demonstrating that tailoring content to specific customer groups can lead to better sales performance. Furthermore, the moderation analysis shows that A/B testing data analysis accuracy strengthens these relationships. In particular, accurate data analysis enhances the positive effects of both variation design attractiveness and call-

to-action effectiveness on online sales, underscoring the importance of combining creative strategies with precise analytical practices to achieve optimal results.

Overall, the study highlights the strategic value of A/B testing as a data-driven approach to enhancing online sales performance. The findings indicate that firms tend to achieve stronger results when they focus on innovative design, engaging marketing elements, and personalized content. These benefits become even more pronounced when supported by accurate and reliable data analysis. In this sense, A/B testing is not just a technical tool, but a strategic capability that integrates creativity with analytical precision. Based on these insights, the study suggests that organizations should invest in well-structured A/B testing practices, ensuring both creative experimentation and high levels of data analysis accuracy in order to maintain and strengthen their competitive advantage in the digital marketplace.

Keywords: A/B Testing Variation Design Attractiveness, Call-to-Action Effectiveness, Audience Segmentation Differentiated Effectiveness, A/B Testing Data Analysis Accuracy, Online Sales, Saudi Arabia.

TABLE OF CONTENTS

ABSTRACT	ii
TABLE OF CONTENTS	iv
Chapter 1: Introduction	1
1.1 Research Background.....	2
1.2 Research Motivation	5
1.2.1 Theoretical Motivation	5
1.2.2 Practical Motivation	6
1.3 Research Questions.....	8
1.4 Significance of the Study	10
1.4.1 Theoretical Significance	10
1.4.2 Practical Significance	11
1.5 Thesis Structure	12
Chapter 2: Theoretical Background	14
2.1 Theoretical Framework.....	15
2.1.1 The Technology-Organization and Environment (TOE) Framework.....	15
2.1.2 Evidence-Based Management	17
2.2. Data-Driven Decision Making and Digital Marketing	19
2.3 Online Sales	20
2.3.1 Digital Marketing and Online Sales	21
2.3.2 Measuring the Effectiveness of Online Sales	22
2. 4. A/B Testing	23
2.4.1 Nature of A/B Testing	23
2.4.2. Challenges and Limitations of A/B Testing	26
2.5 Characteristics of A/B Testing.....	27
2.5.1. A/B Testing Variation Design Attractiveness	27
2.5.2. A/B Call-To-Action Effectiveness	29
2.5.3. A/B Audience Segmentation Differentiated Effectiveness	31
2.6. Data Analytics and Interpretation in A/B Testing.....	32
2.8 Research Gaps.....	36
Chapter 3: Research Model and Hypotheses Development.....	40

3.1 A/B Testing Variation Design Attractiveness and Online Sales	41
3.2 Call-to-Action Effectiveness and Online Sales	42
3.3 Audience Segmentation Differentiated Effectiveness and Online Sales	43
3.4 The moderating Role of A/B testing data analysis Accuracy	45
Chapter 4. Research Methodology	51
4.1 Research Design	52
4.2 Target Population and Sampling.....	53
4.3 Data Collection	53
4.3.1 Exploratory Interviews	53
4.3.2 Questionnaire and Measures.....	54
Chapter 5: Research Findings	59
5.1 Qualitative Results: Semi-structured Interviews	60
5.1.1 Interviewees' Profile.....	60
5.1.2 Challenges of Designing A/B Testing in Digital Market in Saudi Arabia.....	62
5.1.3 A/B Testing Variation Design Effectiveness.....	62
5.1.4 A/B Testing Call-to-Action Effectiveness	63
5.1.5 Audience Segmentation Differentiated Effectiveness	64
5.1.6 A/B Testing Data Analysis Accuracy.....	65
5.1.7 Relevance of the A/B testing characteristics in enhancing online sales	65
5.2. Quantitative Findings	66
5.2.1 Response Rate and Respondents' Profile	66
5.2.2. Measurement Model Analysis	68
5.2.3 Structural Model Analysis	71
Chapter 6 Discussion and Conclusion	76
6.1 Key Findings.....	77
6.2 Discussions	78
6.3 Implications	81
6.3.1 Theoretical implications	81
6.3.2 Practical implications	84
6.4 Limitations and Future Research.....	84
6.5 Conclusion	86
References.....	87

List of Appendices.....	96
Appendix 1. Interview Questions	96
Appendix 2: Survey	98
LIST OF ABBREVIATIONS AND ACRONYMS.....	101
ACKNOWLEDGEMENT	102

Chapter 1:

Introduction

1.1 Research Background

Research increasingly emphasizes the importance of the digital economy's innovative features and their influence on business performance (Wager and Athey, 2017; Kohavi et al., 2013; Azevedo et al., 2018). Much of this work, however, has been shaped by the experiences of large technology firms. Companies such as Amazon, Microsoft, Google, and Facebook have historically had the capacity to run extensive experiments using their own infrastructure, along with the financial resources to support research on A/B testing challenges.

More recently, the widespread digitization of business activities and the growing availability of A/B testing tools have significantly reduced the cost of experimentation (Kohavi and Longbotham, 2017; Azevedo et al., 2018). As a result, A/B testing is no longer limited to major corporations; it is increasingly becoming a routine practice among smaller, internet-enabled firms (Koning et al., 2019).

Despite this growing adoption, there is still ongoing debate among scholars and practitioners about the extent of its impact. Some argue that A/B testing mainly leads to incremental improvements, while others suggest it can drive more substantial changes in product strategy and, ultimately, enhance organizational performance (Koning et al., 2022). Given this uncertainty, there is a clear need to examine A/B testing as a distinct and widely used phenomenon within the broader digital economy.

Today, software tools such as Optimizely and Google Optimize allow organizations of all sizes to run controlled experiments and analyze outcomes using ready-made platforms. As a result, the cost of testing alternative ideas has dropped substantially (Brynjolfsson and McAfee, 2016; Kohavi et al., 2007; Kohavi and Longbotham, 2017). A key reason for this shift is that experimenting with digital product features whether on e-commerce sites or other online platforms is far less expensive than making changes in physical production processes. In addition, the scale of digital businesses enables them to run multiple experiments simultaneously and independently. Advances in data analytics further support this process by helping firms interpret results more accurately

(Brynjolfsson and McElheran, 2016). These types of controlled experiments are commonly referred to as A/B testing (Azevedo et al., 2020).

A/B testing creates value in both direct and indirect ways. Directly, it allows firms to evaluate the effectiveness of specific ideas. Indirectly, it strengthens organizational learning by encouraging the generation and implementation of new ideas (Levinthal, 2017; Gans, Stern, and Wu, 2019). Empirical research suggests that adopting an experimental approach to strategy can improve performance, particularly in younger organizations (Levinthal, 2017; Gans, Stern, and Wu, 2019; Camuffo et al., 2020). For example, Camuffo et al. (2020) argue that A/B testing enables firms to move beyond analyzing past outcomes and instead focus on shaping the future through continuous experimentation and innovation.

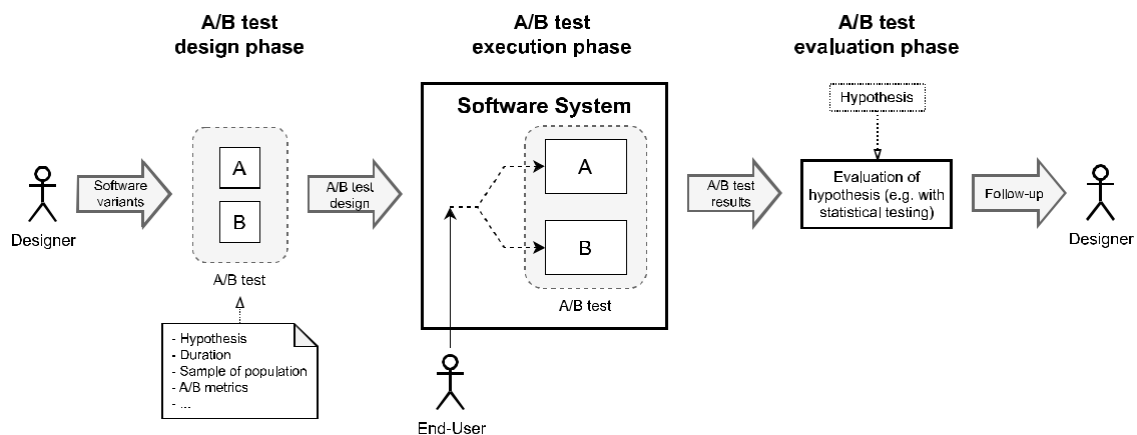


Figure 1.1. General Process of A/B Testing. Adapted from Quin et. al. (2024)

Although relatively few studies have examined the impact of A/B testing across a wide range of firms, both academic and practitioner work highlights its growing importance (Kohavi et al., 2007; Kohavi et al., 2009; Kohavi and Longbotham, 2017; Xu et al., 2015). In practice, online retailers experiment with pricing strategies, product bundles, and display formats (Dubé et al., 2017; Sahni, Zou, and Chintagunta, 2016), while digital platforms test social features, recommendation systems, and content design to increase user engagement (Kumar and Tan, 2015). Although individual A/B tests are often used to assess incremental changes, a broader experimentation strategy can lead to both gradual improvements and more substantial organizational shifts (Browne and Jones, 2017).

Overall, this body of work supports the view that systematic experimentation can play a meaningful role in enhancing firm performance.

Accurately interpreting data from A/B testing is essential for effective decision-making. Advances in data analytics have made it easier for firms to draw meaningful and reliable insights from their experiments. Building on this, the present study argues that the accuracy of data analysis plays a moderating role in the relationship between A/B testing features and a firm’s online sales performance.

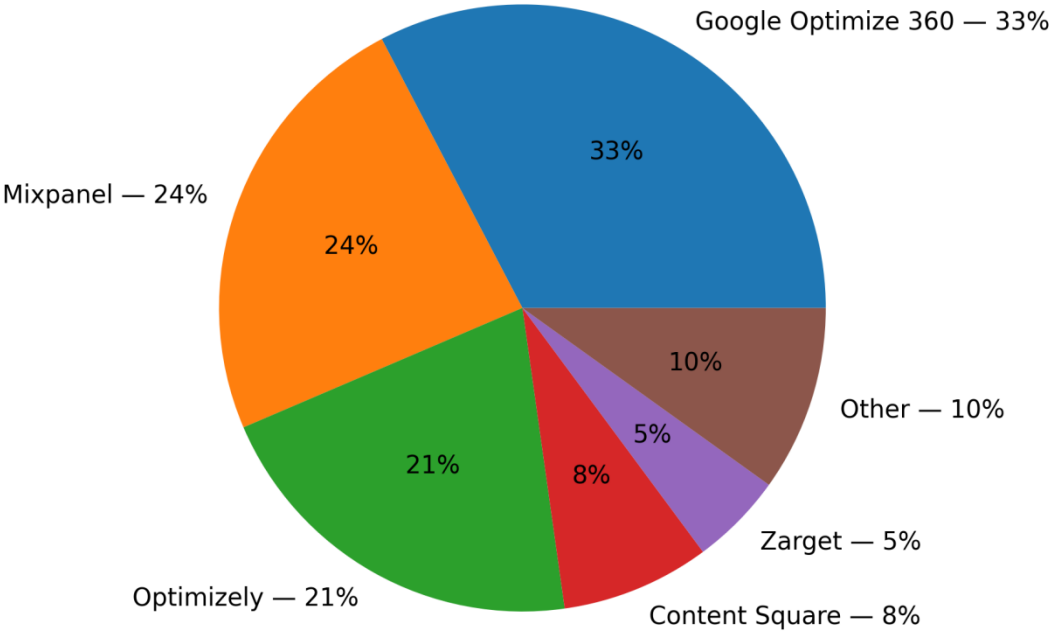


Figure 1.2: Statistics for websites using A/B Testing technologies
Source: (www.trends.builtwith.com, 2025)

Figure 1.2 depicts the distribution of A/B testing tool usage in Saudi Arabia, illustrating the market share of different platforms for experimentation. Google Optimize 360 has the biggest share (33%), showing that it is the most popular tool among users, owing to its smooth interaction with Google Analytics and user-friendly design. Following that, Mixpanel has 24% of the market, making it the second most popular tool. Mixpanel is well-known for its event-driven tracking and robust user statistics. Optimizely is ranked third with a 21% share, thanks to its broad tools for website and product optimization.

Content Square has 8% of the market and is known for its user experience analytics. Zarget, with a 5% market share, provides A/B testing functionalities that are frequently coupled with conversion rate optimization tools. Other tools account for the remaining 10%, which is most likely made up of niche or custom-built solutions. Overall, the data show that Google Optimize 360, Mixpanel, and Optimizely dominate the A/B testing landscape in Saudi Arabia, accounting for 78% of all usage, indicating a strong preference for platforms that combine analytics integration, ease of use, and advanced experimentation features.

The majority of research on the A/B testing method in Saudi Arabia focused on technical challenges and inconsistent results from the application of A/B approaches. For example, Alkhalil et al. (2024) employed empirical A/B testing on an online store and a review of the literature to investigate the important factors influencing consumer trust in e-commerce. With A/B testing proving their impact on user behavior and preferences, the findings indicate that privacy, customer feedback, and ease of use are the most essential factors in generating trust. By meticulously evaluating multiple website versions, the study demonstrates that enhancing data security, including transparent feedback, and maximizing usability significantly enhance consumer confidence. This provides organizations with specific insights into how to increase their e-commerce success.

1.2 Research Motivation

1.2.1 Theoretical Motivation

This study is theoretically motivated by the increasing importance of data-driven marketing in understanding consumer responses in digital contexts. While previous studies have extensively investigated aspects impacting online purchase behavior, less emphasis has been paid to how specific A/B marketing testing characteristics impact online sales performance. In particular, the functions of variety design appeal, call-to-action effectiveness, and audience segmentation differential efficacy have received little attention in empirical studies. Existing research frequently emphasizes broad digital marketing methods over experimentally driven optimization techniques like A/B testing.

As a result, this study aims to contribute to the digital marketing and e-commerce literature by conceiving A/B testing characteristics as significant drivers of online sales outcomes and investigating how these factors influence customer decision-making in online contexts.

Furthermore, this study incorporates the concept of A/B testing data analysis accuracy as a moderating mechanism to improve the effectiveness of marketing experimentation. Although businesses are increasingly relying on analytics to interpret experimental outcomes, the significance of data analysis accuracy in improving the efficacy of A/B testing strategies has gotten little academic attention. By investigating this moderating effect, the study adds to our theoretical knowledge of how analytical precision might increase the value of marketing testing. Furthermore, by focusing on Saudi Arabia's rapidly expanding e-commerce sector, the study contributes empirical evidence from an emerging digital market, where consumer behavior and marketing practices may differ from those observed in more mature e-commerce environments.

This study applies the Technology-Organization-Environment (TOE) framework to examine the impact of A/B testing in digital marketing on the online sales performance of Saudi firms. It is motivated by the need to advance empirical understanding of how A/B testing tools and methodologies contribute to e-commerce optimization in underexplored contexts. While A/B testing is widely adopted in Western e-commerce ecosystems, its effectiveness in Saudi Arabia remains underexamined, particularly given the country's distinctive digital infrastructure, smartphone penetration, and platform preferences. Accordingly, this research investigates how localized user behavior interacts with technological factors such as testing capabilities and data accuracy to influence sales outcomes. From organizational and environmental perspectives, the study addresses existing gaps in how Saudi firms incorporate A/B testing into their operational practices.

1.2.2 Practical Motivation

From a practical perspective, firms operating in digital markets face intense competition and must continuously refine their marketing strategies to enhance online sales performance. Although A/B testing is widely adopted, many organizations lack clear

guidance on which design elements most effectively drive sales outcomes. This study addresses this gap by providing actionable insights for marketing practitioners, e-commerce managers, and data analysts on how variation design, call-to-action strategies, and audience segmentation can be strategically leveraged to improve campaign effectiveness.

Furthermore, the study underscores the critical role of accurate data analysis in maximizing the value of A/B testing. While many firms implement testing practices, they often underutilize insights due to limitations in analytical capabilities. By demonstrating how data accuracy strengthens the effects of key A/B testing characteristics on sales performance, the findings offer practical guidance for integrating creative experimentation with rigorous analytics. Ultimately, this research supports the development of data-driven marketing strategies that enhance customer engagement, optimize campaign outcomes, and sustain competitive advantage in dynamic digital environments.

Existing literature has documented the widespread adoption of A/B testing across large technology companies (Kohavi et al., 2009; Azevedo et al., 2018), software start-ups (Koning et al., 2019), and consumer retailers (Siroker and Koomen, 2013). Within these organizations, A/B testing supports various functional domains, including product innovation and market experimentation (Hampel et al., 2020), user interface design (King et al., 2017), marketing and promotion (Nelson et al., 2020; Miller and Hosanagar, 2020), and software engineering (Gupta et al., 2019). It is commonly applied to decisions such as optimizing webpage design, selecting article headlines, refining call-to-action language, and improving recommendation algorithms (Gallo, 2017).

However, the motivations of decision makers conducting A/B tests differ, leading to heterogeneous effects. For example, software engineers often employ non-inferiority tests to ensure that newly developed features do not adversely affect system performance or user experience (Kohavi and Thomke, 2017). In contrast, marketers and user experience designers typically focus on improving key performance indicators, such as click-through rates, conversion rates, user retention, and revenue generation.

In Saudi Arabia, the rapid expansion of the e-commerce sector, driven by Vision 2030's digital transformation agenda, creates a distinctive organizational and environmental context where data-driven practices intersect with traditional retail models. This study examines how firms structure testing processes, analyze data, and implement insights, all of which influence the effectiveness of A/B testing initiatives. It also considers environmental factors, such as cultural purchasing behaviors and customer privacy expectations, that may moderate the relationship between testing strategies and sales outcomes. By integrating these dimensions within the TOE framework, this study develops a comprehensive model for understanding A/B testing effectiveness in emerging digital economies.

From a practical standpoint, this research addresses the pressing need for Saudi e-commerce firms to adopt data-driven strategies that enhance sales performance. As digital transformation accelerates, online retailers require actionable insights to improve user experience, increase conversion rates, and maximize returns on A/B testing investments. By identifying the most effective testing elements, such as call-to-action design and audience segmentation, and highlighting the role of data analysis accuracy, this study provides concrete guidance for marketers and developers. Ultimately, the findings support more efficient resource allocation, reduce experimentation costs, and enable firms to tailor strategies to local consumer preferences, thereby strengthening competitiveness in the evolving digital marketplace.

1.3 Research Questions

This study aims to examine the impact of three A/B digital marketing testing characteristics: variation design attractiveness, audience segmentation effectiveness, and call-to-action effectiveness, on firms' online sales performance in the Saudi Arabian context. These dimensions are selected due to their critical roles in shaping consumer behavior and influencing sales outcomes. First, variation design attractiveness directly affects user engagement, as visually appealing and well-structured interfaces enhance user experience and increase conversion rates. Second, effective audience segmentation enables firms to tailor marketing efforts to distinct consumer groups, thereby improving

relevance and personalization, which are key drivers of purchase decisions. Third, call-to-action effectiveness plays a pivotal role in directing user actions, as clear and compelling prompts significantly enhance click-through and conversion rates. Collectively, these characteristics capture essential elements of digital marketing visual appeal, targeted communication, and persuasive engagement, making them central to assessing the effectiveness of A/B testing on sales performance.

The inclusion of A/B testing data analysis accuracy as a moderating variable is justified by its role in ensuring the validity and reliability of insights derived from marketing experiments. Even when design, segmentation, and call-to-action strategies are well-developed, inaccurate data analysis can lead to misleading conclusions, suboptimal decision-making, and reduced campaign effectiveness. High levels of analytical accuracy strengthen the relationship between A/B testing characteristics and online sales performance by ensuring that observed effects reflect true patterns rather than statistical noise or measurement error.

The first research question is guided by the practical need to understand how the look and feel of different A/B test designs actually influence customers' buying decisions in Saudi Arabia's online market. When a webpage or ad is visually appealing and easy to navigate, it naturally draws attention and makes it easier for users to complete a purchase. However, knowing which design truly works best is not always straightforward. If the data from A/B testing is not analyzed accurately, businesses might end up choosing the wrong design. That's why it's important to also look at data analysis accuracy, as it ensures that the conclusions drawn about design attractiveness are trustworthy and lead to better sales decisions. Hence, this study seeks to address the following research question:

RQ1: What is the impact of the A/B variation design attractiveness on online sales in Saudi Arabia, and how is this relationship moderated by the A/B testing data analysis accuracy?

The second research question is based on the idea that small elements like call-to-action buttons can make a big difference in whether a customer completes a purchase or not. A

well-written and well-placed call-to-action can encourage users to take the next step, such as clicking “Buy Now” or “Sign Up.” In a fast-growing digital market like Saudi Arabia, businesses rely heavily on these elements to improve performance. However, identifying the most effective call-to-action depends on accurately interpreting A/B testing results. If the analysis is flawed, companies might rely on misleading insights. Therefore, considering data analysis accuracy helps ensure that the impact of call-to-action is understood correctly and used effectively. Therefore, the second research question is stated as follows:

RQ2: What is the effect of call-to-action effectiveness in A/B testing on online sales in Saudi Arabia, and how does the A/B testing data analysis accuracy influence this effect?

The third question is driven by the growing importance of personalization in digital marketing. Not all customers behave the same way, so dividing them into different segments and tailoring content to each group can significantly improve engagement and sales. In Saudi Arabia’s diverse online market, this becomes even more important. However, to know whether segmentation strategies are truly working, businesses must rely on A/B testing data. If that data is not analyzed carefully, it can lead to wrong conclusions about which audience groups are responding best. By including data analysis accuracy, this question ensures that segmentation strategies are evaluated properly, helping businesses make smarter and more effective marketing decisions.

RQ3: What is the impact of audience segmentation differentiated effectiveness in A/B testing on online sales in Saudi Arabia, and to what extent does the A/B testing data analysis accuracy influence this effect?

1.4 Significance of the Study

1.4.1 Theoretical Significance

This study contributes to the academic literature by enhancing the understanding of data-driven marketing and A/B testing in digital commerce. First, it provides empirical

evidence on the effects of key A/B testing characteristics variation design attractiveness, call-to-action effectiveness, and audience segmentation effectiveness on online sales performance. By systematically examining these dimensions, the study advances theoretical insights into how experimental marketing elements shape consumer behavior and sales outcomes in digital environments.

Second, the study develops a moderated research model that links A/B testing characteristics to online sales while incorporating data analysis accuracy as a moderating variable. This contribution adds theoretical depth by demonstrating that the effectiveness of marketing experiments depends not only on creative and strategic design but also on the precision of data interpretation. In doing so, it extends existing A/B testing and digital marketing literature by emphasizing the critical role of analytical rigor in digital experimentation.

Third, the research offers context-specific insights from Saudi Arabia's rapidly expanding e-commerce sector, which remains underexplored in prior studies. By situating A/B testing practices within this emerging market, the study broadens the global understanding of digital marketing across diverse economic and technological contexts. Moreover, the adoption of a mixed-methods approach combining exploratory interviews with SmartPLS-based structural modeling demonstrates a robust methodological framework for investigating digital marketing effectiveness, thereby contributing to research design in this field

1.4.2 Practical Significance

From a practical perspective, this study offers valuable insights for digital marketing practitioners, e-commerce managers, and data analysts seeking to optimize online sales strategies. The findings provide empirical guidance on which A/B testing components most strongly influence sales outcomes, particularly emphasizing the roles of design attractiveness and call-to-action effectiveness. These insights support a shift toward evidence-based decision-making, enabling firms to move beyond intuition and rely on data-driven approaches when designing and evaluating marketing campaigns.

Furthermore, the study informs more effective resource allocation. By identifying the marketing elements with the greatest impact on sales performance, organizations can strategically direct their time, budget, and creative efforts toward high-impact design enhancements, persuasive messaging, and targeted segmentation strategies. This facilitates more efficient use of digital marketing investments.

Finally, the study highlights the critical importance of data analysis accuracy in maximizing the benefits of A/B testing. The findings indicate that even well-designed experiments may yield limited value if supported by weak analytical processes. Accordingly, firms are encouraged to strengthen their analytical capabilities and data interpretation skills to fully leverage A/B testing practices. These implications are particularly relevant for businesses operating in Saudi Arabia's rapidly evolving digital marketplace, where robust data-driven strategies are essential for enhancing competitiveness and improving online sales.

1.5 Thesis Structure

This thesis is structured into eight chapters, each addressing a distinct component of the research. A brief explanation of each chapter is presented below.

Chapter One introduces the study by outlining the research background, problem statement, objectives, and research questions. It also presents the theoretical and practical motivations, highlights the study's significance, defines its scope, and provides an overview of the thesis structure.

Chapter Two presents the theoretical background and comprehensive literature review. This research was guided by two theories: Technology-Organization-Environment (TOE) framework and the principles of Evidence-Based Management (EBM), and explains how these perspectives underpin the proposed relationships. The literature review covered the major study constructs, namely, the concept of online sales performance and the key A/B testing characteristics investigated in this study, namely variation design attractiveness, call-to-action effectiveness, and audience segmentation effectiveness, as well as the role of data analysis accuracy. The chapter further identifies gaps in the existing literature.

Chapter three develops the research hypotheses and introduces the conceptual model, providing theoretical justification for each proposed relationship, including the moderating role of data analysis accuracy.

Chapter four details the research methodology, describing the mixed-methods design that combines qualitative exploratory interviews with a quantitative survey. It also outlines the sampling procedures, measurement development, data collection methods, and analytical techniques, including thematic analysis and SmartPLS, as well as considerations of validity, reliability, and ethics.

Chapter five presents the research findings, including qualitative themes, quantitative results, hypothesis testing, and moderation analysis, followed by an integrated discussion in relation to prior studies.

Chapter six presents the discussion of the findings in light of existing literature on A/B testing and online sales, as well as the moderating role of data analysis accuracy. Additionally, chapter six concludes the thesis by summarizing the key findings and contributions. It also discusses the theoretical, managerial, and methodological implications, acknowledges study limitations, and provides directions for future research.

Chapter 2: Theoretical Background

2.1 Theoretical Framework

This research is guided by two fundamental theories: Technology-Organization and Environment (TOE) framework and Evidence-Based Management (EBM) theory. A brief explanation of each framework is presented in subsequent sections.

2.1.1 The Technology-Organization and Environment (TOE) Framework

The Technology-Organization-Environment (TOE) framework, introduced by Tornatzky et al. (1990), provides a useful lens for understanding how firms adopt and implement new technologies. It highlights three key dimensions. The technological dimension refers to the availability and characteristics of technologies, including their perceived benefits, compatibility with existing systems, and level of complexity. The organizational dimension focuses on internal factors such as firm size, structure, available resources, and managerial support, all of which shape a firm's readiness to adopt innovation. The environmental dimension captures external pressures, including industry competition, regulatory conditions, and broader market trends, which influence strategic decisions regarding technology adoption.

In the context of this study, the TOE framework offers a structured way to examine how A/B testing contributes to online sales performance. From a technological perspective, it helps explain how A/B testing tools function, how easy they are to use, and how well they integrate with existing digital marketing systems. The organizational dimension reflects whether firms possess the necessary expertise, financial resources, and leadership commitment to effectively implement A/B testing practices. Meanwhile, the environmental dimension highlights the role of external forces such as increasing competition and evolving customer preferences that encourage firms to adopt data-driven approaches to improve performance. Taken together, these dimensions provide a more comprehensive explanation of why some firms benefit more from A/B testing than others. This study defines A/B testing as a method that compares two versions (A and B) of a webpage, advertisement, or product presentation to determine which performs better in driving outcomes such as conversions, customer acquisition, and revenue. Rather than focusing solely on the technical side, the TOE framework enables a broader analysis that

also considers organizational readiness and market conditions. Specifically, this study examines the effectiveness and implementation challenges of A/B testing tools (technology), the firm's readiness and resource allocation (organization), and the influence of market competition, cultural preferences, and regulatory factors in Saudi Arabia (environment). This integrated perspective strengthens both the theoretical grounding and practical relevance of the research.

Prior research suggests that A/B testing can lead to variability in performance outcomes, particularly when it encourages firms to experiment with a wider range of ideas (Koning et al., 2022). Generating more ideas increases the likelihood of identifying high-quality solutions, but it also makes selection more challenging (Girotra, Terwiesch, and Ulrich, 2010). When firms are able to identify and prioritize the most promising ideas through systematic experimentation, they can achieve substantial gains. Similarly, continuous iteration allows these ideas to be refined further, leading to cumulative improvements over time (Kohavi and Thomke, 2017; Azevedo et al., 2018; Thomke, 2020). In some cases, experimentation may also reveal that certain ideas are ineffective, helping firms avoid costly missteps. Overall, A/B testing supports both exploration and exploitation, enabling firms to learn quickly and optimize performance at scale.

Empirical studies applying the TOE framework reinforce its relevance. For example, Nguyen et al. (2022) found that technological factors (e.g., relative advantage and compatibility), organizational factors (e.g., top management support and strategic orientation), and environmental factors (e.g., government support and market trends) all play a significant role in driving digital adoption and firm performance. Similarly, other studies highlight the importance of technological readiness, organizational capabilities, and environmental conditions in shaping innovation outcomes (Eze et al., 2013).

Despite its strengths, scholars have noted that the TOE framework can be further extended to capture additional contextual factors (Awa et al., 2017). In particular, while TOE effectively explains adoption at the organizational level, it pays less attention to individual-level influences. Nonetheless, existing evidence suggests that technological, organizational, and environmental factors remain the primary drivers of adoption

decisions and performance outcomes.

Building on these insights, the TOE framework helps explain how A/B testing enhances online sales performance in Saudi Arabia. From a technological standpoint, elements such as design attractiveness and call-to-action effectiveness reflect how firms use digital tools to capture customer attention and influence behavior. Audience segmentation further strengthens this process by allowing firms to tailor their offerings to specific customer groups. At the organizational level, firms need the right capabilities, resources, and strategic mindset to translate insights from A/B testing into actionable improvements. In a similar vein, the rapidly evolving and competitive digital environment in Saudi Arabia encourages firms to rely more heavily on data-driven strategies. When these factors align, A/B testing is more likely to generate meaningful improvements in sales performance.

The TOE framework also provides a useful explanation for the moderating role of data analysis accuracy. From a technological perspective, strong data analysis capabilities enable firms to extract reliable insights from A/B testing, making features such as design, call-to-action, and segmentation more effective. At the organizational level, firms with advanced analytical skills and a data-driven culture are better positioned to interpret and apply these insights. From an environmental perspective, increasing competition and the growing importance of digital platforms further push firms to rely on accurate data analysis to remain competitive. Taken together, this explains why data analysis accuracy strengthens the relationships proposed in H4a, H4b, and H4c: it allows firms to fully leverage A/B testing insights and translate them into improved online sales outcomes.

2.1.2 Evidence-Based Management

Evidence-Based Management (EBM), as introduced by Rousseau (2006) and Pfeffer and Sutton (2006), argues that managerial decisions should be grounded in the best available evidence, including empirical data, scientific research, and systematic experimentation, rather than intuition, experience, or tradition. Within digital marketing contexts, this perspective implies that managers rely on experimental methods such as A/B testing and analytical evidence to determine which strategies most effectively enhance performance outcomes.

This study is grounded in the principles of EBM, positioning A/B testing as a structured, evidence-based approach for evaluating alternative digital strategies, including webpage design, call-to-action effectiveness, and audience segmentation. From this perspective, A/B testing functions as a form of managerial experimentation in which decisions are validated through observable data rather than subjective judgment. The accuracy of data analysis is therefore central to this process, as it determines the reliability and credibility of the evidence generated. When analytical accuracy is high, managerial decisions are more likely to be well-informed and effective, ultimately improving online sales performance. In contrast, weak or inaccurate analysis may distort findings and undermine the benefits of evidence-based decision-making.

EBM is particularly relevant to this study for several reasons. First, it frames A/B testing not merely as a marketing technique, but as a managerial decision-making capability. Second, it highlights the importance of data accuracy in ensuring that evidence is valid and actionable. Third, it directly links analytical rigor to improved decision quality and organizational performance, which aligns well with digital, data-driven business environments such as Saudi Arabia's online retail sector.

Empirical and practitioner evidence further supports this view. For instance, cases of failed experiments such as those discussed in "Where A/B Testing Goes Wrong" (2024) demonstrate that flawed implementation or inaccurate analysis can prevent firms from achieving expected performance gains. Conversely, organizations that apply robust statistical methods and carefully interpret experimental results tend to generate more reliable and actionable insights (Kurz et al., 2025). These observations align closely with the core principles of EBM, where decisions are expected to be guided by rigorous evidence rather than assumptions.

Building on this logic, Eapen et al. (2023) highlight that although relevant research evidence is widely available, it is often underutilized by managers due to challenges in accessibility and practical application. To address this gap, they propose structured and digitally supported systems that integrate research findings, organizational data, and expert knowledge into a unified decision-support process. Such systems enable managers

to make more consistent, transparent, and evidence-based decisions rather than relying primarily on intuition.

Within this study, firms with strong analytical capabilities are better able to interpret A/B testing results, identify effective design and messaging strategies, and refine audience segmentation to better align with customer preferences. In contrast, firms with weak analytical accuracy may misinterpret experimental outcomes, leading to ineffective decisions and reduced performance gains. Accordingly, the conceptual model suggests that data analysis accuracy strengthens the relationship between A/B testing characteristics and online sales performance by enhancing the quality and reliability of managerial decision-making.

2.2. Data-Driven Decision Making and Digital Marketing

Recent literature highlights the increasing importance of data-driven decision-making in today's digitized economy as a key driver of firm performance (Brynjolfsson & McElheran, 2016; Camuffo et al., 2020). Similarly, marketing research has long emphasized the value of systematic market testing for improving products, services, and customer outcomes (Runge & Nair, 2021). Within this broader shift toward analytics-based management, A/B testing has emerged as one of the most widely used forms of randomized experimentation in digital business and marketing contexts (Brynjolfsson & McElheran, 2016).

In contemporary digital organizations, data-driven decision-making is no longer optional but increasingly essential. A/B testing technologies allow firms to apply a more scientific approach to evaluating website designs, advertising content, and product interfaces. Often referred to as online controlled experiments, split testing, or bucket testing, A/B testing gained substantial momentum in the 2010s as firms increasingly turned to empirical evidence to guide strategic and operational decisions.

At its core, an A/B test is a randomized experiment that compares two versions of a digital element such as a webpage, advertisement, or application to determine which performs better. Users are randomly assigned to different groups and exposed to different variants, and their responses are then analyzed to identify the more effective option (Kohavi &

Longbotham, 2017). Although conceptually straightforward, this method has had a transformative impact on online marketing practices. Initially adopted by a small number of large technology firms, A/B testing has now become a standard and often indispensable tool across digital industries (Siroker & Koomen, 2013). However, while its benefits are well documented, researchers also caution that poor implementation or flawed analysis can lead to misleading or unreliable conclusions (Kohavi & Longbotham, 2007).

Given the expanding body of research in this area, terminology related to A/B testing is not always consistent, with similar concepts described using different terms across studies. To address this, the present study draws on foundational work by Siroker and Koomen (2013) and Kohavi et al. (2017) to establish a clear and consistent definition. Accordingly, online controlled experiments are understood here as a core component of the digital marketing toolkit, widely regarded as essential for both startups and large-scale enterprises.

The credibility of A/B testing is reinforced by its widespread adoption in practice. A well-known example is the 2012 U.S. presidential campaign of Barack Obama, where A/B testing of donation pages and visual elements reportedly generated an additional \$57 million in donations (Siroker & Koomen, 2013). This case illustrates how a strong commitment to experimentation and a willingness to rely on data-driven evidence, even when it challenges intuition, can lead to substantial performance gains.

2.3 Online Sales

Online sales refer to the exchange of goods and services through electronically connected devices within a digital environment, where ownership or usage rights are transferred online. In recent years, this form of retailing has become increasingly central to global commerce. Available estimates show that e-commerce accounted for nearly USD 5 trillion in 2021, representing about 19% of global retail sales (Global E-Commerce Share of Retail Sales 2026, 2022). By 2023, this share had risen to approximately 20.8%, reaching around USD 6.3 trillion, with continued growth expected in the coming years (Global E-commerce Sales Growth Report for 2021–2026, 2022).

The rapid expansion of online sales was further accelerated during the COVID-19

pandemic, when restrictions on movement and social interaction significantly altered consumer behavior (Gregurec, Tomičić Furjan & Tomičić-Pupek, 2021). As many physical stores reduced operating hours or closed altogether, consumers increasingly shifted toward online platforms to meet their purchasing needs (Gregurec et al., 2021). This shift contributed to a surge in the number of online retailers, which in turn intensified competition and market saturation in the digital space. As a result, firms now face increasing pressure to differentiate themselves by delivering superior value to customers, as value creation has become a key determinant of success in highly competitive online markets.

2.3.1 Digital Marketing and Online Sales

From a digital marketing standpoint, firms typically rely on a set of key performance indicators (KPIs) to understand how well they are doing. These often include metrics such as conversion rates, goal completions, types of users, traffic sources, non-branded keyword traffic, and keyword rankings (Saura et al., 2017). In practice, KPIs are widely seen as essential because they highlight the most important aspects of a company's operations those that influence both short-term results and long-term growth (Muntean et al., 2016). In other words, KPIs translate complex business activities into clear, measurable signals of performance.

For KPIs to be useful, they must be measurable and clearly linked to the organization's goals. They also help define responsibilities across teams and create a consistent way to evaluate performance using data rather than assumptions. Setting clear, realistic, and well-defined KPIs is therefore a critical step in effective performance management (Tsai & Cheng, 2012). By monitoring these indicators, online businesses can track sales trends and identify where improvements are needed, particularly in customer-facing areas.

The process usually begins with defining clear objectives and understanding which parts of the business have the greatest impact on those goals. For example, a company focused on increasing sales may prioritize different metrics than one aiming to improve service quality or optimize marketing campaigns. By examining these factors, firms can select the most relevant KPIs to increase sales, attract more website visitors, or reduce pressure

on customer support. Although boosting sales is often the main goal for online businesses, success typically depends on several connected elements, such as generating enough traffic, offering competitive pricing, and reducing cart abandonment (Ahmed et al., 2017).

2.3.2 Measuring the Effectiveness of Online Sales

Performance measurement is generally understood as the ongoing process of gathering, analyzing, and reporting data to see how well an organization is meeting its goals. It involves regularly tracking results against predefined targets (Franceschini et al., 2018) and assessing both how efficiently and how effectively activities are carried out (Baki, 2020). In practice, it plays a key role in decision-making across all levels of the organization. It helps managers better understand business models and customer behavior, and supports more informed decisions about website design, product placement, and promotional strategies ultimately contributing to higher sales and more opportunities for cross-selling (Keenan, 2022).

In the context of online businesses, performance metrics provide a concrete way to capture what is happening across key activities. Internally, they are used to evaluate performance and guide strategic choices; externally, they allow firms to compare themselves with competitors and industry benchmarks. Without such metrics, it would be difficult to assess whether a strategy is working or to ensure progress toward organizational goals (Baki, 2020).

Much of the existing research evaluates online performance through key performance indicators (KPIs). Commonly used sales-related KPIs include shopping cart abandonment rate, Average Revenue per Visitor (ARV), Conversion Rate (CR), Average Number of Products per Order (APO), and Average Order Value (AOV). In addition, content-related metrics are frequently examined, such as bounce rate, time spent on a page (or average session duration), and pageviews, which reflect how often a page is visited. Together, these indicators offer a comprehensive view of both user behavior and website effectiveness.

Because KPIs can be tailored to the specific context of a company's website, systematically tracking them allows organizations to evaluate overall business

performance as well as the effectiveness of their online presence (Turner, 2010). Whether a website is transactional or informational, it plays a central role in shaping visibility and engagement (Roy & Sharma, 2021). For transactional websites in particular, the primary goals are typically to convert visitors into customers and to increase overall traffic.

More broadly, KPIs are a set of measurable indicators used to assess how well different parts of a business are performing (Sultan, 2022). In practice, marketers tend to rely heavily on quantitative data (Gräve, 2019). For businesses operating across multiple channels, widely used indicators include sales volume, average unit price, profit margins, expense ratios, inventory turnover, and asset utilization measures. Other studies have also compared web analytics tools such as Google Analytics and SimilarWeb, focusing on metrics like total visits, unique users, bounce rate, and session duration across different industries and countries.

Finally, recent work has emphasized the importance of e-fulfillment in evaluating online retail performance. This includes activities such as warehousing, order processing, delivery, and returns management, all of which directly influence customer satisfaction and operational efficiency (Bressolles & Lang, 2020). Building on this foundation, the present study seeks to examine whether A/B testing can meaningfully improve online sales performance, with a specific focus on the Saudi Arabian context.

2. 4. A/B Testing

2.4.1 Nature of A/B Testing

A/B testing often called split testing or online experimentation is a method used to compare different versions of a webpage, email, or digital feature to see which one performs better. The idea is simple: present alternative versions (for example, version A and version B) to different groups of users at the same time, and then measure which version leads to better results based on key performance indicators (Kohavi & Longbotham, 2017; Esteller-Cucala et al., 2019). By observing how users interact with each version, companies can make more informed decisions about what works best (Mullin, 2022). This approach is widely used to improve user engagement and satisfaction

with digital products and services.

In practice, A/B testing usually focuses on comparing two versions A and B, by exposing them to real users and evaluating their performance (Siroker & Koomen, 2013). The process typically starts with a clear hypothesis, followed by designing an experiment to test it. Unlike traditional testing methods, A/B tests are conducted in live environments, which makes the results more reflective of actual user behavior. The process generally unfolds in three stages. First is the design phase, where researchers define the hypothesis, select the target audience, determine how long the test will run, and choose the metrics to track such as click-through rates, number of clicks, or session activity (Drutsa et al., 2015b). Next comes the execution phase, where both versions are launched and users are randomly assigned to each one while data is collected in real time. Finally, in the evaluation phase, statistical techniques (such as t-tests) are used to determine whether the observed differences are meaningful (Gupta & Chokshi, 2020). Based on the findings, organizations can adopt the better-performing version or continue experimenting.

Because it relies on real user interactions, A/B testing has become a highly valuable tool in digital settings (Kohavi & Longbotham, 2017). It allows organizations to directly assess how specific changes affect user behavior and conversion outcomes (Esteller-Cucala et al., 2019). More importantly, it provides clear, data-driven insights that help improve website design, optimize performance, and enhance the overall user experience factors that can ultimately build customer trust and strengthen brand value (Gallo, 2017). Building on this, the present study explores how A/B testing can contribute to improving online sales performance, particularly within the Saudi Arabian context.

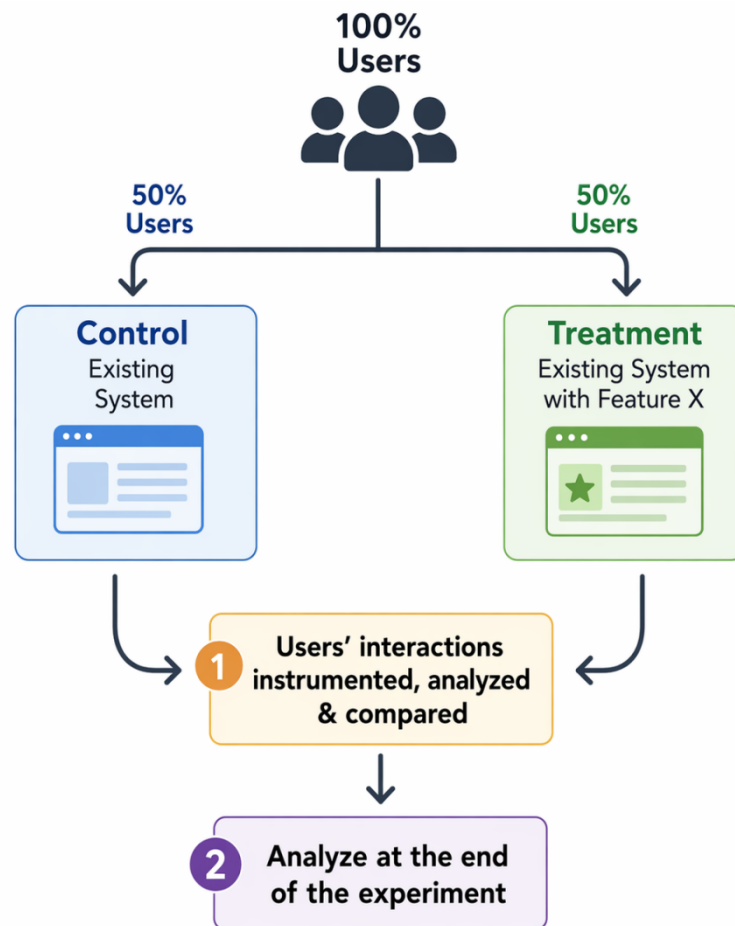


Figure 2.1. A simple controlled experiment: An A/B Test

In real-world settings, many large digital companies, including Airbnb, Amazon, Booking.com, eBay, Facebook, Google, LinkedIn, Lyft, Microsoft, Netflix, Twitter, Uber, Yahoo!/Oath, and Yandex, rely heavily on controlled online experiments to guide their decisions (Gupta et al., 2019). These organizations run experiments at a massive scale, often conducting thousands and sometimes tens of thousands of tests each year, involving millions of users. The scope of these experiments is broad, covering areas such as user interface design, search and recommendation systems, personalization, platform performance, content management, and even customer support. Testing is also not limited to a single channel; it typically spans websites, mobile and desktop applications, and email platforms.

A good example comes from experiments on the Bing platform, where user interactions are carefully tracked and recorded. This data is then used to calculate performance metrics, allowing researchers to compare how different versions perform. At its core, a controlled experiment is quite straightforward: users are divided into groups and shown different versions of the same feature, a control version (A) and a modified version (B). By comparing how users respond to each version, companies can identify which option delivers better results, as illustrated in Figure 2.1.

2.4.2. Challenges and Limitations of A/B Testing

A/B testing is a valuable tool, but it is not without its challenges. One of the main difficulties is that it often requires a large sample size and sufficient time to produce reliable results, something that can be hard for smaller businesses or websites with limited traffic. In addition, if the test is poorly designed, for example, due to flawed randomization or stopping the experiment too early, the results can be misleading (Kohavi et al., 2019). Another limitation is its focus on short-term metrics, such as clicks or conversions, which may overlook longer-term outcomes like customer loyalty or brand perception. Results can also be affected by external factors, including seasonal patterns or sudden market shifts. There are also increasing concerns around user privacy and the ethics of experimenting on customers, and findings from one context do not always generalize well to others.

In a similar vein, A/B testing remains a central tool in modern marketing analytics. As Kumar (2021) notes, it supports more accurate, data-driven decision-making by allowing marketers to rely on experimental evidence rather than intuition. This can help refine strategies and improve overall effectiveness. However, certain decisions, such as offering discounts, require careful consideration, as higher sales volumes may come at the cost of reduced profit margins. In such cases, combining controlled experiments with automated optimization and real-time data analysis can strengthen decision-making, especially when guided by clear evaluation criteria (Gallo, 2017).

Despite the increasing interest in A/B testing, existing review studies tend to focus on specific elements, such as experimental design or testing techniques, and often lack a comprehensive, systematic perspective. This leaves important questions insufficiently addressed, including what exactly is being tested, who is involved in the process, and how results are applied in real-world settings (Quin et al., 2024).

For online businesses, evaluating marketing effectiveness is essential for maintaining a competitive edge. Companies commonly rely on key performance indicators such as conversion rate, number of orders, cart abandonment rate, bounce rate, time spent on site, and unique visitors to better understand customer behavior and sales outcomes (Esteller-Cucala et al., 2019). These metrics provide clear signals about whether marketing efforts are working, and A/B testing offers a structured way to generate and interpret this evidence.

2.5 Characteristics of A/B Testing

This study focuses mainly on three major characteristics that are highly cited in the academic literature: A/B Testing Variation Design Attractiveness, A/B Call-To-Action Effectiveness, and A/B Audience Segmentation Differentiated Effectiveness.

2.5.1. A/B Testing Variation Design Attractiveness

Mandić et al. (2023) examined how A/B testing can be used to evaluate online sales performance, particularly in identifying which user interface (UI) changes most effectively improve key performance indicators. Their results show that design changes do not always lead to better outcomes; in some cases, such as in their first experiment, the original website version actually performed better and was worth retaining. However, other experiments (specifically the second, fourth, and fifth) demonstrated that carefully targeted UI modifications can significantly improve performance metrics. Consistent with earlier studies (Martín et al., 2021; Kern, Hoek, & Hienert, 2023; Bleier & Eisenbeiss, 2015), these findings highlight the importance of personalization in e-commerce and digital advertising. In a similar vein, a clear gap remains in understanding how

personalized UI design directly translates into improved online sales performance. The authors also note that the third experiment emphasized the importance of repeatedly testing the same interface element to fully understand its true effect.

Designing effective websites requires attention to multiple interrelated factors. Khlaisang (2017) identifies five key dimensions: web technology, content, design and presentation, interaction, and creativity. Common interface components include logos, search bars, language settings, login features, navigation menus, banners, and interactive tools such as modal windows and accordion panels (Li et al., 2022). The behavior of these elements, their size, color, placement, and information density strongly influence user experience and emotional responses (González-Mena et al., 2022). In addition, interface flow plays an important role in ensuring that users can move smoothly through a website (Cai et al., 2018). A well-designed e-commerce platform must also combine user-facing features with back-end systems that support overall business operations.

Prior research consistently emphasizes usability and user-centered design as core principles. Effective interfaces should align with users' cognitive processes and feel intuitive and effortless to use (Chen & Soo, 2018; Lu et al., 2011). Eye-tracking and behavioral studies further show that visual design elements such as layout, color, and feature placement significantly affect how quickly and efficiently users complete tasks (Zhou et al., 2022; Foris et al., 2020; Reyes et al., 2020). However, designing interfaces for complex systems with large volumes of user interaction data remains a significant challenge (Więckowska & Rudnicka, 2023).

A user-centered design approach is therefore essential. Sukendar and Susena (2022) emphasize the importance of understanding user characteristics, goals, and preferences when designing interfaces. Effective websites should be simple, consistent, and visually balanced, with careful attention to typography, layout, and readability. They should also provide clear feedback to users regarding actions, errors, and system status (Miraz et al., 2016). Similarly, Swasty and Adriyanto (2017) argue that websites must be both visually appealing and easy to navigate in order to attract and retain users. Beyond usability, successful platforms should also deliver an engaging and distinctive user experience,

while avoiding excessive visual complexity that could overwhelm users (Bortko et al., 2020).

Interactive design features further enhance engagement. For example, hover effects can improve users' perceived ease of use and usefulness in online shopping environments (Chen & Zhai, 2023). Usability factors such as navigation, search functionality, and browsing structure are also critical. Clickable elements should be clearly recognizable through visual cues such as color, size, shape, and placement (Bartoš & Habarta, 2019). Likewise, well-written link text that is clear and action-oriented can significantly influence user behavior, while familiar conventions such as using blue for hyperlinks can further improve usability. Optimizing these elements is essential for improving conversion rates and overall website effectiveness.

From a design perspective, key principles such as clarity, structured layout, visual hierarchy, appropriate use of color, and intuitive navigation are central to creating effective web experiences (Swasty & Adriyanto, 2017). In e-commerce contexts, elements such as branding, textual content, background design, and even motion graphics play an important role in shaping user emotions and engagement during browsing. Despite these advantages, A/B testing still faces several limitations and practical challenges.

2.5.2. A/B Call-To-Action Effectiveness

In recent years, marketing research has increasingly shifted toward studying consumer behavior directly on digital platforms, using built-in A/B testing tools provided by companies such as Facebook and Google Ads. These approaches enable researchers to observe real user behavior in natural settings, for example, whether individuals click on an advertisement or complete a purchase rather than relying on artificial laboratory environments. As a result, the findings tend to be more realistic and practically relevant. However, a key limitation is that these platforms do not always ensure true random assignment of users to experimental conditions, which complicates causal interpretation (Boegershausen et al., 2025).

Because of this, research conducted on online platforms involves an inherent trade-off.

On one hand, it offers strong ecological validity and access to large, diverse populations. On the other hand, it often lacks the level of experimental control required for strong causal inference (Boegershausen et al., 2025). Since researchers depend on platform-based ad delivery systems, hidden allocation mechanisms may introduce bias and weaken internal validity. Consequently, scholars have called for greater transparency from major platforms such as Google and Meta, as well as improved tools that allow genuine randomization in ad assignment (Boegershausen et al., 2025; Braun et al., 2024). In addition, practical challenges remain, including difficulties in defining appropriate evaluation metrics for A/B testing, which can limit the effectiveness of experimental insights (Esteller-Cucala et al., 2019). Another constraint is the need for very large sample sizes to achieve reliable results, sometimes ranging from 500 to 250,000 users per condition, which is not always feasible for all organizations (Zhou et al., 2023).

Digital behavior research is also shaped by the sheer amount of time individuals spend online. On average, people spend about 6 hours and 58 minutes per day engaged in internet-connected activities (We Are Social, 2022). As a result, marketing research has increasingly followed consumer attention into digital environments. Facebook A/B testing one of the most widely used forms of platform-based experimentation, typically involves running advertising campaigns with multiple ad variations shown to mutually exclusive user groups within newsfeeds on Facebook or Instagram. Researchers then compare performance using metrics such as click-through rate (CTR), which measures the likelihood that users exposed to an ad click on a call-to-action such as “shop now,” “learn more,” or “contact us.”

Similarly, Google Ads-based experiments test different display ads (such as banner ads across websites) or sponsored search results that appear when users enter specific queries. In these studies, different ad versions are shown to separate user groups at the same time, and CTR is commonly used as the main outcome measure (Boegershausen et al., 2025). Overall, these platform-based experiments provide valuable behavioral insights, but their methodological limitations highlight the need for more controlled and transparent experimental designs in digital marketing research.

2.5.3. A/B Audience Segmentation Differentiated Effectiveness

Guha et al. (2021) emphasize that customers vary widely in their needs and purchasing behaviors, which highlights the importance of A/B testing as a tool for more precise market segmentation. By running controlled experiments, marketers can identify high-value and loyal customer groups, better understand their purchasing patterns, and estimate customer lifetime value. These insights ultimately help improve the targeting and effectiveness of marketing campaigns (Guha et al., 2021).

In digital advertising environments, targeting systems are often designed to maximize the likelihood of achieving specific campaign goals, such as clicks or conversions. Importantly, this optimization is typically applied separately to each advertisement, resulting in what is known as “divergent delivery,” where different ads are shown to different user groups based on predicted responses (Braun & Schwartz, 2024; Johnson, 2023). As a consequence, the audiences exposed to each ad may differ substantially in characteristics such as demographics, interests, or browsing behavior. This creates an important methodological challenge: differences in performance between ads may not reflect the true effectiveness of the creative content alone, but rather differences in the underlying audiences. This issue can weaken internal validity and make causal interpretation more difficult (Eckles et al., 2018). For instance, an ad highlighting environmental benefits may appear more effective simply because it is shown more often to users who are already environmentally conscious.

User-generated engagement signals further complicate this environment. Reactions such as likes, hearts, angry responses, and user comments are visible to others and can influence subsequent user behavior, including click-through and conversion rates (Agarwal et al., 2024). In this context, Braun et al. (2024) argue that many online platform studies labeled as A/B tests are more accurately described as “multiple-ad studies without holdout groups,” since algorithmic systems continue to shape exposure even after initial user assignment.

Although the lack of strict randomization and the presence of divergent delivery are widely recognized as threats to internal validity (Braun & Schwartz, 2024; Eckles et al.,

2018; Johnson, 2023), there remains debate about how severely these issues distort findings. Some researchers suggest that algorithmic targeting does not fully explain observed differences in consumer responses, particularly when examining moderation effects (Matz et al., 2018). However, due to limited transparency in platform algorithms, it is often difficult to verify whether proposed corrective approaches, such as optimizing for reach or applying moderation-based designs fully address these biases.

Despite these concerns, targeting algorithms are now an unavoidable feature of modern digital advertising systems (Braun & Schwartz, 2024). In practice, advertisers depend on these mechanisms for efficient audience selection and campaign optimization, meaning that divergent delivery reflects real-world advertising conditions. From this perspective, platform-based experiments may be seen as highly ecologically valid, since they capture how advertising actually functions in practice. However, they also measure the combined effect of both advertisement content and algorithmic targeting, rather than isolating the pure causal impact of the ad itself. Because post-randomization optimization reduces experimental purity, these methods have sparked ongoing debate about whether they should be considered true experiments or interpreted more cautiously (Braun et al., 2024; Braun & Schwartz, 2024; de Langhe & Puntoni, 2021). As their use continues to expand, there is a growing need to better understand their limitations and appropriate applications. In typical platform-based A/B testing systems, users are first randomly assigned to mutually exclusive groups exposed to different ad variants (e.g., ad A or ad B). Within each group, machine learning algorithms may further optimize delivery by selecting users most likely to respond to a given variant (Braun & Schwartz, 2024). In addition, within the broader literature on A/B testing in interface design, several studies highlight how specific design features significantly influence user behavior and performance outcomes (Esteller-Cucala et al., 2020; Jylhä & Hamari, 2020; Jankowski et al., 2019; Lestari et al., 2019; Chen & Zhai, 2023).

2.6. Data Analytics and Interpretation in A/B Testing

In today's digital marketing environment, understanding and shaping consumer behavior has become increasingly complex (Kumar, 2021). While traditional analytical methods

still offer useful insights, they often fall short in capturing the level of detail required in modern, data-driven settings. This has led to the growing importance of A/B testing as a core tool in marketing analytics. As Siroker and Koomen (2015) explain, A/B testing provides a structured way to experiment with different marketing strategies and has become one of the most effective approaches for turning user engagement into actual conversions. By comparing alternative versions of digital assets such as webpages, emails, or app interfaces, firms can identify which option performs best in achieving specific goals like clicks, sign-ups, or purchases.

The reliability of A/B testing, however, depends on the use of appropriate statistical methods. Sound data analysis is essential for drawing valid conclusions from experimental results (Kumar, 2021; Ott & Longnecker, 2016). In addition, a strong experimental design is necessary to reduce bias and control for confounding factors that might otherwise distort the findings (Viglia et al., 2021). As digital marketing continues to evolve, maintaining this level of methodological rigor becomes increasingly important for producing trustworthy insights.

In practice, marketers frequently rely on online advertising platforms to compare how users respond to different types of ad content (Braun et al., 2025). However, a key limitation is that these platforms do not expose all users to identical conditions. Instead, ads are delivered to distinct and continuously optimized audience segments, which often differ across experimental variants even within the same A/B test. As a result, observed differences in outcomes reflect not only the effect of the ad itself, but also the influence of algorithmic targeting and user selection. This introduces a challenge for causal interpretation, as the estimated effects may differ from those that would emerge under fully random assignment.

Despite these limitations, A/B testing remains a highly valuable method for optimizing digital products and services and for gaining deeper insights into user behavior. Conducting a robust A/B test requires a clearly defined hypothesis, careful planning, and rigorous statistical evaluation. Key steps include selecting appropriate primary metrics, ensuring adequate sample size, and performing validity checks to confirm the reliability

of results. Interpreting outcomes also requires considering statistical significance, effect size, behavioral patterns, practical relevance, and replicability. Ultimately, decisions about whether to adopt a tested variation should be based on a balanced assessment of these factors, along with business constraints and potential risks (Agarwal, 2023).

From a statistical standpoint, several elements are critical to successful A/B testing. One of the most important is determining an adequate sample size to ensure sufficient statistical power while avoiding unnecessary cost or effort (Guo et al., 2013). Researchers must also distinguish between statistical significance and practical significance, as statistically meaningful results do not always translate into meaningful business impact. As Viglia et al. (2021) highlight, well-designed experiments not only support better managerial decision-making but also contribute to the broader development of marketing analytics as a field.

2.7 A/B Testing and Firm Performance

Existing literature suggests that taking an experimental approach to strategy can meaningfully improve startup performance (Camuffo et al., 2020; Gans, Stern & Wu, 2019). In this context, A/B testing has been associated with measurable gains, including an approximate 10% increase in startup page visits (Koning et al., 2019). Evidence from both researchers and practitioners also shows that A/B testing tools are not limited to small, isolated experiments. Instead, they can support more advanced experimentation strategies that help firms evaluate substantial product changes, not just minor adjustments. As a result, A/B testing contributes to organizational learning, encourages product innovation, and ultimately enhances firm performance (King, Goldfarb & Simcoe, 2019). Empirical findings further indicate that firms using A/B testing introduce new products at rates that are 9% to 18% higher than those that do not, and they are also more likely to experience extreme performance outcomes, both very low and very high levels of page visits reflecting a greater degree of experimentation and learning (Koning et al., 2019). Overall, experimentation strengthens organizational learning processes, which in turn supports improved firm performance (Koning et al., 2019).

Although relatively few firms systematically adopt A/B testing, those that do often report

substantial performance improvements, sometimes ranging from 30% to 100% within a year. Both qualitative and quantitative evidence suggest that experimentation helps startups develop and refine new products more effectively, scale successful ideas, and quickly discontinue underperforming ones. In today's data-driven environment, many organizations are therefore increasingly turning to optimization techniques such as A/B testing to improve decision-making and outcomes (Agarwal, 2023). Historically, experimentation was relatively costly compared to learning from experience. However, advances in digital technologies and the widespread availability of A/B testing platforms have significantly reduced the cost and complexity of running experiments (Kohavi, Henne & Sommerfield, 2007; Kohavi et al., 2009; Kohavi & Longbotham, 2017; Azevedo et al., 2020). This has enabled firms of all sizes to conduct frequent and rapid experiments to guide strategic decisions. The trend has also been reinforced by accelerators, venture capitalists, and entrepreneurs, who increasingly encourage systematic testing across business functions. A key element of this process is selecting an appropriate primary or success metric that aligns with the firm's strategic objectives (Sheng et al., 2023). For instance, revenue-oriented goals may focus on total sales or revenue per user, while engagement-focused goals may rely on metrics such as time on site or page views per session.

Despite these advantages, it remains unclear whether A/B testing consistently leads to stronger organizational learning and improved firm performance. For experimentation to generate real value, firms must not only run tests but also generate meaningful alternatives and ensure that data-driven insights actively inform decision-making (Brynjolfsson & McElheran, 2016). Moreover, the effectiveness of A/B testing may depend on the nature of the ideas being evaluated. Some scholars argue that A/B testing primarily drives incremental improvements, such as small interface changes like button color or text size (Kohavi et al., 2007; Azevedo et al., 2020; Deniz, 2020). While such incremental optimization can improve efficiency, it may also limit a firm's ability to recognize and respond to broader structural changes in the market.

In contrast, other research suggests that A/B testing can also enable more radical

innovation, including the launch of new products, the implementation of recommendation systems, and even the development of new business models (Luca & Bazerman, 2021). In addition, the low cost of experimentation may foster a culture of continuous testing, which can support more transformative innovation over time (Thomke, 2020). For example, firms can decompose large and risky changes such as a full product redesign into smaller, testable components, thereby reducing uncertainty and improving decision quality. Against this background, the present study investigates whether A/B testing contributes to improved online sales performance, with a specific focus on firms operating in Saudi Arabia.

2.8 Research Gaps

A review by Boegershausen et al. (2025) identified 133 studies across 99 published articles that used A/B testing features on major advertising platforms, with most published in marketing journals. The increasing popularity of this approach is largely due to its ease of implementation, its ability to capture real consumer behavior in natural settings (Morales et al., 2017), and its usefulness in tracking conversion outcomes across different stages of the purchase funnel (Colicev et al., 2019). Despite these advances, an important gap remains in the literature: relatively few studies have examined how content attributes perform in non-Western contexts. Most existing research is concentrated in Western markets, which may differ significantly in cultural values and consumer behavior compared to regions such as the Middle East. In this regard, Saudi Arabia represents a particularly relevant context for understanding how cultural factors shape digital engagement. In addition, limited research has explored how combinations of content attributes such as tone and visual design jointly influence user behavior. This study addresses this gap by applying conjoint analysis and multivariate testing.

Most prior research has relied on quantitative methods such as A/B testing, regression analysis, and conjoint analysis to examine user engagement. Although there are relatively few large-scale studies systematically evaluating the organizational benefits of A/B testing across firms, both academic and practitioner-oriented literature emphasize its importance as an internal decision-making tool (Kohavi, Henne & Sommerfield, 2007; Kohavi et al., 2009; Kohavi & Longbotham, 2017; Xu et al., 2015). For example, online retailers frequently experiment with pricing strategies, product bundles, and display formats to

improve performance (Dubé et al., 2017; Sahni, Zou & Chintagunta, 2016), while digital platforms test features such as recommendation systems, social interactions, and content variations to enhance user engagement (Kumar & Tan, 2015).

Azevedo et al. (2018) highlight the growing role of A/B testing and causal inference in shaping modern managerial decision-making. Similarly, researchers across fields such as entrepreneurship, corporate strategy, criminology, and international law increasingly advocate for experimental approaches to evaluate competing policies and strategies (Chatterji et al., 2016). Consistent with the broader shift toward data-driven decision-making (Brynjolfsson & McElheran, 2016), A/B testing has become a widely adopted managerial practice across industries. However, its effectiveness depends strongly on the rigor of data analysis. Studies by Larsen et al. (2022) and Masoero et al. (2023) show that issues such as low statistical power, omitted variables, and analytical errors can distort results and lead to suboptimal decisions. Real-world cases of failed experiments (e.g., “Where A/B Testing Goes Wrong,” 2024) further demonstrate how flawed analysis or biased test delivery can undermine expected performance gains. In contrast, organizations that apply robust statistical methods and evidence-based interpretation are more likely to achieve reliable and meaningful outcomes, consistent with Evidence-Based Management principles (Kurz et al., 2025).

Taken together, the literature positions A/B testing as both a technological capability and a managerial practice that can enhance firm performance when supported by strong analytical rigor. Its effectiveness particularly in relation to factors such as design attractiveness, call-to-action effectiveness, and audience segmentation depends largely on how accurately firms interpret and act on experimental data. Building on this foundation, the present study examines the impact of A/B testing characteristics on online sales performance in Saudi Arabia, while also investigating the moderating role of data analysis accuracy. In doing so, it integrates technological, analytical, and managerial perspectives into a unified framework.

Overall, A/B testing is a powerful tool for data-driven decision-making and marketing optimization. For instance, firms launching new websites often test alternative designs, layouts, or content to identify versions that generate higher engagement and conversion rates. Similarly, marketing campaigns can be refined by experimenting with different messages, offers, or calls to action to determine the most effective approach (Agarwal, 2023). Importantly, these experiments should account for the entire user journey from

initial exposure to final conversion since design variations may influence different stages of the customer funnel in distinct ways (Agarwal, 2023; Sheng et al., 2023).

Table 2.1. Summary of Related Studies on A/B Testing							
Study	Research Focus	Data Source	Theoretical Lens	Major Findings	Theme: A/B & Online sales	Multi-Methods Analysis	Scope: Saudi Arabia
Braun and Schwartz (2025)	Factors affecting the validity of A/B testing.	Experimental and platform Data	Causal Inference Theory	The validity of A/B testing is influenced by nonrandom exposure, divergent delivery, causal inference challenges, and platform incentives.	N. A	N. A	N. A
Mandić et al. (2023)	Impact of A/B testing on improvement of key performance indicators	Experimental data	controlled experiment theory	The performance indicators can be improved by the existing version of the website should be kept, or changes to a website's user interface.	N. A	N. A	N. A
Masoero et al. (2023)	The use of covariate-adjusted estimators to improve the precision of causal effect estimates in online A/B testing	Real-World A/B Testing Data and Covariates	Causal Inference and Statistical Estimation	Provides a scalable framework for implementing covariate-adjusted estimators in real-world online A/B testing platforms.	N. A	N. A	N. A
Cruz-Benito et al. (2018)	Employing A/B testing to evaluate different versions of adaptive web forms	User Interaction Data and A/B testing Data	Human-Computer Interaction	Adaptability can be achieved by detecting users' behaviors, preferences, and profiles using A/B testing.	N. A	N. A	N. A
Esteller Cucala et al. (2020)	Identifying common mistakes and challenges in designing, implementing, and interpreting A/B tests.	A/B testing data and Simulated Data	Causal Inference and Experimental Design	A comprehensive list of common A/B testing errors in the context of online personalization	N. A	N. A	N. A
Kohavi and Longbot ham (2017).	The role of A/B testing in evaluating the impact of changes to websites, apps.	Experimental data and case studies.	Causal Inference and Statistical Theory	provide valuable resource to implementing A/B testing in organizations	N. A	N. A	N. A
Koning et al. (2022)	Experimentation as entrepreneurial strategy and the impact of A/B testing adoption	Startup Performance and A/B Testing Adoption	Entrepreneurial Strategy and Causal Inference	Provides empirical evidence on how A/B testing impacts startup performance.	N. A	N. A	N. A

This study	A/B Testing and firms' Online sales	Qualitative: Semi-structured interviews and Quantitative: Surveys	TOE EBM	Effective design variations, clear call-to-action, and personalized audience segmentation can significantly boost sales, while accurate data analysis enhances these effects.	Yes	Mixed Method: Interviews & Survey.	Yes
------------	-------------------------------------	---	---------	---	-----	------------------------------------	-----

Table 2.1: Summary of Literature Review

Recent studies have also pointed to several important directions for future research, particularly the need to strengthen statistical methods, improve the overall A/B testing process, and expand the use of automation in experimentation (Quin et al., 2024). In practice, A/B testing is now widely embedded in the operations of major technology companies such as Google, Meta, LinkedIn, and Microsoft (Kohavi & Thomke, 2017; Gupta et al., 2019). Within these organizations, segmentation plays a central role, as it helps marketers better understand diverse customer needs and design more targeted campaigns for high-value user groups (Guha et al., 2021).

In a similar vein, interpreting A/B test results requires more than simply identifying statistically significant differences. It involves a careful evaluation of multiple dimensions, including statistical significance, effect size, patterns of user behavior, practical business relevance, and the replicability of findings (Agarwal, 2023). Together, these considerations ensure that experimental results are not only statistically sound but also meaningful and actionable in real-world decision-making.

Chapter 3: Research Model and Hypotheses Development

3.1 A/B Testing Variation Design Attractiveness and Online Sales

According to Mohammad et al. (2024), visual design in marketing advertisements has a positive effect on key user engagement metrics such as click-through rate, conversion rate, and bounce rate. A broader stream of research on content marketing also highlights how different design elements shape user behavior. For example, Lee and Koubek (2010) show that web design features are closely linked to user experience, noting that minimalist and visually balanced layouts are generally associated with lower bounce rates and higher conversion rates. In a similar vein, Kim and Kuljis (2007) find that culturally relevant messaging tends to perform better in non-Western contexts, as users respond more positively to content that aligns with their cultural expectations. Despite the value of these contributions, most existing studies are still largely concentrated in Western markets, leaving a clear gap in understanding how content attributes influence user behavior in the Middle East, particularly in Saudi Arabia.

In digital marketing practice, elements such as headlines, message tone, and visual presentation are widely used to enhance content effectiveness and attract user attention (Jeganathan & Szymkowiak, 2020). Message tone whether formal, informal, or culturally adapted plays an especially important role in shaping engagement outcomes. Aririguzoh (2022) demonstrates that culturally aligned messaging is more effective in non-Western markets, where users tend to engage more strongly with content that reflects their values and norms. Supporting this view, evidence from Jordan shows that localized message tone and calls-to-action can significantly improve conversion rates (Benmamoun et al., 2019). Similarly, Ahmadi et al. (2022) find that culturally tailored headlines and visuals increase the likelihood of user sharing on platforms such as Facebook and Instagram, reinforcing the importance of culturally sensitive digital marketing strategies.

Building on this evidence, the present study argues that design attractiveness plays a key role in improving online sales performance when compared to original or less optimized versions. More specifically, improved design variants are expected to better capture user attention and encourage desired actions such as clicks, add-to-cart behavior, and completed purchases. This is achieved through more appealing layouts, engaging color

schemes, optimized calls-to-action, and enhanced user experience features. When users interact with a more visually appealing and user-centered design, their likelihood of engaging positively with the website increases, leading to higher conversion rates and improved sales performance compared to the baseline design.

H1: A/B testing variation design attractiveness is positively associated with online sales.

3.2 Call-to-Action Effectiveness and Online Sales

Strong and well-crafted calls-to-action consistently generate higher sales compared to weak or poorly designed ones, a pattern repeatedly confirmed through A/B testing. By experimenting with different elements such as wording, placement, color, size, and urgency businesses can identify which combinations are most effective in capturing attention and prompting users to act. Recent research positions A/B testing as a core tool for data-driven decision-making and performance optimization in digital commerce (Kurz et al., 2025; Quin et al., 2024). Firms increasingly rely on controlled experiments to test variations in design, call-to-action, and audience targeting strategies in order to improve engagement and conversion outcomes.

There is also growing recognition that A/B testing is more than just a tactical marketing tool; it is a strategic capability. For instance, Koning, Hasan, and Chatterji (2022) show that firms with structured experimentation practices tend to achieve higher growth and stronger innovation performance. Similarly, systematic reviews (Quin et al., 2024) find that A/B testing can significantly improve key performance metrics such as conversion rates, purchase frequency, and revenue, particularly when it is embedded within continuous learning systems. Supporting this, Vafainia et al. (2022) report that effective call-to-action in direct marketing campaigns positively influence customer purchasing behavior. Taken together, these findings suggest that the ability to design and implement A/B testing effectively through appealing design variations, persuasive call-to-action, and targeted segmentation can lead directly to measurable improvements in online sales.

Similarly, there are important methodological challenges to consider. Braun and Schwartz (2025) highlight issues such as non-random exposure, uneven delivery of test variants, and platform-driven biases, all of which can complicate causal interpretation and affect

the reliability of results. These limitations make careful experimental design and thoughtful interpretation essential. Despite these concerns, empirical evidence continues to show meaningful benefits. For example, Koning et al. (2022) find that A/B testing is associated with a 5-15% increase in website visits, along with improvements in product innovation and development processes.

Building on this evidence, the present study argues that effective call-to-action plays a central role in driving user action. Strong call-to-action clearly communicate value, create a sense of urgency, and guides users smoothly toward completing desired actions, such as making a purchase. Findings from A/B testing consistently demonstrate that websites with clear, compelling call-to-action outperform those with vague or passive prompts, leading to higher click-through rates, stronger engagement, and ultimately greater sales conversions.

H2: Call-to-Action Effectiveness is positively associated with online sales.

3.3 Audience Segmentation Differentiated Effectiveness and Online Sales

Many online advertising platforms, such as Meta and Google, provide built-in tools that allow marketers to test how users respond to different advertisements. In theory, these tools appear to randomly assign users to different ad versions. In practice, however, ad exposure is not purely random in a way that allows researchers to isolate the true causal effect of ad content from the influence of platform algorithms (Braun et al., 2025). This is largely because ad delivery is shaped by underlying auction dynamics (Lambrecht & Tucker, 2019) and by platform-specific relevance scores that adjust bids and determine which ads are shown. Since the exact mechanisms behind these algorithms are not fully transparent and may not even be fully reproducible the resulting outcomes reflect both ad content and algorithmic targeting (Gordon et al., 2019).

Closely connected to this is the concept of personalization, which refers to tailoring web content and design to deliver the right message to the right user at the right time (Sundar & Marathe, 2010; Ghorab et al., 2013; Krouska et al., 2022; Ismail et al., 2023). Personalization helps businesses meet individual user needs, improve competitiveness, and manage information more effectively (Desai, 2021).

Well-designed personalization strategies can significantly enhance user experience and increase the likelihood that users return to a website. In practice, personalized websites adapt elements such as content structure, interface design, and navigation based on user profiles and behavioral data (Bunt et al., 2009). Techniques such as targeted banners, wish lists, and customizable interfaces have also been shown to support higher sales performance (Heinz et al., 2017). Similarly, researchers face a methodological dilemma when studying consumer behavior in these environments. As Braun et al. (2023) note, there are situations where researchers may accept some imbalance in order to study effects within realistic advertising conditions, where content and targeting are naturally intertwined. However, when the goal is to isolate the impact of ad content itself, maintaining balance across experimental conditions becomes essential. Some have suggested that certain platform settings such as optimizing campaigns for “reach” instead of “conversions” might reduce bias. Yet, available evidence suggests that these approaches do not fully eliminate imbalances in how ads are delivered across user groups. Because targeting algorithms continuously adjust exposure based on perceived relevance, claims that platform configurations can guarantee balanced experiments remain uncertain. Alternative strategies, such as creating separate audience groups for different campaigns or tests, have also been explored. However, prior research shows that these approaches do not fully resolve the underlying challenges and may still produce biased results (Eckles et al., 2018). Despite these limitations, audience segmentation remains a powerful tool in digital marketing. By dividing users into groups based on demographics, behavior, past purchases, or interests, firms can tailor their messages and offers more effectively. This targeted approach allows businesses to create more personalized experiences that resonate with specific segments, leading to higher engagement and improved conversion rates. In fact, evidence from A/B testing consistently shows that strategies tailored to segmented audiences outperform one-size-fits-all approaches. Firms that customize their marketing efforts based on user characteristics tend to achieve stronger sales performance, build better customer relationships, and increase the overall relevance of their offerings. While methodological challenges remain, combining A/B testing with thoughtful segmentation

continues to be a highly effective approach for optimizing online marketing outcomes.

H3: *Audience segmentation differentiated effectiveness is positively associated with online sales.*

3.4 The moderating Role of A/B testing data analysis Accuracy

Accurate data is at the heart of effective A/B testing in digital marketing, as it allows firms to draw reliable conclusions about consumer behavior and campaign performance. The process typically begins by deciding what to test, followed by determining how success will be measured. Users are then randomly assigned to two versions of a webpage or feature: version A (the control), which reflects the current design, and version B (the treatment), which includes a specific modification. By comparing how users interact with each version using relevant performance metrics, marketers can identify which option better supports their objectives.

According to Kohavi et al. (2009), if one version shows a statistically significant improvement, it is adopted, while the other is discarded. However, A/B testing can be time-consuming often requiring several weeks, which increases both cost and complexity (Masoero et al., 2023). In addition, Esteller-Cucala et al. (2019) highlight a common challenge: many firms lack a clear framework for selecting appropriate evaluation metrics. Without well-defined measures, the value of testing can be undermined. When done properly, however, accurate data analysis reduces uncertainty and helps marketers identify which elements such as ad copy, design features, or calls-to-action actually drive conversions, ultimately improving return on investment.

More importantly, advanced data analysis strengthens the link between design improvements and online sales performance. Rather than relying only on surface-level metrics, deeper analysis allows firms to understand how specific visual and functional changes influence user engagement, trust, and purchasing decisions. In this way, A/B testing becomes not just a tool for comparison but a way to generate meaningful insights that support more effective and evidence-based decision-making.

H4a: *A/B testing data analysis strengthens the relationship between A/B Testing Variation Design Attractiveness and online sales.*

A/B testing is often seen by marketers as a straightforward way to run real-world experiments on ad copy, visuals, and messaging. However, in targeted online advertising environments, the reality is more complex. As Braun et al. (2025) point out, these experiments differ from traditional randomized controlled trials because researchers do not fully control who sees which version of an ad. When using platform-based tools, subject selection and treatment assignment are influenced by the platform's own algorithms, which have raised concerns across multiple fields (Johnson, 2023).

This becomes especially important when marketers try to generalize findings beyond the specific platform where the test was conducted. Because ad delivery is shaped by relevance-optimization algorithms, the results reflect not only the effect of the ad content but also the platform's targeting logic. As a result, these tests should not be treated as clear causal evidence of what works universally. Instead, analytics teams need to be transparent about these limitations and clearly communicate that A/B test outcomes in such environments are context-dependent and potentially confounded (Braun et al., 2023).

Similarly, improving the quality of A/B testing data analysis can significantly strengthen insights, particularly when evaluating call-to-action effectiveness. Advanced analytical approaches allow firms to move beyond surface-level metrics and better understand how specific call-to-action elements, such as wording, placement, color, or urgency, influence user behavior and decision-making. By carefully accounting for confounding factors and ensuring statistical validity, firms can more accurately isolate the impact of call-to-action on engagement and conversions.

Ultimately, stronger data analysis enables organizations to continuously refine their marketing strategies and validate the effectiveness of optimized call-to-action. This not only improves short-term campaign performance but also supports sustained growth in online sales by ensuring that decisions are grounded in reliable, evidence-based insights.

H4b: A/B testing data analysis strengthens the relationship between call-to-action effectiveness and online sales.

Scholars recommend caution in the interpretation of results from online platform studies and, in particular, in the interpretation of differences in Click-Through Rate (CTRs)

across ads (Braun & Schwartz, 2024). Indeed, statistically significant differences in CTRs across ads are not necessarily managerially meaningful. Managerial meaningfulness means that CTRs need to be large enough to affect “real” mid and lower-funnel outcomes like completing a tutorial or buying a product (Lin et al., 2013). Indeed, the page views resulting from users’ clicks on an ad are merely the starting point of an extended online purchase funnel with significant drop-offs in later stages (Colicev et al., 2019; Gordon et al., 2023). Indeed, buyers are a small subset of people who clicked on the ad, who are themselves a small subset of people exposed to the ad. Instead, it is critical to provide a business interpretation of the value of a click (Johnson, 2023).

Deeper insights into how various customer segments react to customized messaging, offers, and experiences are provided by improved A/B testing data analysis, which increases the impact of audience segmentation differentiated effectiveness in increasing online sales. Businesses can discover hidden trends, pinpoint which versions are most effective for particular market segments, and spot minute behavioral deviations that may go unnoticed with simple reporting by using sophisticated analysis. Businesses may maximize their segmentation strategy thanks to this accuracy, which guarantees that each target group receives the most compelling and pertinent information. Stronger data analysis, hence increases the efficacy of segmentation, resulting in improved customer satisfaction, increased engagement, and eventually improved online sales success.

H4c: A/B testing data analysis strengthens the relationship between call-to-action effectiveness and online sales.

Theoretical Framework and Research Model

Based on the above-mentioned theoretical framework including TOE and EBM, and research objectives, the study's conceptual framework was developed. The proposed research model has 5 variables; 3 variables are independent 1 variable is moderator. Online sales performance is dependent variable. The independent variables represented by A/B testing characteristics influence firms’ online sales.

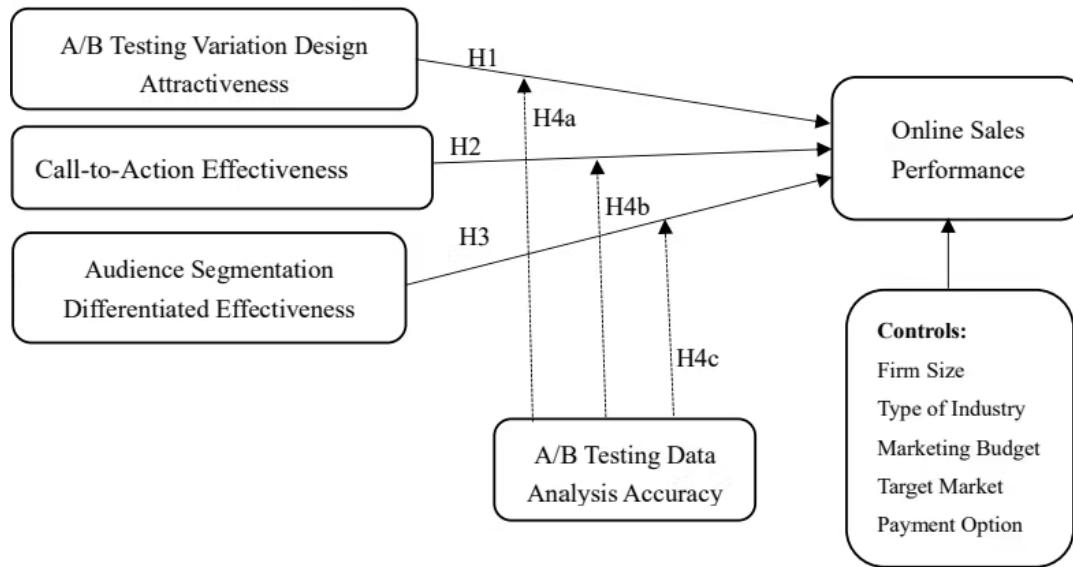


Figure 3.1: Research Model

Table 3.1 basically shows a simple story of how A/B testing leads to better online sales in Saudi Arabia using the TOE framework. On the technology side, it's about the actual tools and features like good design, clear call-to-action messages, smart audience targeting, and strong data analysis which all help improve sales, especially when the data is interpreted correctly. On the organizational side, it highlights that firms also need the right people, skills, resources, and a culture that supports using data, otherwise even good tools won't be used effectively. The environmental side reminds us that competition, digital growth, and changing customer expectations in Saudi Arabia push companies to keep improving and adopting these strategies. In short, the results show that better online sales don't come from one thing alone, but from the combination of strong tools, capable firms, and a competitive market that encourages continuous improvement.

Table 3.1 Relevance of Theory to Research Model

Theory	Dimension	Study Constructs	Role in the Study	Link to Hypotheses
	Technology	A/B Testing Variation Design Attractiveness, Call-to-Action Effectiveness, Audience Segmentation, Data Analysis Accuracy	Represents the firm's use of digital tools and analytics to design, test, and optimize customer interactions	Directly supports H1, H2, H3 (main effects); A/B testing data analysis accuracy strengthens these relationships in H4a, H4b, H4c

TOE	Organization	Firm capabilities (skills, expertise, data-driven culture, resources)	Enables effective implementation and interpretation of A/B testing and data insights	Explains how firms translate A/B testing features into improved sales; strengthens moderation effects (H4a-H4c)
	Environment	Market competition, digital adoption, customer expectations (Saudi context)	Drives the need for firms to adopt data-driven strategies and continuous experimentation	Reinforces the importance of A/B testing and data analysis in achieving a competitive advantage and higher sales
EBM	Best Available Evidence (Data & Analytics)	A/B Testing Data Analysis Accuracy	Ensures decisions are based on reliable, valid, and properly interpreted data from A/B tests	Strengthens relationships in H4a, H4b, H4c by improving decision quality
	Practitioner Expertise	A/B Testing Variation Design Attractiveness, Call-to-Action Effectiveness, Audience Segmentation	Reflects managers' ability to design appealing interfaces, craft persuasive messages, and target the right audience	Drives direct effects in H1, H2, H3 by applying expertise to optimize sales performance
	Organizational Context & Resources	Firm size, A/B testing budget, internal capabilities	Provides the support and infrastructure needed to run experiments and act on insights	Enhances the effectiveness of A/B testing implementation and outcomes
	Stakeholder Values & Market Needs	Customer preferences, engagement behavior, Saudi market dynamics	Ensures that A/B testing outcomes align with customer expectations and market demands	Reinforces why optimized designs, call-to-action, and segmentation lead to higher online sales

Additionally, Table 3.1 shows how Evidence-Based Management helps explain the role of A/B testing in improving online sales. It suggests that better decisions come from combining four key elements: reliable data, managerial expertise, organizational support,

and customer needs. Data analysis provides the evidence needed to make informed choices, while managers use their skills to design effective variations, craft strong calls-to-action, and target the right audiences. Similarly, firms need enough resources and a supportive environment to run tests properly and act on the results. Finally, understanding customer preferences and market conditions ensures that all these efforts lead to meaningful improvements in online sales.

Chapter 4. Research Methodology

4.1 Research Design

This study used a mixed research method. The qualitative part uses interviews to collect preliminary information about the nature and benefits of A/B testing in boosting company's online sales. The quantitative relies on the employment of convenience sampling to distribute self-administered questionnaires to collect data from Digital marketing managers, e-commerce managers and executives, online stores managers, data analysts' specialists in Saudi Arabia. The selection of convenience sampling techniques was based on the time and resource limitations and the nature of data required to undertake this research. The target population of this study includes: Digital marketing managers, e-commerce managers and executives, online stores managers, data analysts' specialists. They were chosen on the ground of their understanding, executing or influencing the A/B testing to boost the company's online sales.

Table 4.1. Summary of Research Methodology

Research Method	Mixed (Quantitative and Qualitative).
Sampling approach	Convenience Sampling
Target Population	Digital marketing managers, e-commerce managers and executives, online stores managers, data analysts' specialists.
Data collection method	Interviews and self-administered questionnaires
Sample size	Surveys: 250 valid surveys Interviews: 6 respondents
Data Analysis	Thematic Analysis (Interviews) SEM using Smart-PLS (Surveys)
Sample scope	Jeddah and Riyadh Cities.

The sample scope is confined to two major cities in Saudi Arabia: Jeddah and Riyadh. The choice of these two cities goes back to their suitability and validity as geographical scope for our research, as both cities are considered well-urbanized and have more companies employing A/B digital marketing testing. Further, this study collected 250 questionnaires to fulfil the research objectives and used Smart-PLS software to analyze the hypothetical relationships among the constructs. Table 4.1 summarizes the key points of the research methodology.

4.2 Target Population and Sampling

The target population for the qualitative phase consisted of professionals directly involved in digital marketing and e-commerce activities in Saudi Arabia. These individuals were selected due to their hands-on experience with A/B testing and online sales operations, enabling them to provide informed and practical insights for the exploratory stage of the study. A purposive sampling technique was employed to ensure that participants possessed the necessary expertise. Accordingly, six interviewees were selected, including digital marketing managers, e-commerce managers and executives, an online store manager, and a data analysis specialist. This approach ensured that the data collected were rich, contextually relevant, and closely aligned with the objectives of the qualitative inquiry.

For the quantitative phase, the target population comprised individuals actively involved in managing, designing, or analyzing A/B marketing tests within Saudi-based firms engaged in online sales. To ensure the relevance of responses, a non-probability purposive sampling method was again applied, focusing on participants with sufficient knowledge of A/B testing practices. A total of 250 valid responses were collected, meeting the recommended sample size requirements for Structural Equation Modeling (SEM) using SmartPLS. This sample size provided adequate statistical power to test the proposed research model and examine the relationships between A/B testing characteristics, data analysis accuracy, and online sales.

4.3 Data Collection

4.3.1 Exploratory Interviews

This study adopted an exploratory approach to gain deeper insights into the role and significance of A/B testing in driving firms' online sales performance. To achieve this, semi-structured interviews were conducted with six participants, including two digital marketing managers, two e-commerce managers, one online store manager, and one data analyst specialist. Depending on participants' availability and location within Saudi Arabia, the interviews were conducted either in person or online via Zoom or Microsoft

Teams. Each interview lasted approximately 30 to 45 minutes and was audio-recorded with participants' consent to ensure accurate transcription and analysis.

Participants were purposively selected based on their direct experience in designing, implementing, or evaluating A/B testing strategies in digital marketing contexts. Prior to each interview, participants were provided with a brief overview of the study's objectives and assured of confidentiality. The interviews followed a semi-structured format, using open-ended questions to explore key areas such as variation design, call-to-action effectiveness, and audience segmentation. This approach allowed for in-depth insights while maintaining flexibility for probing and follow-up questions. Following each session, key points were summarized, and participants were contacted when clarification was required. The collected data were then analyzed using thematic analysis to identify recurring patterns and insights related to how A/B testing practices influence online sales performance in the Saudi market. The target population for this phase included digital marketing professionals, e-commerce managers, performance marketing specialists, data analysts, and digital transformation leaders with hands-on experience in A/B testing within Saudi Arabia.

4.3.2 Questionnaire and Measures

This research study has 3 independent variables: A/B testing variation design attractiveness, call-to-action effectiveness, and audience segmentation differentiated effectiveness on the online sales while considering the moderating role of A/B testing data analysis accuracy in the context of Saudi Arabia. All constructs were measured using a 5-point Likert scale (1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree). Appendix 2 shows the components of the survey questionnaire. Table 4.2 shows the measurement scales for the study variables.

Table 4.2 Constructs and Measurements

Construct	Measurement Items	Source
A/B Testing Variation Design Attractiveness	1 The variations in A/B testing are well-designed to better align with our brand identity than the original. 2 The A/B testing variations are more visually appealing and user-friendly than the original. 3 The differences between A/B testing versions were clear and meaningful. 4 The A/B testing variation's technical performance and speed and technical are superior to the original. 5 The A/B testing variation's messaging is more persuasive in driving conversions than the original.	Kohavi & Longbotham, (2017).
Call-to-action Effectiveness	1. The call-to-action was clearer and more compelling than the original. 2. The call-to-action placement improves visibility and click-through rates compared to the original. 3. The call-to-action language was compelling and action-driven than the original. 4. The call-to-action design (color, size, wording) was more effective in increasing clicks than the original.	Vafainia et al. (2019).
Audience Segmentation Differentiated Effectiveness	1. Demographic segments (e.g., age/gender/location) 2. Behavioral segments (e.g., past purchases/page views) 3. Device-type segments (mobile/desktop/tablet) 4. Customer status segments (new vs. returning vs. lapsed users)	Guha et al. (2021).
Online Sales	1. Our online sales revenue has increased since implementing A/B digital marketing testing 2. We have observed a higher conversion rate from website visitors to paying customers after using A/B testing strategies 3. The number of completed transactions on our website has significantly improved due to A/B testing 4. A/B testing has contributed to a noticeable growth in repeat online purchases by customers 5. Higher overall revenue per visitor during the test period.	Mandić et al. (2023).
A/B Data Analysis Accuracy	1. The analysis of A/B testing results accurately reflected the differences between the A/B test variations. 2. The interpretation of A/B testing outcomes was precise and free from analytical errors. 3. The A/B data analysis correctly identified significant effects between experimental groups. 4. The conclusions drawn from A/B testing were based on accurate analytical procedures.	Ott and Longnecker (2016).

The A/B Testing Variation Design Attractiveness variable measures the efficacy of several design variations in A/B testing. Examples include: "The variations in A/B testing were well-designed to effectively compare different versions," and "The A/B testing design assisted in identifying the better-performing version." Call-to-Action effectiveness assesses the persuasiveness and impact of call-to-action in increasing user engagement and conversions. Respondents will evaluate the clarity, placement, and design of call-to-action. A couple of examples: "The call-to-action was clear and compelling" or "The call-to-action design (color, size, wording) was effective in increasing clicks." The Audience Segmentation Differentiated Effectiveness variable assesses the accuracy and utility of audience segmentation in personalizing marketing operations. Respondents will be requested to express their perception on whether segmentation enhanced personalization and relevance. A few examples include: "The audience was segmented effectively based on relevant criteria" or "Audience segmentation improved the relevance of the marketing messages."

The dependent variable, online sales, assesses the perceived effect of A/B testing tactics on sales results. Respondents will be required to disclose if the testing resulted in increased conversions, engagement, and income. Some examples include: "The A/B testing strategy led to an increase in online sales" or "The conversion rate improved due to the implemented variations."

The moderating variable, A/B Testing Data Analysis Accuracy evaluates the reliability and accuracy of A/B testing data interpretation. Respondents will be asked to evaluate the statistical analysis' correctness and the reliability of the insights. Examples of measurement items include: "The data collected from A/B testing was accurate and reliable" along with "The insights from A/B testing were actionable and trustworthy." This measurement approach allows a structured examination of how A/B testing strategies influence sales success in Saudi Arabia, while also taking into account data analysis quality.

To account for contextual influences on A/B testing outcomes in Saudi Arabia, this study measures five control variables tailored to the local business environment. Firm size

follows Saudi SME Authority (Monsha'at) classifications, categorizing businesses by employee count (1-5 for microbusinesses, 6-249 for SMEs, and above 250 for large companies) and annual revenue in SAR (under 1 million to over 200 million), shown in Table 4.3.

Table 4.3 Description of Control Variables

Controls Variables	Categorizations
Firm Size (No. of employees)	• Freelance/Micro (1-5) • Small (6-49) • Medium (50-249) • Large (250 and above)
Type of Industry	• E-commerce • Oil/Gas/Energy • Retail & Shopping Malls • Banking/FinTech • Tourism/Hospitality • Others
Target Market (you may choose more than one)	• Saudi Nationals • Expatriates in KSA • GCC Regional Customers • B2B (Business Clients)
Marketing Budget (% of total marketing budget)	• <10% • 10-30% • 31-50% • >50%
Annual revenue in Saudi Riyal (SAR)	• Below 1 million • 1-10 million • 10-50 million • 50-200 million • Above 200 million
Payment Method (you may choose more than one)	• Credit/Debit Cards • Cash on Delivery • Mobile Wallets • Buy Now, Pay Later • Mada • Bank Installments

5.4 Data Analysis Procedures

The qualitative phase of the study employed a straightforward thematic analysis approach.

During the interviews, detailed notes were taken, and all participants were guided by a consistent set of questions to ensure comparability across responses. Following the interviews, the notes were carefully reviewed and compared to identify recurring ideas, patterns, and key insights related to A/B testing practices and online sales performance. These recurring elements were then organized into broader themes that captured the shared experiences and perspectives of the participants. This process generated meaningful qualitative insights that informed and supported the development of the quantitative research model.

For the quantitative phase, data collected from 250 survey respondents were analyzed using Structural Equation Modeling (SEM) with SmartPLS. The analysis followed a two-step procedure, consisting of measurement model evaluation and structural model assessment. The measurement model was used to examine the reliability and validity of the constructs, ensuring their statistical robustness. Subsequently, the structural model was tested to evaluate the proposed hypotheses. Bootstrapping techniques were applied to determine the significance of path coefficients, allowing for robust estimation of both direct and moderating effects. Overall, this SEM approach provided strong empirical support for the study's conceptual research framework.

Chapter 5: Research Findings

5.1 Qualitative Results: Semi-structured Interviews

This study set out to examine how three key A/B digital marketing testing characteristics, namely A/B variation design attractiveness, call-to-action effectiveness, and audience segmentation differentiated effectiveness influence firms' online sales performance, as well as the moderating role of A/B testing data analysis accuracy. In the first phase of data collection, exploratory interviews were conducted to gain deeper insights into how A/B testing is applied in practice and how it shapes online sales outcomes in Saudi Arabia. Participants were carefully selected based on their direct experience in digital marketing and e-commerce, ensuring they could provide informed and relevant perspectives. The interview sample included two Digital Marketing Managers, two E-Commerce Managers, and one Online Store Manager, all of whom possess substantial hands-on experience in designing and implementing A/B testing strategies, monitoring campaign performance, and enhancing customer engagement.

Given their roles, these professionals regularly make data-driven decisions regarding variation design, call-to-action effectiveness, and audience targeting, making them particularly well-positioned to reflect on the practical importance of these factors. This purposive selection approach ensured that participants had the expertise and responsibilities closely aligned with the study's objectives, thereby enhancing the depth, relevance, and credibility of the qualitative findings.

5.1.1 Interviewees' Profile

The characteristics of the interviewees and their respective companies are given in Table 5.1. This included the interviewee's occupation, the number of full-time workers, the yearly income, the industry or area of expertise, the target market, the expected budget for A/B testing, and the typical payment methods.

The six interviewees, representing a range of e-commerce and digital marketing roles across diverse industries in Saudi Arabia, reported that their organizations employed between 20 and over 300 full-time staff and generated annual revenues ranging from SAR 6 million to SAR 120 million. Their businesses spanned multiple sectors, including multi-category retail, fashion, electronics, home décor, health and beauty, and gourmet foods.

Marketing budgets allocated 10-25% to A/B digital marketing testing, highlighting the growing emphasis on data-driven strategies to boost sales. Depending on the sector, target customers were typically tech-savvy individuals aged 18 to 45, with some companies focusing on niche markets such as food enthusiasts or homeowners. Popular payment methods included Visa/Mastercard, Apple Pay, Mada, and increasingly, buy-now-pay-later services like Tamara and Tabby, alongside cash-on-delivery for select categories.

Table 5.1. Interviewees' Profile

	Profession	Full-time Employees	Annual Revenues (SAR)	Industry	A/B Testing Budget	Target Customers	Payment Options
1	Digital Marketing Manager	45	18 Million	Fashion retail, e-commerce	Approximately 20% of the total digital marketing budget	Young adults aged 18-35, primarily female shoppers	Credit/debit cards (Mada, Visa, Mastercard), Apple Pay, and COD (cash on delivery)
2	Digital Marketing Manager	120	50 Million	Consumer electronics, retail	Roughly 15% of the digital marketing budget	Tech-savvy individuals aged 25-45, both genders, middle to upper income groups	Mada, Visa/Mastercard, Apple Pay, installment payments (Tamara, Tabby)
3	E-Commerce Manager	30	12 Million	Home décor, lifestyle e-commerce	10% of the total digital marketing spending	Homeowners and interior enthusiasts, primarily 28-45 years old	Credit/debit cards, Apple Pay, bank transfers
4	E-Commerce Manager	75	35 Million	Health and Beauty Products	25% of the digital marketing budget	Female consumers aged 20-40, mostly urban regions	Mada, Visa, Apple Pay, COD
5	Online Store Manager	20	6 Million	Food, beverage, gourmet e-commerce	12% of digital marketing expenditure	Foodies and health-conscious individuals, 30-50 years old	Online banking, credit/debit cards, Apple Pay

6	Data Analyst Specialist	300+	120 Million	Multi-category e-commerce (fashion, electronics, beauty)	18% of the digital marketing spend	Broad audience across Saudi Arabia, aged 18-45	Mada, Visa/Mastercard, Apple Pay, buy-now-pay-later services (Tamara, Tabby)
---	-------------------------	------	-------------	--	------------------------------------	--	--

5.1.2 Challenges of Designing A/B Testing in Digital Market in Saudi Arabia

Interviewees highlighted three main challenges in developing and executing A/B testing in Saudi Arabia. First, creating variations that are both meaningful and effective proved difficult. Often, differences between versions are either too minor to produce measurable results or too extreme, risking customer dissatisfaction. As one manager explained, “Even if our adjustments look different on paper, customers may not notice enough to change their behavior.” Striking the right balance is critical to ensure that tests provide actionable insights rather than ambiguous outcomes.

Second, reliable segmentation and data accuracy remain major hurdles. Conducting statistically sound tests becomes challenging when customer insights are scattered and analytics tools have limitations. A marketing officer noted, “Sometimes the data we rely on is insufficient, which makes it hard to trust the test results or know if a variation actually worked.” These limitations undermine confidence in data-driven decisions.

Finally, organizational support and resources pose significant obstacles. Securing leadership buy-in, sufficient time, and budget for testing can be difficult. As one manager observed, “It’s often challenging to convince upper management to invest in testing because they see it as a delay rather than a long-term advantage.” This highlights how internal resistance and limited resources can slow the adoption of A/B testing and broader data-driven experimentation.

5.1.3 A/B Testing Variation Design Effectiveness

Participants emphasized that the effectiveness of A/B testing variations largely depends on how well the design and language align with customer expectations. One respondent explained, “Even a slight color change in the call-to-action button had a big impact on

clicks when we tested two ad designs.” Another noted, “Not all variations add value; sometimes both versions perform almost identically, making it harder to justify the effort.” A third participant added, “The real effectiveness comes when variations are guided by customer insights, not just random changes.” They also highlighted that adapting content to Saudi cultural standards, ensuring high-quality Arabic translations, and managing even modest photographic requirements often slows campaign deployment and adds complexity to testing.

Marketing managers stressed that successful online sales depend heavily on thoughtful variation design. Well-crafted variations help attract customers, encourage engagement, and ultimately increase conversions. As one manager observed, “We immediately see higher click-through and purchase rates when the design change is meaningful and resonates with our customers.” Another added, “A strong variation not only looks different but also guides the customer journey more smoothly, which translates into better sales outcomes.” These insights show that the value of variation design goes beyond aesthetics it strategically shapes consumer behavior to improve performance.

5.1.4 A/B Testing Call-to-Action Effectiveness

Interviewees highlighted that optimizing call-to-action features through A/B testing is essential for boosting user engagement and conversions. One respondent explained, “We found that simply changing the wording from ‘Buy Now’ to ‘Shop Today’ noticeably increased our click-through rates.” Another added, “call-to-action placement matters a lot; users responded faster when we moved the button higher on the page.” Despite these gains, challenges remain, as one participant noted: “It takes several testing cycles before we see significant results, and sometimes the differences are subtle.” Another shared an example of a dramatic improvement in engagement and completed sales after replacing a generic ‘Shop Now’ button with a more compelling ‘Grab This Deal’ phrasing.

Marketing managers emphasized that the design and placement of call-to-actions are critical levers for increasing conversions. Even small changes in wording, color, or positioning can substantially influence consumer behavior. As one manager remarked, “A simple change in the call-to-action from ‘Submit’ to ‘Shop Now’ made our customers

much more willing to click.” Another added, “We see a direct lift in sales performance when the call-to-action is clear, visible, and action-oriented.” These insights demonstrate that systematically testing call-to-action variations provides actionable guidance for motivating customers to complete purchases, ultimately enhancing online sales outcomes.

5.1.5 Audience Segmentation Differentiated Effectiveness

Respondents highlighted that segmenting audiences before running A/B tests greatly enhances both the clarity and relevance of the results. One participant noted, “Testing all customers together gave confusing results, but once we segmented by age group, the winning variation became clear.” Another added, “Segmenting helps us avoid one-size-fits-all conclusions; what resonates with younger buyers often doesn’t work for older ones.”

However, challenges remain. As one respondent explained, “It can be difficult to gather enough data for smaller segments, and sometimes we don’t reach the statistical confidence we need.” Interviewees also observed clear differences in preferences across demographics: older users favor traditional layouts with information-rich pages, while younger users (ages 18-25) respond better to bold visuals, mobile-first designs, and faster interfaces. Urban customers, for example, showed greater interest in digital features such as flexible payment options and live chat support.

Marketing managers stressed that sales outcomes improve noticeably when A/B test variations are tailored to specific consumer segments. Rather than relying on a “one-size-fits-all” approach, firms can customize designs and messaging to match the needs of different audiences. As one manager explained, “The impact was minimal when we tested the same variation across all customers, but much stronger when we segmented by age and shopping behavior.” Another added, “Audience segmentation allows us to speak the customer’s language, which directly boosts conversions and sales.” These insights underscore that the strength of audience segmentation lies in its ability to generate more precise insights and enhance performance through smarter, targeted marketing.

5.1.6 A/B Testing Data Analysis Accuracy

Respondents agreed that the reliability and usefulness of A/B testing results hinge on the accuracy of data analysis. One participant explained, “Without proper analysis, the test results can be misleading, and we might end up choosing the wrong variation.” Another emphasized the importance of careful data interpretation, noting that “speed is less important than accuracy, as rushed analysis often leads to poor decisions.” A third respondent pointed out resource constraints: “We don’t always have data specialists, so sometimes results are interpreted based on assumptions rather than solid analytics.”

When asked how they handle unexpected or contradictory A/B test outcomes, interviewees described several strategies: extending the testing period, introducing a new variant for clarity, or reviewing the experiment setup for potential flaws. Rather than dismissing anomalous results, they treat them as learning opportunities that can reveal overlooked market patterns or emerging consumer preferences. Participants also identified conversion rate as their primary success metric, while monitoring bounce rate, average order value, and click-through rate to gain a fuller understanding of how changes affect user behavior and revenue.

Marketing managers underscored that precise data analysis forms the foundation of effective A/B testing, ensuring actionable insights and sound decision-making. As one manager stated, “If the data isn’t clean or the analysis isn’t precise, we can’t trust the results, and that puts our sales strategies at risk.” Another added, “We can scale the winning variation confidently because we know it will improve performance.” These reflections highlight that rigorous, accurate data analysis is essential not only for validating A/B test results but also for driving sustained improvements in online sales.

5.1.7 Relevance of the A/B testing characteristics in enhancing online sales

Interviewees consistently highlighted that the overall impact of A/B testing on online sales depends on the interplay between variant design, call-to-action effectiveness, audience segmentation, and precise data analysis. Eye-catching and thoughtfully designed variants were considered the first step in engaging customers. One participant

noted, “A creative design or layout can grab attention instantly, but it won’t drive sales if it doesn’t resonate with our audience.” Another added, “The right call-to-action not only encourages clicks but also creates a sense of urgency without it, even the most appealing designs lose momentum.” This underscores that call-to-action effectiveness is a critical next step in boosting conversions.

Audience segmentation was seen as a powerful tool for maximizing A/B testing outcomes through personalization. “When we target customers more precisely, the winning variation becomes obvious; audience segmentation differentiated effectiveness is what separates average results from real sales growth,” one respondent explained. However, challenges such as limited data and small sample sizes were also acknowledged.

Finally, accurate data analysis was identified as the crucial link connecting all these elements. One participant emphasized, “You can have the perfect design and call-to-action for the right segment, but if the analysis is wrong, the decision will be wrong, and sales will suffer.” Another added, “What turns experiments into actionable business strategies is accurate analysis.” Collectively, these insights illustrate that A/B testing achieves its full potential only when creative design, compelling call-to-actions, targeted segmentation, and rigorous data analysis work together seamlessly.

5.2. Quantitative Findings

5.2.1 Response Rate and Respondents’ Profile

The study employed purposive sampling to select survey participants, focusing on owners and managers of online shops, e-commerce businesses, and digital stores in Jeddah and Riyadh. These individuals were targeted because they are directly involved in both operational and strategic decision-making related to online sales performance. Their roles require active engagement with A/B testing and digital marketing strategies, making them well-suited to evaluate the impact of variation design, call-to-action effectiveness, and audience segmentation on sales outcomes. Out of 370 distributed questionnaires, 250 valid responses were received, yielding a response rate of 67.5%, a level considered robust for survey research within the e-commerce context. This purposeful selection

ensured that participants possessed the relevant experience and knowledge to provide meaningful insights, while the strong response rate enhances the reliability of the data.

The sample provides valuable perspectives on Saudi Arabia's digital commerce landscape, which is largely dominated by small and medium-sized enterprises (87.6% employ 1-49 people, and 95.2 % operate in e-commerce) with modest annual sales (66.8% under 10 million SAR). The data also highlights a conservative approach to experimentation: 98% of companies allocate less than 30% of their marketing budgets to A/B testing, signaling untapped potential for data-driven optimization. Payment methods reflect adaptation to local consumer preferences, with all respondents offering card or Mada payments and 72.4% providing cash-on-delivery options. The market focus is strongly domestic, serving primarily Saudi nationals (100%) and a limited number of expatriates (98.4%), with minimal outreach to the broader GCC region (1.2%).

Table 5.2 Respondents Descriptions

Characteristics	Categorization	Frequency	Percent
Firm Size	Freelance/Micro (1–5)	69	27.6
	Small (6–49)	150	60.0
	Medium (50–249)	23	9.2
	Large (250 and above)	8	3.2
Annual revenue in Saudi Riyal (SAR)	Below 1 million	60	24
	1–10 million	107	42.8
	10–50 million	52	20.8
	50-200 million	18	7.2
	Above 200 million	13	5.2
Industry	E-commerce	238	95.2
	Oil/Gas/Energy	2	0.08
	Retail & Shopping Malls	3	0.12
	Banking/FinTech	3	0.12
	Tourism/Hospitality	3	0.12
	Others	1	0.04
A/B Testing Budget (% of total marketing budget)	<10%	106	42.4
	10-30%	139	55.6
	31-50%	4	1.6
	>50%	1	0.4
Target Market	Saudi Nationals	250	100
	Expatriates in KSA	246	98.4
	GCC Regional	3	1.2

	Customers		
	B2B (Business Clients)	13	5.2
Payment Option	Credit/Debit Cards	250	100
	Cash on Delivery	181	72.4
	Mobile Wallets	168	67.2
	Buy Now, Pay Later	194	77.6
	Mada	250	100
	Bank Installments	15	6

These characteristics carry important strategic implications. Businesses, particularly those with limited resources, should prioritize low-cost, high-impact A/B tests focusing on key areas such as payment methods and audience segmentation. The findings highlight a clear need for greater experimentation literacy within the market. While larger organizations (only 3.2% of the sample) may require tailored approaches, the results are especially relevant for smaller, resource-constrained firms aiming to optimize digital sales in Saudi Arabia's growing e-commerce sector. The study underscores opportunities in market expansion, payment innovation, and the adoption of A/B testing, shedding light on both current practices and future growth potential in the Kingdom's digital economy. The results suggest that, given restricted testing budgets and a strong focus on the domestic market, Saudi e-commerce firms should concentrate on cost-effective experiments targeting high-impact features such as audience segmentation and payment options. The combination of localized targeting and hybrid payment preferences (cash and digital) indicates potential to streamline checkout processes for Saudi customers, while also preparing for cautious regional expansion. Additionally, the limited investment in A/B testing signals a need to educate firms on the return on investment (ROI) of experimentation, especially for smaller enterprises seeking a competitive edge. Overall, this diversity in the sample enhances the generalizability of the findings, ensuring a representative view of the factors affecting online sales performance in Saudi Arabia.

5.2.2. Measurement Model Analysis

This study used SmartPLS 3.3.5 because it is well-suited for conducting moderation analysis, with small-to-moderate sample sizes, and non-normally distributed data.

Table 5.3. Reliability and Convergent Validity

Construct	Mean	ST.DEV	CA	CR	AVE
ABT-VDA	3.71	0.71	0.90	0.92	0.87
CTAE	4.29	0.68	0.89	0.90	0.83
ASDE	4.15	0.62	0.88	0.89	0.79
ABT-DAA	4.28	0.65	0.89	0.90	0.77
OSP	4.37	0.73	0.90	0.91	0.82
Note: ABT-VDA= A/B Testing Variation Design Attractiveness, CTAE= Call-to-Action Effectiveness, ASDE= Audience Segmentation Differentiated Effectiveness, ABT-DAA= A/B Testing Data Analysis Accuracy, OSP= Online sales Performance, CA= Cronbach Alpha, CR= Composite Reliability, AVE= Average Variance Extraction.					

The study's constructs have good validity and reliability, with all Cronbach's Alpha (α) values above 0.75, indicating high internal consistency (Hair et al., 2019). The Average Variance Extracted (AVE) for each construct exceeds the 0.50 level, indicating good convergent validity, as shown in Table 5.3.

Table 5.4. Loadings and Cross-Loadings

Item	ABT-VDA	CTAE	ASDE	OSP	ABT-DAA
ABT-VDA 1	0.92	0.36	0.29	0.41	0.33
ABT-VDA 2	0.86	0.37	0.30	0.42	0.34
ABT-VDA 3	0.91	0.36	0.29	0.40	0.33
ABT-VDA 4	0.86	0.38	0.31	0.43	0.35
ABT-VDA 5	0.87	0.39	0.32	0.44	0.36
CTAE1	0.34	0.87	0.29	0.45	0.31
CTAE2	0.35	0.92	0.30	0.43	0.32
CTAE3	0.33	0.90	0.31	0.46	0.34
CTAE4	0.36	0.86	0.32	0.47	0.35
ASDE1	0.32	0.34	0.91	0.41	0.33
ASDE2	0.31	0.33	0.92	0.42	0.34
ASDE3	0.34	0.35	0.88	0.44	0.36
ASDE4	0.33	0.36	0.87	0.43	0.35
ABT-DAA1	0.39	0.42	0.40	0.90	0.37
ABT-DAA2	0.40	0.43	0.41	0.92	0.38
ABT-DAA3	0.41	0.44	0.42	0.90	0.39
ABT-DAA4	0.42	0.45	0.43	0.91	0.37
OSP1	0.43	0.46	0.44	0.92	0.38
OSP2	0.32	0.30	0.32	0.36	0.91
OSP3	0.30	0.31	0.33	0.37	0.87
OSP4	0.34	0.32	0.34	0.38	0.90
OSP5	0.31	0.33	0.35	0.39	0.88

Note: ABT-VDA= A/B Testing Variation Design Attractiveness, CTAE= Call-to-Action Effectiveness, ASDE= Audience Segmentation Differentiated Effectiveness, ABT-DAA= A/B Testing Data Analysis Accuracy, OSP= Online sales Performance.

The loadings and cross-loadings findings are shown in Table 5.4. All items have high loadings on their respective constructs (Henseler et al, 2016), indicating that they strongly represent the factor they were intended to assess. These constructs have high convergent validity. Similarly, the lower cross-loadings indicate discriminant validity, which means that the items are not overly linked with constructs they are not intended to measure. This pattern indicates that the constructs are separate. Overall, the results show that the measurement model is reliable. Each construct is well assessed by its own indicators, and there is a strong separation between constructs, implying convergent and discriminant validity.

The Heterotrait-Monotrait Ratio results revealed that all construct pairs had values less than the acceptable threshold of 0.85 (Henseler et al., 2016), indicating good discriminant validity. This demonstrates that A/B Testing Variation Design Attractiveness, Call-to-Action Effectiveness, Audience Segmentation Differentiated Effectiveness, A/B Testing Data Analysis Accuracy, and Online Sales Performance are conceptually distinct constructs, making it possible that the structural model results can be confidently interpreted.

Table 5.5 Heterotrait-Monotrait Ratio

Constructs	ABT-VDA	CTAE	ASDE	ABT-DAA
ABT-VDA				
CTAE	0.61			
ASDE	0.58	0.64		
ABT-DAA	0.47	0.52	0.55	
OSP	0.57	0.63	0.69	0.57
Note: ABT-VDA= A/B Testing Variation Design Attractiveness, CTAE= Call-to-Action Effectiveness, ASDE= Audience Segmentation Differentiated Effectiveness, ABT-DAA= A/B Testing Data Analysis Accuracy, OS= Online sales.				

Table 5.6 illustrates the results of discriminant validity. The results support discriminant validity employing the Fornell-Larcker criterion, as the square root of the AVE for each construct (ABT-VDA = 0.93, CTAE = 0.91, ASDE = 0.88, ABT-DAA = 0.87, OSP = 0.90) is greater than the correlations with other constructs. This suggests that each construct has greater variance with its own items than with other constructs, indicating that they are conceptually unique. Similarly, Cronbach's Alpha values (all above 0.70) and AVE values (all above 0.50) show strong reliability and convergent validity, indicating that the measurement model is reliable and valid. Overall, these findings verify the measurement model, providing confidence in upcoming analyses that will evaluate the links between A/B testing characteristics and online sales.

Table 5.6 Discriminant Validity

Construct	ABT-VDA	CTAE	ASDE	ABT-DAA	OSP
ABT-VDA	0.93				
CTAE	4.29	0.91			
ASDE	4.15	0.62	0.88		
ABT-DAA	4.28	0.65	0.65	0.87	
OSP	4.37	0.73	0.73	0.60	0.90
<i>Note:</i> ABT-VDA= A/B Testing Variation Design Attractiveness, CTAE= Call-to-Action Effectiveness, ASDE= Audience Segmentation Differentiated Effectiveness, ABT-DAA= A/B Testing Data Analysis Accuracy, OSP= Online sales Performance.					

5.2.3 Structural Model Analysis

Figure 5.1 illustrates how the PLS regression model reveals that A/B testing features have a significant impact on online sales performance. In particular, the biggest positive effect is shown for A/B Variation Design Attractiveness ($\beta = 0.385$, $t = 6.54$, $p < 0.001$), demonstrating that attractive design significantly improves online sales. A highly significant influence is also shown by Call-to-Action Effectiveness ($\beta = 0.297$, $t = 5.04$, $p < 0.001$), underscoring the significance of a persuasive message in eliciting reactions from customers. Additionally, Audience Segmentation Differentiated Effectiveness ($\beta = 0.194$, $t = 3.14$, $p < 0.01$) makes a significant contribution, suggesting that performance is improved by customizing offerings for particular customers. These results suggest that H1, H2, and H3 are supported.

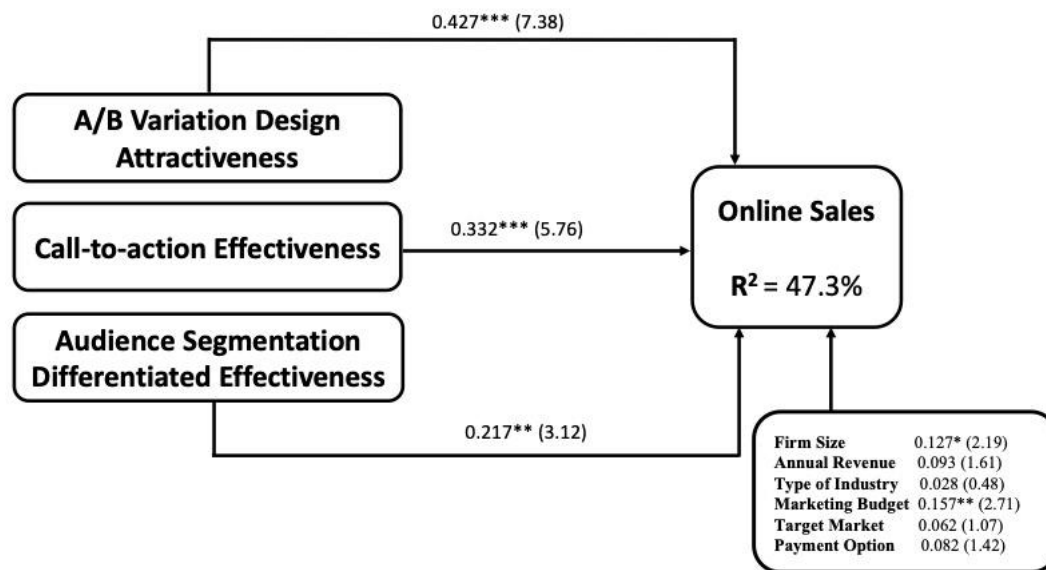


Figure 5.1 Results of Hypothesis Testing: Direct Regression Analysis

The direct regression findings paint a clear picture: what customers see, read, and how well it speaks to them all work together to drive online sales. The strongest impact comes from design attractiveness, which shows that people are naturally drawn to visually appealing and well-structured pages. When a website looks clean, modern, and easy to navigate, it builds trust and encourages users to stay longer and explore. Similarly, persuasive call-to-action messages play a crucial role in turning that interest into action, guiding customers on what to do next in a clear and compelling way. On top of that, tailoring content to specific audience segments further strengthens results, as customers are more likely to engage with offers that feel relevant to their needs and preferences.

From a business standpoint, these results highlight the importance of taking a holistic and data-driven approach to optimization. Companies should not rely on a single element but instead continuously test and refine design, messaging, and targeting together. This means investing in better A/B testing design, crafting clear and engaging call-to-actions, and using customer data to personalize experiences. By doing so, businesses can create more meaningful interactions with their customers, improve conversion rates, and ultimately achieve stronger and more sustainable online sales performance.

Although Data Analysis Accuracy does not have a significant direct effect on online sales

performance, it is positioned as a potential moderator in the model. Among the control variables, Firm Size ($\beta = 0.135$, $t = 2.48$, $p < 0.05$) and A/B Testing Budget ($\beta = 0.169$, $t = 2.89$, $p < 0.01$) show positive relationships with sales performance, while Revenue, Industry, Target Market, and Payment Option exhibit no significant effects. Overall, the model explains 43% of the variance ($R^2 = 0.43$) in online sales performance, indicating moderate explanatory power. This underscores that variation design attractiveness, effective Call-to-actions, and audience segmentation differentiated effectiveness remain the primary drivers of improved online sales, supported by firm-level resources such as size and testing budget.

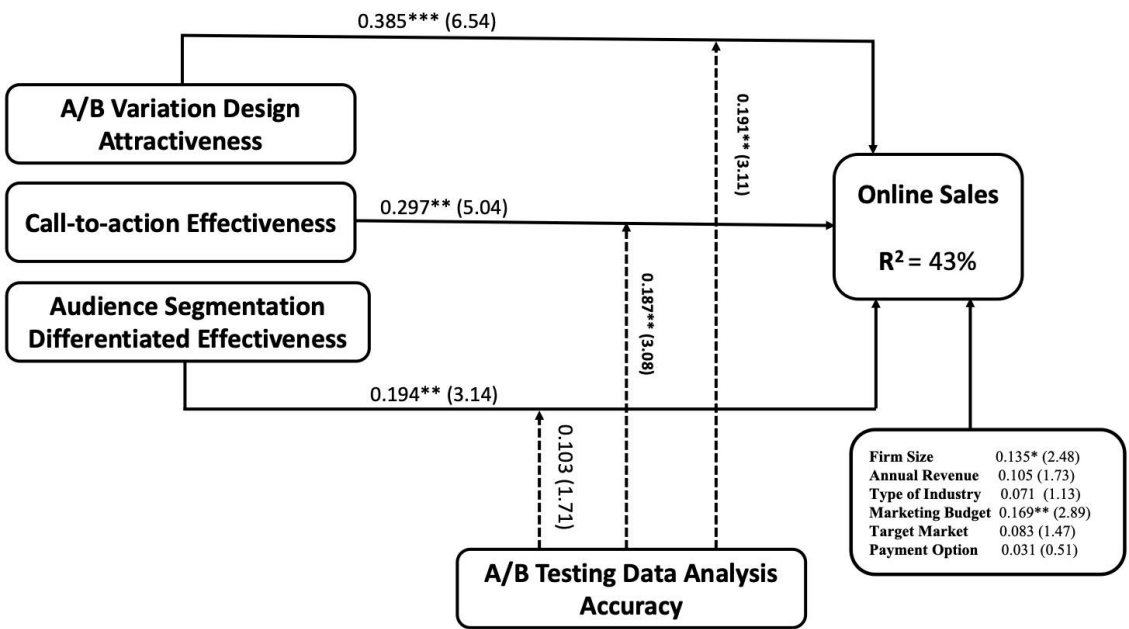


Figure 5.2: Results of Hypothesis Testing: Moderating Regression Analysis

Notably, the explained variance decreased slightly from $R^2 = 47.3\%$ in the direct regression analysis to $R^2 = 43\%$ in the moderating analysis. This suggests that while including the moderating effects slightly reduces the direct explanatory power, the moderation captures additional nuance, emphasizing the importance of accurate data analysis in shaping the relationships between A/B testing features and online sales outcomes.

Table 5.7. Summary of Results

Hypotheses	Results
H1: A/B testing variation design attractiveness is positively associated with firms' online sales.	Supported
H2: A/B testing Call-to-Action Effectiveness is positively associated with firms' online sales.	Supported
H3: Audience segmentation differentiated effectiveness is positively associated with firms' online sales.	Supported
H4a: A/B testing data analysis strengthens the relationship between A/B Testing Variation Design Attractiveness and online sales.	Supported
H4b: A/B testing data analysis strengthens the relationship between call-to-action effectiveness and online sales.	Supported
H4c: A/B testing data analysis strengthens the relationship between audience segmentation differentiated effectiveness and online sales.	Not Supported

The moderating analysis revealed that two A/B testing features remain, significant drivers of online sales performance even when A/B testing data analysis accuracy is considered. Specifically, Variation Design Attractiveness ($\beta = 0.191$, $t = 3.11$, $p < 0.01$) has the strongest positive effect, followed closely by Call-to-Action Effectiveness ($\beta = 0.187$, $t = 3.08$, $p < 0.01$). In contrast, Audience Segmentation Differentiated Effectiveness was not significant when the moderator was included ($\beta = 0.103$, $t = 1.71$), suggesting that while targeting is useful, its impact is less pronounced once data analysis accuracy is accounted for.

Although the path coefficients of the independent variables slightly decreased with the inclusion of the moderating variable, their effects on online sales performance remain meaningful (Becker et al., 2018). This indicates that while accurate data analysis moderates the strength of these relationships, it does not eliminate them. In practice, this means that appealing designs and persuasive call-to-actions continue to drive sales, but their effectiveness is maximized when supported by precise and reliable data analysis. These results show that data analysis accuracy doesn't directly boost sales, but it still plays an important behind-the-scenes role. It acts more like a "quality enhancer," making A/B testing results more reliable and helping businesses choose the right designs, messages, and segments to scale. In simple terms, even great ideas won't perform well if

the data behind decisions isn't accurate. In a similar vein, firms with larger size and higher A/B testing budgets tend to achieve better sales outcomes, likely because they can run more experiments and learn faster. The model's moderate explanatory power ($R^2 = 0.43$) confirms that while design, call-to-actions, and segmentation are the main drivers, success also depends on how well companies support these efforts with resources and strong analytics. Overall, businesses should not only focus on testing what works, but also on interpreting results correctly to turn insights into real growth

Chapter 6

Discussion and Conclusion

6.1 Key Findings

The qualitative analysis indicates that the sample is largely composed of small and micro e-commerce firms, with 60% classified as small enterprises and 27.6% as freelance or micro businesses, while medium and large firms make up only a small fraction. Most companies report annual revenues between 1-10 million SAR, reflecting a moderate financial profile. The industry distribution is heavily skewed toward e-commerce (86.8%), highlighting the digital-first nature of the participating organizations. In terms of marketing investment, most firms allocate a modest 10-30% of their budgets to A/B testing. The target market is predominantly domestic, with nearly all firms serving both Saudi nationals and expatriates. The widespread adoption of diverse payment options credit/debit cards, Mada, buy-now-pay-later services, and cash on deliver underscores the sector's focus on consumer convenience and digital transaction readiness.

Guided by the Technology-Organization-Environment (TOE) framework and Evidence-Based Management (EBM), the study examined how key A/B testing components variation design attractiveness, call-to-action effectiveness, and audience segmentation affect online sales performance. Results show that design attractiveness is the strongest predictor of improved digital outcomes, highlighting the critical role of visual and structural interface quality in engaging customers. A call-to-action effectiveness is also a significant driver, demonstrating the impact of persuasive, action-oriented messaging. Audience segmentation positively contributes to performance, confirming that tailored content creates more relevant and resonant customer experiences.

The moderating analysis reveals that A/B testing data analysis accuracy amplifies the effects of design attractiveness and call-to-action effectiveness, underscoring the importance of reliable data interpretation for maximizing the strategic value of digital experiments. In contrast, segmentation effectiveness does not moderate these relationships, suggesting that while segmentation provides baseline benefits, it does not further enhance outcomes when interacting with data quality. Overall, the findings highlight design, messaging, personalization, and analytical rigor as complementary factors that drive online sales performance. Moreover, there is also a group of papers

pointing to the merits of A/B testing as a key tool for testing web interfaces. Kohavi et al. (2009) outlined A/B testing as a powerful technique for evaluating the success of a specific design element, particularly when it comes to live web-sites. Li, Zhou, Luo, and Dong (2022) noted on the crucial importance of A/B testing in the software industry and its use in combination with machine learning to improve user experience. Hagar and Diakopoulos (2019) highlighted A/B testing as a key tool in content optimization, with click-through rate being a fundamental key performance indicator.

6.2 Discussions

The first phase of data collection involved exploratory interviews, aimed at gaining in-depth insights from participants. These interviews highlighted three main challenges for implementing A/B testing in Saudi Arabia: designing variations that are sufficiently distinct to influence user behavior without causing dissatisfaction, obtaining reliable audience segmentation and accurate data to ensure statistically valid results, and overcoming internal organizational barriers, such as limited resources, lack of management support, and resistance to allocating time and budget for experimentation. Collectively, these challenges make it difficult for firms to generate clear insights and fully embrace data-driven decision-making.

Interviewees emphasized that the effectiveness of A/B testing largely depends on thoughtful variation design, optimized call-to-action elements, targeted audience segmentation, and precise data analysis. Well-crafted variations that align with customer expectations, culturally appropriate content, and high-quality Arabic localization can significantly boost engagement and conversions, whereas poorly differentiated versions provide little benefit. Small, strategic changes in Call-to-action wording, placement, and visibility were also reported to have substantial effects on user actions and sales outcomes. Segmenting audiences by demographics, behavior, or location further enhances the clarity and usefulness of results, helping firms avoid misleading “one-size-fits-all” conclusions. Accurate data analysis was seen as the key link that ensures trustworthy decision-making; unreliable or insufficient data can lead to incorrect interpretations and limit the impact of

A/B testing initiatives.

Prior research suggests that performance variability often increases when A/B testing drives significant organizational changes (Koning et al., 2022). Because A/B testing generates numerous ideas, the quality of these ideas varies (Girotra et al., 2010), making it challenging to identify the most promising ones. Firms that prioritize ideas supported by rigorous experimentation are better positioned for rapid growth. Incremental improvements through A/B testing allow firms to refine high-potential concepts, while continuous testing at scale accelerates optimization and exploitation of top ideas. Even small adjustments can accumulate over time, producing substantial performance gains (Kohavi & Thomke, 2017; Azevedo et al., 2018; Thomke, 2020).

This study adopts a unique multi-method approach, combining qualitative interviews with quantitative surveys. While prior research has largely focused on causal inference and statistical theory, this study leverages the Technology-Organization-Environment (TOE) framework. Earlier studies emphasized testing validity, improving performance metrics, avoiding design flaws, and informing strategic decisions. In contrast, the current research demonstrates that well-designed A/B tests with clear Call-to-action and audience segmentation can significantly boost online sales, and that precise data analysis further strengthens these outcomes. Moreover, none of the existing studies focused specifically on Saudi Arabia, making this geographical context a unique contribution.

The present research examined how A/B testing characteristics variation design attractiveness, Call-to-action effectiveness, and audience segmentation impact online sales performance, while considering the moderating role of A/B testing data analysis accuracy in Saudi Arabia. A/B testing enables firms to make data-driven marketing decisions, identify the strategies that resonate with Saudi consumers, and optimize designs, Call-to-action, and segmentation to maximize conversions and engagement. Segmenting audiences by local demographics or tailoring Call-to-action to cultural preferences enhances these effects. Accurate data analysis is critical for ensuring that these insights are actionable; without it, firms risk drawing incorrect conclusions and making suboptimal marketing choices.

E-commerce companies have widely adopted web personalization techniques, including recommendation systems, to influence user behavior and achieve customer retention (Jankowski et al., 2019). Research by Ferretti et al (2017) heavily relied on personalization techniques for adapting web content to user habits, mainly with the aim of offering suitable products and services. Personalization is based on user preferences and needs (Jylhä & Hamari, 2020), which is confirmed by with Chung et al. (2016) who suggested that adaptive personalization may work better than customer customization (De Bellis et al., 2019) in autonomously tailoring a product. Adaptive personalization has several key characteristics (Chung et al., 2016). It is performed automatically using algorithms, does not require proactive effort from users, observes customer behavior and adjusts the product over time. It is assumed that adaptive personalization will yield better results over time as it learns more about the user.

The lack of representativeness means results of this randomized test are applicable only to the types of users who were selected for the experiment (i.e., targeted), and the criteria used to select those targeted users are unobservable to the experimenter (de Langhe and Puntoni 2021; Braun et al. 2023). In practice, divergent delivery in targeting makes causal inference about the effect of ad content impossible because the comparison in outcomes between ads is not “apples to apples”. This imbalance occurs because the algorithm considers relevance of ad content when targeting users. Without balance, the ad effects are confounded by the selection of users seeing those ads (Hardisty and Weber 2020).

But there is no way for the experimenter to separate how much of the reported difference is due to differences in ad content from how much is due to having different types of users see each ad. The methods for targeting ads to user types are proprietary, and the criteria that drive how the targeted sets of users for each ad differ from one another and from the intended population are not disclosed to the experimenter. The criteria may not even be saved by the platform for subsequent analysis (Gordon et al. 2019, p. 220).

In Saudi Arabia’s fast-growing digital market, continuous improvement based on precise A/B test results can enhance user experience, boost customer satisfaction, and increase online sales. The accuracy of data analysis acts as a crucial moderator: it allows firms to

determine whether variations in design, call-to-actions, or audience segmentation genuinely improve performance. Conversely, inaccurate analysis may lead to flawed decisions that negatively affect sales. Given the region's cultural and market-specific nuances, high data accuracy is essential to maximize the effectiveness of A/B testing, ensuring that changes align with consumer behavior and drive meaningful business outcomes.

6.3 Implications

6.3.1 Theoretical implications

This study advances the understanding of digital experimentation within the Technology-Organization-Environment (TOE) framework by examining how specific A/B testing components influence firms' online sales performance. First, the strong effect of variation design attractiveness underscores the technological dimension of TOE, highlighting that high-quality interface design is a critical technological resource shaping customer response. Second, the significance of call-to-action effectiveness extends insights in the organizational dimension, showing that persuasive messaging functions as an organizational capability that directly enhances customer engagement. Third, the positive impact of audience segmentation supports the environmental dimension of TOE, illustrating how firms' responsiveness to diverse customer groups improves alignment between digital content and market needs. Additionally, the findings of my study approve the applicability of the Evidence-Based Management (EBM) theory (Eapen et al., 2023) because they clearly show that decisions grounded in accurate and well-analyzed data lead to better outcomes. In my research, A/B testing characteristics such as audience segmentation, call-to-action effectiveness, and design attractiveness all influenced online sales, but their impact depended strongly on how accurately the data was analyzed. This supports the core idea of EBM that simply having data is not enough; what really matters is how well that evidence is interpreted and used in decision-making. In this way, my study brings the EBM concept to life by showing that when managers rely on properly analyzed experimental data, they can make more effective, consistent, and evidence-driven decisions, especially in a real-world context like Saudi Arabia's online market.

In their study combining eye-tracking techniques with A/B split testing for social media marketing optimization, Nichifor et al. (2021) identified effective strategies for capturing user attention through increased ad personalization. Their findings suggest that A/B-tested advertisements are particularly effective for brand awareness campaigns when they incorporate visually engaging elements such as happy human faces highlighting the importance of variation design attractiveness in shaping user responses. Moreover, the study shows that audience preferences differ by demographic group: emotionally driven visuals are more effective for women, while technically detailed content is more appealing to men. This reinforces the role of audience segmentation, as tailoring content to specific user groups enhances relevance and engagement (Nichifor et al., 2021).

In today's highly competitive digital landscape, especially for firms operating online, evaluating the effectiveness of marketing activities has become crucial for sustaining and improving performance. However, this process is often challenging due to the complex and dynamic nature of consumer behavior. By examining key performance indicators such as conversion rates, number of orders, cart abandonment rates, bounce rates, time spent on the website, and the number of unique visitors, firms can gain meaningful insights into both sales performance and customer interactions (Esteller-Cucala et al., 2019). These insights allow organizations to determine whether their marketing decisions are driving positive or negative outcomes. One of the most effective methods for generating such insights is A/B testing.

A/B testing has become a widely used and increasingly important method of conducting controlled experiments with real users (Tamburrelli & Margara, 2014). It plays a vital role in assessing how changes to digital interfaces influence user behavior (Esteller-Cucala et al., 2019). At its core, A/B testing involves presenting different versions of a webpage or digital element to separate groups of users and comparing their responses to identify which version performs better in achieving desired outcomes, such as conversions (Siroker & Koomen, 2015; Kohavi & Longbotham, 2017; King et al., 2017). Beyond simply generating data, A/B testing offers actionable insights that enable firms to refine website elements, improve operational efficiency, and enhance the overall user

experience. In doing so, it supports more informed decision-making while also helping to build customer trust and strengthen overall brand performance (Gallo, 2017).

The rapid expansion of digital marketing and social media has fundamentally transformed how companies interact with customers (Saura et al., 2020). In this environment, communication strategies particularly the way messages are framed and presented directly influence customer decision-making and business performance (Batra & Keller, 2013). This highlights the importance of call-to-action effectiveness, as clear, persuasive, and action-oriented messaging plays a key role in guiding users from attention to conversion.

While attracting visual attention is essential, it is not sufficient on its own to convert users into customers. Leading companies such as Facebook, LinkedIn, Airbnb, Microsoft, Uber, eBay, and Amazon address this challenge by systematically using A/B testing as part of their marketing strategies (King et al., 2017). Recognized as one of the most effective methods for large-scale online experimentation, A/B testing enables firms to optimize design elements, call-to-actions, and segmented targeting strategies simultaneously. By doing so, companies can identify which combinations of visual appeal, messaging, and audience targeting are most effective in transforming user engagement into actual conversions (Siroker & Koomen, 2013; Jahandideh et al., 2013).

Additionally, the moderating role of A/B testing data analysis accuracy contributes to evidence-based management research by demonstrating that data quality is not merely an operational requirement but a strategic enabler, amplifying the effects of design and messaging decisions. Interestingly, audience segmentation differentiation did not exhibit a moderating effect, indicating that while segmentation provides baseline benefits, it may not further interact with analytical accuracy to generate additional value. This finding highlights potential boundary conditions for segmentation strategies in digital experimentation, offering practical guidance for firms seeking to optimize their A/B testing efforts.

6.3.2 Practical implications

The findings of this study offer several practical implications for managers seeking to adopt or improve A/B testing practices.

First, prioritize visually appealing and well-structured A/B testing design variations. Design attractiveness was found to have the strongest impact on online sales, so firms should invest in high-quality visuals, intuitive layouts, and consistent aesthetics when creating A/B test variants.

Second, optimize call-to-action wording and placement. Call-to-actions should be clear, concise, and action-oriented, strategically positioned to guide users toward desired behaviors. Even small improvements in call-to-action clarity or visibility can lead to measurable gains in engagement and conversions.

Third, leverage audience segmentation thoughtfully. Segmenting audiences meaningfully enhances performance by delivering tailored content. However, over-segmenting or excessively refining audiences does not further amplify the benefits of accurate data analysis, suggesting diminishing returns for hyper-detailed segmentation.

Finally, ensure high accuracy in data analysis to support evidence-based decisions. The strengthening effect of analysis accuracy highlights the importance of reliable data pipelines, robust statistical procedures, and skilled analysts. Firms should adopt standardized A/B testing protocols, maintain clean and consistent data collection, and employ statistical validation tools to improve decision quality and maximize the impact of digital experiments.

6.4 Limitations and Future Research

While this study provides valuable insights into the role of A/B testing in enhancing online sales performance in Saudi Arabia, several limitations should be acknowledged.

First, the sampling approach and context may limit generalizability. Both the interviews and surveys relied on purposive, non-probability sampling, targeting individuals already involved in digital marketing and e-commerce. As a result, the findings may not represent all firms or industries using A/B testing. Additionally, the quantitative data were collected

solely within Saudi Arabia, meaning cultural, economic, and technological factors specific to this region may influence results and limit applicability to other markets or digital ecosystems.

Second, reliance on self-reported, cross-sectional survey data introduces potential biases and constraints. Respondents may have overestimated their A/B testing practices or provided socially desirable answers. Moreover, the cross-sectional design restricts the ability to identify long-term trends or causal relationships, as the impact of A/B testing can evolve over time. The study focused on three key A/B testing characteristics and one moderating variable, which, while significant, do not capture the full complexity of digital marketing optimization. Factors such as personalization algorithms, platform-specific testing constraints, consumer behavior dynamics, and organizational capabilities were not examined.

Third, industry differences and sample size warrant caution. The sample included firms from different sectors, which helps contextualize findings but may limit comparisons with industries where experimentation is more resource-intensive, such as manufacturing or life sciences (Gans et al., 2019; Pillai et al., 2020). Furthermore, the qualitative phase included only six interviewees, providing useful exploratory insights but not capturing the full range of industry perspectives.

These limitations highlight opportunities for future research. Replicating this study across different industries or cultural contexts could reveal how user preferences, digital behaviors, and design expectations vary globally. Longitudinal or experimental designs would strengthen understanding of causality and capture the evolving nature of A/B testing practices. Future research could also incorporate additional moderating or mediating factors, such as digital maturity, organizational agility, analytics capability, or competitive pressures, to offer a more comprehensive view of experimentation effectiveness.

Granular examination of design and call-to-action elements including color, typography, framing, layout, or call-to-action tone could help identify the specific components with the greatest impact. Additionally, the finding that audience segmentation differentiation

does not moderate outcomes suggests the need to explore alternative or advanced segmentation strategies, such as AI-driven clustering, behavioral segmentation, psychographic profiling, or real-time personalization. These approaches may uncover nonlinear or context-dependent effects, advancing understanding of how personalized digital marketing strategies drive online sales.

6.5 Conclusion

This thesis examined the role of A/B marketing testing as a data-driven tool for improving online sales performance in Saudi Arabia. By combining qualitative insights from industry interviews with quantitative survey data, the study explored how three key A/B testing characteristics, variation design attractiveness, call-to-action effectiveness, and audience segmentation differentiated effectiveness impact online sales outcomes. Additionally, the research investigated the moderating role of A/B testing data analysis accuracy. The mixed-methods approach provided a holistic perspective, capturing both practical, hands-on experiences from professionals and broader empirical patterns from the survey sample, resulting in a well-rounded understanding of A/B testing implementation and its effects on sales performance.

The findings indicate that visually appealing and well-structured variation designs have the strongest positive influence on online sales, highlighting the value of creativity, usability, and user-focused design in digital marketing. Action-oriented and strategically placed call-to-actions were also found to be crucial in motivating customer responses. While audience segmentation contributes positively to performance, its effect was somewhat smaller than design and call-to-action effectiveness. Importantly, the accuracy of data analysis was shown to amplify the impact of both design attractiveness and call-to-action effectiveness, emphasizing that reliable analytics are essential for transforming experimental results into actionable business decisions. Collectively, these findings underscore the strategic importance of thoughtfully designed, precisely executed, and analytically robust A/B testing practices for driving online sales in the Saudi digital marketplace.

References

- [1] Agarwal, A, Shun-Yang, L, Andrew B. W. (2024). The Effect of Popularity Cues and Peer Endorsements on Assertive Social Media Ads. *Information Systems Research* 36(2):1167-1182.
- [2] Agarwal, S. (2023). Optimizing Product Choices through A/B Testing and Data Analytics: A Comprehensive Review. *International Journal of Advanced Research in Science, Communication and Technology*, Available at SSRN: <https://ssrn.com/abstract=4656782>
- [3] Ahmadi, I., Waltenrath, A., & Janze, C. (2022). Congruency and Users' Sharing on Social Media Platforms: A Novel Approach for Analyzing Content. *Taylor & Francis*, 52(3), 369-386.
- [4] Ahmed, H., Jilani, D. T. A., Haider, W., Abbasi, M. A., Nand, S., & Kamran, S. (2017). Establishing Standard Rules for Choosing Best KPIs for an E-Commerce Business based on Google Analytics and Machine Learning Technique. *International Journal of Advanced Computer Science and Applications*, 8(5).
- [5] Al-Natour, S. and Benbasat, I. (2009). The adoption and IT artefacts: a new interaction-centric model for the study of user artefact relationships. *Journal of Association for Information Systems*, 10 (9), pp. 661-685.
- [6] Alkhalil, B. F., Zhuang, Y., Mursi, K. T., & Aseeri, A. O. (2024). E-Commerce Trust in Saudi Arabia: An Empirical Investigation Using A/B Testing. *International Symposium on Innovative Approaches in Smart Technologies*, 1–7. Gallo, A. (2017). A refresher on A/B testing. *Harvard Business Review*, pp 2–6.
- [7] Aririguzoh, S A. (2022). Communication competencies, culture and SDGs: effective processes to cross-cultural communication. *Palgrave Macmillan*, 9(1).
- [8] Awa, H. O., Ojiabo, O. U., & Orokor, L. E. (2017). Integrated technology-organization-environment (T-O-E) taxonomies for technology adoption. *Journal of Enterprise Information Management*, 30(6), 893–921.
- [9] Azevedo, E. M., Deng, A., Montiel Olea, J. L., Rao, J., & Weyl, E. G. (2020). A/B testing with fat tails. *Journal of Political Economy*, 128(12), 4614.
- [10] Baki, R. (2020). Evaluating hotel websites through the use of fuzzy AHP and fuzzy TOPSIS. *International Journal of Contemporary Hospitality Management*, 13(12), 3747-3765.
- [11] Bartoš, P., & Habarta, F. (2019). Does the form of text play a role in click through the button on websites?. *Trendy v Podnikání*, 9(1), 3-8.
- [12] Batra, R.; Keller, K.L. Integrating Marketing Communications: New Findings, New Lessons, and New Ideas. *J. Mark.* 2016, 80, 122–145.
- [13] Becker, J.-M., Ringle, C. M., and Sarstedt, M. (2018). Estimating Moderating Effects in PLS-SEM and PLSc-SEM: Interaction Term Generation x Data Treatment. *Journal of Applied Structural Equation Modeling*, 2(2): 1-21.
- [14] Benmamoun, M., Singh, N., & Sobh, R. (2019). How Advertisers Can Target Arab E-Consumers More Effectively. *Advertising Research Foundation*, 59(2), 171-184.

- [15] Blanchard, S. J., Goldenberg, J., Pauwels, K., & Schweidel, D. A. (2022). Promoting data-richness in consumer research: How to develop and evaluate articles with multiple data sources. *Journal of Consumer Research*, 49(2), 359–372.
- [16] Bleier, A., & Eisenbeiss, M. (2015). Personalized online advertising effectiveness: The interplay of what, when, and where. *Marketing Science*, 34(5), 669-688.
- [17] Boegershausen, J, Cornil, Y, Yi, S, Hardisty, D. (2025). On the persistent mischaracterization of Google and Facebook A/B tests: How to conduct and report online platform studies. *International Journal of Research in Marketing* 42 (2025) 886–903.
- [18] Boegershausen, J., Datta, H., Borah, A., & Stephen, A. T. (2022). Fields of gold: Scraping web data for marketing insights. *Journal of Marketing*, 86(5), 1–20.
- [19] Bortko, K., Jankowski, J., Bartków, P., Pazura, P., & Śmiałkowska, B. (2020). Attracting user attention to visual elements within website with the use of Fitts's law and flickering effect. *Procedia Computer Science*, 176, 2756-2763.
- [20] Boswell, P. (2011). *Google Analytics: Measuring Content Use and Engagement*. Proceedings of 58th Annual Conference Society for Technical Communication. Sacramento, CA: Society for Technical Communication.
- [21] Braun, M., & Schwartz, E. M. (2025). Where A/B Testing Goes Wrong: How Divergent Delivery Affects What Online Experiments Cannot (and Can) Tell You About How Customers Respond to Advertising. *Journal of Marketing*, 89(2), 71-95.
- [22] Braun, M., De Langhe, B., Puntoni, S., & Schwartz, E. M. (2024). Leveraging digital advertising platforms for consumer research. *Journal of Consumer Research*, 51(1), 119–128.
- [23] Bressolles, G., & Lang, G. (2020). Key Performance Indicators for performance measurement of e-fulfillment systems in multi-channel retailing. *International Journal of Retail & Distribution Management*, 48(1), 35-52.
- [24] Browne, W and Mike, S. J. (2017). What works in e-commerce: a meta-analysis of 6700 online experiments.” Qubit Digital Ltd 21.
- [25] Brynjolfsson, E, and McElheran, K. (2016). The rapid adoption of data-driven decision- making. *American Economic Review*, 106(5):133–39.
- [26] Bunt, A., Conati, C., & McGrenere, J. (2009). Mixed-Initiative interface personalization as a case study in usable AI. *AI Magazine*, 30(4), 58-64.
- [27] Cai, L., He, X., Dai, Y., & Zhu, K. (2018). Research on B2B2C E-commerce website design based on user experience. *Journal of Physics: Conference Series*, 1087(6), 62043.
- [28] Camuffo, Arnaldo, Alessandro Cordova, Alfonso Gambardella and Chiara Spina. (2020). A scientific approach to entrepreneurial decision making: Evidence from a randomized control trial.” *Management Science* 66(2):564–586.
- [29] Cao, R, Rembrand, K and Ramana, N. (2020). Biased sampling of early users and the direction of startup innovation. *Harvard Business School Entrepreneurial Management Working Paper* (21-059).

- [30] Chatterji, K. A, Findley, M, Jensen, N. M, Meier, S, and Nielson, F. (2016). Field experiments in strategy research. *Strategic Management Journal*, 37(1):116–132
- [31] Chen, C., & Soo, K. C. (2018). The research on human engineering-based design of the website interface of smart cellphone. *Journal of Physics: Conference Series*, 1098(1), 12012.
- [32] Chen, C., & Zhai, W. (2023). The effects of dynamic prompts and background transparency of hover feedback design on the user interface of shopping websites. *Asia Pacific Journal of Marketing and Logistics*, 35(4), 809-827.
- [33] Christian Hampel, Markus Perkmann, and Nelson Phillips. (2020). Beyond the lean start-up: experimentation in corporate entrepreneurship and innovation. *Innovation*, 22(1):1–11.
- [34] Chung, T. S., Wedel, M., & Rust, R. T. (2016). Adaptive personalization using social networks. *Journal of the Academy of Marketing Science*, 44(1), 66-87.
- [35] Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.
- [36] Colicev, A., Kumar, A., & O'Connor, P. (2019). Modeling the relationship between firm and user generated content and the stages of the marketing funnel. *International Journal of Research in Marketing*, 36(1), 100–116.
- [37] de Bellis, E., Hildebrand, C., Ito, K., Herrmann, A., & Schmitt, B. (2019). Personalizing the customization experience: A matching theory of mass customization interfaces and cultural information processing. *Journal of Marketing Research*, 56(6), 1050-1065.
- [38] De Langhe, B. and S. Puntoni (2021). Does Personalized Advertising Work as Well as Tech Companies Claim? *Harvard Business Review*. url: https://bit.ly/PdeL_HBR.
- [39] Desai, D. (2021). Website personalization: Strategy for user experience design & development. *Turkish Journal of Computer and Mathematics Education*, 12(12), 3516-3523.
- [40] Dubé, J, Zheng, F, Nathan, F, and Xueming, L. (2017). Competitive price targeting with smartphone coupons.” *Marketing Science* 36(6):944–975.
- [41] Durmaz, Y., & Efendioglu, I. H. (2016). Travel from traditional marketing to digital marketing. *Global journal of management and business research*, 16(2), 34-40.
- [42] Eapen, B. R., Baba, V. V., & Sohail, M. (2023). Evidence-Based Management: A Design Theory, Template, and Technology for a Knowledge Delivery Platform. *Metamorphosis: A Journal of Management Research*, 22(1), 73-84.
- [43] Eckles, D., Gordon, B. R., & Johnson, G. A. (2018). Field studies of psychologically targeted ads face threats to internal validity. *Proceedings of the National Academy of Sciences*, 115(23), E5254.
- [44] Esteller-Cucala, M., Fernandez, V., & Villuendas, D. (2019). Experimentation Pitfalls to Avoid in A/B Testing for Online Personalization. *27th Conference on User Modeling, Adaptation and Personalization*, 153-159.
- [45] Eze, S., Awa, H., Okoye, J., Emecheta, B. and Anazodo, R. (2013). Determinant factors of information communication technology adoption by government-owned universities in Nigeria: a qualitative approach. *Journal of Enterprise Information Management*, 26, 4, pp. 427-443.
- [46] Fabijan, A., Olsson, H.H., Bosch, J. (2016). The Lack of Sharing of Customer Data in Large

- Software Organizations: Challenges and Implications. In: Sharp, H., Hall, T. (eds) *Agile Processes, in Software Engineering, and Extreme Programming*. XP 2016. Lecture Notes in Business Information Processing, vol 251. Springer, Cham. https://doi.org/10.1007/978-3-319-33515-5_4
- [47] Ferretti, S., Mirri, S., Prandi, C., & Salomoni, P. (2017). On personalizing web content through reinforcement learning. *Universal Access in the Information Society*, 16(2), 395-410.
- [48] Foris, D., Tecau, A. S., Hartescu, M., & Foris, T. (2020). Relevance of the features regarding the performance of booking websites. *Tourism Economics: The Business and Finance of Tourism and Recreation*, 26(6), 1021-1041.
- [49] Franceschini, F., Galetto, M., & Maisano, D. (2018). *Designing Performance Measurement Systems: Theory and Practice of Key Performance Indicators*. Boston, MA: Springer.
- [50] Gallo, A. (2017). A Refresher on A/B Testing. Harvard Business Review Home. Retrieved from: <https://hbr.org/2017/06/a-refresher-on-ab-testing>
- [51] Gans, Joshua S, Scott Stern and Jane Wu. 2019. Foundations of entrepreneurial strategy. *Strategic Management Journal* 40(5):736–756.
- [52] Ghorab, M. R., Zhou, D., O'Connor, A., & Wade, V. (2013). Personalized information retrieval: Survey and classification. *User Modeling and User-Adapted Interaction*, 23(4), 381-443.
- [53] Girotra, Karan, Christian Terwiesch, and Karl T Ulrich. 2010. "Idea generation and the quality of the best idea." *Management Science* 56(4):591–605.
- [54] Global e-commerce share of retail sales 2026. (2022, July). Statista. Retrieved from: <https://www.statista.com/statistics/534123/e-commerce-share-of-retail-sales-worldwide/>
- [55] Global E-commerce Sales Growth Report for 2021-2026. (2022). Shopify. Retrieved from: <https://www.shopify.com/blog/global-ecommerce-sales>
- [56] Gomez-Uribe, Carlos A and Neil Hunt. (2016). The Netflix recommender system: Algorithms, business value, and innovation." *ACM Transactions on Management Information Systems (TMIS)* 6(4):13.
- [57] González-Mena, G., Del-Valle-Soto, C., Corona, V., & Rodríguez, J. (2022). Neuromarketing in the digital age: The direct relation between facial expressions and website design. *Applied Sciences*, 12(16), 8186.
- [58] Gordon, B. R., F. Zettelmeyer, N. Bhargava, and D. Chapsky (2019). A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook." *Marketing Science*, 38(2):193–225.
- [59] Gordon, B. R., Moakler, R., & Zettelmeyer, F. (2023). Close enough? A large-scale exploration of non-experimental approaches to advertising measurement. *Marketing Science*, 42(4), 637–837.
- [60] Gräve, J. (2019). What KPIs are key? evaluating performance metrics for social media influencers. *Social Media + Society*, 5(3).
- [61] Gregurec, I., Tomičić-Pupek, K., & Kutnjak, A. (2021). Disruption caused business transformations in tourism-related industries. 2021 IEEE Technology & Engineering Management Conference – Eu-rope, 1-6.

- [62] Guha, P, Echagarruga, C, Tian, E. (2021). Optimising marketing strategies by customer segments and lifetime values, with A/B testing. *Applied Marketing Analytics*, 7, (2), pp 144–153
- [63] Guo, H., Pohl, E., & Gerokostopoulos, A. (2013). Determining the right sample size for your test: theory and application. In *Reliability and Maintainability Symposium*.
- [64] Gupta, S, Chokshi, S. (2020). Digital marketing effectiveness using incrementality. In: Singh, Mayank, Gupta, P.K., Tyagi, Vipin, Flusser, Jan, Ören, Tuncer, Valentino, Gianluca (Eds.), *Advances in Computing and Data Sciences*. Springer Singapore, Singapore, pp. 66–75.
- [65] Gupta, S, Kohavi, R, Tang, D, Xu, Y, Andersen, R, B, Eytan, C, Niall, Chandran, S, Chen, N, Coey, D. (2019). Top challenges from the first practical online controlled experiments summit. *ACM SIGKDD Explor. Newsl.* 21 (1), 20–35.
- [66] Hagar, N., & Diakopoulos, N. (2019). Optimizing content with A/B headline testing: Changing newsroom practices. *Media and Communication*, 7(1), 117-127.
- [67] Hardisty, D. J. and E. U. Weber (2020). Impatience and Savoring vs Dread: Asymmetries in Anticipation Explain Consumer Time Preferences for Positive vs. Negative Events. *Journal of Consumer Psychology*, 30(4):598–613.
- [68] Heinz, S., Linxen, S., Tuch, A. N., Fraßbeck, L., & Opwis, K. (2017). Is it still where I expect it? Users' current expectations of interface elements on the most frequent types of websites. *Interacting with Computers*, 29(3), 325-344.
- [69] Henseler, J., Ringle, C. M., & Sarstedt, M. (2016). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science*, 43(1), 115–135.
- [70] Ismail, H., Hussein, N., Harous, S., & Khalil, A. (2023). Survey of personalized learning software systems: A taxonomy of environments, learning content, and user models. *Education Sciences*, 13(7), 741.
- [71] Jahandideh, S.; Asefzadeh, S.; Barzegari, E. (2013). Dental services marketing: A case of dental clinics in Iran 46 Publications 134 Citations See Profile. *Dataset International Health Management*, 5, 12–18.
- [72] Jankowski, J., Hamari, J., & Wątróbski, J. (2019). A gradual approach for maximizing user conversion without compromising experience with high-visual-intensity website elements. *Internet Research*, 29(1), 194-217.
- [73] Jansen, B. J., Jung, S., & Salminen, J. (2022). Measuring user interactions with websites: A comparison of two industry-standard analytics approaches using data of 86 websites. *PloS One*, 17(5), e0268212-e0268212.
- [74] Jeganathan, K., & Szymkowiak, A. (2020). Social Media Content Headlines and Their Impact on Attracting Attention. *Journal of Marketing and Consumer Behaviour in Emerging Markets* 1 (10):49-59
- [75] Johnson, G. A. (2023). Inferno: A guide to field experiments in online display advertising. *Journal of Economics & Management Strategy*, 32(3), 469–490. <https://doi.org/10.1111/jems.12513>.

- [76] Jylhä, H., & Hamari, J. (2020). Development of a measurement instrument for visual qualities of graphical user interface elements (VISQUAL): A test in the context of mobile game icons. *User Modeling and User-Adapted Interaction*, 30(5), 949-982.
- [77] Keenan, M. (2022). 15 Critical Ecommerce Metrics You Must Track in 2022. Shopify. Re-trieved from: <https://www.shopify.com/blog/basic-ecommerce-metrics>
- [78] Kern, D., Hoek, W., & Hienert, D. (2023). Evaluation of a search interface for preference-based ranking - measuring user satisfaction and system performance. Ithaca: Cornell University Library.
- [79] Khlaisang, J. (2017). Proposing a new pedagogy-based website design: A usability test with life-long learners. *Education and Information Technologies*, 22(4), 1713-1735.
- [80] Kim, H S., & Kuljis, J. (2007). Cultural Manifestations on Website Design. 29th International Conference on Information Technology Interfaces, 19, 251-256.
- [81] King, R., Churchill, E. F., & Tan, C. (2017). Designing with data: Improving the user experience with A/B testing. O'Reilly Media, Inc. ISBN 9781449334833.
- [82] Kohavi, R., & Longbotham, R. (2015). Online controlled experiments and A/B tests. *Encyclopedia of machine learning and data mining*, 1-11.
- [83] Kohavi, R., & Longbotham, R. (2017). Online Controlled Experiments and A/B Testing. In C. Sammut & G. I. Webb (eds.), *Encyclopedia of Machine Learning and Data Mining* (pp. 922-929). Boston, MA: Springer.
- [84] Kohavi, R., Longbotham, R., Sommerfield, D., & Henne, R. M. (2009). Controlled experiments on the web: Survey and practical guide. *Data Mining and Knowledge Discovery*, 18(1), 140-181.
- [85] Kohavi, R., Tang, D., & Xu, Y. (2020). Trustworthy online controlled experiments: A practical guide to a/b testing. Cambridge University Press.
- [86] Koning, R, Sharique, H, and Aaron, C. (2019). Digital Experimentation and Startup Performance: Evidence from A/B Testing." Harvard Business School Working Paper, No. 20-018.
- [87] Koning, R., Hasan, S. and Chatterji, A. (2022). Experimentation and Startup Performance: Evidence from A/B Testing. *Management Science* (68:9), 6434-6453.
- [88] Krouska, A., Kabassi, K., Troussas, C., & Sgouropoulou, C. (2022). Personalizing environmental awareness through smartphones using AHP and PROMETHEE II. *Future Internet*, 14(2), 66.
- [89] Kumar, A and Yinliang, T. (2015). The demand effects of joint product advertising in online videos." *Management Science* 61(8):1921–1937.
- [90] Kumar, S. (2021). The Role of A/B Testing in Advancing Marketing Analytics: A Systematic Approach. *Journal of Scientific and Engineering Research*, 2021, 8(12):323-330
- [91] Kurz, Aaron & Kampik, Timotheus & Pufahl, Luise & Weber, Ingo. (2024). Business process improvement with AB testing and reinforcement learning: grounded theory-based industry perspectives. *Software and Systems Modeling*. 24. 87-109.
- [92] Lambrecht, A. and C. Tucker (2019). Algorithmic Bias? An Empirical Study of Apparent Gender Based Discrimination in the Display of STEM Career Ads. *Management Science*, 65(7):2966–2981.

- [93] Lee, S., & Koubek, R. J. (2010). The effects of usability and web design attributes on user preference for e-commerce websites. *Elsevier BV*, 61(4), 329-341.
- [94] Lestari, R. A., Muslim, E., & Moch, B. N. (2019). User interface evaluation of the official store for FMCG (fast-moving consumer goods) products in an e-commerce website using user experience approach. *IOP Conference Series: Materials Science and Engineering*, 505(1), 12079.
- [95] Levinthal, D. A. (2017). Mendel in the C-Suite: Design and the Evolution of Strategies.” *Strategy Science* 2(4):282–287.
- [96] Li, W., Zhou, Y., Luo, S., & Dong, Y. (2022). Design factors to improve the consistency and sustainable user experience of responsive interface design. *Sustainability*, 14(15), 9131.
- [97] Lin, M., Lucas, H. C., & Shmueli, G. (2013). Too big to fail: Large samples and the P-value problem. *Information Systems Research*, 24(4), 906–917.
- [98] Lu, Y., Sebe, N., Hytten, R., & Tian, Q. (2011). Personalization in multimedia retrieval: A survey. *Multimedia Tools and Applications*, 51(1), 247-277.
- [99] Luo, P., Ngai, E. W. T., Li, Y., & Tian, X. (2021). Effects of visit behavior on online store sales performance: Personal computer (PC) versus mobile channels. *Internet Research*, 32(3), 728-767.
- [100] Mandić, M., Gregurec, I. and Vujović, U. (2023). Measuring the effectiveness of online sales by conducting A/B testing’. *Market-Tržište*, 35 (2), 223–249.
- [101] Martín, M., Jiménez-Martín, A., Mateos, A., & Hernández, J. Z. (2021). Improving A/B testing on the basis of possibilistic reward methods: A numerical analysis. *Symmetry*, 13(11), 2175.
- [102] Masoero, L., Hains, D., & McQueen, J. (2023). Leveraging covariate adjustments at scale in online A/B testing. Ithaca: Cornell University Library.
- [103] Matz, S. C., Kosinski, M., Nave, G., & Stillwell, D. J. (2018). Reply to Eckles et al.: Facebook’s optimization algorithms are highly unlikely to explain the effects of psychological targeting. *Proceedings of the National Academy of Sciences*, 115(23), E5256.
- [104] Matz, S. C., M. Kosinski, G. Nave, and D. J. Stillwell (2017). “Psychological Targeting as an Effective Approach to Digital Mass Persuasion.” *Proceedings of the National Academy of Sciences*, 114(48): 12714–12719.
- [105] Miraz, M. H., Excell, P. S., & Ali, M. (2016). User interface (UI) design issues for multilingual users: A case study. *Universal Access in the Information Society*, 15(3), 431-444.
- [106] Mohammad, A. A. S, Al- Daoud, K. I, Al- Daoud, S, Samarah, A, Vasudevan, A., & Li, M. (2024). Content Marketing Optimization: A/B Testing and Conjoint Analysis for Engagement Strategies in Jordan. *Journal of Ecohumanism*, 3(7), 3086.
- [107] Mookerjee, S., Cornil, Y., & Hoegg, J. (2021). From waste to taste: how “Ugly” labels can increase purchase of unattractive produce. *Journal of Marketing*, 85 (3), 62–77.
- [108] Morales, A. C., Amir, O., & Lee, L. (2017). Keeping it real in experimental research understanding when, where, and how to enhance realism and measure consumer behavior. *Journal of Consumer Research*, 44(2), 465–476.
- [109] Mullin, S. (2022). A/B Testing: The Complete Guide with Expert Tips. Retrieved from:

<https://www.shopify.com/blog/the-complete-guide-to-ab-testing>

- [110] Muntean, M. I., Tarnaveanu, D., & Ion, A. R. (2016). E-Commerce Performance. Shopping Cart Key Performance Indicators. *Informatica Economica*, 20 (1), 62-71.
- [111] Nelson, L. Simester, D and K Sudhir. (2020). Introduction to the special issue on marketing science and field experiments. *Marketing Science*, 39(6):1033–1038.
- [112] Nguyen, T.H.; Le, X.C.; Vu, T.H.L. (2022). An Extended Technology-Organization-Environment (TOE) Framework for Online Retailing Utilization in Digital Transformation: Empirical Evidence from Vietnam. *Journal of Open Innovation: Technology, Marketing and Complexity*, 8, 200.
- [113] Nichifor, E.; Lix˘andriou, R.C.; Chit,u, I.B.; Br˘atucu, G.; Sumedrea, S.; Maican, C.I.; Tec˘au, A.S. (2021). Eye Tracking and an A/B Split Test for Social Media Marketing Optimization: The Connection between the User Profile and Ad Creative Components. *Journal of Theoretical and Applied Electronic Consumer Research*, 16, 2319–2340.
- [114] Orazi, D. C. and A. C. Johnston (2020). Running Field Experiments Using Facebook Split Test. *Journal of Business Research*, 118:189–198.
- [115] Ott, R. L., & Longnecker, M. (2016). An introduction to statistical methods and data analysis. Cengage Learning.
- [116] Pillai, Sandeep D, Brent Goldfarb and David A Kirsch. (2020). The origins of firm strategy: Learning by economic experimentation and strategic pivots in the early automobile industry.” *Strategic Management Journal* 41(3):369–399.
- [117] Priya, V., Subha, S., & Balamurugan, B. (2017). Analysis of performance measures to handle medical e-commerce shopping cart abandonment in the cloud. *Informatics in Medicine Unlocked*, 8, 32-41.
- [118] Quin, F, Weyns, D, Galster, M, Silva, C. (2024). A/B testing: A systematic literature review. *Journal of Systems and Software*, 211, 112011.
- [119] Runge, Julian and Harikesh Nair. (2021). Exploration in action: The role of randomized control trials in online demand generation.” Stanford GSB Working Paper.
- [120] Sahni, Navdeep S, Dan Zou and Pradeep K Chintagunta. (2016). Do targeted discount offers serve as advertising? Evidence from 70 field experiments.” *Management Science* 63(8):2688– 2705.
- [121] Saura, J.R.; Palos-Sanchez, P.; Herráez, B.R. (2020). Digital Marketing for Sustainable Growth: Business Models and Online Campaigns Using Sustainable Strategies. *Sustainability*, 12, 1003.
- [122] Sheng J, Liu H, Wang B. (2023). Research on the Optimization of A/B Testing System Based on Dynamic Strategy Distribution. *Processes*, 11(3):912.
- [123] Siroker, D., & Koomen, P. (2015). A/B testing: The most powerful way to turn clicks into customers. John Wiley & Sons.
- [124] Thomke, S. H. (2020). Experimentation Works: The Surprising Power of Business Experiments. Harvard Business Press.
- [125] Tornatzky, L.G.M.; Tchell, F.; Alok, K.C. (1990). The Process of Technological Innovation; Lexington Books; The Free Press: Cochranon, PA, USA.

- [126] Vafainia, S., Breugelmans, E., & Bijmolt, T. (2022). Calling Customers to Take Action: The Impact of Incentive and Customer Characteristics on Direct Mailing Effectiveness. *Journal of Interactive Marketing*, 45(1), 62-80.
- [127] Viglia, G., Zaefarian, G., & Ulqinaku, A. (2021). How to design good experiments in marketing: Types, examples, and methods. *Industrial marketing management*, 98, 193-206.
- [128] We are Social. (2022). Digital 2022 Global Overview. Retrieved from: <https://datareportal.com/>.
- [129] www.trends.builtwith.com (2025). Audience Measurement Analytics and Tracking Web Usage Distribution in Saudi Arabia. Retrieved from: <https://trends.builtwith.com/analytics/audience-measurement/country/Saudi-Arabia>
- [130] Xu, Y, Nanyu, C, Addrian, F, Omar, S, Anmol, B. (2015). From infrastructure to culture: A/b testing challenges in large-scale social networks. In *Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM pp. 2227–2236.

List of Appendices

Appendix 1. Interview Questions

Section A: General Questions

1. Please identify the number of full-time employees in your organization
2. Please share your organization's average Annual revenue in Saudi Riyal (SAR).
3. What industry or industries does your company operate in?
4. Please determine the percentage of A/B digital marketing testing in relation to the total marketing budget.
5. Who are your target customers?
6. What are the common payment options used by your customers?

Section B. Relevance of proposed A/B testing characteristics in enhancing online sales

7. Do you think that these A/B testing characteristics have a significant positive effect on boosting online sales: a/b testing variation design attractiveness, call-to-action effectiveness, and audience segmentation differentiated effectiveness? If yes, please explain how?
8. Does your marketing department pay attention to the A/B Testing Data Analysis Accuracy?

Section C: Specific questions for each construct

9. In your A/B tests, how do you usually go about creating variations (e.g., layout, color, content changes)? Could you provide an example of a test in which the variation design had a major effect on sales?
10. What difficulties did you encounter when developing various iterations of a digital marketing component (such as advertisements or landing pages) for A/B testing in the Saudi market?

11. How do you choose which call-to-actions to test in your campaigns such as "Buy Now," "Learn More," and "Sign Up"? Have you observed cultural preferences in the way Saudi customers react to various calls to action?
12. Could you give an instance when altering a call to action in an A/B test resulted in a discernible shift in sales or conversion rates?
13. For A/B testing in Saudi Arabia, how do you divide up your audience based on demographics, behavior, and geography? Which segmentation strategy has increased online sales the most?
14. Regarding various Saudi client segments (e.g., age groups, gender, urban vs. rural), have you noticed variations in A/B test results?
15. Which important metrics such as conversion rate, average order value, and bounce rate do you monitor to assess how well an A/B test is generating online sales?
16. Before making modifications, how can you make sure that an A/B test's findings are statistically significant?
17. In comparison to other markets, what particular difficulties have you faced while performing A/B testing for digital marketing in Saudi Arabia?
18. What effects can cultural or religious considerations have on the design and results of A/B tests in Saudi Arabia? Examples of these aspects include campaign timing, messaging, and imagery.
19. How do you make sure that A/B test findings are reliable and accurate before making data-driven decisions?
20. How would you respond if the findings of an A/B test were unexpected or contradicting?

Appendix 2: Survey

Dear Participant,

This is Bin Shaaib Hamzah Ahmed Ali, a postgraduate student at Midocean university. Currently, I am undertaking my master thesis: “The Role of A/B Digital Marketing Testing for Boosting the Online Sales in Saudi Arabia”. I would like to invite you to participate in this study. I appreciate you taking the time to complete this survey. The purpose of this study is to investigate the effects of three distinct A/B testing techniques on online sales performance in the Saudi Arabian market: variation design, call-to-action effectiveness, and audience segmentation. Your knowledge of digital marketing will be useful in identifying best practices and obstacles when it comes to optimizing A/B tests for higher conversion rates. It should take ten to fifteen minutes to complete the survey, and all answers will be kept private and used only for scholarly study. Your input to this significant study is much appreciated, and your experience is highly respected.

Section 1: Firm’s Profile

1. Firm Size (No. of employees)

- ☐ Freelance/Micro (1-5)
- ☐ Small (6-49)
- ☐ Medium (50-249)
- ☐ Large (250 and above)

2. Annual revenue in Saudi Riyal (SAR)

- ☐ Below 1 million
- ☐ 1–10 million
- ☐ 10–50 million
- ☐ 50-200 million
- ☐ Above 200 million

3. Type of Industry

- ☐ E-commerce
- ☐ Oil/Gas/Energy
- ☐ Retail & Shopping Malls
- ☐ Banking/FinTech
- ☐ Tourism/Hospitality
- ☐ Others

4. A/B Testing Budget (% of total marketing budget)

- ☐ <10%
☐ 10–30%
☐ 31–50%
☐ >50%

5. Target Market: (Select all that apply)

- ☐ Saudi Nationals
☐ Expatriates in KSA
☐ GCC Regional Customers
☐ B2B (Business Clients)

6. Payment Option: (Select all that apply)

- ☐ Credit/Debit Cards
☐ Cash on Delivery
☐ Mobile Wallets
☐ Buy Now, Pay Later
☐ Mada
☐ Bank Installments

Section 2: Please share your professional perceptions about three key aspects: variation design, call-to-action effectiveness, audience segmentation, data analysis accuracy, and the level of the firm's online sales. Please respond using the provided 5-point Likert scale where 1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, and 5 = Strongly Agree.

A/B Testing Variation Design Attractiveness	1	2	3	4	5
1. The variations in A/B testing are well-designed to better aligns with our brand identity than the original.					
2. The A/B testing variations are more visually appealing and user-friendly than the original.					
3. The differences between A/B testing versions were clear and meaningful.					
4. The A/B testing variation's technical performance and speed and technical are superior to the original.					
5. The A/B testing variation's messaging is more persuasive in driving conversions than the original.					

Call-to-Action Effectiveness	1	2	3	4	5
1. The call-to-action was clearer and more compelling than the original.					
2. The call-to-action placement improves visibility and click-through rates compared to the original.					

3. The call-to-action language was compelling and action-driven than the original.					
4. The call-to-action design (color, size, wording) was more effective in increasing clicks than the original.					

Audience Segmentation Differentiated Effectiveness	1	2	3	4	5
1. Demographic segments (e.g., age/gender/location)					
2. Behavioral segments (e.g., past purchases/page views)					
3. Device-type segments (mobile/desktop/tablet)					
4. Customer status segments (new vs. returning vs. lapsed users)					

A/B Testing Data Analysis Accuracy	1	2	3	4	5
1. The analysis of A/B testing results accurately reflected the differences between test variations.					
2. The interpretation of A/B testing outcomes was precise and free from analytical errors.					
3. The A/B data analysis correctly identified significant effects between experimental groups.					
4. The conclusions drawn from A/B testing were based on accurate analytical procedures.					

Online Sales	1	2	3	4	5
1. Our online sales revenue has increased since implementing A/B digital marketing testing.					
2. We have observed a higher conversion rate from website visitors to paying customers after using A/B testing strategies.					
3. The number of completed transactions on our website has significantly improved due to A/B testing.					
4. A/B testing has contributed to a noticeable growth in repeat online purchases by customers.					
5. Higher overall revenue per visitor during the test period.					

Thank You

LIST OF ABBREVIATIONS AND ACRONYMS

ABT-VDA	A/B Testing Variation Design Attractiveness
CTAE	Call-to-Action Effectiveness
ASDE	Audience Segmentation Differentiated Effectiveness
A/B-DAA	A/B Testing Data Analysis Accuracy
OS	Online sales
TOE	Technology-Organization-Environment
EBM	Evidence Based Management
SR	Saudi Riyal (Local currency)

ACKNOWLEDGEMENT

First and foremost, I would like to express my deepest gratitude to my beloved parent, whose unwavering encouragement, motivation, and prayers have always been my greatest source of inspiration.

I would also like to convey my heartfelt thanks to my respected supervisor, Mohamed Ewies, and teacher Heba Huzien for their dedicated guidance, patience, and invaluable academic expertise. Their insightful feedback and consistent support have been instrumental in shaping and completing this research. This thesis could not have been accomplished without their committed supervision.

Furthermore, I sincerely appreciate the help and companionship of my friends and classmates, who stood by me and offered assistance whenever I needed it. Their presence made this journey both meaningful and manageable.

Lastly, I would like to acknowledge all the participants of this study. Their willingness to contribute their time and share valuable insights played a vital role in making this research possible. I am truly grateful for their participation.

بحث مقدم للحصول على درجة الماجستير في التسويق الرقمي

دور اختبار التسويق الرقمي أ\ب في تعزيز المبيعات عبر الإنترنت في المملكة العربية السعودية: دراسة ميدانية

إعداد الطالب:
حمزة أحمد علي بن شعيب
251000062

إشراف:
د. محمد العويس