Athens Journal of Business & Economics –
Volume 10, Issue 4, October 2024 –Pages 253-270

Evaluation and Further Development of the B2B Startup Experimentation Framework: Application in B2C E-Commerce

By Carsten H. Hahn[*]*, Vili Sotirova*[±]*, Patrick Brecht*[°] *& Anja Ströbele*[•]

Different challenges and uncertainty arise from digital transformation as managers are forced to find new channels or make alternative investments decisions. Companies can use experiments as knowledge-generating resources to mitigate the uncertainty surrounding the adapted business model. The B2B Startup Experimentation Framework (B-SEF) was developed for startups in the B2B environment to discover and validate a business model's desirability through business experiments. This study uses the criteria completeness, consistency, plausibility, accuracy, and feasibility to investigate to what extent the B-SEF can be adapted to conduct business experiments in a B2C environment. Based on the B-SEF approach, two experimentation rounds are conducted regarding multiple online advertising channels and their efficiency in generating new customers in a B2C cosmetic online shop. Findings show that the B-SEF's generic structure is also suitable for conducting business experiments in the B2C environment when two adjustments are made regarding the sales funnel in the macro-level of the framework. First, the funnel levels required reordering to represent the customer journey better. Second, the new funnel level "awareness" needs to be added, as tracking awareness is relevant in the success of an e-commerce store. This research contributes by providing a guideline for entrepreneurs who want to conduct similar business experiments and extracting the company's most and least efficient advertising channels to acquire new profitable customers. The study's originality lies in assessing the B-SEF's suitability in the B2C context and providing a tool for its application to conduct business experiments comprehensively and successfully, especially regarding documentation and data collection.

Keywords: *digital business models, validation, validation framework, business experiments, B2C e-commerce*

Introduction

McKinsey (2021) analyzed digital strategies for companies in the post-pandemic era as a significant increase in digitization was recorded. The study revealed nine out of ten companies believed they needed a digital business model transformation or have already implemented one to remain economically viable until 2023. Furthermore, 64% of the respondents believed in the need to build new digital

[*]Professor, Karlsruhe University of Applied Sciences, Germany.
[±]Research Assistant, Karlsruhe University of Applied Sciences, Germany.
[°]Research Associate, Karlsruhe University of Applied Sciences, Germany.
[•]Research Associate, Karlsruhe University of Applied Sciences, Germany.

https://doi.org/10.30958/ajbe.10-4-1 doi=10.30958/ajbe.10-4-1

businesses while 21% said they needed to integrate digital technologies into their current business model. Merely 11% claimed their current business model's profitability can be ensured until 2023 (McKinsey 2021). Accordingly, the COVID-19 pandemic has intensified the need to adopt digital technologies in business processes and tasks and to fundamentally integrate digital strategies into business models (Seiler 2021). Although transforming or introducing new business models fosters profitability, associated uncertainties and risks must be regarded. Due to evolving market dynamics, technological advancements, or changing customer demands, companies must adapt their business models, find new sales channels, or make new investments. Thus, managers are increasingly confronted with making business-relevant decisions in an uncertain environment. In the absence of historical data, decision-making might be based on intuition and experience. If this data type is available, it may not reflect current consumer behavior or interests accurately, and possibly turn out to be irrelevant data (Thomke and Manzi 2017). In the case of e-commerce, research has shown digital business models were also impacted by COVID-19. For instance, Chevalier (2022) reported 40% of e-commerce decision-makers in North America and Europe experienced tougher competition levels. Bhatti et al. (2020) observed strategies, focusing on expediting the supply chain and logistics, had to be adapted to accommodate changed consumer behavior and preferences. This change in consumer behavior was confirmed by Van Gelder (2023) investigating the United Kingdom (UK) who reported a 75% growth in online shopping between March 2020 and February 2021. Unsurprisingly, a significant decline in offline shopping was registered (Van Gelder 2023).

Business experiments pose a potential solution to navigate in uncertain environments (Bland and Osterwalder 2020). Experiments intentionally reproduce processes or events to generate knowledge about dependencies or relationships (Aityan 2022). Potential benefits include risk reduction, improved competitiveness, and facilitated and efficient decision-making (Mind Tools Content Team n.d.). However, business experiments can also be a waste of time or resources when unfruitful ideas are relentlessly pursued or results are misinterpreted (Toolshero n.d.). In the pursuit of a universal tool to conduct experiments, attention was drawn to the B2B Startup Experimentation Framework (B-SEF) by Brecht et al. (2021). The framework helps business-to-business (B2B) startups to conduct business experiments and quickly and cost-effectively validate a business model's attractiveness. It comprises a macro and a micro level, each consisting of four sequential steps. The macro level provides a general overview of the business model and the customer journey, while the micro level focuses on specific components and conducts experiments to validate different assumptions regarding the business model (Brecht et al. 2021).

This research is set out to investigate whether the B-SEF is also suitable for experimentation in the business-to-consumer (B2C) environment, while considering the known differences between the two markets. The B2C environment has a larger potential customer base than the B2B market and focuses on end-consumers as clients rather than other companies. Since the nature of customers, decision-making, relationships, transaction types, and sales processes differ between B2B and B2C markets (Werani 2012), it could be assumed that different approaches to experimentation

for the respective business contexts may be necessary. This research's main goal is to apply and initially evaluate the B2B validation framework in a B2C context by validating the business model of a chosen target company. To accomplish this, the model is examined under the premise of conducting validation experiments to improve the company's business model and identify efficient channels for online advertising. Therefore, the following research question is formulated: *To what extent can the B-SEF be used to conduct business experiments to improve a company's business model in the B2C environment?*

After applying the framework and evaluating it according to five criteria derived from the acceptance indicators by Gerberich (2011), this study concluded the B-SEF's generic structure is suitable for conducting business experiments in the B2C environment. The case study revealed a possible efficient application when two minor adjustments are considered in the macro-level of the framework. The funnel levels need to be reordered, which will improve the B2C business model representation in the customer journey. Furthermore, adding the new funnel level "awareness" showed to be relevant as tracking awareness plays a significant role in the success of an e-commerce store.

The remaining paper is structured as follows. Next, the next section provides a brief review of theoretical foundations. The third section describes the methodological approach chosen to answer the research question. In the consecutive parts, the results and findings are reported. This paper continues with discussing findings and deriving limitations and suggestions for future research. This paper concludes with final remarks.

Literature Review

B2C – B2B in E-Commerce

Regarding the differences between the B2B and B2C environment, one can highlight that B2B e-commerce encompasses electronic sales transactions between companies, who can either be suppliers and consumers. In contrast, B2C e-commerce companies are only suppliers, who sell goods or services to end customers via online shops (Aichele and Schönberger 2016). B2C markets have a larger potential customer base with lower sales per customer. The purchase decision is usually based on the emotional decision-making processes of an individual, entrepreneurs face less risk, and payments are likely made after the transaction. In contrast, B2B entrepreneurs must bear higher risks, transactions are more complex, and purchase decisions are typically made by a buying center. The latter two characteristics create a more elaborate sales process. Lastly, unlike end consumers, whose consumption is guided by emotions, business representatives typically make purely rational decisions based on the company's actual needs (Rėklaitis and Pilelienė 2019). Table 1 summarizes the main differences between the two markets:

Table 1. *Differences between B2C and B2B Markets*

Criterion	**B2C**	**B2B**
Target customer	End consumers	Companies
Market size	Large	Smaller
Sales volume (per customer)	Low	High
Decision maker	Individual	Group
Risk	Low	High
Purchase process	Short	Long
Immediate payment	Often	Not required
Transaction	Simple	Complex
Decision	Emotional	Rational
Demand based on	Desire	Need
Use of mass media	Essential	Avoidable

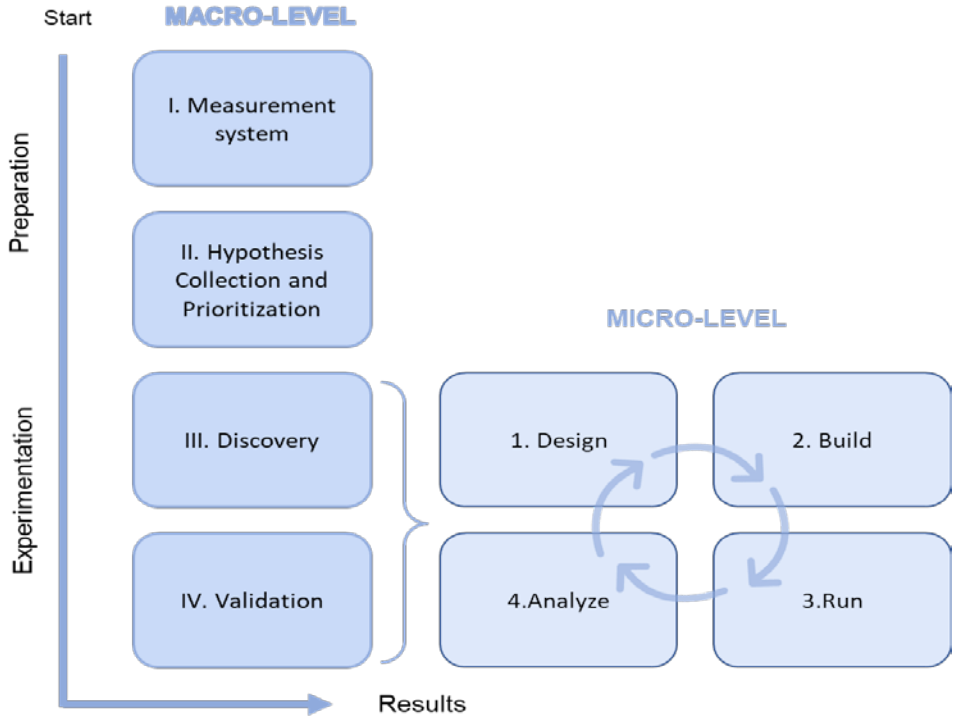
(Validation) Experiments

In research, two terms are essential for understanding the nature of experiments, namely validation and verification. Validation attempts to prove whether the correct product (or model) has been developed. The product or model is deemed valid if the usage goals and requirements are achievable. Verification, in comparison, assesses if the model meets the requirements enabling its performance while focusing on whether the model was constructed correctly (Mügschl-Scharf 2021). Experiments are carried out observations, during which data is collected and analyzed. These observations are normally conducted under artificial conditions (Aityan 2022). In experimentation, data is collected in multiple iterative cycles and interpreted with the help of hypotheses, which increases the reliability of findings (Thomke 2003). Therefore, experiments can support companies and decision makers in various business aspects, for instance, in attracting new customers or increasing customer satisfaction by testing and validating new products or business ideas or determining specific cause-effect relationships. Unlike intuitive decision-making, business experiments generate realistic data and new knowledge to validate products, processes, or ideas. Thus, facilitating a more sophisticated decision-making to minimize risks, reduce costs, and avoid uncertainty (Thomke 2020).

B2B-Startup Experimentation Framework (B-SEF)

This framework supports business-to-business (B2B) startups in conducting business experiments to validate a business model's attractiveness quickly and cost-effectively. As illustrated in Figure 1, the B-SEF framework consists of a macro and a micro level, each containing four sequential steps. The macro level is divided into two phases – preparing an experiment (preparation) and applying it (experimentation). The micro-level, inspired by Thomke's four-step iterative cycle, loops the startup iteratively through the design, build, run, and analyze phase to derive insights (Brecht et al. 2021).

Figure 1. *B2B-Startup Experimentation Framework*



The Macro-Level

In the first step of the framework, a measurement system is developed based on the McClure's customer acquisition funnel. The funnel visualization helps companies analyze and allocate resources efficiently in different stages of the sales process. McClure's funnel, also known as the Pirate Funnel, describes the customer journey from initial contact to revenue generation. The stages are acquisition, activation, retention, referral, and revenue (AARRR) (McClure 2007). There are various variations of the funnel, such as AIDA model, purchasing funnel, marketing funnel, sales funnel, or conversion funnel, depending on the business model and goals. Regardless of the chosen variantion, the funnel approach should support the effective experimentation in business settings (Arney 2021). To develop the measurement system, different metrics are assigned to each phase of the funnel and company-specific threshold values are defined, which play a relevant role for the later analysis of the results (Brecht et al. 2021). The second step focuses on formulating and prioritizing different hypotheses with the help of the business model canvas (BMC), which divides a business into nine relevant categories. After the canvas is filled out with all relevant content, assumptions are made for each field. In this context, potential business risks, which are extracted in the process, can be converted into hypotheses. Consecutively, these hypotheses are prioritized, and the most essential subjects transferred to the framework's second phase – the experimentation. Experimentation involves discovery and validation experiments. Discovery experiments verify a concept's core idea, while validation experiments

provide more meaningful evidence but require more resources such as time, personnel, and money (Brecht et al. 2021, Bland and Osterwalder 2020).

The Micro-Level

The micro-level consists of a four-step iterative cycle, inspired by Thomke (2003). First, the conceptual design of the experiment takes place, which is based on the existing findings and observations from the preparation phase (design). The tests are then designed next (build) and executed in a real-world environment to achieve increased external validity (run) (Brecht et al. 2021). Finally, the results are analyzed, and potential insights are derived (analyze). The previously developed measurement system is applied at this cycle stage to draw a comparison between the expected and the actual results. If the hypotheses are verified or falsified with sufficient accuracy, the experimental cycle is concluded. Otherwise, the newly gained knowledge should be used to modify the experiment accordingly and begin the cycle anew.

Methodology

Sample Choice

A real-life e-commerce store selling natural cosmetics was used to evaluate and further develop the validation framework. The company is suitable for the case study as it is a digital business model from the B2C industry operating from Austria. The cosmetic products for skin, hair, and oral care are suitable for men and women of all ages and are sold directly to end consumers via an online store. Therefore, they have been offered to the public via mass media.

Method

To answer the research question, the B-SEF was evaluated in a case study focusing on a B2C e-commerce shop. The framework was applied to conduct representative validation experiments, which were conducted in a natural environment with scenarios corresponding to the online store's actual work processes. Thus, field experiments with uncontrollable environments and potential disturbing factors interfering might be encountered. The experimentation derived two types of findings. On the one hand, the findings gave new insights into the current business models of the B2C e-commerce store and revealed potential improvement possibilities. Due to the scope of this study, hypotheses were formulated regarding prioritized advertising channels. On the other hand, the findings produced improvement potential for the framework itself. To extract this potential, the application was evaluated based on five evaluation criteria (Rabe et al. 2008, Gerberich 2011). The criteria originated from the work by Rabe et al. (2008) and were further enhanced by Gerberich (2011), creating the so-called acceptance indicators. The evaluation aimed to check the completeness, consistency, plausibility,

accuracy, and feasibility of the theoretical model. Table 2 shows the conception of the evaluation process.

Table 2. *Conception of the Evaluation Criteria*

Evaluation criteria	Description
Completeness	Evaluation in terms of absent relevant factors for the design of validation experiments.
Consistency	Evaluation of the coherence regarding interrelationships of individual components and the consistency of the framework's terminology.
Plausibility	Evaluation of the traceability of the entire framework
Accuracy	Evaluation of the level of detail and the granularity of the framework
Feasibility	Evaluation of feasibility/practicability in implementing the framework

Research Execution

The research is executed by applying the macro and micro-level of the B-SEF framework. Beginning with *Preparation*, a measurement system with the key performance indicators is developed to measure the success and failure of the experiments. Next, the main questions and hypotheses are derived from the business model canvas of the example company in *Experimentation*. Finally, the micro-level is assessed by looping through the four-step cycle.

Macro-Level: Preparation
Measurement System

When developing the measurement system based on McClure's CAF, the first peculiarity of the B2C online shop's business model was identified (McClure 2007). The pirate funnel and its sequence did not fully match the company's funnel. In this digital business model, customers typically made a paid purchase first before recommending the product or returning for a repeat purchase. The original funnel was developed for a different type of business model, namely software as a service (SaaS) and therefore could not be applied for the purposes of this experiment without adaptation (Winter 2022). To adapt it to the company's business model in the case study, the order of the funnel level was changed to acquisition, activation, revenue, retention, and referral. Customers generate revenue shortly after being activated, with repeat orders or referrals occurring, if they are satisfied with the product. The difference in the funnel order was not necessarily due to the B2B or B2C difference, but due to the type of business model. Additionally, a new level – *Awareness* – was introduced as an initial step, allowing potential customers to discover the product solving their problem (Jansen and Schuster 2011). This phase was measured and analyzed to determine factors that generated high awareness instead of revenue, which is typical for the e-commerce business model.

Table 3 provides an overview of all metrics from the B2C online shop's customer journey, which together represent the measurement system for the experiment. The

table shows six funnel levels. For each step in the customer journey, metrics have been defined to analyze successes and failures. Across most funnel levels, two metrics were considered. First, the absolute number of people who have reached this step in the customer journey. Second, this number is related to the total expenses to calculate the average cost incurred to achieve this event.

Table 3. *Measuring System of the Example Company*

Funnel level	Metrics	Description
Awareness (TOF)	Impressions	The frequency ads appearance among the target audience.
	CPM	The average cost per 1,000 impressions. Calculated by dividing the total amount spent on an advertising campaign by the number of impressions and multiplying by 1,000.
Acquisition (MOF)	Website Visitors	The total number of website visitors, corresponding to the number of link clicks as all links lead to the website.
	CPC	The average cost per link clicks. Calculated by dividing the total expenses by the number of clicks on the link.
Activation (BOF)	Adds to Cart	The number of attributed events where visitors have added something to the shopping cart.
	CPATC	The average cost per add to cart. Calculated by dividing the total expenses by the number of add to cart events.
Revenue (BTF)	Results	The number of successful events based on campaign goals; in this case, the number of orders.
	CPR	The average cost per order. Calculated by dividing the total expenses within the selected timeframe by the number of results.
	ROAS	The overall return on advertising spent from website purchases. Calculated by dividing the conversions from website purchases by the total amount of expenses.
Retention	Returning Customers	Percentage of recurring customers out of the total number of customers.
	CLV Customer Lifetime Value	Equivalent to customer lifetime value, representing an investment-based customer worth.
Referral	NPS Net Promoter Score	The percentage of customers likely recommending the company.

Hypotheses and Prioritization

The second step in the preparation phase was formulating hypotheses, which were tested in the experiment. Following the suggestion of the B-SEF, the hypotheses

were derived from the company's BMC (Brecht et al. 2021). In this study, the canvas element *channels* were given the highest priority, more precisely, online advertising channels due to their potential impact on the company's cost structure. The largest cost factor with potential for optimization was identified as the advertising costs for online ads on platforms such as Facebook, Instagram, Google, and YouTube. The case company needed to find an efficient way to use these channels for the different customer journey phases to optimize their resource allocation and revenue. The management's assumption that Facebook was the only profitable social media platform for customer acquisition was tested using different advertising channels. Therefore, the following hypotheses were derived and tested in parallel during the first experiment:

H1: Profitable customer acquisition can be achieved through advertising on Facebook.
H2: Profitable customer acquisition can be achieved through advertising on Instagram.
H3: Profitable customer acquisition can be achieved through advertising on Google.
H4: Profitable customer acquisition can be achieved through advertising on YouTube.

Customer acquisition was considered profitable if the costs and ROAS stay within the threshold values specified in the next step of the experimentation.

Micro-Level

To verify or falsify the hypotheses with sufficient accuracy, the micro experimentation cycle ran twice. Based on the insights of the first iteration, the second experimentation cycle ran with modified parameters.

Iteration 1

(1) *Design and Build:* The channels relevant for the experiment were selected, considering three main aspects: most common social media channels, the platforms targeted at an active audience, and the results of a competitive analysis. Facebook, Instagram, Google, and YouTube were the selected channels for the experiment's design. For each channel, several customer segments were identified based on the customer journey phases from the measurement system. Different advertising types were included to detect or neutralize potential side effects in the channel's performance. The content selection was based on a competitive analysis aimed to find widely spread ad types and content by various companies in the same industry generating high engagement. To increase the results' validity, another variable was considered in the experiment. The same ad types were displayed with content for two different products. More precisely, two carousels for two different products were shown for each customer segment on Facebook. This approach can increase the reliability of the hypothesis's falsification or verification if both products showed the same trend in the results.

Next, the relevant metrics for an experiment to verify or falsify the hypotheses were defined. The first four levels of the measurement system were used to track the customer journey, and all the metrics were measured and analyzed. Threshold values were set for each metric to evaluate the results in a later stage. The retention

and referral levels were excluded from the experiment due to time and budget constraints. Historical and predicted data were used to determine the threshold values for each metric, the calculations are summarized in Table 4.

Table 4. *Calculation of Threshold Values*

1. Historical Data (Jan 2022-May 2022)		**Threshold**
Customer Lifetime Value (CLV)	Average values from the internal database of the online store	45.00 €
Average Order Value (AOV)		31.00 €
Average Cost per Order (ACPO)		17.00 €
Click-Through-Rate (CTR)		2.35 %
Added-to-Cart-Rate (ATCR)		6.47 %
Conversion Rate (CR)		3.13 %
2. Calculation of Relative Key Figures (+36% deviation at BEROAS*)		
Break-Even-ROAS (BEROAS)	= (AOV / ACPO) × 0,64	1.2
CPM	= (CTR × CR × 1000 × CLV / BEROAS)	28.36 €
CPC	= (CR × CLV / BEROAS)	1.21 €
CPATC	= (ATCR × CLV / BEROAS)	2.49 €
CPR	= (CLV / BEROAS)	38.56 €
3. Calculation of the Absolute Key Figures		
Planned budget (BDG)		2,200.00 €
Impressions	= BDG / CPM × 1000	77,569
Website Visitors	= BDG / CPC	1,823
Adds to Cart	= BDG / CPATC	882
Results	= BDG / CPR	57
*To compensate for the average discrepancy resulting from tracking limitations of the Ads Manager tools, an additional 36% is added in the calculation of threshold values.		

This experiment aimed at profitable customer acquisition, therefore *cost per result (CPR)* and *return on ad spend (ROAS)* were deemed the most significant metrics since they measure average costs and profitability. For initial customer acquisition, CPR corresponded to customer acquisition costs (CAC). If the average CPR for a channel was higher and profitability was lower than the threshold, a business model selling through that online advertising channel was not validated, and alternative channels should be sought.

(2) Run and Analyze: During the first experimental cycle, all advertising materials were launched simultaneously without any modifications during the execution. After running for seven days with the originally planned budget, the results were analyzed using a developed Excel template to consolidate data from different channels. The template could potentially support the B-SEF's feasibility and applicability in practice. The hypotheses were not verified or falsified in the first cycle, prompting to retest certain factors under modified conditions in the second cycle.

Iteration 2

(1) Design and Build: The results from the first iteration cycle helped build an understanding of advertising channels. Therefore, the first experimentation cycle served as discovery experiment. To verify or falsify the hypothesis with a higher probability, an identical experiment with minor modifications was conducted for the second cycle.

(2) Run and Analyze: During the second cycle, ads were run continuously for another seven days with a higher budget. Unlike the first round, the performance of ads was monitored, and adjustments were made if necessary. Ads showing negative profitability for more than two days were turned off, and the budget was reallocated to better-performing ads. This approach made the second experiment more realistic and allowed more budget to be allocated to profitable channels and customer segments.

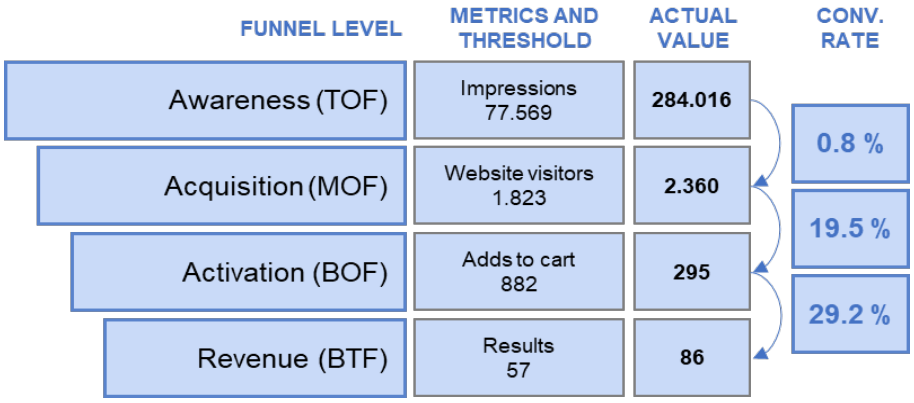
Results & Findings

Iteration 1

The analysis shows Facebook acquired the largest customer share via the ad type *blogpost*. However, these results cannot be considered valid as the meta-algorithm did not invest in the other available ad types, and thus prevented them from generating impressions. Therefore, in the next iterative cycle, all other ad types on Facebook were included in a new ad set to avoid side effects and analyze their performance separately. Instagram acquired customers with a ROAS below 1.0 for almost all ad types and customer segments, indicating that this advertising channel may not be profitable. It should be noted that the *blogpost* ad type, which produced above-average results on Facebook, was not available on Instagram. Therefore, it is recommended to further test Instagram, especially as a channel with a supporting role in a multichannel strategy. The last two channels, YouTube, and Google did not acquire any customers. Some ads did not generate impressions as no budget was spent. The reason was possibly because the algorithm did not collect any data and information on potential customers in the previous months, and the allocated budget could not be invested. Modifications can be made for these two channels in the second experiment cycle before excluding them from future advertising campaigns.

Figure 2 summarizes the results from the first experimental cycle in terms of the customer acquisition funnel. With a budget of 2,200 EUR, the advertising generated more than three times the expected number of impressions (282,016). The metric of website visitors also exceeded expectations, reaching a total number of 2,360 visitors. However, the expected add to cart events were not achieved due to inadequate conversion rates between these two funnel stages. Two of the four channels incurred costs without generating income, but the expected order number was exceeded. In the first experiment cycle, 86 new customers were acquired, exceeding the expected 57 orders.

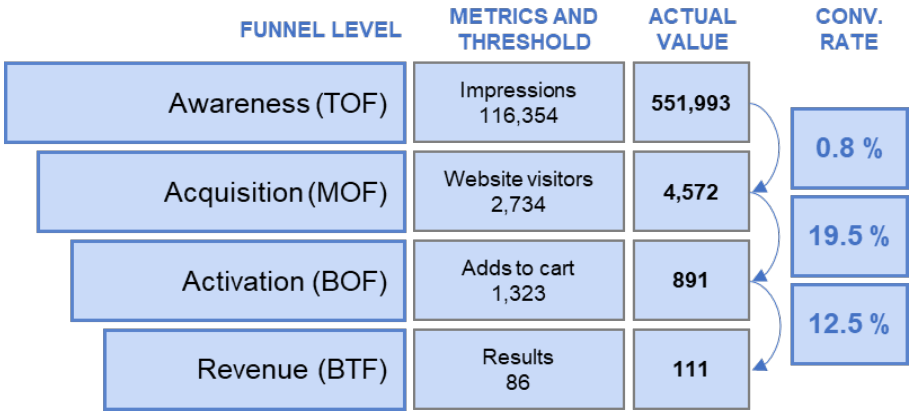
Figure 2. *Experiment Cycle 1, Results*



Iteration II

Based on the first cycle, four factors were identified that negatively affected the validity of the results, prompting to introduce four modifications. (1) Blog posts were published in a separate ad set on Facebook, (2) automatic placement was used to allocate budget on Instagram, for the (3) Google and (4) YouTube campaigns targeting clicks instead of conversions were chosen to make budget allocation for customer acquisition easier. Figure 3 shows the results of the second experiment cycle that was conducted with a 30% budget increase. With the new budget of about 3,300 EUR more than half a million users were reached, resulting in almost doubling the number of website visitors (4,572), 891 add to cart events and 111 new customers.

Figure 3. *Experiment Cycle 2: Results*



In conclusion, several insights for the company were gained through the two experimental cycles. Regarding hypothesis 1, which identified Facebook as a channel for profitable customer acquisition, was verified. However, the remaining three hypotheses, addressing Instagram, YouTube, and Google for profitable customer acquisition – were falsified. The B2C online shop should use these insights for further experiments by testing these channels using other ad content, customer segments, or methods. Despite the many variables influencing the field experiments,

it can be stated that Facebook, especially with the blogpost advertisement format, is suitable for profitable customer acquisition for the target company.

Evaluation of the Framework

The B-SEF is a generic model that includes all the necessary elements for conducting experiments in the B2C context effectively. Applying the model in the case study confirmed that it is ***complete*** regarding its contents. Separating the framework into macro and micro levels is useful for conducting experiments iteratively and in a structured manner. Although experiments could be conducted solely based on the micro level, the macro elements give a holistic view of the experiments and the business model. It enables the example company to prioritize and allocate its limited resources to relevant experiments. Even though it is not necessary to separate the experiments in the case study, separating discovery and validation experiments appears beneficial for tailoring experiment types, contents, and criteria to a specific company.

The evaluation criterion ***consistency*** was largely fulfilled, however, there is room for improvement in one aspect. The framework uses a coherent and understandable terminology applicable to the B2C context of the example company. The differences between the macro-level measurement system metrics and thresholds of the micro-level were not directly apparent. Specification regarding main differences with additional descriptions is recommended.

The B-SEF was found to be logical and intuitive (***plausible***) in the case study. The framework includes an iterative experimentation cycle, which was shown to be useful for testing and modifying successful customer acquisition channels in a B2C online shop. The structured testing and the resulting insights can motivate entrepreneurs and help them make informed decisions about their business model. The B-SEF is a plausible and structured framework that provides a clear and understandable connection between its elements.

The ***accuracy*** criterion was largely fulfilled as the framework has sufficient detail to be applied to the B2C context. Some elements of the framework needed to be adapted due to the specificities of the e-commerce business model, such as using a modified customer acquisition funnel. It is recommended to offer a selection of different customer acquisition funnels for different business models to make the B-SEF more efficient. The framework's flexibility allowed for business experiments with different scenarios, however further research was necessary to fully implement the framework. Therefore, the description of the elements and their application could be further refined.

The ***feasibility*** criterion has been largely met, but the initially developing certain elements, such as the measurement system or BMC, was time-consuming. Startups face challenges in designing business experiments to uncover underlying causality. For very young companies applying the B-SEF may be limited, and therefore potentially not feasible. In the case of the B2C online shop, extensive data was available, which made it easier to determine suitable thresholds. For startups with fewer data available, guidance for determining thresholds could be helpful as it is a time-consuming but important step in obtaining valid results. The practical

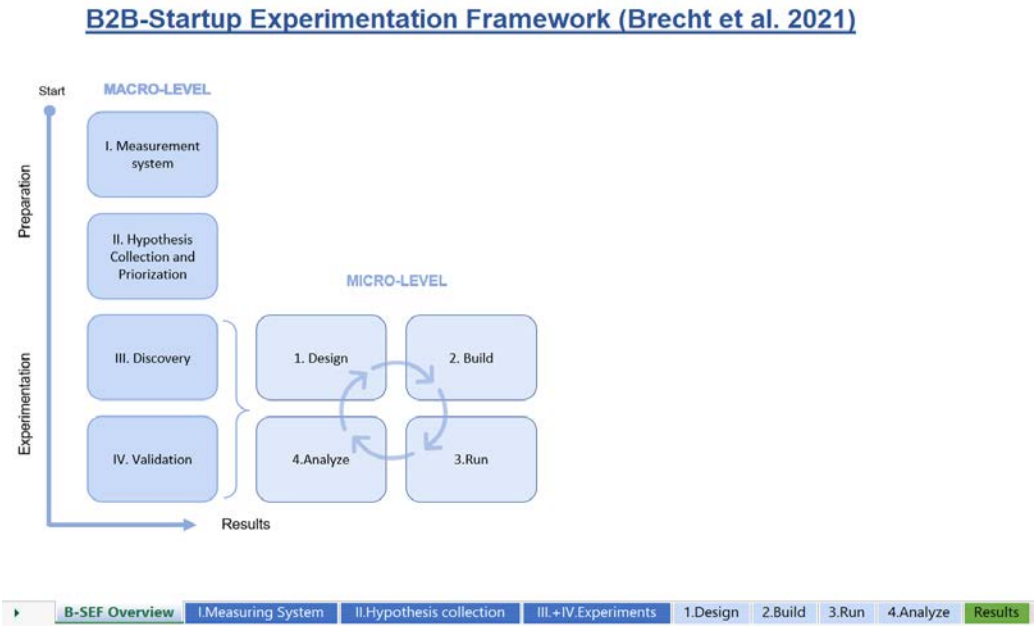
relevance of the B-SEF for B2C startups is undisputed, but some elements may limit its generalizability. A tool covering all relevant steps of the framework and enabling a structured evaluation of results could increase feasibility.

Further Development of the Framework

The B-SEF framework was applied in a B2C online shop, revealing significant potential for development. To address this, an Excel tool inspired by the B-SEF was created to aid in experiment planning, documentation, and analysis. The tool facilitates data collection, comparison, and evaluation, supporting decision- making for cross-channel advertising. Its universal applicability and user-friendly features should allow its usage for various companies and business areas. Figure 4 displays an excerpt of the Excel template, a tool encompassing multiple worksheets, each representing a step from the framework.

The first worksheet provides an overview of the B-SEF and its elements with hyperlinks facilitating easy navigation to each step's respective worksheet. Templates were created for both the macro and micro levels, simplifying each execution and documentation step. For example, the second and third worksheets contain templates of the customer acquisition funnel and business model canvas with instructions, supporting preparing the experiments on the macro level. Regarding the micro level, it allows users to conveniently summarize and consolidate prioritized hypotheses, chosen methodology, and success criteria using test cards. It facilitates effective planning and documenting the experiment design in the first step. In the second step, the tool enables users to develop and capture the necessary framework conditions for analysis, such as defining variables and their respective values. It ensures consistency and organizes information, which is crucial for accurate and reliable analysis. In the third step, the Excel tool serves as a platform for executing experiments, allowing seamlessly consolidating data from various advertising reports. It streamlines the data collection process and provides efficiency. Finally, in the fourth step, the tool simplifies data analysis by automatically transferring relevant information to the designated analysis worksheet. It saves time and reduces the potential for errors, enabling smoother result evaluations and interpretations. Overall, the framework steps are interconnected, with information from previous phases frequently required for subsequent phases. Linking relevant cells across worksheets simplifies information transfer between steps and reduces time and potential errors. The Excel tool serves as a comprehensive and practical resource, providing essential functionalities and capabilities to facilitate the execution, documentation, and analysis of experiments at each micro-level stage.

Figure 4. *Excel Tool for Operationalizing the B-SEF*



Discussion

In this paper, the research question was centered around examining to what extent the B-SEF can be used to conduct business experiments to improve a company's business model in the B2C environment. It was assumed that with minor adjustments to accommodate the differences between B2B and B2C markets, mainly in the supporting theoretical models, B-SEF can be effectively applied in a B2C context for conducting business experiments. The initial validation of the B-SEF involved analyzing the results from the conducted experiments, which provided significant insights into the business model and customer journey of the example company at the macro level. This helped identify a relevant problem regarding the efficiency of advertising channels. Building upon this, two experiment cycles were designed and executed at the micro level to identify profitable customer acquisition channels, revealing that only Facebook was effective out of the tested channels (Facebook, Instagram, Google, and YouTube). The experiments reached approximately 830,000 potential customers with a budget of 5,000 EUR, resulting in 167 new customers and an average customer acquisition cost (CAC) of around 31 EUR, below the threshold of 38 EUR, making the customer acquisition profitable.

Upon evaluation of the results and analysis of the data, the initial assumption was confirmed. The B-SEF framework shows promise in providing a structured and holistic approach to conducting experiments in the B2C context, offering valuable insights for startups and entrepreneurs in optimizing their business models. In terms of consistency, accuracy, and feasibility, the case study suggested that the framework is largely suitable for conducting experiments with room for improvement in structured execution and documentation. Detailed descriptions of individual elements

and using software solutions could enhance the framework's accuracy and feasibility. In this light, it is important to note that this evaluation represents an initial assessment of the framework within a single case study. Further research is required to validate and generalize these findings. Areas for improvement include clarifying the distinctions between macro- and micro-level elements, refining the elements and their application, and addressing challenges related to data availability and determining suitable thresholds. Overall, the B-SEF framework exhibits potential as a practical and relevant tool, but its generalizability and feasibility may benefit from further development and refinement.

Other limitations within this study relate to the available data and experimental conditions. The reliability of data collected by advertising platforms can be questioned because results from the internal system differ from those of the meta-ads manager as they only represent statistically aggregated estimates. Furthermore, interpretation of the results relied heavily on assumptions due to uncontrolled variables in field experiments. Future experiments are needed to increase the degree of validation of the findings. The evaluation of the B-SEF framework was limited to a specific B2C e-commerce company. It is recommended to apply and evaluate the framework in other experiment types and with other B2C business models to increase the validity of the results. Other suggestions for future research are to apply a quantitative approach, in which several companies from different industries and with different business models apply and evaluate the B-SEF. Another research area that should be examined in more depth is the evaluation of the framework through comparison with other comparable approaches. It might generate impulses for expanding the B-SEF. The developed tool for operationalizing the framework can also be used and tested in further studies and expanded for other types of experiments. Additionally, the suitability of the B-SEF for other business models should be investigated.

The findings of this study have significant implications for other managers and business owners in the field. Applying and evaluating the framework provided valuable insights into its potential uses and highlighted the benefits of its generalizability. Managers can leverage this knowledge to enhance their understanding and application of the framework, particularly in the context of identifying efficient online advertising channels. The case study presented in this research serves as a practical example, demonstrating how the framework can be effectively applied. Overall, the findings of this study contribute to this research field, enabling others to harness the advantages of the framework and its applications for improved business outcomes.

Conclusion

This work confirms the practical relevance of the B-SEF by describing another successful application and supporting its applicability in other case studies. The presented adaptation and application of the validation framework in an B2C e-commerce business model can provide guidance for similar businesses with comparable experiment goals and types. Especially in the areas of data collection and documentation, the case study showed the importance of working through the

process steps thoroughly to ensure success and extract the benefits provided by the framework. Future research should focus on more applications by startups operating in different industries to test the limits of the framework regarding the business model type and further confirm its applicability in the B2B and B2C environment.

References

Aichele C, Schönberger M (2016) *E-Business: Eine Übersicht für erfolgreiches B2B und B2C*. (E-Business: An overview for successful B2B and B2C). Wiesbaden: Springer-Verlag.

Aityan S (2022) *Business Research Methodology: Research Process and Methods*. Cham: Springer International Publishing.

Arney N (2021) *Seeing Digital Piracy in a New Light: Effects Causes and Reframing of Online Piracy from a Marketing Perspective*. Available at: https://papers.ssrn.com/sol 3 papers.cfm?abstract_id=3834761.

Bhatti A, Akram H, Basit HM, Khan AU, Raza SM, Naqvi MB (2020) E-commerce trends during COVID-19 Pandemic. *International Journal of Future Generation Communication and Networking* 13(2): 1449–1452.

Bland DJ, Osterwalder A (2020) *Testing Business Ideas: A Field Guide for Rapid Experimentation (Strategyzer)*. Hoboken, New Jersey: Wiley.

Brecht P, Hendriks D, Stroebele A, Hahn CC, Wolff I (2021) Discovery and Validation of Business Models: How B2B Startups can use Business Experiments. *Technology Innovation Management Review* 11(3): 17–31.

Chevalier S (2022) *Market competition evaluation according to e-commerce executives 2020*. Statista. Available at: https://www.statista.com/statistics/1173855/evalu ate-market-competition-ecommerce-companies-worldwide/.

Gerberich T (2011) *Lean oder MES in der Automobilzulieferindustrie: Ein Vorgehensmodell zur fallspezifischen Auswahl*. (Lean or MES in the automobile supplier industry: a process model for case-specific selection). Wiesbaden: Gabler Verlag.

Jansen BJ, Schuster S (2011) Bidding on the Buying Funnel for Sponsored Search and Keyword Advertising. *Journal of Electronic Commerce Research* 12(1): 1–18.

McClure D (2007) *Product Marketing for Pirates*. Available at: https://500hats.typepad. com/500blogs/2007/06/internet-market.html.

McKinsey (2021). *Companies Digital Business Model Changes*. Statista. Available at: https:// www.statista.com/statistics/1248716/companies-digital-business-model-changes/.

Mind Tools Content Team (n.d.) *Business Experiments - Taking Intelligent Risks*. Available at: https://www.mindtools.com/a26g880/business-experiments.

Mügschl-Scharf M (2021) *Validierung*. (Validation). Available at: https://medteconline.de/ info/magazine/wiki/validierung.

Rabe M, Spieckermann S, Wenzel S (2008) *Verifikation und Validierung für die Simulation in Produktion und Logistik: Vorgehensmodelle und Techniken*. (Verification and validation for the simulation in production and logistics: process models and techniques). Wiesbaden: Springer Publishing.

Rėklaitis K, Pilelienė L (2019) Principle Differences between B2B and B2C Marketing Communication Processes. *Management of Organizations: Systematic Research* 81(1): 73-86.

Seiler D (2021) *The new digital edge: Rethinking strategy for the postpandemic era*. McKinsey Digital Survey. Available at: https://www.mckinsey.com/capabilities/mc

kinsey-digital/our-insights/the-new-digital-edge-rethinking-strategy-for-the-postpandemic-era.

Thomke SH (2003) *Experimentation Matters: Unlocking the potential of new technologies for innovation*. Boston, MA: Harvard Business School Press.

Thomke SH (2020) *Experimentation Works: The Surprising Power of Business Experiments.* Boston: Harvard Business Review Press.

Thomke S, Manzi J (2017) *Versuch macht klug. Tests auf der Basis wissenschaftlicher Prinzipien können für Manager bei ihren Entscheidungen sehr hilfreich sein. So entwickeln Sie gute, praxisgerechte Experimente.* (Experiment makes smart. Test based on scientific principles can be very helpful for manager in decision-making. How to develop good and practice relevant experiments). Harvard Business Manager.

Toolshero (n.d.) *Business Experiments theory, forms and examples*. Available at: https://www.toolshero.com/decision-making/business-experiments/.

Van Gelder K (2023) *Percentage change in online purchases due to the coronavirus (COVID-19) pandemic in the United Kingdom from March 2020 to February 2021*. Statista. Available at: https://www.statista.com/statistics/1230225/changes-in-online-buying-among-uk-consumers-since-covid-19/.

Werani T (2012) B*usiness-to-Business-Marketing. Ein wertbasierter Ansatz.* (Business-to-Business Marketing. A value-based approach). Stuttgart: Kohlhammer (Kohlhammer Edition Marketing).

Winter B (2022) *Why activation is the most important pirate metric for SaaS growth. Appcies.* Available at: https://www.appcues.com/blog/pirate-metric-saas-growth.