

Doctor of Business Administration Thesis

UNIVERSITY OF LIVERPOOL

Linking consumer satisfaction to consumer behavior and retailer revenues: an empirical analysis of e-commerce performance influencers

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor of Business Administration by

Mark Kenneth James Chrystal

Date : 1 April 2017

First Supervisor: Dr. Pascale Hardy

Second Supervisor: Dr. Victoria Hanna

Word Count : 53011

Abstract

The aim of this research was to develop a validated set of actions that would lead to sustainable improvements in business performance and consumer attitudes for a production e-commerce website in the United States. In order to achieve the research objective a participatory action research methodology was utilized to generate quantitative and qualitative data and analysis through a mixed-methods approach. At completion, this effort was responsible for the development of new organizational knowledge and a series of managerial and business practice changes. The resulting business practice changes were credited with both short and longer-term business performance improvements. From an academic perspective, this research is one of the first to test a number of literature supported theories at real-world scale on a production e-commerce website. Of particular note are findings that extend the extant research into website consumer satisfaction measurement and the relationship between consumer feedback recency and issue identification. This thesis also highlights the importance of action research as a catalyst for organizational and individual meta-learning.

Declaration

I hereby certify that this Thesis constitutes my own work, that where the language of others is set forth, quotation marks so indicate, and that appropriate credit is given where I have used the language, ideas, expressions, or writings of another. I declare that the Thesis describes original work that has not previously been presented for the award of any other degree of any institution.

Table of Contents

CHAPTER 1: INTRODUCTION	5
1.1 PRACTICE-BASED PROBLEM OVERVIEW.....	7
1.1.1 <i>Firm and Industry Research Context</i>	8
1.1.2 <i>Preliminary Diagnosis and Approach</i>	8
1.2 RESEARCH AIMS AND OBJECTIVES.....	9
1.3 THESIS STRUCTURE.....	10
CHAPTER 2: LITERATURE REVIEW	12
2.1 UNDERSTANDING HOW CONSUMERS ASSESS ASSORTMENTS.....	13
Section Summary: Assortment Perceptions.....	15
2.2 UNDERSTANDING HOW CONSUMERS MAKE PURCHASE DECISIONS.....	16
2.2.1 <i>Consumer Costs</i>	16
2.2.2 <i>Purchase Determinants</i>	17
Section Summary: Purchase Decisions.....	19
2.3 CONSUMER SATISFACTION.....	20
2.3.1 <i>Connecting Perception Measurement to Behaviors</i>	20
2.3.2 <i>Satisfaction Measurement</i>	21
2.3.3 <i>Moderators of Satisfaction</i>	21
2.3.4 <i>Satisfaction, Value Perceptions, and Consumer Costs</i>	22
2.3.5 <i>Satisfaction and Service Quality Perceptions</i>	22
2.3.6 <i>Methods of Perception Measurement</i>	23
2.3.7 <i>SERVQUAL</i>	24
2.3.8 <i>Alternative Methods</i>	25
2.3.9 <i>Other Perception Measurement Considerations</i>	26
2.3.10 <i>Implementation of Findings</i>	27
Section Summary: Consumer Satisfaction.....	28
2.4 ELECTRONIC COMMERCE.....	29
2.4.1 <i>Connecting Perception Measurement to Behaviors</i>	29
Unique research into online shopping.....	29
Consumer Engagement.....	29
Influencers of Purchase Behavior.....	30
e-satisfaction Influencers and Measurement.....	31
Section Summary: Electronic Commerce.....	32
2.5 ACTION RESEARCH AND LITERATURE REVIEW CONCLUSION.....	33
CHAPTER 3: CONCEPTUAL MODEL AND RESEARCH HYPOTHESES	34
3.1 CONCEPTUAL INTERVENTION MODEL.....	35
3.2 CONCEPTUAL MODEL.....	35
3.3 RESEARCH HYPOTHESIS DEVELOPMENT.....	39
3.4 MODEL AND HYPOTHESIS SUMMARY.....	42
CHAPTER 4: METHODOLOGY	43
4.1 INTRODUCTION.....	44
4.2 APPLICATION OF ACTION RESEARCH METHODS.....	44
4.2.1 <i>Research Group Formation and Dynamics</i>	46

4.2.2 Research Group Working Practices.....	47
4.3 APPLICATION OF MIXED METHODS ACTION RESEARCH.....	50
4.4 QUALITATIVE DATA COLLECTION AND ANALYSIS METHODS	52
4.4.1 Data Collection	52
4.4.2 Data Analysis.....	53
4.5 QUANTITATIVE DATA COLLECTION AND ANALYSIS METHODS	55
4.5.1 Data Collection	55
First On-Site Survey Development.....	56
Second On-Site Survey Development	58
Post-Visit Survey Development.....	58
Click Data Collection Methodology	59
Colorization Test.....	60
Clickable Top Nav Test.....	61
Product Description Test	61
Video Test	61
Fly-Out Menu Test.....	62
Size Display Test.....	62
Click Data Methods of Measurement	62
4.5.2 Data Analysis – Survey Data.....	64
Scale Reliability and Dimensionality.....	65
Hypothesis Testing	65
4.5.3 Data Analysis – Clickstream Data	65
Hypothesis Testing	66
4.6 RESEARCH ACTIVITY TIMELINE.....	66
4.7 METHODOLOGY SUMMARY.....	67
4.7.1 Applicability of the Methodology to the Intended Outcomes	68
CHAPTER 5: RESULTS AND FINDINGS.....	69
5.1 ANALYSIS AND RESULTS	70
5.1.1 Hypothesis 1: Analysis	70
Clickable Top Nav Experiment	70
Fly-Out Menu Experiment v1	72
Fly-Out Menu Experiment v3	74
5.1.2 Hypothesis 1: Results.....	77
5.1.3 Hypothesis 2: Analysis	78
Colorization Experiment.....	78
Fly-Out Menu v2 Experiment	80
5.1.4 Hypothesis 2: Results.....	82
5.1.5 Hypothesis 3: Analysis	83
Qualitative Insights.....	87
5.1.6 Hypothesis 3: Results.....	88
5.1.7 Hypothesis 4: Analysis	88
Qualitative Insights.....	90
5.1.8 Hypothesis 4: Results.....	90
5.1.9 Hypothesis 5: Analysis	91
Qualitative Insights.....	93
5.1.10 Hypothesis 5: Results.....	93
5.1.11 Hypothesis 6: Analysis.....	93
5.1.12 Hypothesis 6: Results.....	96

5.1.13 Hypothesis 7: Analysis.....	97
Qualitative Insights.....	98
5.1.14 Hypothesis 7: Results.....	99
5.1.15 Hypothesis 8: Analysis.....	100
Product Description Experiment.....	100
Video Experiment.....	103
Size Display Experiment.....	105
5.1.16 Hypothesis 8: Results.....	109
5.1.17 Hypothesis 9: Analysis.....	109
5.1.18 Hypothesis 9: Results.....	114
5.2 PRINCIPAL RESEARCHER SUMMARY OF THE FINDINGS.....	115
5.2.1 Consumer Decision Costs.....	115
5.2.2 Purchase Justification.....	117
5.2.3 Product Information.....	118
5.2.4 Customer Satisfaction.....	119
CHAPTER 6: ACTION RESEARCH AND PERSONAL REFLECTIONS	121
6.1 PROJECT INITIATION PHASE	122
Meta-learning.....	124
6.2 EXPERIMENTATION PHASE	125
Meta-learning.....	126
6.3 RECOMMENDATIONS PHASE	127
Meta-learning.....	129
6.4 ACTION RESEARCH OUTCOMES – CONTRIBUTION TO PRACTICE	130
Meta-learning.....	131
CHAPTER 7: CONCLUSION.....	134
REFERENCES AND APPENDICES.....	139
REFERENCES.....	139
APPENDIX	152
Appendix 1: Initial Literature Supported Web Survey Design.....	152
Appendix 2: First On-Site Questionnaire	153
Appendix 3: Second On-Site Questionnaire	154
Appendix 4: Post-Visit Questionnaire.....	155
Appendix 5: Side-by-Side Survey Comparison and Academic Literature Support.....	156
Appendix 6: Colorization A/B Test Treatments	157
Appendix 7: Product Description A/B Test Treatments.....	158
Appendix 8: Video A/B Test - Control Treatment.....	159
Appendix 9: Fly-Out Nav A/B Test Treatments	160
Appendix 10: Size Display A/B Test Treatments.....	161
Appendix 11: Clickstream Data Definitions	162
Appendix 12: Survey Response Summary.....	163
Appendix 13: Charts, Tables and Figures Summary.....	165
Appendix 14: Ethical Approval	167

Chapter 1: Introduction

Ever since the earliest merchants decided to start stocking more than one item, there has been a debate about what and how much product to carry in an assortment. In fact, the term *retailer* is derived from the French word *tailleur*, meaning to cut-off or pare (Harper, 2014). From such beginnings, retailers have had to develop methods for determining which items to include and how best to convey their assortments in order to appeal to consumers. Due to recent changes in the way consumers want to shop for merchandise the retail industry is now teetering, according to Stacey's (2011) definition, at *the edge of chaos*. The sizeable change in consumer shopping preferences means that only the fittest organizations will adapt and survive (Stacey, 2011; Rigby, 2011).

The primary cause of the current turmoil in the retail industry comes from the increasing level of electronic commerce adoption by consumers (Rigby, 2011). The adoption of ecommerce has resulted in the exponential growth in information flow to consumers, which has increased product and price competition (Rigby, 2011). In just the last fifteen years, ecommerce has grown to become at least ten-percent of total retail sales revenue (Rigby, 2011). In this evolving environment, retailer survivability requires an organizational understanding of the new omnichannel landscape, and with it, a deeper understanding of customer needs and wants (Xia and Zhang, 2010; Rigby, 2011). Accordingly, it has been found that retailers who respond to customer feedback significantly outperform those who do not (Xia and Zhang, 2010). In order to do this, retailers must adapt from being product-centric organizations to customer-centric organizations. Retail organizations need to develop internal mechanisms that enable an understanding of customer preferences and then create the ability to act upon them (Wilson, 2002; Aubrey and Judge, 2012; Xia and Zhang, 2010; Rigby, 2011).

Organizations must adapt to changes in their environment in order to remain relevant (Grenier, 1972). Yet, the customer response and broader impact of any changes can be highly unpredictable (Stacey, 2011). It is within the context of adaptation causing unexpected results, that this thesis research has been conducted. Within the scope of the research was the need to develop a resolution to the change in customer behaviors that arose from the implementation of an updated website design. In order to develop a sustainable solution, the research effort adopted a mixed method action research approach. The findings from this research provide compelling insights into the benefits of a mixed method action research approach for creating organizational knowledge and sustainable change. The research also provides insight into how certain website design elements influence consumer behavior. Ultimately, the research led to the identification of a set of website design actions that were able to generate a positive financial outcome for the

organization. In addition, and more importantly, the action research process encouraged meta-learning and the recognition that specific organizational capabilities were in need of improvement.

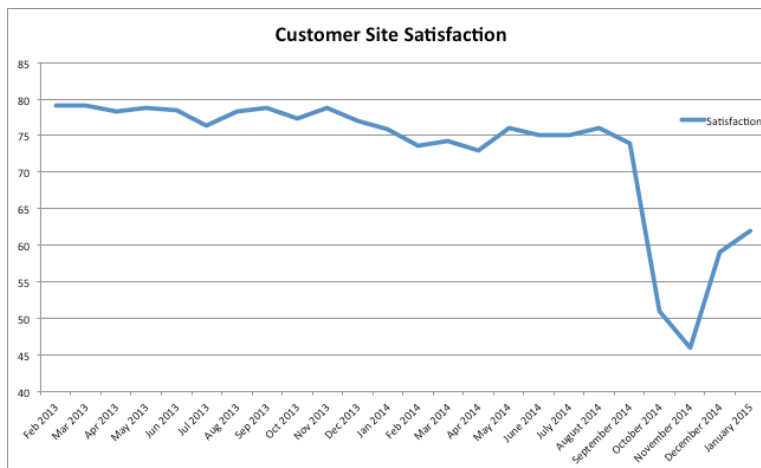
While this study was not intended to produce generalizable knowledge, it does provide practitioners with a conceptual basis for the development of a similar effort in their own organizations. For the academy, this thesis provides insight into the empirical and large-scale testing of several key consumer behavior and electronic commerce theories that were previously only tested under laboratory conditions. This research also provides an important contribution to the practice of mixed methods action research.

1.1 Practice-Based Problem Overview

Birling Gap Clothiers (BGC) is the fictitious name of the actual company that was researched for this thesis. Late in 2013, the company decided to move its website to an updated software platform. A website “re-platform” is a change in the software that underpins the site, which usually results in changes to the design, usability, and performance of the site (Kocher, 2009). According to the VP of Digital, the objective of the re-platform was to provide the foundation for continued growth of the ecommerce business, which had been growing at an annual rate in excess of twenty percent for the prior several years. The new software platform was determined to possess capabilities and scalability that the existing platform did not. The new software platform was implemented in mid-October, 2014. Due to the nature of the website re-platform, the implementation required a new website design to be implemented that could be supported by the new software platform. As part of the required design changes, BGC decided to try to incorporate some new functional elements made available through the software update.

Immediately upon implementation of the new website, and software platform, it became apparent that performance expectations were not being met. Commerce through the site dropped more than forty percent for the first few days, and then stabilized at around a thirty percent decline compared to the previous year. Customer satisfaction with the site also declined precipitously. This decline in customer site satisfaction was measured via an on-site survey that asked customers for their overall sentiment toward the website. Satisfaction measures are difficult to generalize across organizations; the best method for benchmarking perceptions is to compare changes across time, rather than across organizations (Parasuraman and Zeithaml, 1994). Data illustrating the change in customer satisfaction is illustrated in CHART 1-1, based on eighteen thousand combined responses over a two-year time period.

Chart 1-1: Customer Site Satisfaction (Feb '13 – Jan '15)



1.1.1 Firm and Industry Research Context

BGC is a multi-national apparel retailer that, according to its management team, holds a ninety-five percent level of brand awareness with its target consumer. The company also boasts a thirty-five percent market share within its core product-offering category. Despite possessing such a large share of the market, its share has declined over the last few years from a peak of forty percent. The management team has also divulged that the decline in market share coincided with a private equity acquisition of BGC, the replacement of the old management team, and the implementation of a new product pricing strategy.

At the time of writing, the retailer operated hundreds of physical store locations, in addition to an ecommerce website. The combined multi-channel retail organization delivered in excess of seven hundred million dollars in annual revenue, with approximately ten percent of this revenue being generated directly by the website. The management team believed that their specific segment of the industry was not growing, and was generally stable year over year. Therefore, in order to grow the business additional market share gains were required. The next largest competitor held about six percent market share, and the balance of the industry was comprised of thousands of individual “mom-and-pop” locations. In the years leading up to the point of research, the management team cited an increase in online competition. The increased competition was mostly coming from internationally based retailers with no physical presence in the United States, BGC’s home market.

1.1.2 Preliminary Diagnosis and Approach

In February of 2015, I was asked to join BGC as an external consultant, with the goal of rectifying the poor performance of the website and establishing a basis for future revenue growth. From an initial assessment of the website, it was clear that there were a large number of software bugs that needed to be remedied in order to return it to a basic operating standard. However, given the recent market share declines, which began over a year before the website re-platform, I suspected that other problems might also exist.

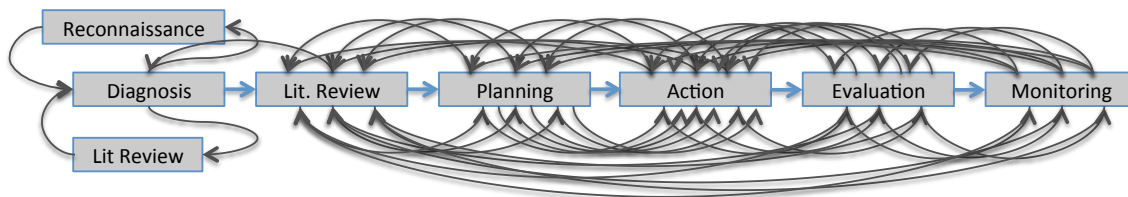
My preliminary diagnosis was that the performance decline stemmed from a failure to understand how their old website influenced customer perceptions, customer satisfaction, and customer demand. Therefore, as the company made changes to the design of the website, they may not have understood where and how they were impacting upon their ability to maintain revenue generation and customer satisfaction. This initial diagnosis stemmed from the relative lack of quantitative and qualitative site usage data that BGC’s digital team had access to. In addition, the VP of Digital stated that the new design was developed in support of the broader repositioning strategy, which on the surface did not appear to be resonating with customers, given the declining market share data. This initial diagnosis led me to believe that I needed to provide a research intervention that developed insights into the current performance. This intervention would also need to focus on organizational development, so that BGC would be better able to diagnose similar problems in the future.

The research presented by this thesis document is based on the effort that I undertook to further understand this problem and develop additional competencies within the business. This thesis document is intended to provide insight into how I was able to utilize scholarly and practitioner-based skills in order to improve the performance of BGC's business. In doing so, I will illustrate my academic research skills and my ability to translate those skills into academic quality work that has made a contribution to practice. Given the need to build a context-specific solution to this problem, it was ideally suited to the usage of an action research (AR) methodology (Coghlan and Brannick, 2010; Greenwood and Levin, 2007). In line with this, I have organized the structure of the thesis according to the AR methodology that was utilized during the research effort. More specifically, I adopted the action research methodology suggested by Ivankova (2015), which was based on Kurt Lewin's (1948) original framework. A synthesis of the research literature suggests a six-step cyclical approach to action research studies, as follows:

1. Diagnosis: Initial problem statement and conceptualization
2. Reconnaissance: Literature review and initial data collection
3. Planning: Definition of expectations, research and transformation design
4. Action: Data collection and co-research group meetings
5. Evaluation: Analysis of the data, reflection on results and transformations
6. Monitoring: Determination of research changes, continuations, or conclusions

While the literature purports an organized research process around these six steps, in practice, I found the process to be altogether messier, and far less linear in nature than I appreciated at the outset of the research. The practical reality of the action research process is that it is far closer to the process displayed in Figure 1-1, than the six-step process from the literature. However, I now appreciate the challenge of articulating such a dynamic and reflective process. Figure 1-1 illustrates the many cycles of reflection and the movement back and forth between the steps in the process that actually occurred during the research process.

Figure 1-1: Action Research Process in Practice



1.2 Research Aims and Objectives

At the conclusion of the research, the goal was to have co-developed a practically validated set of actions that could improve business performance and consumer attitudes. In addition to developing a better understanding of the problem and associated improvement actions, it

was also essential to connect those actions to practical changes in BGCs knowledge and business procedures. In accordance with an Action Research based methodology, the goal was to ensure that the research provided an organizational catalyst for new knowledge building and laid the foundation for sustainable improvements in operating practices.

While the primary goal of the research was to generate a sustainable performance improvement for the BGC organization, there were secondary aims related to the testing and validation of academic theories and retail practices. One of these secondary aims was to provide the basis for an improved understanding of the factors that may be influencing the performance of similar retailers. While generalizability was intentionally limited, it was hoped that the outcomes from this research would prove to be insightful for any practitioners dealing with similar practice-based problems. For the academy, this research may provide useful insights into the application of theories previously untested on a production ecommerce website, and into the usage of mixed-method and action research methodologies.

1.3 Thesis Structure

In order to understand the research outcomes from this thesis, it is important to have an understanding that a great many cycles of action and reflection on action occurred during the research process. This reflective process, and the meta-learning that occurred, ultimately provided the catalyst to longer-term benefits in the organization. The hypothesis testing and specific actions that are detailed here led to immediate benefits, but the process of generating those benefits proved to be far more important than the near-term outcomes. Yet, as with many of the noted academic authors, I found it difficult to write a thesis that sufficiently illustrates such a dynamic meta-learning process. For that reason, I have structured the thesis for ease of reader understanding, rather than as a chronological representation of the evolving, cyclical and, at times, chaotic nature of the practice of an action research project (illustrated in Figure 1-1).

Consistent with the action research literature, the structure of this thesis paper follows the general path undertaken by the actual research effort. The research effort began with an initial understanding of the problem and the determination of objectives, as just discussed. Next, a detailed review of the literature was conducted in order to build a strong theoretical understanding of the research area. Using insights from the literature, a conceptual model for conducting the research was co-developed with BGC, followed by the establishment of hypotheses. Once the conceptual framework and hypotheses were understood, a detailed method for undertaking the research was developed. This methodology included how each action, measure, and analysis method would be utilized to generate practical insights for the business. The results generated from the research were then analyzed, reflected upon, and often generated changes in the remaining experimental actions or business practices. After all experimentation was complete, final sets of analysis and personal and co-research reflections were completed. As the Principal Researcher, and in accordance with action

research methods, this included my own reflections on the research methodology and process. These reflections resulted in a final set of findings, actions, and implications for the business.

In summary, the thesis structure matches a very broad view of the utilized action research methodology (*shown below in parentheses*) but does not fully capture its dynamic reality, or the meta-learning cycles that underpin it. The written structure for the thesis is as follows:

1. Introduction (*Diagnosis*)
2. Literature Review (*Reconnaissance*)
3. Conceptual Model and Hypotheses (*Planning*)
4. Methodology and Action Plan Development (*Planning*)
5. Results, Findings and Reflections (*Action, Evaluation and Monitoring*)
6. Reflections (*Evaluation and Monitoring*)
7. Conclusions (*Monitoring*)

Chapter 2: Literature Review

This chapter provides a synthesis and discussion of the academic literature that informed the research process. The objective of the literature review was to provide the theoretical scaffolding for the research methodology and to improve the interpretability of the data that would eventually be collected. In this regard, the literature proved to be central to the framing of the problem, the development of potential solutions and to building researcher and business-based knowledge. The literature also served to challenge my personal assumptions about, and emotional responses to, the problem and the research environment.

Based on my initial diagnosis of the problem, I started the literature review with the aim of generating a better understanding of the problem areas that appeared to surround the issue. Given the strong decline in customer satisfaction, I conducted a broad review of the customer satisfaction literature, with a focus on how satisfaction connects with behaviors. I also felt that it was important to understand how previous research into consumer satisfaction had been conducted, including the identification of any generally accepted measurement methodologies. Given that the root of the problem may have been caused by trying to reposition the overall brand, it was also important to understand how consumers generally assess assortments. It was hoped that the literature in this area could provide insights into the types of assortment changes that could lead to changes in customer satisfaction and purchase propensity. With BGC's ultimate objective being to improve revenue generation, literature on how consumers make purchase decisions was also extensively reviewed.

Due to the fact that electronic commerce is a relatively new phenomenon, I found that the consumer behavior and satisfaction literature was not as developed in this area. As a consequence, I did not limit the aforementioned reviews to website studies only, but I did conduct research into electronic commerce so that I could understand how and where the literature review findings may be influenced by this particular shopping channel, as compared to others.

2.1 Understanding How Consumers Assess Assortments

Janiszewski (1998) and Moe (2003) found that consumers have two broad typologies: goal-directed and exploratory. Goal-directed consumers tend to approach a shopping environment with a specific task in mind (Janiszewski, 1998; Moe, 2003). These tasks generally include intent to purchase and specific search and product comparison activities (Moe, 2003). In contrast, exploratory consumers tend to browse a retailer in order to learn more about a general category of product or just for pure enjoyment (Moe, 2003). Moe (2003) further refined the broader typologies by stratifying the exploratory searchers into hedonistic and knowledge building consumers, and the goal-directed searchers into directed-buying and search/deliberation consumers. These typologies have been found to fundamentally impact how consumers perceive an assortment, the retail environment and the level of support they receive (Gupta and Kabadayi, 2010; Ha and Lennon, 2010; Moe, 2003; Janiszewski, 1998).

Research conducted by Hoch et al. (1999) suggests that location, regardless of consumer typology, is the primary driver of store patronage, followed by assortment variety and then price. However, Handelsman and Munson (1985) cite numerous consumer research studies that indicate that assortment size is the primary driver, followed by perceptions of assortment quality and then price. Regardless of the exact decision-making order, there is a significant interplay between each of these consumer patronage drivers where assortment variety, price, and quality perceptions all affect patronage preferences (Betancourt and Gautschi, 1990). As part of this decision-making process, Diehl, et al., (2003) found that consumers heavily consider the perceived ease with which they will be able to accomplish their goal-directed or exploratory objectives. As such, consumers are likely to consider assortment location (online, a physical store, a catalog, etc.) and their prior shopping experiences with that location as part of their location selection decision (Betancourt and Gautschi, 1990).

Once a preferred location has been selected, consumers choose a retailer based on perceptions of their assortment variety (Hoch, et al., 1999). Ultimately they choose a retailer with an assortment that they expect will provide them with the best opportunity to find a match for their needs (Mantrala et al., 2009; Hoch, et al., 1999). Three factors most commonly underpin a consumer's preference for assortment variety: desire to find a particular product type that satisfies an expected future need (Baumol and Ide, 1956; Mantrala et al., 2009), changes in tastes (Hoch, et al., 1999; Mantrala, et al., 2009), or a desire to try alternative product options within a particular category (Hoch, et al., 1999). Consequently, consumers tend to assess an assortment based on the likely availability of their preferred item and how readily they expect to find that item from amongst the balance of the assortment (Broniarczyk, et al., 1998). On this basis, consumers do not consider the sheer size of the assortment as the primary assortment attractor (Broniarczyk, et al., 1998). Instead, the key assortment desirability determinant is the likelihood to find a desirable purchase option.

One element that is central to consumer perceptions about the likelihood to find a desirable purchase option is the perceived quality of the assortment (Diehl, et al., 2003). A consumer research study by Chernev and Hamilton (2009) indicated that consumers tend to choose a smaller assortment when the options in that assortment are more attractive than those in a competing, but larger assortment (Chernev and Hamilton, 2009). Conversely, the lower the perceived quality of an assortment, the larger the assortment needs to be in order to attract customers (Chernev and Hamilton, 2009). Supporting these findings Chernev and Hamilton (2009) also found that expansions to a high-quality assortment, through additional SKUs or product features, led to less improvement in assortment attraction than the same expansions did in lower quality assortments. Therefore, the perceived quality of an assortment directly impacts upon the need for assortment variety.

Based on the importance of perceived option quality within an assortment, it appears that the ideal strategy is to select only the most preferred products within each product category. This is evidenced by Wal-Mart's strategy to focus their category assortments on well-known brands within each category (Pellegrini, 2010), and by Apple's objective to limit the overall SKUs in their assortment to only those that are most appealing. These findings also stress the importance of research into the competitive breadth, types of SKUs and prices being offered in the marketplace. Such research is required in order to determine how customers perceive the assortment and how those perceptions compare competitively (Meeker et al., 2009). By conducting customer research, retailers can develop differentiation strategies that can directly lead to improvements in customer assortment perceptions (Meeker et al., 2009).

One consistent finding from the academic research is that the availability of product information plays a significant role in perceptions of assortment quality (Diehl et al., 2003; Mogilner et al., 2008; Johnsen, 2009). One approach for improving assortment information availability comes from providing a natural method of product search and filtering (Diehl et al., 2003). This appears to be especially important for large assortments (Diehl et al., 2003). In this regard, consumers find highly organized assortments to be more attractive than those without apparent organizing characteristics (Hoch et al., 1999). In support of this, Mogilner et al. (2008) conducted research that suggests consumers perceive assortment variety based on the number of clearly delineated categories in the assortment, rather than on how many total SKUs are actually offered. Hoch et al. (1999) also found that creating a perception of variety, through improved assortment communication and merchandising, could somewhat mitigate other assortment deficiencies. Based on these findings, the general conclusion is that consumers are attracted to assortments that make their search and decision-making process as easy as possible (Mogilner, et al., 2008; Johnsen, 2009; Broniarczyk, et al., 1998; Hoch et al., 1999).

Field research by Betancourt and Gautschi (1990) shows that promotions on particular items have a more beneficial impact on overall assortment demand than when other items are

promoted. They found that when considering the use of promotions to drive assortment interest, the ideal traffic driving and assortment appeal scenario is to promote items that are most complementary to the balance of the assortment. However, the items selected for promotion must also have high levels of demand when promoted (Betancourt and Gautschi, 1990). In this context, substantial knowledge of consumer product preferences is critical for promotional item selection (Meeker et al., 2009). This is because items that hold a relatively large percentage of the consumer's discretionary budget were also found to be more appealing on promotion than items that rank lower in total expenditure (Betancourt and Gautschi, 1990). However, there are caveats with implementing such an approach, as it has also been shown that retailers who focus on price promotions tend to attract customers that are highly price sensitive (Diehl et al., 2003). Therefore, trying to generate assortment interest through price promotion activities can have an adverse impact on the perception of assortment quality, which in turn can shift the profile of the customer-base to one that is more price conscious than it may have been previously (Diehl et al., 2003).

Similar to how specific price promotions can increase overall assortment appeal, assortment variety can also have the same effect. Putting all other factors aside, Johnsen (2009) found that product assortment variety was directly connected to additional purchase occurrences in the balance of the assortment. However, Hoch et al. (1999) found that simply duplicating elements of the assortment, as a means of improving variety perceptions, has the reverse effect. It is the proximity of complementary products that has been shown to directly improve purchase likelihood in the balance of the assortment (Dass et al., 2013). This affinity influence is especially strong when an item is price promoted or introduced into the assortment (Dass et al., 2013). Therefore, demand for each item in the assortment is a function of its associations (proximity, pricing, affinity) with each other item in the assortment (Dass et al., 2013). This means that demand for an item within an assortment is contingent on the pricing and positioning of itself relative to other complementary or competing choices (Dass, et al., 2013). In addition to these intra-assortment elasticity's are impacts from other categories within the broader assortment offering, or from other retailers (Dass, et al., 2013). On this basis, assortment attractiveness is also a function of how well the retailer merchandises the assortment for the consumer and is, therefore, not always dependent on assortment breadth expansion (Johnsen, 2009).

Section Summary: Assortment Perceptions

Synthesis of the literature in this area provided some important consumer behavior insights. For example, the specific objectives of a customer are central to their perceptions of a retailer and a retailer's assortment. The literature highlights that not all customers are seeking to purchase, many may be seeking to gather assortment information in anticipation of a future purchase. How well the customer's tasks are supported appears to be critical. In particular, navigational ease, pricing, promotions, assortment organization, assortment variety and assortment quality all seem to play an important role. These insights provide opportunity areas where BGC should assess customer perceptions.

2.2 Understanding How Consumers Make Purchase Decisions

2.2.1 Consumer Costs

While the research shows that consumers prefer retailers who offer assortment variety, there is also evidence that consumers are limited in their ability to process information and decide amongst broad numbers of alternatives (Broniarczyk and Griffin, 2014). In essence, this means that high levels of assortment information and product variety attract consumers, but ultimately both of these factors impede the ability to make a purchase decision (Broniarczyk and Griffin, 2014). Iyengar and Lepper (2000) put this perspective to the test and developed convincing evidence that consumers prefer larger assortments but purchase far less frequently from them than they do from a more limited assortment. Subsequent research has also validated these findings (see Broniarczyk and Griffin, 2014; Chernev and Hamilton, 2009).

With increases in the number of assortment choices come increases in purchase avoidance, delay, and product browsing (Broniarczyk and Griffin, 2014; Chernev and Hamilton, 2009). This is due to the additional choice complexity that larger assortments create (Broniarczyk and Griffin, 2014; Chernev and Hamilton, 2009). Betancourt and Gautschi (1990) posit that consumer purchase decision-making should be considered as a cost. Where the higher the decision-making cost, through lack of assortment filters or through a large number of comparable items, the higher the likelihood that consumers will find the cost too high to make a purchase decision (Betancourt and Gautschi, 1990). In fact, when thought of in these terms a number of factors infer costs upon consumers that affect their propensity to

Table 2-1 : Consumer Costs

	Consumer Costs	
	High	Low
Price	Expensive	Cheap
Assortment Size	Large	Small
Location	Distant	Nearby
Service Level	Low	High
Comparable Products	Mostly Well-Known	Mostly Unknown Brands
Assortment Knowledge	High	Low
Assortment Filtering Support	Poor	Strong

purchase: price, selection, location, service, brand and assortment knowledge (Betancourt and Gautschi, 1990; Krishnan, et al., 2002; Mogilner et al., 2008).

In a study of consumer actions prior to purchase, Urbany, et al (1989) found that uncertainty about how to choose between many similar assortment options leads to purchase deferrals due to the need to gather more product information before making a selection. Conversely, consumers with limited knowledge of an assortment, but presented with fewer product choices were found to be more likely to purchase than those with higher store familiarity or those presented with more choices (Urbany, et al., 1989). Interestingly, this implies that a customer that finds a suitable product from a limited assortment is more likely to purchase that product if they are unfamiliar with the store or brand. This study also showed that store and brand-familiar

customers tended to make faster purchase decisions when presented with clear assortment choices, but when presented with larger assortments it took far longer for this same customer to make a purchase decision (Urbany, et al., 1989).

Table 2-1 summarizes the research into consumer purchase behavior by illustrating how certain retail characteristics can impact upon a consumer's propensity to purchase. In this context, high costs are generally associated with purchase avoidance or purchase deferral. Conversely, retail characteristics that have low consumer costs generally lead to a higher likelihood of purchase. Most of these consumer purchase costs actually have time investment at their source. Consumers have a tendency to treat time as a limited resource with an inherent cost (Betancourt and Gautschi, 1990). Therefore, increasing or decreasing the amount time a customer must invest in order to navigate the store and assortment, prior to making a purchase, will have a direct impact on their likelihood to purchase (Betancourt and Gautschi, 1990). Betancourt and Gautschi (1990) provide an example from France where many local grocers went out of business as the wealth of the population increased. Primarily this was due to the negative correlation between time available to purchase groceries and wealth, which resulted in grocery purchasers opting for one-stop shopping destinations (large supermarkets) as a means of minimizing time costs (Betancourt and Gautschi, 1990).

The French grocery example is consistent with Urbany et al.'s (1989) research, which shows that an easing of consumer purchase costs can significantly improve perceptions of an assortment. One method for reducing these costs is to improve the level of assortment filtering that is available to the consumer. Mogilner et al., (2008) conducted testing that highlights the importance of product delineation through categorization. These authors show that categorization and filtering of products can be used as an approach for mitigating consumer purchase costs that come with the navigation of large assortments. They also found that these improvements increase the perception of assortment variety in smaller assortments. Interestingly, such categorization does not need to be implemented perfectly, as even abundant but imprecise categorization was discovered to improve perceptions over having few clearly communicated product categories (Mogilner et al., 2008). Such findings have strong implications for store merchandising, marketing and product navigation in both brick-and-mortar and electronic store types.

2.2.2 Purchase Determinants

Through the usage of behavioral modeling Liang (2010) posits that the level of utility certainty a consumer assumes about a product directly influences their likelihood of purchase and re-purchase. In other words, consumers buy items based on their belief that it will satisfy their needs, either in the short or long term (Lai, 1994). It seems that the more certainty that a consumer has about their probable level of satisfaction, possibly through prior experience or the recommendations of others, the more likely they are to purchase an item. Mantrala et al. (2009) cite research findings that indicate that the selection of particular assortment choices also leads to increased purchasing in other segments of the

assortment. Such affinities are derived because consumers buy products that are complementary to other products that they will or have purchased, not as separate activities (Lai, 1994). In line with this Huber et al. (1982) state that less dominant assortment choices can draw consumers to the dominant choices. This phenomenon is a critical component to consider in assortment construction and is often the reason that retailers adopt 'good, better, best' strategies. This type of tiered assortment strategy involves the selection of limited quantities of aspirational 'best' products to drive interest in the balance of the assortment.

Lai (1994) theorizes that consumers conceptualize their needs based on a network of interrelated product complements. Lai (1994) uses the example of dining room furniture, where the purchase of a new table is likely to lead to interest in new chairs, cabinets, drapes and carpet at some point in the future. In this example, it can be construed that the consumer would not be satisfied with just replacing the table. In addition, even though the table was purchased first, it is likely that these other items were evaluated and influenced the purchase of the table in some way. This concept is important for retailers to recognize as part of their merchandising, marketing, and promotional strategies. Overall, these findings also challenge the notion that simple Pareto charts can be used as a means of identifying SKUs to rationalize out of the assortment. Huber et al. (1982) state that the best method for determining such product affinity relationships is through a systematic program of testing.

The research findings related to product affinity relationships are probably best explained by a process of rationalization that consumers undertake when evaluating a purchase. More specifically, consumers tend to make decisions based on 'good- reasons' instead of 'good- options' (Sela et al., 2009). Accordingly, the literature suggests that purchase decision-making is based on a rationalizing process. This tendency to rationalize also appears to influence how consumers respond to certain types of assortments. Sela et al. (2009) found that larger assortments encourage consumers to purchase more utilitarian options than smaller assortments do. The hypothesis is that larger assortments make it harder to make a purchase decision; ultimately this leads to an increased tendency to select the choice that is easiest to rationalize (Böhm and Pfister, 1996; Sela et al., 2009). This is because guilt stems from having fewer 'good reasons' for the purchase of products that provide pleasure and not utility, especially when the purchase can be considered extravagant (Böhm and Pfister, 1996).

When consumers feel that they can 'reward' themselves for their prior utilitarian actions, larger assortments lead to increased purchasing of less utilitarian options (Sela et al., 2009; Böhm and Pfister, 1996). Consistent with these findings, Simonson (1999, pp.368) identified three psychological elements that underpin how consumers evaluate purchase choices: "ease of justification, ease of information processing and the activation of decision rules." The "activation of decision rules" describes how consumers are influenced in their purchasing by prior decisions and experiences. This research showed that consumers tend to

alternate their hedonistic and utilitarian purchase activity, rather than always selecting a middle of the road option.

Simonson (1999) conducted research, which validated the following about consumer purchase tendencies:

- Consumers can only evaluate a specific sub-set of the assortment at any particular time. As a result, retailers can target these sub-sets to improve purchase likelihood.
- Purchasing will increase for a high-priced, high-margin item if another item that is priced even more highly is introduced to the assortment.
 - This rule also works in reverse for items on promotion, where promoting one item in the assortment will increase the justification to buy that item, but promoting all items will remove that reason and likely lead to a deferred purchase, as decision cost is increased.
- A similar effect can be achieved by presenting the original high-priced item in a separate location from the balance of the assortment because it minimizes feature and price comparison with the other lower priced items
 - This works best if the separated item is a well-known brand name, which the consumer believes they have justification to purchase.

Once a purchase has been made, post-purchase regret is higher and satisfaction lower for products purchased from a large assortment of comparable choices (Broniarczyk and Griffin, 2014; Diehl and Poyner, 2010). This is because large assortments tend to reduce confidence that the purchaser was able to select the best option due to an inability to compare all of the choices (Broniarczyk and Griffin, 2014; Diehl and Poyner, 2010). Consumers also perceive larger assortments more negatively than smaller assortments if exact preference matches are not found (Diehl and Poyner, 2010). As a result, post-purchase satisfaction is lower when associated with larger versus smaller assortments (Diehl and Poyner, 2010).

Consequently, Diehl and Poyner's (2010) research supports Betancourt and Gautschi's (1990) findings that prior experience with a retailer will impact upon a consumer's location choice at the beginning of their next shopping effort.

Section Summary: Purchase Decisions

According to the literature, consumers are influenced in their purchase decision-making by two main factors: decision-making costs and choice rationalization. Accordingly, the lower the time investment required to find a desirable product, and the easier a product item can be rationalized, the higher the likelihood that it will be purchased. There appear to be many components to decision-making costs and choice rationalization. However, the literature suggested the need to consider how these two focus areas could be improved at BGC.

2.3 Consumer Satisfaction

2.3.1 Connecting Perception Measurement to Behaviors

Coelho and Esteves (2007) suggest that a key component and antecedent of many consumer behaviors is the level of satisfaction that is derived during the consumption process. Other research shows that it is quality, value and satisfaction perceptions that directly affect consumer behaviors (Cronin, et al., 2000). The literature suggests a general order for how these perceptions lead to behavioral reactions. For example, Udo, Bagchi and Kirs (2010) found that service quality is a direct antecedent of satisfaction. Kuo, et al. (2009) supports this view by showing that consumer satisfaction derives from perceptions of service quality and purchase value. Rajic and Dado (2013), in a retail industry study, found that customer satisfaction was the only direct and significant antecedent of consumer behavior.

Despite some slight variations in the literature, a reasonable conclusion from the literature is that consumer behavior has satisfaction as a direct antecedent. In turn, it appears that satisfaction has perceptions of quality and value as its direct antecedents (Cronin, et al., 2000). Albeit, that there is some debate about how heavily and consistently value perceptions contribute to levels of satisfaction (Udo, et al., 2010; Cronin, et al., 2000). Cronin, Brady and Hult's (2000) multinational empirical study found that value perceptions contributed more to quality perceptions than to satisfaction, except where value is the primary point of competition. As such, these authors conclude that service quality perceptions are more important to satisfaction than value perceptions are; yet, service quality perceptions are almost always influenced by value perceptions. As a result, it should be concluded that only through the collection of data regarding quality, value, and satisfaction is it possible to fully understand behavioral influences (Kuo, et al., 2009; Cronin, et al., 2000).

Despite the debate about how intertwined satisfaction, service quality and value perceptions actually are, there is considerable evidence to suggest that, combined, these perceptions are highly correlated with business performance (Naik and Srinivasan, 2015; Olsen, et al., 2014; Morgan, et al, 2005; Wicks and Roethlein, 2009; Udo, et al., 2010; Cronin, et al., 2000). In particular, it is consumer satisfaction that is believed to be the most indicative of the performance of a company (Diehl and Poynor, 2010; Newman and Werbel, 1973). Morgan, Anderson and Mittal (2005) found that the utilization of customer satisfaction feedback directly led to improvements in business performance through improvements in customer loyalty and a reduction in price sensitivity. Wicks and Roethlein (2009) also found causal relationships between consumer satisfaction perceptions and customer retention and overall business profitability. As a consequence, being able to measure consumer perceptions is essential for being able to determine where performance improvement opportunities exist (Naik and Srinivasan, 2015; Olsen, et al., 2014).

2.3.2 Satisfaction Measurement

The goal of customer satisfaction measurement is to understand perceptions and attitudes (Coelho and Esteves, 2007). Satisfaction is an emotion generated by one's experience versus expectations (Woodruff, et al., 1983). The associated emotion can either be positive, negative or indifferent (Devaraj, et al., 2002). Preceding this emotion is a cognitive appraisal of the experience based on a direct interaction (Woodruff, et al., 1983; Cronin, et al., 2000). Said another way, reflection upon how an experience differs from one's expectations leads to a level of satisfaction that influences subsequent behavior (Woodruff, et al., 1983; Cronin, et al., 2000).

In order to measure satisfaction, it is important to understand the cognitive appraisal process at its base (Parasuraman, et al., 1994). Consistent with the service satisfaction literature, two main factors have been found to moderate assortment satisfaction perceptions: the estimated quality of the assortment (Chernev and Hamilton, 2009; Diehl et al., 2003) and the relative consumer costs incurred for selecting the assortment (Hoch et al., 1999; Handelsman and Munson, 1985; Betancourt and Gautschi, 1990). Given that assortment and service satisfaction both appear to derive from the same antecedents, satisfaction measurement tends to consistently draw on these two factors (e.g. Parasuraman, et al., 2005; Parasuraman, Zeithaml and Berry, 1994; Cronin, et al., 2000).

2.3.3 Moderators of Satisfaction

In order to develop a measurement approach, it is important to note that consumer products fall into either *experiential* or *search* types (Lim, et al., 2015). Definitionally, *experiential* products must be interacted with in order to understand their attributes. In contrast, *search* products can be assessed without direct interaction (Lim, et al., 2015). As a result, when purchased online, search products tend to have higher levels of satisfaction than experiential products (Huang, et al., 2009). Most likely, this is because it is difficult to assess the quality and value associated with products that must be experienced first-hand, when shopping online. Therefore, service quality is likely to be more important for online consumers when they are shopping for experiential products than for other product types (Kim and Jackson, 2009).

Mogilner, et al. (2014) found, in a study of U.S. consumers, that individual satisfaction is influenced by temporal outlook. Their study indicates that individuals who are focused on the future tend to be more satisfied than those focused on present day needs. Much as search and experiential products moderate satisfaction levels, satisfaction appears to be further moderated by product usage intent. For example, a product that is associated with a new stage of life, like a wedding dress, is likely to have higher satisfaction than a product that is associated with the current stage of life (Mogilner, et al., 2014). As a result, it is reasonable to assume that consumers view different product types with different satisfaction perspectives (Mogilner, et al., 2014; Lim, et al., 2014).

2.3.4 Satisfaction, Value Perceptions, and Consumer Costs

Consistent with the literature on consumer costs and value perceptions, Szymanski and Hise (2000) found that online shopping satisfaction is highly correlated to convenience. Two factors probably contribute to this finding. First, at the time of this particular study, online shopping was still relatively new and offered a sharp contrast to the usual (physically intensive) shopping methods of the time. Second, and as discussed earlier, shopping convenience has been found to be intrinsic to consumer costs (Betancourt and Gautschi, 1990). In support of this, Lim et al. (2015) found that satisfaction and purchase intent were highly correlated with the quality of information flow to prospective consumers. As a consequence, both satisfaction and propensity to purchase can be improved by increasing the amount of relevant content on a website (Lim, et al., 2015). This is somewhat counter to the research of Urbany, et al. (1989) who showed that increased levels of assortment information actually led to purchase deferrals due to increases in consumer decision costs. This may indicate a difference between online and offline information processing costs.

2.3.5 Satisfaction and Service Quality Perceptions

Wicks and Roethlein (2009) point out that there is no universal definition of quality, and that it differs across and within industries. Consequentially, perspectives on quality have been found to be multidimensional in nature, especially with regard to service and e-service (Zeithaml, et al., 2002; Petnji Yaya, et al., 2012). Despite the varying definitions, measures of service quality have been found to be highly explanatory of changes in consumer satisfaction. For example, Udo, Bagchi and Kirs (2010) found that measures of e-service quality were able to explain 86% of variances in e-satisfaction ratings for websites. In line with this, Devaraj, Fan and Kohli (2002) found that 82% of e-satisfaction was attributable to service quality metrics. What is critical to this explanatory power is the ability to tailor the measure and method of capturing service quality perceptions to the industry and business being studied (Gaur and Agrawal, 2006; Zeithaml, et al., 2002).

2.3.6 Methods of Perception Measurement

Most of the academic research in this area is conducted under the ‘service quality’ moniker. However, there is debate as to whether service quality perception measurement is really different from satisfaction measurement (Wicks and Roethlein, 2009). The academy has generally concluded that it is difficult to measure perceptions of satisfaction, service quality and value separately (Kuo, et al., 2009; Cronin, et al., 2000; Petnji Yaya, et al., 2012). Therefore, almost all of the literature in this area involves capturing perceptions across each of these dimensions, despite the more limited moniker assigned to the research area. In further support of this, the International Organization for Standardization (ISO) requires the measurement of customer satisfaction as part of their quality management accreditation (see http://www.iso.org/iso/qmp_2012.pdf).

Table 2-2: Survey Approaches in the Literature

Empirical Customer Quality/Satisfaction Research: Survey Approach

<u>Author(s)</u>	<u>Scale Used</u>	<u># Questions</u>	<u>Valid Responses</u>	<u>Response Rate</u>	<u>Audience</u>	<u>Survey Distribution</u>
Kim & Jackson (2009)	5-point Likert	40	252	Unknown	Consumers	Online (web survey)
Changsoo & Tadisina (2008)	5-point Likert	25	204	13.5%	Cosumers	Mail or Online (web survey)
Diehl & Poynor (2010)	10-point Likert	1	120	N/A	Students	In-person
Kuo, Wu & Deng (2009)	5-point Likert	33	387	N/A	Students	In-person
Lim, Al-Aali & Heinrichs (2015)	7-point Likert	9	210	N/A	Students	In-person
Chang & Chen (2009)	7-point Likert	27	314	Unknown	Students & Consumers	Online (web survey)
Naik & Srinivasan (2015)	7-point Likert	15	1000	N/A	Consumers	In-person
Szymansk & Hise (2000)	7-point Likert	13	1007	48.0%	Consumers	Online (email)
Udo, Bagchi & Kirs (2010)	7-point Likert	21	211	N/A	Students	In-person
Devaraj, Fan & Kohli (2002)	7-point Likert	39	134	86%	Students	In-person

The most common method of capturing consumer information is via a survey (See Table 2-2). This is because consumer researchers seek behavioral data, but given difficulties observing large numbers of consumers they must resort to the surveying of attitudes and opinions (Coelho and Esteves, 2007). Consumer attitudes towards satisfaction, service quality, and value have traditionally been ascertained by asking consumers about their experiences compared to their expectations (Parasuraman, 1988; Parasuraman, et al., 1994; Woodruff, et al., 1983). The literature calls this expectancy-disconfirmation theory. However, empirical research by Parasuraman and Zeithaml (1994) showed that capturing consumer expectations was not as important as capturing consumer perceptions. This 1994 work validated prior research by Parasuraman, Zeithaml and Berry (1985), which has proved to be seminal in the measurement of consumer perceptions (Ya Lan, et al., 2015). The paper by Parasuraman, et al. (1985) provided the basis for the SERVQUAL method of measuring perceptions. This method has become one of the most widely used across the service industries (Naik and Srinivasan, 2015; Ya Lan, et al., 2015; Gaur and Agrawal, 2006; Kuo, et

al., 2009; Changsoo and Tadisina, 2008; Parasuraman, et al., 1994; Parasuraman, 2013; Changsoo and Tadisina, 2008; Dabholkar, et al., 1996).

2.3.7 SERVQUAL

SERVQUAL has become the generally accepted basis for customer satisfaction measurement across a broad array of different industries and settings (Parasuraman, et al., 1994; Parasuraman, 2013). Ya Lan, et al. (2015) found 367 SERVQUAL related articles in their review of the literature from 1998 to 2013. The SERVQUAL methodology involves the collection of consumer responses to 22 questions across five dimensions: reliability, responsiveness, assurance, empathy and tangibles. The dimensions capture information on the direct service experience, and the environment and methods used for service delivery. However, the SERVQUAL model is more appropriate for traditional (brick and mortar) service quality measurement than online or purely transactional processes (Parasuraman, et al., 2005; Kim and Jackson, 2009; Changsoo and Tadisina, 2008). Because websites rely on virtual service delivery, the standard SERVQUAL method is inappropriate for collecting consumer opinions (Kim and Jackson, 2009; Changsoo and Tadisina, 2008; Parasuraman, et al., 2005).

In response to the web usage concerns, Parasuraman, et al (2005) reframed the SERVQUAL model for service quality measurement on websites and called it E-S-QUAL. Adapting the traditional SERVQUAL model to an electronically focused model was essential due to the increasing prevalence of technology as the intermediary between employees and customers (Parasuraman, 2013). In fact, modern consumers no longer required human interaction, and actually prefer faster, streamlined, methods of achieving their consumption goals (Parasuraman, 2013). The reframing of SERVQUAL was accomplished by studying the dimensions that consumers considered to be the antecedents of online service quality. Their web-based research found that website design and reliability are the primary dimensions that affect website service quality perceptions (Parasuraman, et al., 2005). To a lesser extent, security/privacy and customer service were also found to be contributors to service quality perceptions (Parasuraman, et al., 2005). It was found that these lesser dimensions only became relevant in the case of a significant service issue (Parasuraman, 2013). In such cases, it was found that customer satisfaction becomes entirely dependent on the ability to interact with an employee in order to resolve their issue efficiently (Parasuraman, 2013). Due to this phenomenon, Parasuraman, et al. (2005) developed a separate measurement scale called E-RecS-Qual, which focuses more on problem resolution management (Petnji Yaya, et al., 2012).

The core E-S-QUAL satisfaction measurement scale focuses on four main components of e-satisfaction: information security, website performance, order fulfillment, and ease of site navigation (Kim and Jackson, 2009). Even though service quality is highly correlated with consumer satisfaction, the omission of consumer satisfaction from the E-S-QUAL and

SERVQUAL methods should be noted. Given that it is satisfaction that is most predictive of consumer behaviors and business performance an alternative or adjusted method is probably necessary (Gaur and Agrawal, 2006). However, there is considerable research in the retail industry specifically that suggests that such adjustments are absolutely necessary for that environment (Gaur and Agrawal, 2006; Dabholkar, et al., 1996).

2.3.8 Alternative Methods

In the E-S-QUAL literature, three dimensions have been consistently validated as being worthy of inclusion: information security, website performance and ease of navigation (Petnji Yaya, et al., 2012). As such, the E-S-QUAL order fulfillment dimension has consistently failed to find empirical support. It is hypothesized that this inadequacy stems from the fact that E-S-QUAL does not distinguish between retailers who sell services, those that sell products and those that sell both (Petnji Yaya, et al., 2012). The core dimensions have proven to be effective for measuring e-service quality, but further dimensions are required to improve context specific applications (Gaur and Agrawal, 2006; Dabholkar, et al., 1996; Petnji Yaya, et al., 2012).

One particular adjustment to the SERVQUAL model has become popular within the retail industry: the Retail Service Quality Scale (RSQS) (Gaur and Agrawal, 2006). RSQS uses SERVQUAL as a conceptual basis, but modifies its measures toward the physical appearance of a retail store (Gaur and Agrawal, 2006). RSQS retains seventeen of the original SERVQUAL survey items, and adds eleven additional items (Dabholkar, et al., 1996). The additional eleven items were based on a review of the shortcomings of SERVQUAL uncovered through a review of the literature and the further empirical validation (Dabholkar, et al., 1996). These new items primarily focus on physical and transactional service support. For example, store layout and ease of product location and store navigation were included (Dabholkar, et al., 1996). In a separate study, Rajic and Dado (2013) also found that physical retail “store atmospherics” were strongly related to service quality perceptions.

There is a considerable research into the antecedents of customer satisfaction in an online environment. As further support for the exclusion of the fulfillment dimension, Kim and Jackson (2009) found that it was not directly related to website satisfaction. A consistent conclusion is that website content, and ease of use are the primary drivers of online satisfaction (Green and Pearson, 2009; Devaraj, et al., 2002). In addition, product purchase value and percentage discount are also highly correlated with purchase satisfaction (Devaraj, et al., 2002).

Reflection on the literature suggests that in order to understand customer behavior, satisfaction should be the primary measurable. Parasuraman and Zeithaml’s (1994) study of 800 consumers found that direct assessment of customer perceptions was a more predictive method of determining customer satisfaction than trying to measure the difference between

expected service and actual service quality. However, their results suggest that augmenting any satisfaction data with an understanding of customer expectations would allow for the development of a targeted response (Parasuraman and Zeithaml, 1994). Therefore, in order to increase purchasing behavior, it is critical to understand where perceptions are sub-optimal, in either service quality or value. This supports the need for a survey method that captures satisfaction, quality, and value data directly.

2.3.9 Other Perception Measurement Considerations

The literature highlights a number of other factors to consider when collecting consumer perception data. One consideration is for the construction of the survey itself. For example, Coelho and Esteves (2007) studied whether five-point or ten-point scales were better for customer satisfaction measurement. These authors concluded that there was no significant difference between the two scales in terms of response rates, but that the larger scale provided improved explanatory power for the variables being measured. A review of the E-S-QUAL literature shows that two-thirds of researchers prefer a 5-point Likert scale (Petnji Yaya, et al., 2012). This preference in the E-S-QUAL research was also captured separately through my own summary of the empirical service quality research (see Table 2-3).

In terms of the length of the surveys, all seem to stay under forty questions. Across the ten articles summarized in Table 2-3, the average survey length is 22 questions. The surveys tend to be larger, when trying to cover a broad number of dimensions, including satisfaction. At the other end of the spectrum, satisfaction only surveys tend to have fifteen or fewer questions. In fact, Diehl and Poyner (2010), in their study of consumer satisfaction and assortment size, used a single question to measure satisfaction. In large part, control of the survey size is due to a focus on maximization of valid surveys and overall response rates. For mail-based surveys, in the area of online satisfaction measurement, response rates average between 8% and 15% (Changsoo and Tadisina, 2008). However, there is an increasing tendency to distribute web surveys (see Kim and Jackson, 2009; Chang and Chen, 2009; Changsoo and Tadisina, 2008), whose response rates are not easily calculated, and seldom published.

Table 2-3 : Theoretical Research Summary

Customer Quality/Satisfaction Research: Theoretical & Analysis Summary

<u>Author(s)</u>	<u>Theoretical Basis</u>	<u>Perceptual Dimensions Studied</u>	<u>Analysis Type</u>
Kim & Jackson (2009)	E-S-QUAL	information security, website performance, order fulfillment, ease of use	CFA
Changsoo & Tadisina (2008)	E-S-QUAL	Trust, communication, ease of use, assortment availability, website content, site functionality	CFA
Diehl & Poynor (2010)	Expectancy-Disconfirmation	Overall Satisfaction Rating	ANCOVA
Kuo, Wu & Deng (2009)	SERVPERF	ease of use, content quality, performance, communication	CFA / SEM
Lim, Al-Aali & Heinrichs (2015)	Expectancy-Disconfirmation	Post-purchase overall satisfaction rating	CFA
Chang & Chen (2009)	Expectancy-Disconfirmation	Post-purchase overall satisfaction rating	CFA / SEM
Naik & Srinivasan (2015)	SERVQUAL	Overall satisfaction ratings and influencers	CFA / T-tests
Szymanski & Hise (2000)	Expectancy-Disconfirmation	Convenience, Merchandising, Site Design, Security	EFA
Udo, Bagchi & Kirs (2010)	Expectancy-Disconfirmation	Convenience, content, security, ease of use, overall satisfaction	ERA / SEM
Devaraj, Fan & Kohli (2002)	SERVQUAL	Ease of use, content, value, time efficiency, empathy, reliability, responsiveness assurance	CFA

Empirical findings from the literature suggest that it is important to tailor the SERVQUAL concept to the specific industry and firm being researched (Gaur and Agrawal, 2006). This is because the results generated using direct application of both SERVQUAL and RSQS has produced unreliable research outcomes (Gaur and Agrawal, 2006). Zeithaml, et al. (2002) believes that researchers should focus measurement on areas that customers perceive to be important regarding the specific product or service (Zeithaml, et al., 2002). For example, a number of studies have found that employee responsiveness and reliability are not usually correlated with service quality perceptions (Zeithaml, et al., 2002; Devaraj, et al., 2002). So, the inclusion of such measures needs to be fully considered when developing a survey for implementation in a specific business environment.

2.3.10 Implementation of Findings

Once the survey data has been collected, there is only limited research on how businesses should respond to what they learn (Olsen, et al., 2014). As a result, there is little guidance and depth of knowledge for organizations on how best to do so (Parasuraman, 2013). This is compounded by the fact that measures of customer satisfaction are difficult to generalize, especially across different types of service provider (Changsoo and Tadisina, 2008). Therefore, given the limited academic research in this area, any insights will likely need to be modified to fit the specific application being considered (Changsoo and Tadisina, 2008). Even the ability to compare satisfaction ratings across organizations is limited and organizations are better off measuring changes in their own performance than basing any decisions on other comparisons (Parasuraman and Zeithaml, 1994).

Possibly due to the lack of academic research support, most organizations fail to respond to the consumer insights they generate (Olsen, et al., 2014). One of the main reasons that customer perception data is not heavily utilized as the basis of action in businesses is

because it is notoriously unreliable (Olsen, et al., 2014). Further to this point, more than 70% of retail firms report that they are dissatisfied with their ability to act upon insights derived from customer satisfaction measurement (Wilson, 2002). This likely stems from the usage of generic survey methods that are unsuitable to their specific environments, and to corporate cultures that do not favor action from survey results (Olsen, et al., 2014). In part, businesses may not be generating useful consumer perception data because they have become more focused on the measurements than on the actions to be taken (Wilson, 2002).

Olsen, Witell and Gustafsson (2014) describe *customer orientation* as being critical to an organization's ability to measure and act on consumer satisfaction data. In this regard, there are four phases of customer orientation: data collection, data analysis, information distribution and implementation of actions (Olsen, et al., 2014; Morgan, et al., 2005). This supports the need for internal reflection and interpretation of the results by the specific business unit being researched. So, not only is it essential to construct an appropriate survey, but also to create an environment where the data will be internalized and acted upon. In order to utilize the data to improve performance, it has been found that operational strategies that target service quality are the most effective (Udo, Bagchi and Kirs, 2010). As stressed earlier in this chapter, only through collection and analysis of data that covers satisfaction, service quality and value can appropriate operational strategies be developed (Wicks and Roethlein, 2009; Udo, Bagchi and Kirs, 2010).

Section Summary: Consumer Satisfaction

In combination, the literature in this area provides for a theoretical understanding of consumer satisfaction and how to measure and respond to consumer opinions in practice. The literature provides insight into the antecedents of consumer behavior, and how those antecedents may be measured and influenced. Specifically, a focus on consumer satisfaction influencers and methods of measurement are important for gaining an understanding of customer behaviors. Temporal outlook, service quality, perceived value, convenience and customer expectations all appear to contribute to consumer assortment and experience response. The literature also provides strong support for the measures and methods for understanding each of these areas, and for measuring consumer satisfaction overall. Also highlighted, are the problems many organizations face when attempting to translate satisfaction measures into actionable knowledge. The accumulation of these insights led directly to many of the tests and measurement constructs used in the research, all of which are discussed in more detail later in the thesis.

2.4 Electronic Commerce

2.4.1 Connecting Perception Measurement to Behaviors

Unique research into online shopping

Lee and Koubek (2010) define a website as a source of information, which is a single complete system connected by consistent functional attributes. Such functional attributes are the performance, design, usability and products provided by the website (Lee and Koubek, 2010). There are four main types of websites: “entertainment, information, communication, and commerce” (Lee and Koubek, 2010, pp. 330). Zhenhui, et al. (2010) cite the statistic that commerce websites are visited more than four million times per minute on average (pp.35), highlighting their importance to business (Tezza, et al., 2011). These e-commerce visits now generate annual sales in excess of \$400 billion dollars, in just the United States and China combined (Demangeot and Broderick, 2010; Peng and Kim, 2014). Other than travel, clothing and apparel have become the most purchased products on the internet (Ha and Lennon, 2010).

Commerce websites have proved to be a boon not only for commercial reasons, but also in the study of consumer behavior (Demangeot and Broderick, 2010). Online research provides the researcher with huge quantities of behavioral data, such as clickstream and page view data (Demangeot and Broderick, 2010). Clickstream data is a record of where and how often visitors click on any webpage. Page view data measures which pages are viewed, for how long, and the entry and exit points from the page. One notable piece of research, on 3,059 internet shoppers, found that there are high levels of similarity between online and offline shopper typologies (Ganesh, et al., 2010). The primary exception is that there is a higher prevalence of hedonistic browsing behavior online than in traditional brick-and-mortar retail (Ganesh, et al., 2010). Such findings can be generated through the usage of clickstream data, and its ability to highlight different customer typologies (Wang et al., 2010) Clickstream and page view data can also be useful for personalizing the experience to the specific typologies or even to individual consumers (Wang, et al., 2010).

From the available behavioral data, it has been found that a website’s homepage should appeal to a broad set of potential and current customers (Treiblmaier and Pinterits, 2010). This design aim comes from data that shows that websites attract both exploratory and goal-directed visitors (Treiblmaier and Pinterits, 2010). Therefore the primary entry point to a website should appeal to both typologies (Treiblmaier and Pinterits, 2010). However, Wang, et al (2010) found that retailers with strong market share, should assume a greater percentage of goal-directed visitors than otherwise, and tailor their web design toward task completion, not aesthetics, in order to better support these consumers.

Consumer Engagement

Given that shopping on a website is done without interaction with human sales associates, how consumers are presented with information on the website is extremely important to

their decision-making and to their perception of the company they are shopping with (Demangeot and Broderick, 2010). Demangeot and Broderick (2010) found, in a 301-person internet shopper study, that consumer engagement depended on how easily they can explore and make sense of the website. Specifically, the following site attributes have been found to be important: quality of page content including the clarity of product messaging (sense-making) and the navigability and visual displays on the site (ease of exploration) (Ha and Lennon, 2010).

Loyalty toward a website can be measured by repeat visits and the time spent on the website (Gupta and Kabadayi, 2010). Trust is an important component of website loyalty and requires confidence in the usability, security, and integrity of the site (Gupta and Kabadayi, 2010). Gupta and Kabadayi (2010) found that site loyalty influences are different depending on the consumer's typology (exploratory or goal-directed). Goal-directed consumers prefer websites that support their shopping needs through usability and integrity. Conversely, consumers who are "just browsing" prefer sites that provide more than just a transaction-oriented experience (Gupta and Kabadayi, 2010).

Influencers of Purchase Behavior

Much of the empirical research into online consumer behavior draws on Stimulus-Organism-Response theory (S-O-R) (Peng and Kim, 2014). The S-O-R consumer psychology framework was developed by Mehrabian and Russell (1976). The construct is that stimuli (shopping environment), affect organisms (consumers), which create a response (purchase or avoidance behaviors) (Peng and Kim, 2014; Zhenhui, et al., 2010; Wang, et al., 2010). The S-O-R model theorizes that changes in stimuli will impact an individual's cognition and therefore their behavior (Zhenhui, et al., 2010; Peng and Kim, 2014). Such stimuli changes could include differing the product information displays or the navigational elements (Zhenhui, et al., 2010; Wang, et al., 2010). As an example, Zhenhui, et al.'s (2010) S-O-R based study divided the test subjects between groups that received strong navigational cues and those that received limited cues. Using the S-O-R framework, it was found that providing improved navigation control to consumers increased their cognitive engagement and their purchase intent (Zhenhui, et al., 2010).

There is also considerable academic coverage of the usability of technology in the Human Computer Interaction (HCI) literature (Green and Pearson, 2009; Lee and Kozar, 2012). HCI forms much of the starting point for theories on website usability (Lee and Kozar, 2012; Green and Pearson, 2009). Lee and Kozar (2012) define website usability as "the extent to which web sites can be used by specified users to achieve specified goals" (pp.451). Websites with high usability allow their users to efficiently find relevant content and accomplish their desired tasks (Lee and Kozar, 2012). As a result, usability has been found to influence sentiment toward a website and purchase intent directly (Peng and Kim, 2014; Lee and Kozar, 2012).

Some competing points of view regarding content relevance have, however, been generated by the literature. Belanche et al. (2012) believe that simple product content is more important for commerce website sites than the use of rich content, like multi-media. In contrast, Lee and Kozar (2012) and Zhenhui, et al. (2010) found that rich multi-media content (videos or 360-degree product views) actually led to higher levels of cognitive engagement and purchase intent. It is important to note that the Belanche, et al. (2012) study was focused on the purchase of bus tickets. In comparison, Zhenhui et al. (2010) studied the purchase of non-fiction books and greeting cards. The literature suggests that customer, product and stimulus types all influence satisfaction and behavior (Ha and Lennon, 2010; Gupta and Kabadayi, 2010; Janiszewski, 1998; Moe, 2003). In this case, it could be argued that the purchase of bus tickets compared to the purchase books or greeting cards have very different information requirements and customer expectations. As such, finding a difference in multi-media perspectives is probably to be expected. These different research findings further reinforce the conclusions of Zeithaml's et al. (2002) and Changsoo and Tadisina (2008). It is difficult to generalize behavioral findings, and any measures, analysis, and conclusions need to consider the specific context and customer types being researched.

e-satisfaction Influencers and Measurement

Ladhari (2010) contends that there is no commonly accepted measure of customer satisfaction on a website. In part, this is due to the fact that this type of online research is relatively new (Ladhari, 2010). This is also due to the specific nature of each website and the need to tailor satisfaction measures based on the needs of that specific customer (Fassnacht and Kose, 2007; Ladhari, 2010; Finn, 2011). This finding is consistent with the broader research on consumer satisfaction and service quality, as discussed earlier (e.g., Olsen, et al., 2014 and Zeithaml, et al., 2002). In further support of this point, a study of 1,634 internet shoppers found that satisfaction with a website varies based on each customer's site usage objectives (Ha and Lennon, 2010). For example, consumers that are goal-directed tend to be more satisfied when a website provides them with information that is relevant to their purchase desires. Whereas browsing consumers prefer visual stimuli more associated with the overall brand experience than with specific product purchasing tasks (Ha and Lennon, 2010). All of which make it difficult to develop broad methodologies for e-satisfaction measurement.

While there is no universally acceptable way to measure satisfaction (online or offline), the electronic commerce literature does uncover some common themes. In particular several key factors are common to e-satisfaction: website performance, trust, ease of use, data security and relevant site content, site functionality and aesthetic appeal (Green and Pearson, 2009; Ladhari, 2010; Belanche, et al., 2012; Wang et al., 2010; Wang, et al., 2011; Peng and Kim, 2014). The literature also suggests that many of these same factors impact the time spent on a website, the purchase intent and the desire to return (Green and Pearson, 2009; Belanche, et al., 2012). Peng and Kim (2014) posit that it is because these

factors elicit a cognitive and emotional response in website visitors, that they directly influence behaviors.

Wang, et al. (2010) conducted an empirical study of some of the aforementioned factors and their impact upon customer e-satisfaction. This study was conducted on a 2x2 basis with high/low on one axis and aesthetic approach (how to display) and functional approach (what to display) on the other. In this case, aesthetics refers to whether rich content (e.g., pictures and videos) is presented on the site or not, whereas functional display involves adjustments to font types, button colors, and navigational elements. It was found that aesthetic appeal was more important for customer satisfaction than website functionality was (Wang, et al., 2010). In addition, strong aesthetics provided deeper cognitive engagement in consumers who fit the exploratory typology, than otherwise. Functionality was found to be far more important for goal-directed consumers (Wang, et al., 2010).

Wang, et al., (2010) go on to conclude that aesthetic appeal is always a consumer preference, as long as it does not impede on task accomplishment by the goal-directed consumer. When aesthetics do impede task completion, it is likely to cause a negative impact on the satisfaction of goal-directed consumers (Wang, et al., 2010). Conversely, low aesthetic appeal and high functionality attributes also caused dissatisfaction in consumer's who are seeking information rather than an immediate purchase (Wang, et al., 2011). Yet, aesthetic appeal is almost certainly in the eye of the beholder, and therefore, these are findings that are difficult to generalize. As such, the literature continues to highlight important theoretical considerations that need to be tailored to the specific research context.

From an e-satisfaction measurement perspective, there is broad support for the usage of online survey methods (Belanche et al., 2012; Ladhari, 2010; Ha and Lennon, 2010; Ganesh, et al., 2010). Ganesh, et al. (2010) argues that the best means of understanding the online consumer is through online research. At one extreme, it is believed that e-satisfaction can be measured by simply asking how satisfied a consumer is with their experience, using a 5-point Likert scale (Ha and Lennon, 2010; Diehl and Poynor, 2010). At the other extreme are broader web survey's, which combine a number of e-satisfaction measures from the literature with business and customer specific measures (Ladhari, 2010). In additional benefit of broader web survey usage is the ability to collect information on customer typology, as this is a proven moderator of e-satisfaction (Ganesh, et al., 2010). Additional benefits from the web survey approach come from their ability to reduce interviewer bias and allow for large numbers of participants to be generated (Ganesh, et al., 2010).

Section Summary: Electronic Commerce

Websites have been of great benefit for generating commerce and for studying consumer behavior. Digitally based studies of consumer behavior are easier than they were in

traditional shopping venues due to the large amount of behavioral data that is routinely collected by websites. While many of the factors that influence consumer behavior on websites have been found to be similar to offline influencers, there are also significant differences. The literature highlighted many of these key differences, and how consumers behave and perceive their experience on a website differently than they do in a brick-and-mortar environment. The collected research findings provide a theoretical understanding for how consumers shop online, and how those behaviors are influenced and measured. Therefore, the electronic commerce literature provided the conceptual scaffolding required to understand how BGC's website may be positively and negatively affecting consumer behavior.

2.5 Action Research and Literature Review Conclusion

Action research is focused on practical and context-specific problem resolution (Kernstock and Brexendorf, 2012; Coghlan and Brannick, 2010; Ivankova, 2015). This is accomplished through a process of researcher and stakeholder collaboration (Ivankova, 2015; Coghlan and Brannick, 2010; Kernstock and Brexendorf, 2012; Kirwan and Conboy, 2009). According to Kernstock and Brexendorf (2012), the goal of action research is to develop immediate problem solutions and also to foster organizational learning so that improved management practices provide longer-term organizational benefits. When the research is focused on technology-centric areas, action research is particularly useful as an aid to the organization being studied (Kirwan and Conboy, 2009).

Action research has been shown to deliver solutions to complex context specific problems and to provide superior organizational change adoption (Ivankova, 2015; Kirwan and Conboy, 2009). However, the approach is also prone to researcher biases and does not often produce generalizable results (Kirwan and Conboy, 2009). In order to promote validity and minimize researcher bias, there is a need for constant reinforcement of the action research methodology and its cycles of planning, action, and reflection (Coghlan and Brannick, 2010; Kirwan and Conboy, 2009). The key components of which are consistent co-research team discourse and reflection (Greenwood and Levin, 2007).

With the desire to construct a practical long-term solution to the problem, the literature on action research highlights the importance of a collaborative learning process. On this basis, the goal of the literature review was to develop new knowledge and insights that could be synthesized and shared within BGC. The dialogue and reflection that ensued after completion of the literature review formed the basis of the conceptual research model and research hypotheses that follow in the next chapter.

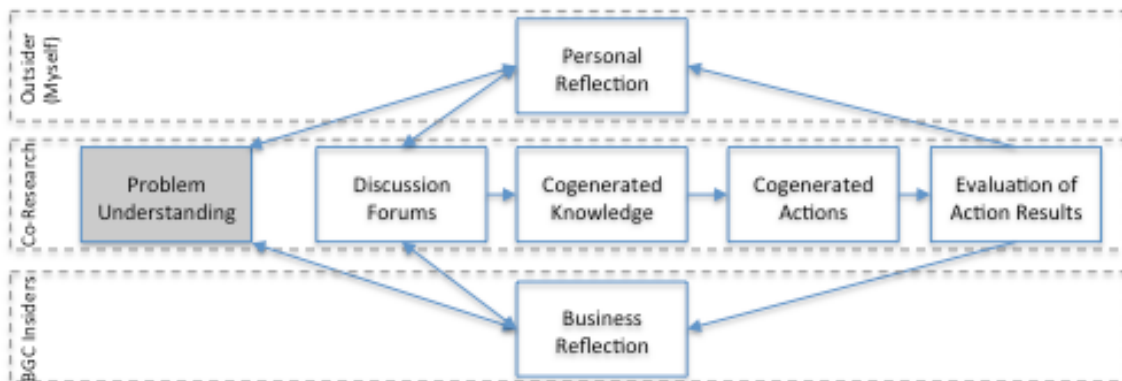
Chapter 3: Conceptual Model and Research Hypotheses

It is upon the literature review findings that the conceptual approach and hypotheses for this thesis research have been developed. The literature review provided insights into BGCs problem, but those insights needed to be discussed and reflected upon in order to determine additional areas for research. This chapter outlines the conceptual approach that was undertaken to develop immediate action-oriented improvements in the business and to generate long-term sustainable change. This conceptual approach was developed through a combined scholar-practitioner and business focused series of discussions and reflections. Following the conceptual model, specific hypotheses are developed in order to guide an action plan and methodology aimed at delivering the desired improvements to the business. Chapter 4 provides deeper insights into the full methodology that was utilized to generate the hypotheses and conceptual research model.

3.1 Conceptual Intervention Model

According to Greenwood and Levin (2007), pragmatic action research is enabled by the creation of arenas where mutual learning and discourse can occur. As such, it is cogenerated research that embodies AR and provides its transformative power (Kernstock and Brexendorf, 2012). Ultimately, the AR effort ceases when the participants are satisfied with the results, or when they decide to pursue other interests (Coghlan and Brannick, 2010). With these points in mind, my initial research goal was to establish a plan for intervention and the creation of regular discussion forums. Using Greenwood and Levin's (2007, pp.94) Cogenerative Action Research Model as a guideline, I constructed a conceptual intervention model, as illustrated in Figure 3-1.

Figure 3-1: Conceptual Intervention Model



The creation of a conceptual intervention model was a useful guide for the formulation of discussion forums, meeting agendas, and points of reflection. I used this conceptual model to inform dialogue with BGC's Executive Leadership Team (ELT). In this dialogue I stressed the importance of regular meetings and points of discussion so that we could formulate a mutually agreed response to the problem. I also shared that I expected new knowledge and revised business practices to be an explicit output of the overall effort. There was a strong consensus on these points from the ELT. My feeling was that this agreement was motivated by the ELT's desire to control the outcome. Despite any ulterior motives, this agreement enabled the creation of discussion forums, the selection of a co-research team, and the platform to develop a conceptual research model. The regular ability to discuss the problem also led to the co-development of research hypotheses and the ultimate plan and actions that were deployed. How this conceptual intervention model was brought to life is described in more detail in Chapter 4.

3.2 Conceptual Model

One of the first group meetings focused on the ELT's desire to effect positive improvements to the web business as quickly as possible. During this discussion, the CEO decided that the change experiments would need to be conducted on the production website rather than

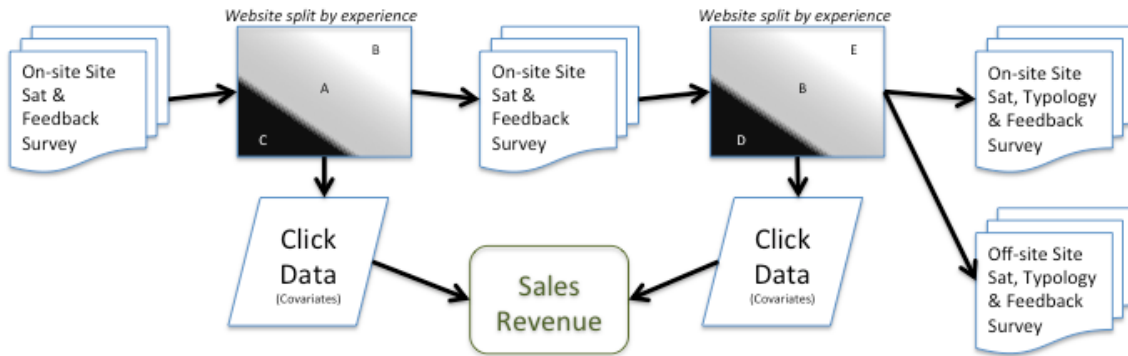
through focus group experimentation. Primarily, this decision was based on the need to realize immediate performance improvements and the belief that this would be best accomplished by 'real-world actions.'

The conceptual research design implication of the CEO's decision was a significantly reduced ability to measure individual customer satisfaction and behavioral stimuli. As a result, the measurement of consumer attitudes and behaviors needed to be accomplished across the population of the website's users. A positive from this type of approach is that the resulting measurements are not as prone to the errors that arise from focus group convenience sampling (Cresswell, 2014; Easterby-Smith, et al., 2012; Oyeyemi, et al., 2010). In addition, by opening participation to the entire population of customers, the research effort increased in organizational importance and visibility, an outcome that the CEO appeared to be trying to engineer through her decision.

Following the CEO's decision, a group comprised of myself and other key stakeholders convened to develop an approach to testing site improvements. The conceptual research model (see Figure 3-2) is a depiction of the cogenerated approach that this group ultimately settled upon. The idea was to capture site satisfaction data before and after each batch of website changes. In this way, the impact of the site changes could be understood through changes in customer sentiment toward the site. In addition, the site changes could be measured against a control group in order to understand how customer behavior and revenue was impacted.

It was agreed that customer behaviors could be measured by analyzing differences in click data (documented in the conceptual hypothesis model as the 'covariates') and actual product purchasing. The group consensus was to conduct multiple rounds of site changes, where behavioral and satisfaction impacts could be measured. Following the final series of site changes, additional customer data could then be collected that allowed for an analysis of on-website versus off-website, and typology, impacts upon customer sentiment. Due to this conceptual research framework, it would also be possible to monitor sales revenue across the course of the research time period. This monitoring of sales revenue over time would allow for analysis of the macro relationship between the website customer's satisfaction and purchasing.

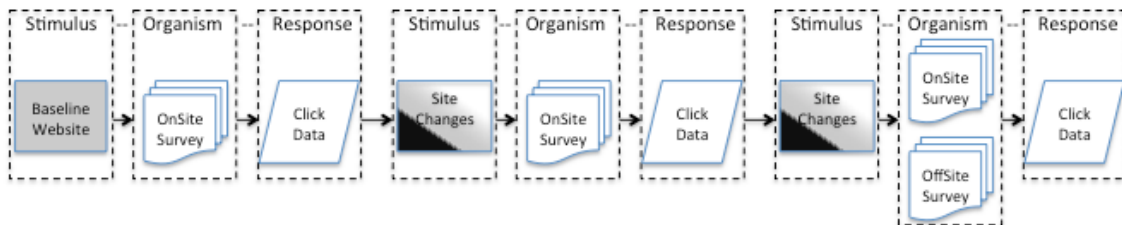
Figure 3-2: Conceptual Research Model



Underpinning the conceptual approach to the research is Mehrabian and Russell’s (1976) Stimulus-Organism-Response (S-O-R) theory. Use of S-O-R theory as a conceptual basis for research is common across the academic literature in this area (see Peng and Kim (2014), Zhenhui, et al. (2010), Wang, et al. (2010), Wu, et al., (2014)). According to S-O-R theory, stimuli (S) in the environment affect organisms/consumers (O), which elicits a response (R). A common interpretation of the S-O-R model is to measure stimuli through both customer perception and actual behavioral measurement (for example, Wu, et al., 2014). The organism impact can be measured through surveying perceptions, and the behavioral response through purchasing intent (Wu, et al., 2014). This theory was discussed with the co-research group in order to generate discussion about how best to construct the approach to site testing. How the research was ultimately conceptualized is shown in Figure 3-3.

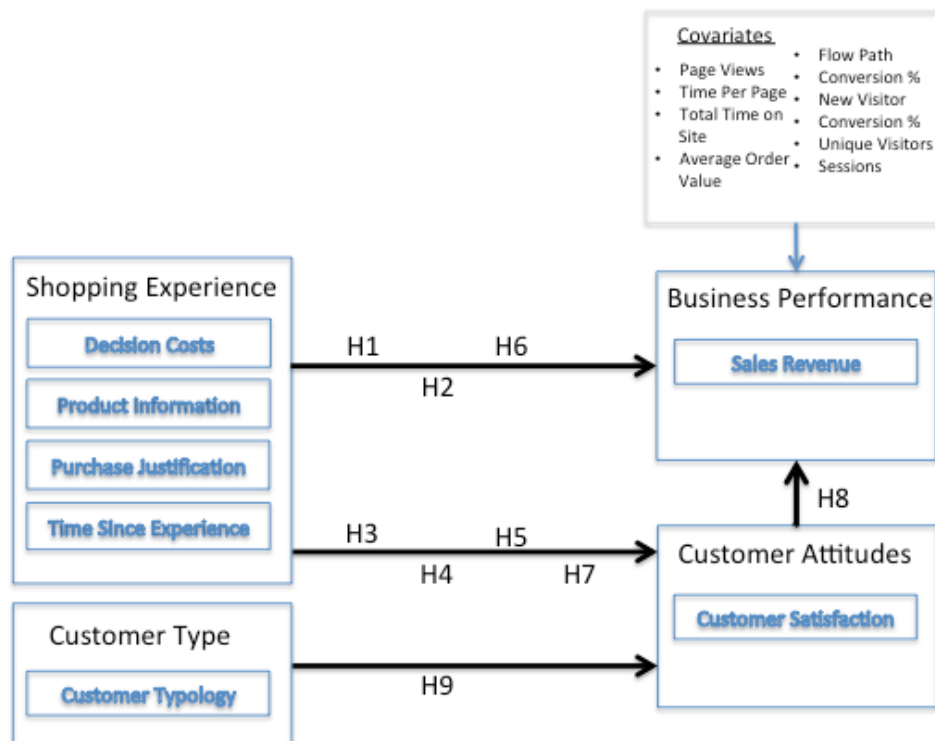
Considering an S-O-R framework, the co-research group discussed how the stimulus generated by the original website affected the attitudes of customers. These attitudes were measured through the initial on-site survey. In turn, we discussed how those attitudes generated a response, which was then measured through initial click data collection. Next, we discussed how we would make further changes to the website (stimulus). Attitudes (organism) and behavioral responses would then be captured through the on-site survey and click data, respectively. Following this, we discussed making further site changes (stimulus), followed by measurement of attitudes (organism) and behavioral changes (response).

Figure 3-3 : Research Conceptualization Using the S-O-R Framework



Following the development of the conceptual research model, the co-research debated the sources of the problems being experienced on the website. The result was a number of hypotheses that were agreed as being important for further research. Based on the output of these discussions, a hypothesis model was developed that draws on similar model development work conducted by Salanova et al. (2005) and Schmitz et al. (2014). A conceptual model of each hypothesis has been advanced to highlight the theoretical research framework (see Figure 3-4). This hypothesis model demonstrates how the broad areas of customer attitudes, behaviors, customer type, shopping experience, and business performance are connected by the research. The model also highlights a number of behavioral covariates that were considered as part of the analysis. These covariates were identified from the literature and further expanded upon and ratified through group discussion. It was agreed that each covariate would be measured in order to provide a deeper understanding of the results generated from each change to the website.

Figure 3-4 : Conceptual Hypothesis Model



3.3 Research Hypothesis Development

Consumer decision costs and sales revenue.

Consumer decision-making costs have been found to be a significant determinant of the desire to purchase from a retailer (Betancourt and Gautschi, 1990). The easier it is for a customer to accomplish their desired tasks, the lower their decision-making costs and the higher their likelihood to purchase (Betancourt and Gautschi, 1990; Urbany, et al., 1989; Broniarczyk and Griffin, 2014; Chernev and Hamilton, 2009). In an online environment, website navigability (ease of use) influences purchase intent (Peng and Kim, 2014; Lee and Kozar, 2012; Demangeot and Broderick, 2010; Ha and Lennon, 2010; Zhenhui, et al., 2010). Therefore, reduced decision-making costs should lead to increases in revenue. The following null hypothesis is the result¹:

H1₀ : $\mu = K$; Consumer decision costs have no impact upon sales revenue

Purchase justification and sales revenue.

Highlighting promotional items can increase demand across the whole assortment (Betancourt and Gautschi, 1990). In order to maximize the impact on demand, research can be conducted to identify how customers respond to different promotional cues (Meeker, et al., 2009). The literature suggests that by highlighting the promotional value of products that it will spur increased revenue across the assortment, and increase the justification to purchase the highlighted products (Betancourt and Gautschi, 1990; Meeker, et al., 2009; Simonson, 1999). Therefore, improvements to how product discounts are communicated should increase revenue. The following null hypothesis is the result:

H2₀ : $\mu = K$; Purchase justification has no impact upon sales revenue

Consumer decision costs and customer satisfaction.

Perceptions of a retailer's assortment are influenced by the perceived likelihood that a desirable purchase option is available to the customer (Hoch, et al., 1999; Broniarczyk, et al., 1998). As a result, reducing the consumer decision costs required to find a desirable product should lead to higher levels of satisfaction with the assortment (Hoch et al., 1999; Handelsman and Munson, 1985; Betancourt and Gautschi, 1990). In an online environment, decision costs and satisfaction are influenced by website content and ease of use (Peng and Kim, 2014; Lee and Kozar, 2012; Zhenhui, et al., 2010; Green and Pearson, 2009; Devaraj, et al., 2002). Therefore, reduced decision-making costs should lead to increased consumer satisfaction levels. The following null hypothesis is the result:

¹ μ = mean of the test group ; K = mean of the control group

H3₀ : $\mu = K$; Consumer decision costs have no impact upon customer satisfaction

Purchase justification and customer satisfaction.

Promoting a select number of items will increase the purchase justification to purchase those items (Simonson, 1999). Increasing the ability of a customer to rationalize their purchase decision improves their satisfaction with that purchase (Sela, et al., 2009; Böhm and Pfister, 1996). Perceived purchase value is highly correlated with satisfaction (Deveraj, et al., 2002). Therefore, increasing the justification to purchase from a product assortment should lead to higher levels of consumer satisfaction with a business. The following null hypothesis is the result:

H4₀ : $\mu = K$; Purchase justification has no impact upon customer satisfaction

Time since experience and customer satisfaction.

Mogilner, et al. (2014) found that satisfaction is influenced by temporal outlook. Based on these findings, consumers that are focused on immediate tasks will have lower satisfaction levels than those shopping for a future task (Mogilner, et al., 2014). A reasonable conclusion is that consumer currently engaged in their shopping tasks will express lower levels of satisfaction than they do otherwise. Therefore, satisfaction levels should be lower when measured during the shopping visit than they are when measured post visit. The following null hypothesis is the result:

H5₀ : $\mu = K$; During visit and post-visit customer satisfaction does not differ

Customer Satisfaction and sales revenue.

Numerous studies indicate that the primary influencer of consumer purchase behavior is satisfaction with the shopping experience (Udo, Bagchi and Kirs, 2010; Kuo, et al., 2009; Rajic and Dado, 2013). In turn, perceived service quality and product value are the primary antecedents of satisfaction perceptions (Cronin, et al., 2000; Kuo, et al., 2009). The combined perception, as measured by satisfaction, has been found to be the customer measure most indicative of the performance of any given company (Diehl and Poynor, 2010; Newman and Werbel, 1973; Wicks and Roethlein, 2009). Therefore, higher levels of satisfaction should correspond to higher levels of revenue, and vice versa (Naik and Srinivasan, 2015; Olsen, et al., 2014; Morgan, et al, 2005; Wicks and Roethlein, 2009; Udo, et al., 2010; Cronin, et al., 2000). The following null hypothesis is the result:

H6₀ : $\mu = K$; Customer satisfaction is not related to sales revenue

Product information and customer satisfaction.

Assortment perceptions are influenced by product information communication (Diehl, et al., 2003). Improving the way the assortment is communicated to the customer will increase

customer satisfaction with the assortment (Mogilner, et al., 2008; Johnsen, 2009; Broniarczyk, et al., 1998; Hoch et al., 1999). In addition, increasing the quality of information provided to consumers increases their satisfaction (Lim, et al., 2015). For example, visual displays, such as product pictures and videos, influence customer satisfaction (Demangeot and Broderic, 2010; Zhenui, et al., 2010). Therefore, changes in product information communication should lead to changes in customer satisfaction levels. The following null hypothesis is the result:

$H7_0 : \mu = K$; Product information displays have no impact upon customer satisfaction

Product information and sales revenue.

Just as satisfaction is influenced, demand for an assortment is also a function of how well a retailer communicates their offering to potential customers (Johnsen, 2009). Improving how products are categorized can reduce consumer decision costs and increase the propensity to purchase (Mogilner, et al., 2008; Simonson, 1999; Lim, et al., 2015). Improvements in product information can reduce product utility uncertainty and increase purchase likelihood (Liang, 2010). As an example, providing rich media (e.g. videos) has been shown to increase purchase intent (Zhenhui, et al., 2010). Conversely, providing too much product information has been associated with purchase deferrals (Urbany, et al., 1989). Therefore, a focus on product information quality over quantity should lead to increased purchase demand and revenues. The following null hypothesis is the result:

$H8_0 : \mu = K$; Product information displays have no impact upon sales revenue

Customer typology and customer satisfaction.

A number of consumer research studies have shown that customer perceptions are significantly impacted by their task orientation, or typology (Gupta and Kabadayi, 2010; Ha and Lennon, 2010; Moe, 2003; Janiszewski, 1998). For example, consumers focused on purchasing for future needs, compared to present day needs, will tend to have higher satisfaction levels (Mogilner, et al., 2014). In addition, goal-directed consumers prefer highly usable websites, whereas hedonistic consumers prefer increased levels of product information (Gupta and Kabadayi, 2010). It has also been concluded that higher levels of satisfaction are associated with providing each of the main customer typologies with support for their specific tasks (Peng and Kim, 2014; Lee and Kozar, 2012). Therefore, it should be possible to identify necessary website improvements through understanding differences in satisfaction levels between each of the customer typologies. The following null hypothesis is the result:

$H9_0 : \mu = K$; There are no differences in satisfaction levels between customer types

3.4 Model and Hypothesis Summary

This chapter of the thesis discusses how the conceptual model was constructed based on the theoretical scaffolding developed by the literature review and subsequent discussions with business insiders. The resulting conceptual construct provided guidance on the types of data to collect and the methods for doing so. This led to the formation of hypotheses that would be used to guide the design of experiments and generate the knowledge necessary to improve the performance of BGCs e-commerce business. In total, nine hypotheses were developed in order to better understand the perceptions and behaviors of BGCs website customer base. Each hypothesis was developed to provide insights into a specific set of consumer behavior influencers and their associated perceptual and purchasing manifestations.

Table 3-1 : Research Hypothesis Summary

Hypotheses		Research Areas	
H1 _o	Consumer decision costs have no impact upon sales revenue	Decision Costs	Business Performance
H2 _o	Purchase justification has no impact upon sales revenue	Purchase Justification	Business Performance
H3 _o	Consumer decision costs have no impact upon customer satisfaction	Decision Costs	Customer Attitudes
H4 _o	Purchase justification has no impact upon customer satisfaction	Purchase Justification	Customer Attitudes
H5 _o	During visit and post-visit customer satisfaction does not differ	Temporal Outlook	Customer Attitudes
H6 _o	Customer satisfaction is not related to sales revenue	Customer Attitudes	Business Performance
H7 _o	Product information displays have no impact upon customer satisfaction	Product Information	Customer Attitudes
H8 _o	Product information displays have no impact upon sales revenue	Product Information	Business Performance
H9 _o	There are no differences in satisfaction levels between customer types	Customer Attitudes	Customer Typology

Chapter 4: Methodology

A key insight from the literature is the importance of context and customer specific knowledge when trying to improve engagement and purchasing. Because it is difficult to generalize findings from other organizations or industries, research must be conducted on the customer and business unit that is seeking improvement (Zeithaml, 2002; Changsoo and Tadisina, 2008). In large part this is because customer perceptions heavily influence attitudes toward a business, and these perceptions are not generalizable (Changsoo and Tadisina, 2008). The BGC ELT and co-research groups, both articulated their support for these insights from the literature. Therefore, any chosen research methods needed to be organizationally relevant and suitable for developing new knowledge. This chapter discusses the methods that were ultimately utilized for the research and the rationale that supports their applicability to the problem.

4.1 Introduction

In order to develop credible and practically useful results from the research, an Action Research (AR) methodology was utilized. At its core, AR is a powerful method for gaining business participation in the development of new business knowledge and new business practices (Coghlan and Brannick, 2010; Greenwood and Levin, 2007). In particular, AR is useful for the development of context-specific knowledge, which is a requirement for the construction of sustainable and practical solutions for any unique business entity (Coghlan and Brannick, 2010; Greenwood and Levin, 2007). Consequently, an action research methodology was ideally aligned with the expected outcomes of the research effort

In order to develop a unique and practical solution, it was determined that both qualitative and quantitative data collection and analysis would be required. Research that considers both types of data allows for deeper insights than quantitative data can provide alone (Cresswell, 2014; Ivankova, 2015). Website analytics tools are excellent at collecting quantitative behavioral data (Huang, et al., 2009; Johnson, et al., 2004), but lack the qualitative insights to explain why the behaviors occurred. To balance this concern, the extant literature suggests the collection of qualitative data directly from customers as a method for understanding where and why business enhancements are required. Therefore, the co-research team agreed that the research effort would need to include both types of data. In order to generate combined insights from such data, mixed methods of research were used. Cresswell (2014) defines mixed methods research as a method of integrating quantitative and qualitative data into a single research effort. Mixed methods have a strong track record of application in academic research (Shah and Corley, 2006; Cresswell, 2014; Ivankova, 2015), albeit that there are a number of notable shortcomings (Howe, 1988; Johnson and Onwuegbuzie, 2004; Donaldson, 2008) that will be discussed later in this chapter.

4.2 Application of Action Research Methods

Action Research specifically aims to create sustainable organizational change through the active participation of the organization in the research process (Coghlan, 2011; Coghlan and Brannick, 2010; Greenwood and Levin, 2007; Shani and Pasmore, 1982). The AR methodology is built on the belief that community participation is required to generate solutions that are appropriate to the community (Ivankova, 2015). Through broad-based community involvement, real learning and capacity development can occur that can lead to a more sustainable post-research environment (Ivankova, 2015; Pedler, 2008). Such developments occur through AR's emphasis on community action, reflection and the subsequent re-framing of the problem and the further actions to be taken (Greenwood and Levin, 2007).

One of the requirements for action research is that the research process is entirely democratic and pursued mutually (Greenwood and Levin, 2007; Ivankova, 2015). Such

participatory research practice requires that the traditional outsider research role is ceded to a model of co-generated knowledge (Greenwood and Levin, 2007; Ivankova, 2015). In such a model, the improvement in the situation being researched, through a collaborative process of action and reflection, is far more important than the application of outside standards of rigor (Greenwood and Levin, 2007). This does not mean that integrity and criticality are to be sacrificed, but that balance is achieved between praxis and scientific research methods (Greenwood and Levin, 2007).

Coghlan and Brannick (2010) propose a five-step approach to Action Research (see Figure 4-1). Their approach is characterized by the inclusion of organizational insiders as “co-researchers” (Figure 4-1, step 2) (Coghlan and Brannick, 2010, pp.9). The combined research group then develops and implements an action plan that seeks to deliver the desired outcomes (step 3 and step 4). This is followed (step 5) by a process of “meta-learning”, where new organizational knowledge is created from reflection on the consequences of the actions (Coghlan and Brannick, 2010, pp.12). In order to be effective, the reflective process needs to occur individually, but also through group discourse (Pedler, 2008). Such a dialogic process should lead to learning across the group and also to a general consensus on the actions that need to be taken and how the problem may need to be reframed (Pedler, 2008). Following the meta-learning process, the planning, action and reflection process is repeated until the desired results are achieved (Coghlan and Brannick, 2010). It is the Coghlan and Brannick (2010) methodology that was utilized as the foundation of the actual action research process. Figure 1-1 on page 9, further illustrates the cyclical and complex nature of this methodology in practice.

Figure 4-1 : Action Research Methodology



At the inception of this research effort, BGC’s CEO and CFO presented the current organizational problem as one of declining revenue without a clear path to remedy. I responded by conducting an initial literature review, collecting some initial performance data, and speaking to a number of different functional experts with responsibility for the website. The discussions were with those responsible for technology, web operations, and web merchandising on the website. My conclusion was that two separate problems existed. First, the new software platform was not sufficiently tested before implementation, and a large number of software bugs existed that were disrupting the desired functionality of the

website. My other belief was that customer satisfaction might have declined because the new site did not cater to customer needs as effectively as it should. Upon reaching this viewpoint, a problem statement and a future state vision were formed (per step 1 of Figure 4-1). The future state vision was as follows:

- The new software platform has been fully tested for bugs, and all customer facing issues have been resolved.
- An organizational capability has been created that can identify site enhancements that have a measurable and positive impact on revenues and customer satisfaction.

4.2.1 Research Group Formation and Dynamics

For the purposes of this thesis, it is the second vision statement that is the focus. The problem statement, detailing the two separate areas of problems, and the future state vision were articulated to BGC's CEO and CFO. With a focus on the latter vision statement, I recommended the formation of a cross-functional group that would participate in the development of a solution to meet this end goal. Through this dialogue, a "co-research" group was agreed (per step 2 of Figure 4-1). In addition, the Executive Leadership Team (ELT) (which includes BGC's CEO, CFO, CMO, Head of Merchandising, Head of Stores, and CIO) was agreed as an oversight group that would review any recommended actions generated by the "co-research" group. In effect, this meant that each function head gained visibility to the recommendations generated by the "co-research" group. Including the ELT also encouraged further strategic dialogue about the recommendations, caused revisions to the "co-research" team thinking, and ultimately resulted in more appropriate actions being taken. Further, the inclusion of the senior management team ultimately improved the organizational adoption of the recommended actions.

Reflection on the formation of these two insider research groups highlighted an important power dynamic. The ELT determined that they would make a final decision on all recommendations, and they hand-selected the group of 'co-researchers' who would work on the problem day-to-day. In practice, this would mean that the co-research group would spend a significant portion of their time deliberating about what and how to inform the ELT, with specific sensitivity to any potentially controversial discussion points. An additional political dynamic came from the reality that the individuals selected for the co-research team were also those involved in the re-platform effort that went so poorly. By choosing this group, the ELT was acknowledging that these individuals were the most knowledgeable about BGC's website. Yet, through confidential smaller group discussions, it became clear that the ELT did not fully trust their opinions. While this lack of confidence was not stated directly to the co-research team members, their behavior suggested that they generally knew that only immediate performance improvements would safeguard their jobs. On a practical basis, this

meant that I needed to serve a political role, where my support of co-research group opinions was required in order to validate those opinions.

Chart 4-1 : Research Groups and Primary Direction of Power

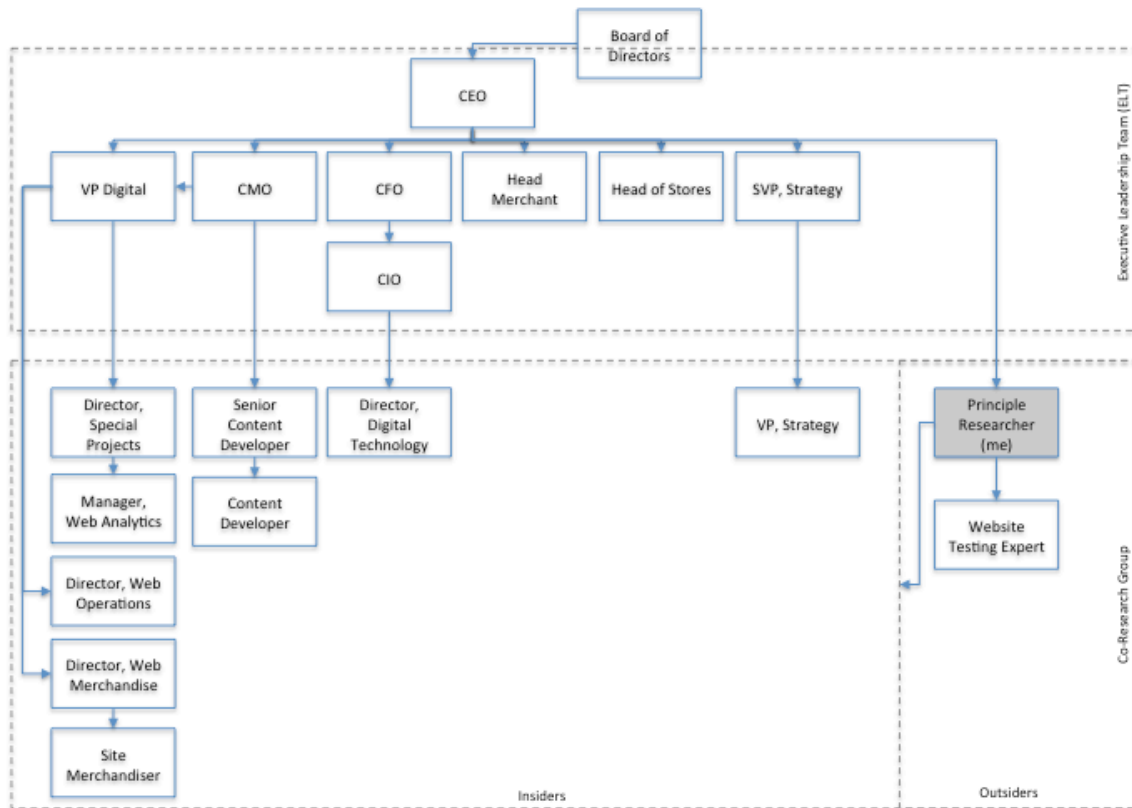


Chart 4-1 shows each of the individuals who comprised the ELT and co-research groups, displayed by organizational role. The directional arrows, in this chart, show the general influence of political power on each individual. For example, prior to the start of the research effort the Chief Marketing Officer (CMO) had direct responsibility for the website, and the VP, Digital reported to the CMO. At the start of the project effort, the responsibility for the digital business was removed from the CMO, with the VP, Digital subsequently reporting directly to the CEO. However, the CMO retained political influence over the VP, Digital. Given that the ELT looked to me to provide validity to the recommendations of the co-research group, I have depicted that this also gave me some power over the rest of co-research group. While I tried to limit this influence over them, it is important to note that this dynamic existed as part of the research process.

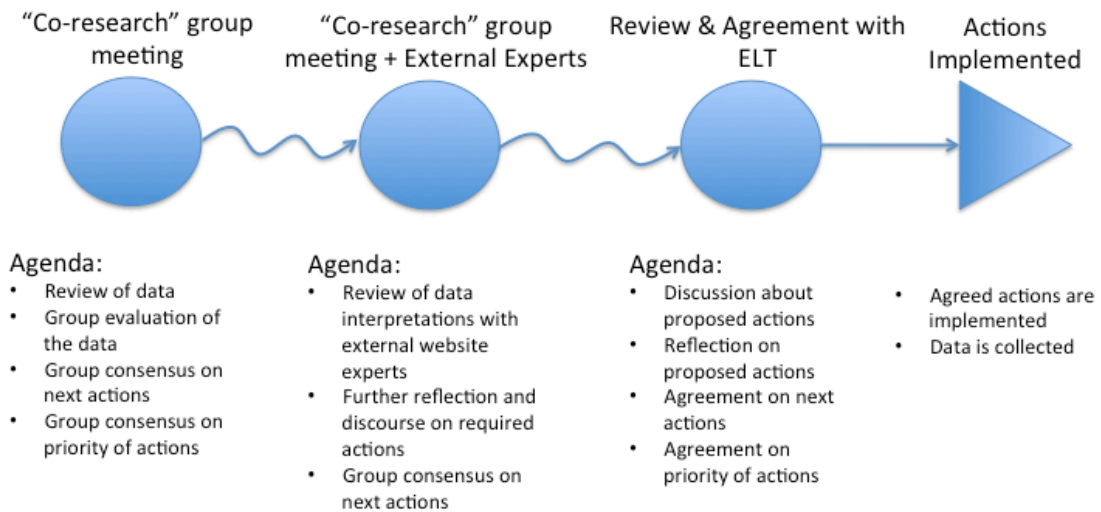
4.2.2 Research Group Working Practices

Following agreement on group formation and working practices, I engaged in a more robust review of the literature. This literature review allowed for the formation of some initial

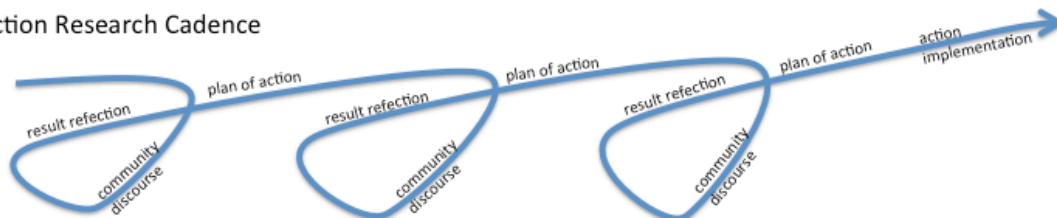
hypotheses, which I reviewed with the co-research group. In parallel with my literature review, the co-research group collected and reviewed customer satisfaction scores and feedback based on the survey methods previously being used by the business. This data was primarily collected via an on-site feedback survey. Based on the preliminary customer data and the literature review findings, we aligned on a set of actions that were believed to offer immediate customer facing and revenue improvements. These actions were presented to the ELT, where discussion led to a general agreement and priority definition.

It should be noted that the hypotheses that form the basis of the thesis research were also the result of my synthesis and presentation of the literature review findings to the co-research and ELT teams, and were based on a cogenerated viewpoint. Based on the outcome of these discussions, experiments were designed. However, not all of the designed experiments could be executed, due to either resource or capability constraints. Therefore, the experiments presented in this thesis are only those that the business agreed to execute. Despite the inability to execute certain experiments, exploration of alternate approaches did generate organizational learning that will be discussed later. Another prerequisite for implementing the experiments was the resolution of the majority of the remaining software bugs. While BGC focused on resolving these bugs, I sought and obtained ethical approval, from the University of Liverpool, to proceed with the thesis research.

Figure 4-2 : Weekly Action Research Cycles



Action Research Cadence



Once all parties were in the position to begin, we started to implement changes on the website, and to collect data. As part of this, we established three meetings per week. On Tuesday mornings we met as a co-research team. In this meeting we would discuss the data and results collected from the previous week, decide on actions we wanted to take based on the data, and we would agree on upcoming priorities. On Tuesday afternoons we would meet again to discuss how best to execute the agreed actions. This Tuesday afternoon meeting also included an external website expert who would express their independent viewpoint on the data. Including an objective outsider caused further reflection amongst the co-research group, and periodically led to changes in our previously agreed actions and refinements to our execution plans. The weekly action research cycle is depicted in Figure 4-2.

The two rounds of “co-research” group dialogue proved to be very effective at building co-research team knowledge and agreement on the actions to be taken, including how best to execute them. Initially, I was looked to as the primary leader of these meetings. However, after about a month, the Directors of Special Projects and Digital Technology felt knowledgeable enough to take over the leadership role. My initial reaction to this change was not positive, as I felt somewhat marginalized by the reduced role. However, upon reflection, I realized that this was a clear sign of progress toward building organizational capabilities. This change in roles also led to improved actions and improved reflections by the business on those actions, as greater process ownership encouraged the business participants to leverage fully their superior knowledge of the company and the customer.

Following the two meetings each Tuesday, we held a meeting with the ELT every Monday. This meeting involved a presentation of the collective interpretations of the data collected and the recommendations of the “co-research” group. In this meeting, further dialogue and interpretation of the data occurred. The outcome was either a re-framing of the actions or agreement to proceed as recommended. The hypotheses proposed in this thesis, the data collected, and the reflections and actions taken were all generated through this weekly organizational discourse and action research cycle (illustrated in Figure 4-2). In order to record the discourse that took place during each round of discussions I kept a notebook and wrote down all key discussion points and conflicting points of view. At the end of each meeting I also reflected on each discussion and took notes on my observations of the individuals and the group dynamic.

Based on the principles of action research, the adoption of a participatory approach naturally resulted in some research design concessions. For example, a collaborative approach to survey design meant that compromises had to be reached regarding the questions that were incorporated. As a consequence, the survey design was informed by the literature, but not solely based on traditional standards of academic rigor and support. This is in fitting with Thorpe and Holt’s (2008) definition of action research, where theory informs action and is used to study the aftereffects. As a methodological validation point, the

surveys used later in the research became more heavily weighted toward academically supported practice. In large part, this was due to group reflection and dialogue about the results generated from the initial surveys. I took this as another sign that the participatory nature of action research was leading to improvements in organizational knowledge. Specific information on each of the surveys used during the research effort is provided later in this chapter.

4.3 Application of Mixed Methods Action Research

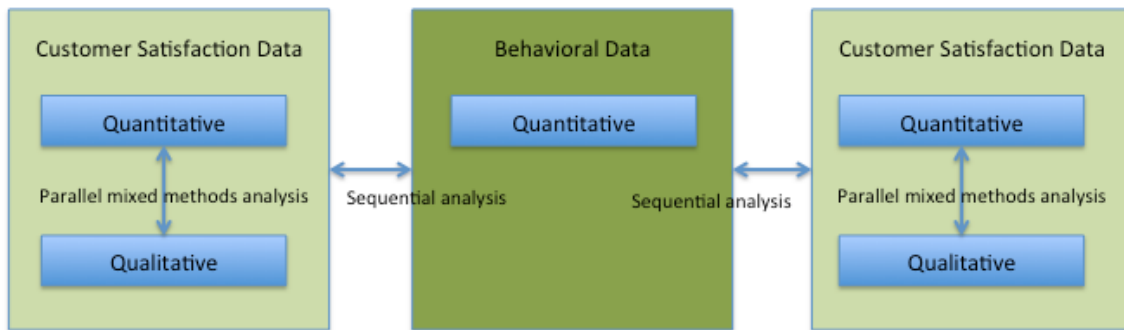
As mentioned in the introduction to this chapter, the literature review identified that both quantitative and qualitative data was required to fully achieve the future state vision as set forth. Given that an action research methodology was also required, this research effort needed to combine both methods. The resulting methodology has been termed Mixed Methods Action Research (MMAR) (Ivankova, 2015). Ivankova (2015) conducted research into MMAR practices in the academy and found 108 peer-review articles that used MMAR as a basis. Generally, it was found that MMAR is simply a combination of mixed methods and action research practices. Accordingly, the primary difference from a traditional action research study is that both quantitative and qualitative data is used in order improve the depth, breadth, and validity of the AR study (Ivankova, 2015). In comparison to a traditional mixed methods study, MMAR requires a higher level of community participation in the research process (Ivankova, 2015).

The literature identifies three main types of mixed methods research. *Sequential* mixed methods research involves making either the qualitative or quantitative data dominant in the research, with the non-dominant form providing insights to the dominant data results (Cresswell, 2014; Ivankova, 2015). Most commonly, this involves collecting quantitative data and then using qualitative data to provide an explanation of the quantitative results (Cresswell, 2014). *Parallel* mixed methods involve collection, integration, and analysis of the two types of data at the same time (Cresswell, 2014; Ivankova, 2015). The idea is to identify disparities between the two data sets in order to identify areas for further study (Cresswell, 2014; Shah and Corley, 2006). Such a method is also useful when trying to gain a deeper understanding of survey results (Easterby-Smith, et al., 2012; Shah and Corley, 2006). *Compensatory* mixed methods use each of the data types to compensate for the weaknesses in the other (Easterby-Smith, et al., 2012). For example, qualitative data lacks generalizability, while quantitative data lacks explanation (Easterby-Smith, et al., 2012). Most often, *compensatory* mixed methods involve finding a statistical relationship between the qualitative and quantitative data (Easterby-Smith, et al., 2012).

For this thesis research study, two of the three main methods of mixed methods analysis were utilized. In order to gain a deeper understanding of the quantitative customer satisfaction data, a parallel strand of open-ended qualitative data was collected and analyzed. A separate strand of quantitative data was collected to measure customer behavior on the website. This data was analyzed sequentially against the combined

satisfaction data. The approach is conceptually illustrated in Figure. 4-3. Not documented in Figure 4-3 are the qualitative insights generated from the action research process, within BGC.

Figure 4-3 : Mixed Method Analysis Type Usage



In the design and utilization of mixed methods, it was important to note some of the potential shortcomings that have been reported in the academic literature. There is a belief that quantitative and qualitative data usage is inextricably aligned with one's epistemological position (Burrell and Morgan, 1979; Howe, 1988). As a result, functionalist and interpretivistic epistemologies should only use research and analysis methods that agree with those research beliefs (Burrell and Morgan, 1979; Johnson and Onwuegbuzie, 2004; Donaldson, 2008; Howe, 1988). Personally, I am more apt to subscribe to Willmott's (1993) and Shah and Corley's (2006) views, where a critically reflective combination of these extreme epistemological views is possible. As a consequence of this outlook, I see no impediment to a combination of rigorous quantitative and qualitative research methods.

Also arising from epistemological concerns is the belief that most researchers lack the necessary skills and experience to leverage both qualitative and quantitative research methods successfully (Howe, 1988). In my opinion, this concern has more merit. Up until my doctoral studies, the majority of my data analysis training was quantitatively focused. Given my current pursuit of a doctorate, my research and analysis skills also do not have the benefit of many years of practice. Further reflection on the concern raised by Howe (1988), caused me to focus the data analysis on methods that I had previous experience with. Later in the thesis, I will discuss why each method was deemed to be appropriate, but underlying the selection of the methods was also a desire to ensure that I did not incorporate a research method that I did not fully understand.

Figure 4-4 illustrates the general MMAR approach that was used. This figure illustrates the general process following on from the development of the problem statement. Highlighted is the influence of the literature review on the development of hypotheses and how the incorporation of customer data led to the development of changes to the website. In parallel with the site changes, behavioral data was collected and compared against additional customer feedback data that was collected post the changes. Underpinning every stage of this was a cyclic action research methodology of planning, action, and reflection.

Figure 4-4 : MMAR Research Approach Overview



4.4 Qualitative Data Collection and Analysis Methods

4.4.1 Data Collection

In order to collect the qualitative data, web-based questionnaires were utilized. A web-based survey method was used because it is the best way of engaging with online customers (Szymanski and Hise, 2000; Ganesh, et al., 2010; Wilkerson, et al., 2014). In support of this, Thorpe and Holt (2008) state that qualitative data generation should be conducted using the methods that offer the best access to the target of study. Web-based survey methods have also been shown to minimize response bias compared to other qualitative survey methods (Greene, et al., 2008). Web-based surveys also produce equivalent data quality compared to more traditional survey methods (Wilkerson, et al., 2014; Lewis, et al., 2009; Hoonakker and Carayon, 2009). In addition, web-based surveys generate similar response rates to non-electronically based surveys, but provide better response times (Hoonakker and Carayon, 2009). Web-based surveys also reduce the analysis errors inherent in capture methodologies that require written responses to be manually entered into a computer for analysis (Easterby-Smith, et al., 2012). Based increasing levels of academic support (Morison, et al., 2015), a web-based mode of customer data generation was considered valid for this research.

As mentioned in the introduction to this chapter, the purpose for collecting qualitative data was to gain deeper insight into the customer satisfaction feedback as a means of identifying problem areas and resolution plans. Parasuraman and Zeithaml (1994) posit that consumer perceptions are essential for understanding customer satisfaction. In order to do this, qualitative data can be collected through open-ended interview or survey questions

designed for collecting individual perspectives (Cresswell, 2014; Romand, Jr., et al., 2003). Therefore, the customer surveys utilized for this research each contained an area for open-ended customer perception feedback.

Three different surveys were deployed in order to collect open-ended qualitative data (see Table 4-1 for an overview of each survey). These questionnaires collected customer data prior to any site changes being made, and then after each major round of site changes were complete. The first two questionnaires were made accessible to all website users. This accessibility was accomplished by providing a “feedback” link on each page of the website.

Table 4-1 : Surveys Used to Gather Customer Feedback

Survey	Site Visit Survey Timing	Research Timing	Qualitative Data Collected	Objective
Pre-Post Site Sat & Feedback	During website browsing	1-3 weeks prior to site changes & 1-7 weeks post initial site changes	Open-ended feedback	To compare feedback themes before and after site change experimentation and better understand site satisfaction ratings
Customer Typology, Site Sat & Feedback	During website browsing	3-8 weeks post secondary site changes	Open-ended feedback	To understand how customer typology influences perspectives and understand site satisfaction ratings
Customer Typology, Site Sat & Feedback	Upon exiting the website	5-8 weeks post secondary site changes	Open-ended feedback	To understand differences between on-site and post visit perspectives

Any customers wishing to provide feedback were able to click on the link, at which point they were presented with a short questionnaire (see appendix 2 and 3). The third questionnaire was provided randomly to two-percent of customers that browsed three or more pages of the website. Once the randomly selected customers exited from the website they were immediately presented with the questionnaire.

All three questionnaires collected both quantitative and qualitative data. From a qualitative perspective, open-ended feedback was collected via an open entry field on each questionnaire. In this entry field, the customer was able to type any feedback that they wished to share. Quantitative feedback was collected using closed-ended questions, and will be discussed in the quantitative collection and analysis portion of this section of the paper.

4.4.2 Data Analysis

In order to meet standards of research rigor and credibility, qualitative research must be grounded in a structured method of analysis (Corbin and Strauss; Cooney, 2011). In addition, qualitative research credibility requires the usage of academically commonplace methods and the demonstration of openness about the process and its limits (Cooney, 2011; Shah and Corley, 2006). Credibility is further enhanced when qualitative interpretations are generated directly from the data (Cooney, 2011), and result in a positive influence on the situation being studied (Corbin and Strauss, 2008). As a best practice, qualitative data should

be organized, reviewed for content and then interpreted into codes and themes (Cresswell, 2013; Wilkerson, et al., 2014; Romand, Jr., et al., 2003). In order to promote validity, the coded data should then have multiple checks for accuracy (Cresswell, 2014).

According to the literature, codes and themes can either be generated from the data, generated from the literature, or by a combination of both methods (Cresswell, 2014; Thorpe and Holt, 2008; Ivankova, 2015). For this research, a combination approach was chosen, according to similar studies conducted by Greaves and colleagues (2014) and Cetin and colleagues (2014). Initial themes were identified based on the literature review findings and the resultant hypotheses, as follows: Decision Costs, Purchase Justification, Customer Centricity, Assortment Quality, Product Delivery. The survey data was then imported into MAXQDA v12.0 for review and coding. Through review of the open-ended feedback, thirteen codes emerged from the data, and were named based on some of the key insights derived from the literature review and through discussions of the feedback with the co-research group: Navigation, Task Support, Product Information, Service Quality, Desired Purchase Timeline, Availability of Desirable Product, Product Quality, Communication of Prices, and Product Value, Software Bugs, Site Performance, Fulfillment, and Security and Privacy.

Based on this inductive method of coding (Shah and Corley, 2006), two new themes emerged from the data: Technology Support and Security. Despite prevalence in the literature, these themes were omitted from the initial list because they were not considered to be central to the hypotheses. However, after group discussion about the data, it was determined that these additional codes were required in order to create a full understanding of the customer feedback. The final set of codes were then associated with a theme for purposes of analysis and understanding (Ivankova, 2014; Cresswell, 2013; Cresswell, 2014), see Table 4-2. As a result of this process, the codes and themes were co-generated from a review of the data and the literature, as suggested by Cresswell (2013, pp.185).

Table 4-2 : Qualitative Analysis Themes and Codes

<u>Theme</u>	<u>Corresponding Codes</u>
Decision costs	navigation; product information
Purchase justification	communication of prices; product value
Technology support	software bugs; site performance
Customer Centricity	service quality; task support
Assortment quality	availability of desirable product; product quality
Product Delivery	fulfillment; desired purchase timeline
Security	Privacy & Security

In order to ensure the validity of the data, each code was re-reviewed to confirm that it matched the context of the customer feedback. In order to accomplish this, MAXQDA was

used to segment each piece of coded text into code blocks so that each code could be reviewed for contextual consistency and coding accuracy. This re-review process found a number of codes that had been mistakenly applied, and those errors were corrected. Following this re-review process, a summary of each code and its associated text was produced and provided to members of the co-research team. The coding of the data was reviewed and discussed, and further code adjustments were made based on the resulting coding consensus. For example, I had initially coded feedback about the prevalence of pop-up windows as a “service quality” issue and code. After discussion, most of the customer feedback about “annoying pop-up windows” was re-coded under the “software bugs” code. This decision was based on the view of the Director, Digital Technology who stated that the pop-up windows were the result of a computer coding issue, creating an undesirable experience.

Following common academic qualitative analysis practices (per Cresswell, 2014; Ivankova, 2015; Shah and Corley, 2006) the final themes and codes were analyzed in parallel with the quantitative satisfaction ratings generated by the surveys. The method allowed for a better understanding of the site satisfactions scores and allowed for more than just a quantitative basis for forming a conclusion as to whether any real progress was garnered from the site changes that were deployed. The resulting interpretations are discussed in the findings chapter of this thesis.

4.5 Quantitative Data Collection and Analysis Methods

4.5.1 Data Collection

At the outset of this methodological discussion about quantitative data collection, it is important to remember that this research effort is based on the core principles of Action Research. The process of AR is highly collaborative and balanced between praxis and scientific research requirements. In accordance with these principles, quantitative survey development followed a ‘non-traditional’ path. Traditional research tends to follow a linear path through theoretical development, research, and findings (Nowotny, et al., 2003). In contrast, AR follows a decidedly less linear and more circular process (Ivankova, 2015; Greenwood and Levin, 2007). Greenwood and Levin (2007) make the point that for a research effort to be considered AR, it must follow AR principles at every stage of the research process. As a consequence, the following discussion about data collection is as much a narrative account of the AR process that was required to build survey consensus, as it is an account of the theory that was used to inform quantitative data collection.

In order to generate quantitative data, three different surveys and a website click data method were created. As previously discussed, the surveys were used to collect qualitative and quantitative data. The collection of both types of data required the usage of both close-

ended (quantitative) and open-ended (qualitative) survey questions (Ivankova, 2015). The goal of the survey method was to gain direct feedback on customer satisfaction so that changes in satisfaction levels could be measured in response to changes in the website. This required the development and distribution of a survey that could measure these changes immediately before and after they were implemented. This effort began with a detailed review of the relevant academic literature.

First On-Site Survey Development

There is much debate in the literature as to the best method for garnering quantitative customer satisfaction data. Over the past thirty years, the most common method is to collect data via a survey derived from Parasuraman, et al.'s (1985) SERVQUAL model (Naik and Srinivasan, 2015; Ya Lan, et al., 2015). In 2005, Parasuraman, et al. reframed this model to fit with the emerging need to measure customer satisfaction in an online environment. Parasuraman, et al.'s (2005) new model was called: E-S-QUAL. While E-S-QUAL gained research adoption in the academic literature, it failed to include a direct measure of customer satisfaction (Gaur and Agrawal, 2006). As a result, E-S-QUAL needs to be modified for utilization within the retail industry in particular (Gaur and Agrawal, 2006). There is also a considerable strength of opinion that survey designs should be tailored to the firm and customer being researched (Fassnacht and Kose, 2007; Ladhari, 2010; Finn, 2011; Olsen, et al., 2014 and Zeithaml, et al., 2002).

The initial survey that I proposed to the co-research group was developed directly from the literature (see appendix 1). This proposal was based on suggestions by Ladhari (2010) to combine e-satisfaction with business and customer specific measures. The intent of such a design is to collect a rounded view of the customer and their opinions about a specific website (Ladhari, 2010). Accordingly, a shopping frequency question was included based on the belief that shopping frequency could be an indicator of customer sentiment and its influence on behavior (Lim, et al., 2015; Verhagen and van Dolen, 2009). Questions drawn from Moe (2003) and Janiszewski (1998) were included in order to capture customer typology. Parasuraman, et al.'s (2005) E-S-QUAL model of online service quality was also included in full. Lastly, given Jiang and Rosenbloom's (2005) and Betancourt and Gautschi's (1990) belief that price and value perceptions are important to customer sentiment, three price perception questions were included. Prior to it being fully discussed with BGC stakeholders, this initial survey design was included in my thesis proposal.

Through discussion amongst the co-research group, it was determined that this survey was too long (at 34 questions) and not completely relevant to the business and customer. In addition, at the outset of the research process, we discovered that there were contractual limits to the number of survey questions (seven) that could be issued through BGC's online survey provider. In my somewhat myopic attempt to base the initial survey design on the literature I had also failed to include a direct customer satisfaction rating question, per the concerns raised by Gaur and Agrawal (2006). Therefore, through a reflexive and dialectic

process, a revised survey design was developed (see appendix 1). In fitting with AR practice, the resulting survey balanced the needs of the business, the research effort, and the contracts that were in place.

The revised survey comprised of seven questions, six of which were targeted at quantitative data collection. The survey included a direct customer satisfaction measure (Gaur and Agrawal, 2006), drawing on the approaches used by Ha and Lennon (2010), Diehl and Poynor (2010) and Belanche, et al. (2012). Based on similar studies by Green and Pearson (2009), Ladhari (2010), Belanche, et al., (2012), Wang et al., (2010) and Peng and Kim (2014), specific questions were included to capture perceptions of site content and site navigation. Perceptions of aesthetic appeal were collected based on Wang, et al. (2010) and Wang, et al., (2011). The intention to return to use the website and to recommend the website to a friend were included based on BGC's belief that these were both important measures of their specific customer's level of satisfaction with the site. Both questions also had support in the academic consumer satisfaction literature as part of the analysis on loyalty intentions. Parasuraman, et al. (2005) included intent to return and recommendation to a friend in their E-RecS-QUAL scale. In addition, Balanche, et al. (2012) and Long and McMellon (2004) found that these questions generated data highly correlated with e-satisfaction.

Omitted from this survey were questions regarding customer typology, shopping frequency, price perceptions, fulfillment, privacy and site performance. A primary determinant of the reason to exclude these question types (other than question number limitations) was that they had the least support in the academic literature. In fact, a number of the omitted question types have been found to generate inconclusive results (Petnji Yaya, et al., 2012; Kim and Jackson, 2009). Instead, per the requirements of the business and the hypotheses, the final survey design focused on the core areas of e-satisfaction identified by the literature (e.g., Green and Pearson, 2009; Deveraj, et al., 2002) and BGC's understanding of what was important to their customer (per Olsen, et al., 2014 and Zeithaml, et al., 2002). This decision was supported by the literature, which indicated that the omitted questions were mostly useful for gaining deeper customer insights (Green and Pearson, 2009; Devaraj, et al., 2002). Due to the ability to capture open-ended feedback for this purpose (Ivankova, 2015; Cresswell, 2014), it was felt that these additional quantitative questions could be eliminated.

In addition to the question development having a basis in the literature, so did the measurement scales. Petnji Yaya, et al., (2012) found that two-thirds of all academic e-satisfaction surveys were based on a five-point scale. Coelho and Esteves (2007) also found few differences between five and ten point measurement scales. The exception was that there was some improved explanatory power associated with larger scales. In acknowledgment of the usage in the literature, and Coelho and Esteves' (2007) findings, we decided to use a five-point scale for the majority of the survey. Based on the potential benefits, we included a ten-point scale for the measurement of loyalty intention.

Second On-Site Survey Development

After implementing the aforementioned survey on the BGC website, the co-research team continued to work with BGC's online survey partner. After several weeks of discussions, it was agreed that the on-site survey could be expanded to eleven questions. This provided the opportunity for the co-research team to review the initial survey results and determine if any adjustments were required. This effort coincided with BGC's Executive Leadership Team deciding that they needed to gain a deeper understanding of their online and store customer base. In large part, the ELT's desire for additional customer knowledge was influenced by the weekly discussions about customer behavior and satisfaction that were being generated by this action research effort.

Preliminary reviews of the qualitative feedback data highlighted areas of site opportunity that would require additional research. Specifically, the qualitative data highlighted that site performance and task support were potential problem areas. Other research refinement changes stemmed from initial result discussions with the ELT and their subsequent desire to generate a deeper understanding of customer typology and visit patterns. As a consequence, a consensus was reached about the need to produce a revised customer survey. In order to maintain continuity, it was decided to leave the main site satisfaction measurement as it was. In addition, the navigation/ease of use and loyalty question types remained.

To determine the exact question types to include in the revised survey the co-research group reassessed the initial survey that I had proposed. We discussed each question type, informed by the literature review findings. Through discussion of Parasuraman et al.'s (2005) E-S-QUAL scale, we included questions on site navigation and site performance. After consideration of Jiang and Rosenbloom (2005), we added a question on price perceptions. Based on findings by Lim, et al. (2015) and Verhagen and van Dolen (2009) we added a question on site visit frequency. Following a review of the preliminary qualitative feedback, we also reached an agreement that customer typology and task support insights were important to understand in order to determine any action requirements in this area. The customer typology and task support questions were drawn from Moe (2003) and Janiszewski (1998) combined with BGC's knowledge of customer preferences. The resulting survey is shown in Appendix 3.

Post-Visit Survey Development

The development of a post-visit survey allowed for the expansion of the survey to thirteen total questions. This expanded question capability allowed for an additional category of questions to be added, compared to the concurrent on-site survey: product information. Two 'product information' questions were added in order to better collect data for hypotheses H7 and H8. The questions were drawn from academic research conducted by Diehl, et al. (2003), Urbany, et al. (1989), Janiszewski (1998), Belanche, et al. (2012) and Zhenhui, et al. (2010). Further modifications compared to the on-site survey occurred, based on continuing group discussions, in already established question categories. Aesthetic

appeal had not been measured since the first on-site survey, and was added back as a number of the group members felt this was a potential customer issue. Due to functionality changes to the site regarding navigation and ease of use, we decided to add an additional question in this area in order to provide a deeper understanding of customer perspectives. Due to a large volume of site performance feedback on the on-site surveys, we also decided to add an additional question in this area.

In order to accommodate additional survey questions, compared to the on-site survey, several questions were also removed. Shopping frequency was removed as the co-research group felt that adequate frequency data could be collected via the on-site survey, and this information was not perceived to be central to our immediate objectives. In addition, we decided to remove shopping event and purchase completion impediments for the same reasons. A side-by-side comparison of the surveys, including academic references and support, is provided in appendix 5.

As briefly mentioned earlier in the methodology chapter, the post-visit survey was provided randomly to two-percent of customers that browsed three or more pages of the website. Once the randomly selected customers exited from the website they were immediately presented with the questionnaire. Random sampling was used to ensure that each customer had an equal probability of being selected. This type of sampling approach is considered to be the best for minimizing sample bias (Easterby-Smith, et al., 2012; Cresswell, 2014).

The only sample selection criterion was that the customer must have visited three pages on the site before they could be included in the pool for random sampling. This limitation was applied so that anyone completing the survey had at least tried to navigate through the site at a basic level. This sampling criterion means that customers who could not navigate the site at all were excluded from participation. This difference in sampling methodology is also clearly different from the on-site methodology. I asked for discussion about the merits of this difference during a weekly group meeting. The strong consensus of the insider group was that this method would produce more informative insights than using a purely random sample of all website visitors could. The group's recommended method was a common practice at BGC, and in order to ensure business support, I ultimately had to agree to this approach. In certain situations like this, I sensed that it was better for the overall success of the project to concede my own views and support strongly felt co-research group recommendations as a means of building trust and confidence amongst the group.

Click Data Collection Methodology

As a way of being able to identify the cause and affect relationships in online environments, several methodologies have been proven to be highly reliable: A/B testing and Multi-Variable Testing (MVT) (Kohavi, et al., 2009; Moe, 2003). A/B testing involves the randomized splitting of traffic entering a website between different site experiences (Kohavi, et al., 2009). By doing so, it is possible to measure changes in behavior through the mining

of click data (Kohavi, et al., 2009; Bucklin and Sismeiro, 2009). For the purposes of this research, A/B testing was utilized to understand how shopping experience directly influenced behavior.

In order to construct the A/B tests a number of influential pieces of literature were referenced. Consideration was given to how to construct each test relative to Simonson’s (1999) model of behavioral psychology, which was also used in the development of each hypothesis. In addition, Mehrabian and Russell’s (1976) S-O-R theory was used to better conceptualize how each test needed to be organized in order to measure the O and R impacts. A number of similar research studies were also reviewed and their results interpreted into six distinct A/B test scenarios (see Table 4-3). I initially interpreted each test, based on the literature, in order to fit BGC’s requirements. Once presented to the co-research and ELT groups these initial interpretations were further modified to fit the operational context and requirements of the business.

Table 4-3 : A/B Test Construct

Test	Behavioral Psychology (Simonson, 1999; Mehrabian & Russell, 1976)	Test Construct References	Behavioral (R)esponse Tests of Hypotheses
Colorization	Ease of Justification, S-O-R	Broniarczyk & Griffin (2014); Wu, et al. (2014)	H2: purchase justification impact on sales revenue
Clickable Top Nav	Ease of Processing, S-O-R	Lee & Kozar (2012); Belanche, et al. (2012); Bart, et al. (2005)	H1: decision cost impact on sales revenue
Product Description	Ease of Processing, S-O-R	Wu, et al. (2014); Gwo-Guang & Hsui-Fen (2005)	H8 & H1: product information and decision cost impact on sales revenue
Video	Ease of Processing, S-O-R	Belanche, et al. (2012); Zhenhui, et al. (2010)	H8: product information impact on sales revenue
Fly-Out Menu	Ease of Justification, Ease of Processing, S-O-R	Broniarczyk & Griffin (2014); Diehl et al. (2003); Zhenhui, et al (2010)	H1 & H2: decision cost and purchase justification impact on sales revenue
Size Display	Ease of Justification, Ease of Processing, S-O-R	Broniarczyk & Griffin (2014); Lee & Kozar (2012)	H2 & H8: product information and purchase justification impact on sales revenue

At the outset of the research process, BGC’s knowledge of how to conduct A/B test was very rudimentary. On a de facto basis this lack of knowledge resulted in numerous unsuccessful trials of each test concept deployment. Each of these unsuccessful trials was actually an important component of the knowledge building process, although they did cause some tension amongst the group. Generally, this meant that there was a one or two-week period of trial and error in order to perfect the implementation and the collection of customer response data associated with each A/B test deployment. In addition to an improved ability to properly execute the tests, these unsuccessful trials also led to an improved ability to measure and understand the results of each test. These improvements stemmed from group reflection on the each unsuccessful attempt, which led to a better understanding of the objectives, the tools being used and the data being captured. Data collected during each unsuccessful trial was excluded from the data analyzed for decision-making and this thesis paper.

Colorization Test

Despite the need to modify the tests used in the literature, I used a theoretical basis as the starting point for discussions on each test construct. A test of the impact of the usage of different colors to emphasize sale messaging and add-to-cart was derived from Simonson (1999) and Wu, et al. (2014). Wu, et al. (2014) found that the usage of color on a website influenced an emotional and behavioral response from site visitors. Simonson (1999)

highlights how purchase intent can be influenced by changes in product emphasis and positioning. Interpretation of these points led to the formulation of a colorization test. In this test, the color of discounted product prices was changed to emphasize that they were on sale. In addition, the color of the add-to-cart button was changed to the same color to emphasize purchase opportunity. The actual A vs. B treatments that were implemented are shown in appendix 6.

Clickable Top Nav Test

Lee and Kozar (2012) found that website navigability and interactivity both have a direct impact on purchase intent. Website usability was also found to directly affect consumer's emotional response and satisfaction levels (Belanche, et al., 2012). The ability to quickly navigate through a website has also been found to increase levels of trust and purchase intent (Bart, et al., 2005). One of the problems identified through preliminary customer data collection was a concern about the ability to navigate the website. In particular, not all elements of the persistent navigation tools (top nav) could be clicked upon, contrary to user expectations. Therefore, based on the literature it was determined that purchase intent and satisfaction could probably be improved by creating the ability to click on all elements of the top nav.

Product Description Test

Empirical testing identified that the layout of a website impacted upon the emotional response to the site, which in turn manifests through satisfaction and purchases (Wu, et al., 2014). It has also been theorized that clarity of communication contributes to purchase intent (Gwo-Guang and Hsiu-Fen, 2005). In addition, each of these elements contributes to decision costs, therefore linking them to ease of decision processing (Broniarczyk and Griffin, 2014). Combined, these views suggest that changes in website layout, where product information is de-emphasized, would lead to behavioral implications. One such change, was identified, where the transactional (e.g. add-to-cart) buttons were displayed below the screen fold on most screen resolutions. The hypothesis was that this made it difficult for the customer to figure out how to purchase the product. Therefore, we devised a test where the transactional buttons were switched with the product descriptions. The experimental and control treatments that were used as shown in appendix 7.

Video Test

Broniarczyk and Griffin (2014) propose the need to conduct research into the usage of videos as an information source and their impact on decision costs. It is believed that rich information sources such as videos allow consumers to generate a much better understanding of the product being considered (Broniarczyk and Griffin, 2014). Zhenhui, et al. (2010) suggest that giving consumers the ability to receive additional product information, such as through the activation of a video, has a direct impact on their purchase intent. Conversely, Belanche, et al. (2012) contends that such information adds a decision-making burden and leads to purchase deferral. Through discussion with BGC's ELT, it

became clear that the production of product videos was a significant financial expense. As a result, it was requested that we understand whether the inclusion of product videos actually lead to increased revenue and improvements in customer perceptions. Consequently, an A/B test was developed where some customers had access to product videos, while others did not. Please reference appendix 8, to see how product imagery, including videos, was displayed to the customer. From a test perspective, the video selection was simply removed, but all other imagery was retained.

Fly-Out Menu Test

Improving product categorization has been shown to improve perceptions of the assortment and the retailer in general (Mogilner, et al., 2008; Broniarczyk, et al., 1998). In addition, better communication of where to find products has been linked to reductions in decision costs and purchase intent (Broniarczyk and Griffin, 2014; Hoch, et al., 1999). Providing better control of the ability to navigate the website also fosters greater engagement with site visitors (Zhenhui, et al., 2010). With these theoretical positions in mind, the co-research group felt that improvements to the website's main navigation tool (the fly-out top nav) could be beneficial. With this in mind, we designed an improved menu, which improved on the readability and categorization of the products within the menu. A subset of the fly-out menu experiment and control changes is shown in appendix 9. Three different versions of this menu were ultimately tested against the control site. These changes were prompted by group reflection on some of the preliminary results data. This will be discussed further in the results and findings chapter of the paper.

Size Display Test

Lee and Kozar (2012) conclude that delivering up-to-date product information provides a significant influence on purchase intent. This finding is consistent that of Jiffeng, et al. (2012), who suggest that product uncertainty also has a strong bearing on customer satisfaction. It appears likely that this is due to the reduction in decision costs that come from provided better information levels (Broniarczyk and Griffin, 2014). Jiang and Rosenbloom (2005) postulate that improvements in product availability information increases repeat purchase justification. These findings were also consistent with preliminary reviews of the qualitative data collected by the first on-site survey. Following discussions about how to respond, the co-research group decided to implement a test that increased product availability information by more prominently displaying the availability of each product size in the experiment, compared to keeping it in a more obscure location in the test. Please see appendix 10 for an actual example of the two test treatments.

Click Data Methods of Measurement

Through discussion with the co-research groups, it was determined that customer behavioral data needed to be measured using 'click data' captured through each of the site usage monitoring tools and *Monetate*. In the early years of electronic commerce, Montgomery, et al. (2004) and Johnson et al. (2004) noted that the usage of this type of data was relatively

uncommon from an academic research perspective. While the usage of click data continues to be in a formative stage, it is becoming more prevalent as the basis for empirical research (see Ayanso and Mokaya, 2013; Chiang, et al., 2013; Bucklin and Sismeiro, 2009). Click data is rapidly becoming one of the richest sources of consumer behavior information (Su and Chen, 2015; Huang and Miegheem, 2014; Rutz and Bucklin, 2012). This type of data is particularly useful for understanding consumer browsing and purchasing patterns (Su and Chen, 2015; Rutz and Bucklin, 2012), but because it is aggregated above the individual consumer level, it is not ideal for understanding demographics or other individual identifying information (Rutz and Bucklin, 2012).

Table 4-4 : Behavioral Measurement Variables

Construct	References
Average Order Value	Huang et al. (2009); Huang & Van Miegheem (2014); Ayanso & Yoogalingam (2009)
Conversion %	Moe & Fader (2004); Moe (2003); Montgomery et al. (2004); Ayanso & Mokaya (2013)
Flow Path	Montgomery et al. (2004); Moe (2003); Su & Chen (2015); Rutz & Bucklin (2012)
New Visitor Conversion %	Buklin & Sismeiro (2003)
Page Views	Huang et al. (2009); Buklin & Sismeiro (2003); Moe (2003); Senecal et al. (2005); Huang & Van Miegheem (2014); Rutz & Bucklin (2012)
Repeat Visitors	Buklin & Sismeiro (2003); Chiang, et al. (2013)
Sales Revenue	Ayanso & Mokaya (2013); Hauser, et al. (2009)
Sessions	Moe (2003); Bucklin & Sismeiro (2009); Montgomery, et al. (2004); Su & Chen (2015); Rutz & Bucklin (2012)
Time per Page	Kwan et al. (2005); Huang et al. (2009); Huang & Van Miegheem (2014); Su & Chen (2015)
Total Time on Site	Szymanski & Hise (2000), Senecal et al. (2005); Huang et al. (2009); Su & Chen (2015)
Unique Visitors	Moe (2003); Moe & Fader (2004)

From reviews of the academic literature, the variable construct detailed in Table 4-4 was developed (see appendix 11 for definitions of these and other related click data variables). These metrics and other common forms of clickstream browsing and purchasing data were collected as part of the A/B testing activities. Each metric was calculated at the individual website user level and then aggregated to a total user group level by day, and by test treatment. For example, conversion % was calculated for each control and experimental treatment, by day, for the duration of the each test treatment. Using this data, it became possible to understand differences between each treatment, via analysis methods that will be covered in the next chapter of this paper.

It should also be noted that the aggregation of data from individual to day level is a routine task of the three different web applications being used for this research (*Google Analytics, Coremetrics, and Monetate*). As a consequence, the co-research group and I played no part in the interpretation of the individual data or in the calculation of the base metrics used for the analysis. The data was pre-aggregated and pre-calculated at the day level for each specific test treatment. For the purposes of this thesis, this pre-aggregation forms the basis of the dataset used for the A/B test analysis.

4.5.2 Data Analysis – Survey Data

In order to analyze the quantitative survey data, methods were selected that met three criteria: suitability to the problem and the data, common usage in the relevant academic literature, and usage within BGC. This is in fitting with an action research methodology, where it is considered crucial to select methods that are considered to be appropriate to both practice and academia. It is only through such a practice that rigorous methods of research can be applied in a manner where practice-based knowledge is created, and transformative change is inspired (Ivankova, 2015).

To prepare the surveys for analysis, the collected data was reviewed for completeness, and any surveys that contain missing or otherwise erroneous data were discarded (e.g., Treiblmaier and Pinterits, 2010). Following this, the closed questions were converted to numerical responses. The Likert responses were converted to a 1 to 5 scale, with a ‘strongly negative’ response being represented by a score of 1, whereas a ‘strongly positive’ response was scored as a 5. In parallel, the open-ended responses were themed and coded, allowing for each response score to be calculated in total and also matched to a code and theme (per Cresswell, 2014 and Ivankova, 2015).

Responses to multiple choice questions, such as “How frequently do you visit this site?” and “What was your primary reason for visiting this site today?”, were used to stratify the quantitative responses. This approach allowed for a comparison of attitudes between different customer typologies and shopping missions. As a specific example, the question responses were stratified according to the customer typologies developed by Janiszewski (1998) and Moe (2003). Janiszewski (1998) created a consumer typology by separating consumers into two types: goal-directed and exploratory searchers. Moe (2003) further developed this typology by stratifying the exploratory searchers into hedonistic and knowledge building consumers, and the goal-directed searchers into directed-buying and search/deliberation consumers. For this research, each customer visit reason type was categorized according to these typologies (see Table 4-5).

Table 4-5 : Consumer Typology Measures

Survey Response	Janiszewski (1998) Typology	Moe (2003) Typology	Intent to Purchase (Moe, 2003)
To make a purchase online	Goal-directed	Directed Buying	Now
To save product favorites for a future store visit		Search/Deliberation	Soon
To find a store			Soon
To schedule a store appointment			Soon
Research for a future online purchase	Exploratory	Knowledge Building	In the future
Research for a future in-store purchase		Hedonistic	In the future
Just to browse			Just browsing
To look for event planning information			Just browsing
Other	Other	Other	Other

Scale Reliability and Dimensionality

In order to assess the reliability of the measurement scales, Cronbach's alpha was calculated for each survey scale (per Cole, 2005; Kim, et al., 2009; Belanche, et al., 2012; Long and McMellon, 2004). To perform this calculation, and all of the other statistical operations, SPSS v.22.0 statistical software was utilized. Any scales not meeting a .7 minimum Cronbach's alpha value were adjusted or omitted from the analysis, per common academic research practices (e.g., Peng and Kim, 2014; Kim, et al., 2009). All of the scales were actually found to meet this criterion, with the exception of the first survey. The first survey was almost entirely based on organizational requirements, rather than support from the literature. After we had discussed this reliability issue, the co-research and ELT teams agreed to reconstruct each subsequent survey with consideration of analytical reliability.

To assess the unidimensionality of the scales a principal components analysis was undertaken. This exploratory factor analysis was conducted on the combined results of each survey in order to determine if there were any latent variables present. The extraction of factors required item loadings greater than 0.5 and eigenvalues of 1.0 or greater. Factor loading values were generated using Varimax rotation with Kaiser Normalization (per Belanche, et al., 2012; Long and McMellon, 2004; Kim, et al., 2009; Rajic and Dado, 2013; Parasuraman, et al., 2005). The resulting factors were also reviewed to ensure that they explained a significant percentage of the total variance (e.g., Belanche, et al., 2012; Parasuraman, et al., 2005). Through this analysis, two additional variables were identified and added to the survey analysis that followed.

Hypothesis Testing

Following validation of the scales and variables, the survey data was ready for hypothesis testing. From a review of the literature, Pearson chi-square analysis is the most common method for understanding the differences between surveys (e.g. Lau and Wong, 2000; Dommeyer and Moriarty, 1999; Li, et al., 2008; Kim, et al., 2009; Cole, 2005; Huang, et al., 2009), Gwo-Guang and Hsiu-Fen, 2005). Other tests of differences, such as *t-tests*, are also commonly used but these methods assume that the data is parametric in nature. Given that the response data was ordinal in nature and based on a Likert scale, it could not be assumed that the distance between each Likert response was consistent. As a consequence, the data was assumed to be non-parametric and required the adoption of non-parametric hypothesis testing methods. Consistent with the aforementioned literature, only chi-square probabilities of <0.05 were considered to be indicative of statistically significant differences.

4.5.3 Data Analysis – Clickstream Data

In contrast to the survey data, the clickstream data was confirmed to be parametric in nature via descriptive analysis conducted using SPSS v.22.0. The clickstream data that was collected measured consumer behaviors on the website across each of the measurement constructs detailed in Table 4-4. The objective of the clickstream data analysis was to

determine if the experiments had any impact on sales revenue. In addition, the other measurement constructs were analyzed in order to better understand changes in the underlying behaviors generated by the experiments, as compared to the behaviors exhibited in the control group.

Hypothesis Testing

In order to compare an experiment versus a control, as in the A/B testing approach that was utilized for this thesis, *t-tests* have been found to be the most common hypothesis testing approach (eg. Kohavi, et al., 2009; Caro, 2012; Verhagen, 2009; Chang and Chen, 2009). Accordingly, in order to compare each measurement construct between the experiment and control groups, two-tailed paired sample *t-tests* were used (e.g. Naik and Srinivasan, 2015). Paired *t-tests* are particularly useful for testing the difference in means across a single group that is being influenced by two different treatments (Ivankova, 2015). According to academic research practice (e.g., Caro, 2012; Kohavi, et al., 2009; Chang and Chen, 2009), statistically significant differences were assumed at a two-tailed paired *t-test* *P* value of <0.05.

To further support the hypothesis testing, stepwise regression analysis was conducted to gain insight into the measurement constructs that were highly correlated with sales revenue in each test scenario (e.g. Gupta and Kabadayi, 2010; Szymanski and Hise, 2000). Each regression model was reviewed for severity of multicollinearity using the Variance Inflation Factor (VIF) test (Szymanski and Hise, 2000; Gupta and Kabadayi, 2010; Studenmund, 1992). Any model having a VIF in excess of 5.0 was considered to have high multicollinearity and was excluded from the analysis (per Studenmund, 1992), albeit some literature suggests a VIF cut-off of 10.0 (Gupta and Kabadayi, 2010; Parasuraman, et al, 2005). In order to be considered for analysis, each model also had to have an R^2 of greater than 0.7. In addition, Pearson correlations were performed and cross-tabulated for each measurement construct. This provided a further understanding of the quantitative relationships that exist across each of the constructs.

4.6 Research Activity Timeline

In order to provide a better understanding of the interrelationship of each research instrument, Table 4-6 was constructed. In total the primary data collection effort took thirty-one weeks, where the first week of data collection with the first survey instrument was recorded as 'Research Activity Week 1'. The organization was actually collecting overall site satisfaction data prior to the beginning of the thesis research. In a few cases, this data was retrieved and used for longitudinal analysis. When referring to such cases, the weeks are noted in relation to the research activity weeks, and are therefore shown as zero or negative week values.

From Table 4-6 one can see that the first survey data was collected over a period of three weeks prior to the implementation of the first four website experiments. Following these

A/B tests, the second survey was implemented and captured data for eleven weeks. Because this second survey overlapped with the next two experiments, it is considered to be the least important measure of customer attitudes. This second survey collected data for eleven weeks in order to collect the same number of responses as the first survey. The first survey was conducted during a period of heavy site usage, whereas the remaining surveys were distributed during relatively lower levels of purchase seasonality. The third on-site survey and the post-visit survey were distributed at the same time, and for a period of seven weeks.

Table 4-6 : Research Activity by Research Week

Activity	Research Activity Week																															
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	
Pre-Site Sat & Feedback																																
Clickable Top Nav Experiment				X	X	X	X	X	X	X	X																					
Colorization Experiment				X	X	X	X	X																								
Video Experiment						X	X	X	X																							
Product Description Experiment				X	X	X	X	X	X	X																						
Post-Site Sat & Feedback																																
Fly-Out Menu Experiment																X ₁	X ₁	X ₁	X ₁	X ₂	X ₂	X ₃	X ₃	X ₃	X ₃							
Size Display Experiment																							X	X	X							
Customer Typology, Sit Sat & Feedback (On-Site)																																
Customer Typology, Sit Sat & Feedback (Post-Visit)																																
Overall Site Satisfaction & Sales Revenue Data																																

In order to test the relationship between overall site satisfaction and sales revenue, the total site satisfaction score from each of the on-site surveys was combined over time. This data was available because the first survey actually collected data from research week 1 to week 11, and the second survey actually collected data from research week 12 to week 24. However, all of the survey response data for weeks 4 to 11 and 22 to 24 were excluded from the analysis, with the exception of the overall site satisfaction rating. In particular, data from weeks 4 to 11 was excluded due to the overlap with the first four site experiments and the potential impact this may have had on the survey results. Other research activities occurred before and after the weeks illustrated in Table 4-6, such as the analysis of the data and discussion and reflection upon the results. As such, this table is intended to illustrate only how the timing of each research instrument is interrelated.

4.7 Methodology Summary

This chapter of the thesis provided a detailed discussion of the methodology employed for this research. Importantly, this included discussion of the rationale for selecting each component of the overall methodology, supported by insights from the extant literature and agreement with the insider stakeholders. As part of the initial methodological discussion, it was explained how action research provided the overall research framework. Within the

action research framework, there was a discussion about the validity of the types of data collected and the methods that would be employed for doing so. Support was also provided for each of the quantitative and qualitative data analysis methods that were selected. Lastly, a research timeline was offered, which provided a bridge from the conceptual model of data collection discussed in chapter 3, to the practice-based methodology that was ultimately deployed.

4.7.1 Applicability of the Methodology to the Intended Outcomes

It is important to note that the support for each element of the methodology was developed prior to and during the data collection process, and not as an ex-post justification for those methods. As a consequence, not all of the elements discussed in this chapter actually provided deep insights into the problem; the research findings will be discussed in detail in the next chapter. However, each selected element of the research method was determined to be important based on the rigorous participatory action research process that was undertaken. A benefit of having undertaken this participatory process is that it can confidently be stated that the resulting methodology was the best one that we could collectively develop. Specifically, the selected methodologies were the most viable methods within the lexicon of the individuals and business involved, as determined through extensive dialogue, debate, trial, and error.

Despite the selected approach being the best co-generated approach to the problem, this does not mean that better paths did not exist. Simply, that the methods described were the best approach that we could collectively agree to, collectively understand, and collectively engage in. I do not believe that academic research integrity was significantly sacrificed in the resulting method selection, but I strongly believe that the selected approach was the best available to us. As a consequence, it was the most applicable method for delivering the intended outcome and in keeping with recommended practices from the academic literature on participatory action research (e.g., Greenwood and Levin, 2007, and Ivankova, 2015).

Chapter 5: Results and Findings

This chapter of the thesis details the analysis, results, and findings from the research into each hypothesis. The objective is to convey the research methods used, and how the results were interpreted. The analysis was conducted systematically against each hypothesis statement, and then collective findings were generated through discourse and reflection. As a consequence, this chapter is presented in a hypothesis-by-hypothesis format and concludes with a findings sub-section that discusses the overall implications of the research effort.

5.1 Analysis and Results

5.1.1 Hypothesis 1: Analysis

For hypothesis 1, the null hypothesis to be tested was²:

$$H1_0: \mu = K ; \text{ Consumer decision costs have no impact upon sales revenue}$$

Given that consumer decision-costs have been identified as being related to the amount of time it takes to complete a desired task (Betancourt and Gautschi, 1990), website ease of use (navigation) was central to the testing of H1 (Lee and Kozar, 2012; Belanche, et al., 2012; Bart, et al., 2005). In order to test this hypothesis three different test treatments were developed and implemented at different times: *Clickable Top Nav Experiment*, *Fly-out Menu Experiment v.1*, and *Fly-Out Menu Experiment v.3*.

Clickable Top Nav Experiment

The *Clickable Top Nav* experiment involved enabling the ‘clickability’ of the main product category headers at the top of each website page. An example of the ‘top nav’ is shown in appendix 9. Prior to implementing this test, each main product category was not clickable, and would result in no response from the website. In the experiment, each of these main product categories were linked to the main Product Landing Page (PLP) for each respective category. The group’s working hypothesis was that significant numbers of customers were confused by the lack of clickability, and that improving this functionality would improve the navigability and ease of use of the website.

The Clickable Top Nav experiment ran for approximately six weeks, from research week 4 to week 11. Using *Monetate’s* A/B testing application, website traffic was randomly split 50/50 between the experiment and the control. In total 3,704,431 website sessions occurred during this period, with 1,852,757 experiencing the Clickable Top Nav experimental treatment, and 1,853,674 experiencing the control treatment. The following clickstream metrics were aggregated from individual user sessions to the total day level, for each treatment, by the *Monetate* application (definitions are provided in appendix 11):

- Conversion rate
- New visitor conversion rate
- Add to cart rate
- Cart abandonment rate
- Bounce rate
- Revenue per session
- Average page views
- Average time on site
- Average order value

² μ = mean of the test group ; K = mean of the control group

In order to test for differences between the treatment and control, paired sample *t*-tests were performed comparing each metric across the 39 days of the experiment. Any metric comparisons that resulted in a two-tailed *P* value of 0.05 or greater were discarded from the analysis. The only exception being revenue, given that it was the main target of the testing effort. A summary of the paired samples *t*-test results is shown in Table 5-1. According to these results, sales revenue does not appear to have been impacted by the experimental treatment.

Table 5-1 : Clickable Top Nav *t*-test Significant Results Summary

Variable	Sessions	Mean	% Change in Mean	N	Std. Deviation	Std. Error Mean	t	df	P
BounceRateExperiment	1852757	.214	-1.4%	39	.023	.004	-5.334	38	.000
BounceRateControl	1853674	.217		39	.021	.003			
AveragePageViewsExperiment	1852757	12.014	1.4%	39	.452	.072	6.950	38	.000
AveragePageViewsControl	1853674	11.851		39	.456	.073			
AverageTimeOnSiteExperiment	1852757	27.888	-0.7%	39	2.400	.384	-2.115	38	.041
AverageTimeOnSiteControl	1853674	28.072		39	2.387	.382			
RevenuePerSessionExperiment	1852757	1.900	-0.2%	39	.372	.059	-.141	38	.888
RevenuePerSessionControl	1853674	1.904		39	.390	.062			

Despite the lack of impact on sales revenue, there did appear to be an impact on bounce rate, average page views, and average time on site. By comparing the difference in the mean value of these variables, several insights can be drawn. First, bounce rate appears to have declined in the experiment, indicating fewer visitors choosing to leave the site immediately. In addition, time on site is lower, indicating a reduction in decision costs. This is further supported by the increase in page views. Therefore, it does appear that the experiment successfully reduced decision costs and improved ease of use.

In support of the *t*-test analysis, an exploratory factor analysis (EFA) was conducted for both the experimental treatment and the control treatment. Further validating the *t*-test results, the factors extracted from the experiment separate time on site and page views from revenue, when compared to the control group. In fact, revenue is grouped with the conversion rate variables and cart abandonment in the experiment, which are all variables that do not appear to have been affected by the experiment.

Table 5-2 : Stepwise Regression Summary – Clickable Top Nav Experiment Treatment

Variable	Partial regression coefficient (B)	Standard error	Standardised regression coefficient (Beta)	t-statistic	VIF	F-statistic / Significance	
						of F	R2 / Adjusted R2
Intercept	-2.074	0.114		-18.127		2047.516	0.991
ConversionRateExperiment	130.596	2.460	0.868	53.095	1.105	.000c	0.991
AverageOrderValueExperiment	0.016	0.001	0.288	17.641	1.105		

Dependent Variable: RevenuePerSessionExperiment

A stepwise regression was performed using the experimental data in order to further understand the variables that most strongly contribute to changes in sale revenue. From this analysis (shown in Table 5-2), it was determined that conversion rate and average order value are the most highly influential variables on sales revenue per session. This again validates the findings of the *t*-test analysis, as neither of these variables was found to have changed significantly from control to experiment.

Fly-Out Menu Experiment v1

The *Fly-Out Menu* experiment involved the redesign of the top nav fly-out menu. The ‘top nav’ refers to the persistent navigational ribbon at the top of a website and serves as the main method of site navigation. An example of the fly-out menu experimental treatment for v1, versus the control treatment, is shown in appendix 9. Prior to implementing this test, there was no organization to the sub-menus below each main top nav category header, which made it difficult to read and understand how to navigate the site. In the experiment, the sub-menu was organized by shopping objective and additional product categorizations were added. The goal was to provide customers with an easier way to narrow down their product preferences and reduce the amount of time (therefore reducing decision costs) that it took to navigate the site.

The Fly-Out Nav experiment ran for approximately four weeks, from research week 14 to week 17. Using *Monetate’s* A/B testing application, website traffic was randomly split 75/25 between the experiment and the control. In total 1,750,088 website sessions occurred during this period, with 1,311,242 experiencing the fly-out menu experimental treatment, and 438,846 experiencing the control treatment. The decision to split the traffic, with 75% receiving the experimental treatment was a decision by the business based on their belief that it was going to be immediately beneficial to revenue generation. For the analysis of each treatment the same set of clickstream metrics³ were aggregated from individual user sessions to the total day level by the *Monetate* application.

Table 5-3 : Fly-Out Menu v1 *t*-test Significant Results Summary

Variable	Sessions	Mean	% Change in Mean	N	Std. Deviation	Std. Error Mean	t	df	P
RevenuePerSessionExperiment	1311242	2.1374	-4.1%	27	.283	.054	-1.582	26	.126
RevenuePerSessionControl	438846	2.2289		27	.405	.078			
AveragePageViewsExperiment	1311242	13.9801	-1.6%	27	1.019	.196	-4.728	26	.000
AveragePageViewsControl	438846	14.2135		27	1.102	.212			
AverageOrderValueExperiment	1311242	127.7770	-6.0%	27	11.427	2.199	-2.721	26	.011
AverageOrderValueControl	438846	135.8719		27	19.604	3.773			

In order to test for differences between the treatment and control, paired sample *t*-tests were performed comparing each metric across the 27 days of the experiment. Any metric

³ See Clickable Top Nav experiment results metric summary

comparisons that resulted in a two-tailed P value of 0.05 or greater were discarded from the analysis. The only exception being revenue, given that it is the main target of this testing effort. A summary of the paired samples t -test results is shown in table 5-3. As with the previous results, sales revenue does not appear to have been impacted by the experimental treatment.

In addition to the lack of statistically significant impact on revenue per session, this treatment also resulted in two further unexpected results: page views and average order value both declined. In fact, sales also declined in real terms, albeit not with a high degree of statistical probability. Based on group discourse on these results, we concluded that the new approach to categorization and navigation was not leading to business benefits, and showed no clear indicators that decision costs were being reduced either.

Table 5-4 : Exploratory Factor Analysis – Fly-Out Menu v1

Exploratory Factor Analysis - Experiment

Factor Items	Factor Loadings by Component			Extracted Communalities
	1	2	3	
ConversionRateExperiment	0.980			0.981
NewVisitorConversionExperiment	0.947			0.928
AddToCartExperiment	0.931			0.931
CartAbandonmentExperiment		0.775		0.659
BounceRateExperiment			0.922	0.853
RevenuePerSessionExperiment	0.672	-0.594		0.921
AveragePageViewsExperiment	0.857			0.829
AverageTimeOnSiteExperiment	0.547	0.640		0.739
AverageOrderValueExperiment		-0.705		0.793

Eigenvalue 7.633
% of the variance explained 84.8

Exploratory Factor Analysis - Control

Factor Items	Factor Loadings by Component			Extracted Communalities
	1	2		
ConversionRateControl	0.839	0.513		0.967
NewVisitorConversionControl	0.819			0.893
AddToCartControl	0.925			0.883
CartAbandonmentControl		-0.733		0.585
BounceRateControl				0.145
RevenuePerSessionControl		0.922		0.908
AveragePageViewsControl	0.916			0.841
AverageTimeOnSiteControl	0.567			0.367
AverageOrderValueControl	-0.699	0.544		0.784

Eigenvalue 6.375
% of the variance explained 70.8

In order to delve deeper into these results, an exploratory factor analysis (EFA) was conducted for both the experimental treatment and the control treatment. The results of

this EFA are shown in Table 5-4. As found in the Clickable Top Nav EFA analysis, sales revenue was factored together with conversion rate and cart abandonment rate. This occurred again in the Fly-Out Menu v1 control group EFA. However, the EFA analysis for the Fly-Out Menu experiment v1 provides a grouping of factors that the co-research group found confusing (e.g. it did not make sense to the group that conversion rate and cart abandonment were in separate factors).

A stepwise regression analysis of the experimental treatment data again supported the conversion rate and average order value variables as having the strongest contribution to sales revenue (see Table 5-5). Based on these results, the co-research group concluded that without seeing an impact on either of these variables it was unlikely that any sales benefits could be derived from the current experimental treatment. As a consequence, additional adjustments to the fly-out menu treatments were developed and implemented. The first adjustment was to add ‘Sale’ categories to the sub-menu structure (as shown in appendix 9 under treatment v2.). The co-research group debated whether the addition of price promoted merchandise to this menu might improve the conversion rate, and therefore lead to increases in sales revenue. However, given the likely impact of this change on price justification, I have chosen to display the results of the Fly-Out Menu v.2 test in the section discussing the testing of Hypothesis 2, rather than Hypothesis 1. However, the results of this experiment were also included in the write-up on conclusions and findings for Hypothesis 1.

Table 5-5 : Stepwise Regression Summary – Fly-Out Menu v1 Experiment Treatment

Variable	Partial regression coefficient (B)	Standard error	Standardised regression coefficient (Beta)	t-statistic	VIF	F-statistic /	R2 /
						Significance of F	Adjusted R2
Intercept	-2.319	0.069	0.000	-33.611		2373.877	0.995
ConversionRateExperiment	135.818	2.120	0.987	64.073	1.132	0.000	0.995
AverageOrderValueExperiment	0.017	0.000	0.704	45.724	1.132		

Dependent Variable: RevenuePerSessionExperiment

Fly-Out Menu Experiment v3

Following reflection on v2 (see hypothesis 2) of this experiment, further adjustments were made to the fly-out menu, and captured under version 3. This adjustment involved some further refinement of the categories included in the sub-menu structure. For example, the version 3 revision did not include ‘Sale’ merchandise in menu, making it more similar to the v1 treatment in this regard. As a result, v3 was directly targeted at decision-costs and not as highly influenced by price justification as v2, hence the decision to include the v3 write-up in this specific section of the paper. An example of the v3 treatment is shown in appendix 9.

The Fly-Out Top Menu v3 experiment ran for approximately six weeks, from research week 19 to week 24. Using *Monetate’s* A/B testing application, website traffic was randomly split 66/34 between the experiment and the control. In total 1,225,647 website sessions

occurred during this period, with 808,321 experiencing the Clickable Top Nav experimental treatment, and 417,326 experiencing the control treatment. The clickstream metrics were aggregated from individual user sessions to the total day level, for each treatment, by the *Monetate* application. In addition to the previously used clickstream data metrics, a number of additional metrics were added in order to improve the explanatory power of the tests (definitions are provided in appendix 11):

- Appointments Created
- PDP Save to Favorites
- PDP Find a Store
- PDP Make an Appointment
- Click Sale in Top Nav
- Click Make an Appointment in Top Nav

In order to test for differences between the treatment and control, paired sample *t*-tests were performed comparing each metric across the 40 days of the experiment. Any metric comparisons that resulted in a two-tailed *P* value of 0.05 or greater were discarded from the analysis. The only exception being revenue, given that it was the main target of this testing effort. A summary of the paired samples *t*-test results is shown in table 5-6. According to these results, sales revenue, again, does not appear to have been impacted to a statistically significant degree.

Table 5-6 : Fly-Out Menu v3 *t*-test Significant Results Summary

Variable	Sessions	Mean	% Change in Mean	N	Std. Deviation	Std. Error Mean	t	df	P
RevenuePerSessionExperiment	808321	2.269	2.9%	40	.320	.051	1.396	39	.171
RevenuePerSessionBControl	417326	2.205		40	.471	.075			
AveragePageViewsExperiment	808321	14.167	-2.2%	40	.864	.137	-4.424	39	.000
AveragePageViewsBControl	417326	14.481		40	.844	.133			
AverageTimeOnSiteExperiment	808321	29.175	-1.0%	40	1.926	.304	-2.137	39	.039
AverageTimeOnSiteBControl	417326	29.472		40	1.983	.313			
PDPMakeanAppointmentClickExperiment	808321	.002	9.1%	40	.001	.000	2.367	39	.023
PDPMakeanAppointmentClickBControl	417326	.002		40	.001	.000			
ClickSaleinTopNavigationDesktopExperiment	808321	.027	-3.1%	40	.007	.001	-2.609	39	.013
ClickSaleinTopNavigationDesktopBControl	417326	.028		40	.008	.001			

Interestingly, time on site seemed to decline in the experiment, suggesting a reduction in time (decision) costs. However, page views also declined, which was a concern given that this metric was positively correlated with the conversion rate, the prime influencer of sales revenue, in both prior H1 experiments. Countering this concern was an increase in appointments made from the product detail pages (PDP). This indicates that while an increase in online purchasing was not generated through the experiment, a significant increase in appointments was generated. The conclusion, of the research group insiders, was that because appointments generally convert to a store purchase at a rate of 60%, increases in appointments could be seen as a good proxy for increases in store sales. A reduction in clicks on sale categories was likely the driver of the reduced number of page

views, but because overall sales were not impacted the group did not view this as a negative change in behavior.

Exploratory factor analysis was undertaken in order to further understand the differences between the experiment and control treatments (see Table 5-7). As can be seen in the table summary, the factors and factor loadings are virtually identical between the two A/B test treatments. The only exception is that average order value appears to be of more relative importance in the experiment, where it accounted for 15% of the total variance, than it was in the control.

Table 5-7 : Exploratory Factor Analysis – Fly-Out Menu v3

Exploratory Factor Analysis - Experiment

Factor Items	Factor Loadings by Component			Extracted Communalities
	1	2	3	
ConversionRateExperiment	0.941			0.986
NewVisitorConversionExperiment	0.865			0.885
AddtoCartExperiment	0.648			0.809
CartAbandonmentExperiment	-0.858			0.761
BounceRateExperiment		-0.813		0.793
RevenuePerSessionExperiment	0.922			0.967
AveragePageViewsExperiment		0.856		0.823
AverageTimeOnSiteExperiment		0.890		0.836
AverageOrderValueExperiment			0.965	0.933

Eigenvalue 7.793
% of the variance explained 86.6

Exploratory Factor Analysis - Control

Factor Items	Factor Loadings by Component			Extracted Communalities
	1	2		
ConversionRateControl	0.951			0.961
NewVisitorConversionControl	0.827			0.774
AddToCartControl	0.738			0.741
CartAbandonmentControl	-0.878			0.773
BounceRateControl		-0.895		0.810
RevenuePerSessionControl	0.892			0.911
AveragePageViewsControl		0.818		0.731
AverageTimeOnSiteControl		0.821		0.748
AverageOrderValueControl				0.419

Eigenvalue 6.868
% of the variance explained 76.3

A stepwise regression analysis provided additional insights by indicating that new visitor conversion rate was the primary influencer of sales revenue in the experimental treatment, where it accounted for .706 out of the .871 r^2 value of the model (Table 5-8). One of the co-research members suggested that this could be a positive indicator that the navigational changes were helping first-time users of the website navigate the website better than they were previously able to. As a group, we generally agreed with this viewpoint but concluded that more supporting data was needed. As with the previous regression models, average order value again proved to be strongly correlated with sales revenue. Add to cart also appeared in this regression model, even though it only contributed 0.055 of the change in r^2 its inclusion was supported by the EFA models.

Table 5-8 : Stepwise Regression Summary – Fly-Out Menu v3 Experiment Treatment

Variable	Partial regression coefficient (B)	Standard error	Standardised regression coefficient (Beta)	t-statistic	VIF	F-statistic / Significance of F	R2 / Adjusted R2
Intercept	-2.707	0.508	0.000	-5.330		81.371	0.871
NewVisitorConversionExperiment	93.315	15.420	0.589	6.052	2.653	0.000	0.861
AverageOrderValueExperiment	0.019	0.003	0.420	6.553	1.153		
AddToCartExperiment	24.293	6.167	0.401	3.940	2.898		

Dependent Variable: RevenuePerSessionExperiment

5.1.2 Hypothesis 1: Results

The results of the analysis indicated that the experimental treatments generally reduced time costs and improved ease of navigation. However, these reductions in decision costs did not definitively result in changes in sales revenue, even though some improvements to store sales look like a potential outcome. As a consequence, the null hypothesis (consumer decision costs have no impact upon sales revenue) must be accepted.

Despite being unable to disprove the null hypothesis, there were sufficient positive indicators to suggest that some of the experimental changes were beneficial to customers. In particular, the reduction in time costs from the Clickable Top Nav and Fly-Out Menu v3 experiments may have longer-term customer perception benefits than was measurable in this research time frame. Also, improvements in bounce rate and the increasing importance of new visitor conversion suggests that the changes may be benefitting visitors with less experience with the site, in particular.

The change implemented with the Fly-Out Menu v1 experiment highlights the importance of consumer testing. The business was extremely confident about implementing the v1 change, yet the results were not generally positive, and this change may have even added some navigational confusion. Without the ability to conduct structured A/B testing through an action research process it is probable that this change would have been implemented

without an adequate check and balance. Therefore, a good point of learning emerged for the business, where an understanding of the need to test and understand changes before they are fully implemented provides incremental value to the business. I presented this point of reflection to both the co-research group and the ELT as part of the dialogue about this first series of experiments. My reflection on BGC's re-platforming effort, and the perspective I put forth, was that they did not test and understand those changes adequately. This did not create as much dialogue as I thought it might, but there seemed to be a tacit agreement that this was evidence that the prior methods were inadequate. It is possible that more robust support was tempered by the accountability of the participants to the prior methods. Despite the political undercurrents, this dialogue, and the evidence, seemed to provide sufficient organizational support to pursue the additional experiments that were proposed.

5.1.3 Hypothesis 2: Analysis

For hypothesis 2, the null hypothesis to be tested was:

$$H_{2_0}: \mu = K ; \text{Purchase justification has no impact upon sales revenue}$$

There is a viewpoint in the academic literature that highlighting the pricing and value of certain products can generate changes in customer purchase intent (Betancourt and Gautschi, 1990; Meeker, et al., 2009; Simonson, 1999). As a consequence, highlighting the value of products can lead to an increase in demand for the assortment (Betancourt and Gautschi, 1990). In order to test this hypothesis two different test treatments were implemented at different times: *Colorization Experiment*, and *Fly-Out Menu Experiment v.2*.

Colorization Experiment

The *Colorization* experiment involved changes to the color of the discounted price of a product and of the 'Add To Cart' button on the product detail page (PDP). An example of these colorization changes is shown in appendix 6. Prior to implementing this test, and in the control group, the color of discounted prices and the Add To Cart button was very similar to the surrounding colors. The idea behind the colorization changes was to highlight the value and purchasability of the product, increasing justification to purchase.

The Colorization experiment ran for approximately three weeks, from research week 4 to week 8. Using *Monetate's* A/B testing application, website traffic was randomly split 50/50 between the experiment and the control. In total 1,580,725 website sessions occurred during this period, with 791,772 experiencing the Colorization experimental treatment, and 788,953 experiencing the control treatment. The analysis involved the same clickstream metrics as was used in the testing for Hypothesis 1.

In order to test for differences between the treatment and control, paired sample *t*-tests were performed comparing each metric across the 20 days of the experiment. Any metric

comparisons that resulted in a two-tailed P value of 0.05 or greater were discarded from the analysis. The only exception being revenue, given that it was the main target of this testing effort. A summary of the paired samples t -test results is shown in table 5-9.

Table 5-9 : Colorization t -test Significant Results Summary

Variable	Sessions	Mean	% Change in Mean	N	Std. Deviation	Std. Error Mean	t	df	P
AddToCartExperiment	791772	.044	2.2%	20	.005	.001	2.532	19	.020
AddToCartControl	788953	.043		20	.005	.001			
RevenuePerSessionExperiment	791772	1.551	5.7%	20	.358	.080	2.285	19	.034
RevenuePerSessionControl	788953	1.467		20	.365	.082			
AverageOrderValueExperiment	791772	128.698	4.2%	20	9.126	2.041	2.244	19	.037
AverageOrderValueControl	788953	123.510		20	7.908	1.768			

Table 5-10 : Exploratory Factor Analysis – Colorization

Exploratory Factor Analysis - Experiment

Factor Items	Factor Loadings by Component			Extracted Communalities
	1	2	3	
ConversionRateExperiment	0.985			0.978
NewVisitorConversionExperiment	0.956			0.944
AddtoCartExperiment		0.795		0.846
CartAbandonmentExperiment	-0.869			0.916
BounceRateExperiment		-0.967		0.963
RevenuePerSessionExperiment	0.983			0.979
AveragePageViewsExperiment		0.950		0.912
AverageTimeOnSiteExperiment			0.861	0.743
AverageOrderValueExperiment			-0.560	0.512

Eigenvalue 7.793
% of the variance explained 86.6

Exploratory Factor Analysis - Control

Factor Items	Factor Loadings by Component			Extracted Communalities
	1	2	3	
ConversionRateControl	0.981			0.974
NewVisitorConversionControl	0.949			0.937
AddToCartControl	0.508	0.731		0.793
CartAbandonmentControl	-0.857			0.885
BounceRateControl		-0.970		0.949
RevenuePerSessionControl	0.991			0.993
AveragePageViewsControl		0.938		0.891
AverageTimeOnSiteControl			0.957	0.916
AverageOrderValueControl	0.634			0.606

Eigenvalue 6.833
% of the variance explained 75.9

According to the t -test results, revenue per session saw a statistically significant change in its mean value (to a $P < 0.05$). In addition, the mean value increased fairly significantly compared to the control, implying a revenue benefit to the business from this experimental treatment. In support of the increase in revenue, was a statistically significant increase in average order value and add to cart rate. The former of which has tended to be a strong

predictor of revenue impacts. The EFA analysis also highlights this same impact, where average order value and add to cart are not picked up in factor 1 in the experiment versus

Table 5-11 : Stepwise Regression Summary – Fly-Out Menu v3 Experiment

Variable	Partial regression coefficient (B)	Standard error	Standardised regression coefficient (Beta)	t-statistic	VIF	F-statistic / Significance	
						of F	R2 / Adjusted R2
<i>Intercept</i>	-1.656	0.122		-13.553		964.309	0.994
ConversionRateExperiment	128.360	2.932	0.858	43.772	1.119	0.000	0.993
AverageOrderValueExperiment	0.012	0.001	0.304	15.370	1.137		
AverageTimeOnSiteExperiment	0.005	0.002	0.054	2.867	1.017		

Dependent Variable: RevenuePerSessionExperiment

the control (Table 5-10).

Stepwise regression analysis continued to highlight the importance of conversion rate and average order value on sales revenue (Table 5-11). However, this was the first regression model to also highlight the importance of time on site. Interestingly, this relationship appears to be positive in nature, meaning more time on site will likely translate to higher sales revenue. In the literature, time is proposed as a customer cost and a detriment to purchase intent, however, in this case, that does not appear to be the case. Group discussion suggested that the primary reason for this was probably because the colorization changes occurred on the product landing (PLP) and product detail pages (PDP). The collective conclusion was that this meant that only customers that navigated from the homepage, onto the product pages, would see the colorization changes. Therefore, any behavioral influence was generated with customers that are likely to spend more time on the site anyway, rather than those that just review the homepage before leaving the site. In addition, some members of the co-research group expressed the opinion that the colorization may have caused customers to deliberate about their actions or to review more products than they did in the control group, albeit that page views did not increase in the experiment over the control.

Fly-Out Menu v2 Experiment

The *Fly-Out Menu v2* experiment involved the addition of ‘Sale’ categories to the fly-out menu structure and the colorization of ‘Sale’ category in the top nav header. An example of these changes is shown in appendix 9. Prior to implementing this test, Sale was not a product category that was represented in the main navigational sub-menu under each category header. As such, if you were interested in sale dresses, you had to navigate to the ‘Sale’ section first, but with the change, you could find sale products more easily within the dress category. In addition, the Sale category header was previously the same color as all of the other category headers, but in the experiment, the color was changed to the same color

that tested positively in the colorization experiment. The objective of these changes was to increase customer awareness of sale merchandise within categories that they were interested in and to highlight sale merchandise in general as a means of increasing purchase justification.

The Fly-Out Menu v2 experiment ran for approximately two weeks, from research week 18 to week 19. Using *Monetate's* A/B testing application, website traffic was randomly split 50/50 between the experiment and the control. In total 588,772 website sessions occurred during this period, with 295,289 experiencing the experimental treatment, and 293,483 experiencing the control treatment. The analysis involved the same clickstream metrics as was used in the testing for Hypothesis 1.

In order to test for differences between the treatment and control, paired sample *t*-tests were performed comparing each metric across the 11 days of the experiment. Any metric comparisons that resulted in a two-tailed *P* value of 0.05 or greater were discarded from the analysis. The only exception being revenue, given that it was the main target of this testing effort. A summary of the paired samples *t*-test results is shown in table 5-12.

Table 5-12 : Fly-Out Menu v2 *t*-test Significant Results Summary

Variable	Sessions	Mean	% Change in Mean	N	Std. Deviation	Std. Error Mean	t	df	P
RevenuePerSessionExperiment	295289	2.387	-4.8%	11	.386	.116	-1.363	10	.203
RevenuePerSessionControl	293483	2.507		11	.573	.173			
AppointmentsCreatedExperiment	295289	0.012	6.9%	11	.001	.000	2.676	10	.023
AppointmentsCreatedControl	293483	0.012		11	.001	.000			

The only statistically significant value change was generated in appointments created. This metric measures the number of appointments created on a per session basis and, as discussed previously, is a good proxy for store sales improvements. Despite the lack of a statistically supported change in revenue generated on the website, on a raw basis revenue declined 4.8%. This decline in sales, regardless of statistical significance, was sufficient for BGC to decide to terminate this test prior to significance being reached. There was a strong belief amongst the co-research team that the experimental treatment was causing customers to purchase more frequently from sale merchandise, and less frequently from the higher priced regular price merchandise. This view was supported by the Manager of Web Analytics who showed data supporting an increase in sale product unit sales in the experiment as compared to the control. This data was provided during a co-research team meeting by a supplementary analysis and is not included here, except anecdotally.

Table 5-13 : Stepwise Regression Summary – Fly-Out Menu v2 Experiment and

Variable	Partial regression coefficient (B)	Standard error	Standardised regression coefficient (Beta)	t-statistic	VIF	F-statistic / Significance of F		R ² / Adjusted R ²	
Intercept	-2.312	0.151	0.000	-15.346		531.974	0.000	0.993	0.991
ConversionRateExperiment	149.927	5.292	0.878	0.108	1.029				
AverageOrderValueExperiment	0.016	0.001	0.642	20.723	1.029				

Dependent Variable: RevenuePerSessionExperiment

Variable	Partial regression coefficient (B)	Standard error	Standardised regression coefficient (Beta)	t-statistic	VIF	F-statistic / Significance of F		R ² / Adjusted R ²	
Intercept	-2.489	0.188	0.000	-13.244		394.272	0.000	0.990	0.987
AverageOrderValueControl	0.017	0.001	0.784	22.087	1.003				
ConversionRateControl	154.768	9.610	0.571	16.105	1.003				

Dependent Variable: RevenuePerSessionControl

A stepwise regression analysis comparing the experiment to the control also suggested an impact on average order value (Table 5-13). In the control treatment, average order value contributed .664 of the .994 R² value. In contrast, average order value only contributes 0.401 of the .993 R² value in the experimental group. Given the shortened testing time frame for this experiment, it is not necessary to share additional analytical results. This decision does highlight the revenue sensitivity at BGC, given the rapid response to curb any activities that could contribute to sales reductions. This also highlights an important nuance of an action research study compared with a traditional mode-1 research effort, where business perspectives receive equal, or even greater, weight with research objectives.

5.1.4 Hypothesis 2: Results

The co-research group concluded that the results of the colorization analysis indicated that the experimental treatments generally improved sales by increasing purchase justification. There was agreement that the increase in purchase justification was manifested through increases in add to cart propensity, and average order values. The data captured in the fly-out menu v2 test did not offer a definitive contradiction to the results of the colorization analysis, but some points of caution emerged from our discussions about the results. Specifically, we learned that that not all purchase justification changes are likely to lead to positive revenue outcomes. This further reinforced BGC's understanding, and the value assessment, of this new testing approach to site changes. Despite not all changes leading to positive outcomes, we agreed that there was sufficient evidence that increased levels of purchase justification do impact upon purchasing behaviors. As a consequence, the null hypothesis (purchase justification has no impact upon sales revenue) is rejected.

5.1.5 Hypothesis 3: Analysis

For hypothesis 3, the null hypothesis to be tested was:

$$H_{3_0} : \mu = K ; \text{ Consumer decision costs have no impact upon customer satisfaction}$$

Reducing the consumer decision costs required to find a desirable product should lead to higher levels of customer satisfaction with the assortment (Hoch et al., 1999; Handelsman and Munson, 1985; Betancourt and Gautschi, 1990). Because customer satisfaction is influenced by website content and ease of use ((Peng and Kim, 2014; Lee and Kozar, 2012; Zhenhui, et al., 2010; Green and Pearson, 2009), a focus on a reduction in decision-making costs should result in higher satisfaction levels. The literature review findings formed the basis of analysis into the relationship between consumer costs and satisfaction.

In order to validate H3₀ four changes in site navigation and website content were implemented and displayed to all website visitors. The implemented changes were the direct result of group evaluations of the A/B test experiments conducted between research week 4 and week 24. To wit, the following A/B test experiments were implemented to all site visitors as follows: *Clickable Top Nav Experiment (week 11)*, *Product Description Experiment (week 11)*, and *Fly-Out Menu Experiment (week 24)*, *Size Display Experiment (week 24)*. In order to measure the impact on customer satisfaction, survey results were collected prior to the A/B experiments being conducted (1st Survey), after the first two experiments were implemented (2nd Survey), and after all four experiments were implemented (3rd Survey). The timing of the surveys compared to the experiments is illustrated in Table 5-14.

Table 5-14 : Survey Data Collection and Site Change Implementation by Research Week

	Research Activity Week																														
Activity	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31
1st Survey																															
Clickable Top Nav Experiment				X	X	X	X	X	X	X	X																				
Colorization Experiment				X	X	X	X	X																							
Video Experiment					X	X	X	X																							
Product Description Experiment				X	X	X	X	X	X	X	X																				
2nd Survey																															
Fly-Out Menu Experiment														X ₁	X ₁	X ₁	X ₂	X ₂	X ₂	X ₃	X ₃	X ₃	X ₃	X ₃	X ₃						
Size Display Experiment																							X	X	X						
3rd Survey																															
Post-Visit Survey																															
Combined On-Site Survey Data																															

Key: Start Stop Implement Collect Data

In order to understand if there was a change in customer sentiment due to the implementation of the website changes, site satisfaction scores were compared between each of the three surveys. Each survey collected a total site satisfaction score and qualitative feedback. In addition, the third survey collected customer typology data and

additional types of satisfaction scores. An overview of the three different on-site surveys is provided in table 5-15. In total 1,162 total survey response were collected across the three on-site surveys. Customers provided overall site satisfaction ratings in 1,055 of those surveys, and also provided open-ended feedback in 963 of the surveys. Survey's without ratings were discarded from the analysis. Any surveys where the closed-ended questions were completed but no open-ended feedback was provided were coded as 'no comment' surveys. Based on this approach, 1,055 surveys were used in this analysis.

Table 5-15 : On-Site Survey Overview

On-Site Survey Type	Research Timing	Type of Data Collected	Survey Responses	Surveys with Ratings	Complete Surveys
1st Survey	Immediately prior to conducting any website experiments	Overall site satisfaction, open-ended feedback	416	348	340
2nd Survey	During website experimentation	Overall site satisfaction, open-ended feedback	428	410	408
3rd Survey	Immediately following website experimentation and full change implementation	Overall and sub-category site satisfaction, customer typology, open-ended feedback	318	297	215

The Methodology chapter of this thesis paper provided a detailed discussion of how the open-ended feedback for each survey was coded and themed. In summary, these codes and themes were initially constructed based on the literature review findings, but some additional codes and themes emerged from the feedback data. The codes and themes were maintained across the analysis of each separate survey so that analytical comparisons could be made between them. Once each code and theme was validated, they were matched to the overall site satisfaction rating that each customer provided related to the specific coded and themed feedback. This methodology allowed a comparison of the mean scores between surveys, and by each code and theme. Because the aggregated data was ordinal and non-parametric in nature, chi-square analysis was utilized as the primary method for calculating mean difference probabilities. In order to challenge the null hypothesis a chi-square (X^2) *P*-value of <0.05 was required.

Exploratory factor analysis of the coded data was also conducted. The goal was to identify any latent variables that existed across the qualitative feedback. The EFA was conducted using the coded data from all of the surveys combined. The objective was to identify additional variables that would be present in each survey and could, therefore, be measured for change impacts. Two factors emerged from this analysis, as shown in Table 5-16. Group reflection determined that the codes associated with component 1 are highly consistent with those measured in Parasuraman, et al.'s (2005) e-SERVQUAL scale, hence the decision

to term this latent variable “Site Quality”. Through co-research group discussion it was surmised that the component 2 factored codes generally related to a customer’s ability to accomplish their desired tasks on the site. On this basis, the “Mission Support” latent variable was constructed.

Table 5-16 : Exploratory Factor Analysis – Qualitative Feedback Codes

Theme	Code	Factor Loadings		Extracted Communalities
		1	2	
Assortment Quality	Availability of Desirable Product	.519		.295
	Product Quality	.930		.865
Customer Centricity	Task Support		-.851	.749
	Service Quality	.910		.996
Decision Costs	Product Information	.753	.626	.958
	Navigation	.545	.721	.817
Product Delivery	Fulfillment	.992		.996
	Desired Purchase Timeline	.715	.645	.928
Purchase Justification	Communication of Prices	.669	.684	.916
	Product Value		-.970	.969
Security	Security & Privacy	.979		.964
Technology Support	Software Bugs	.850		.962
	Site Performance	.655	.659	.863

In order to test H3₀, chi-square analysis was performed on the qualitative coding pertaining to navigation and product information, both contributors to consumer decision-making costs. These codes were aggregated into a ‘Decision Costs’ theme, as shown in Table 5-16, and chi-square analysis was also conducted at the theme level. In addition, both latent variables and overall site satisfaction ratings were tested across each of the on-site surveys. In this analysis the 2nd and 3rd on-site surveys are compared to the 1st on-site survey in order to determine if any satisfaction changes occurred as a potential result of the navigational and product information changes that were implemented on the website. A summary of the 1st Survey responses is shown in Table 5-17. A full summary of all of the survey responses by survey type is provided in appendix 12.

Table 5-17 : 1st Survey Summary – Decision Cost Measures

Theme	Codes	1st Survey - Response Scale *					Scale Descriptives		
		Strongly Negative		Neutral Opinion		Strongly Positive	n	Mean	SD
		1	2	3	4	5			
Decision Costs		31.03	18.97	10.34	24.14	15.52	58	2.74	1.50
	Navigation	27.50	17.50	10.00	30.00	15.00	40	2.88	1.48
	Product Information	38.89	22.22	11.11	11.11	16.67	18	2.44	1.51
Component 1: Site Quality		40.73	21.19	14.90	12.25	10.93	302	2.31	1.40
	Component 2: Mission Support	32.28	21.26	12.60	18.11	15.75	127	2.64	1.49
Overall Site Satisfaction		37.93	20.11	14.66	13.79	13.51	348	2.45	1.44

* Frequency of Responses

P is the chi-square test probability of association

Table 5-18 shows the responses from the 1st and 2nd surveys and the X^2 *p*-value for each variable. From a review of the X^2 probabilities, satisfaction ratings for decision costs are likely to be different between the two surveys ($P = 0.045$), caused by a change in satisfaction regarding navigation ($P = 0.004$). In fact, navigational satisfaction has declined from a mean rating of 2.88 to a mean of 2.48. The only other significant change in satisfaction appears to come from ‘Mission Support,’ where the mean rating increased to 3.01 versus 2.64 in the first survey ($P = 0.029$).

Table 5-18 : 1st Survey and 2nd Survey X^2 Analysis – Decision Costs

Theme	Codes	1st Survey - Response Scale *					Scale Descriptives			2nd Survey - Response Scale *					Scale Descriptives			X^2 Probability
		Strongly Negative		Neutral Opinion		Strongly Positive	n	Mean	SD	Strongly Negative		Neutral Opinion		Strongly Positive	n	Mean	SD	<i>P</i> (1v2)
		1	2	3	4	5				1	2	3	4	5				
Decision Costs		31.03	18.97	10.34	24.14	15.52	58	2.74	1.50	27.12	25.42	22.03	11.86	13.56	59	2.59	1.36	0.045
	Navigation	27.50	17.50	10.00	30.00	15.00	40	2.88	1.48	29.03	29.03	19.35	9.68	12.90	31	2.48	1.35	0.004
	Product Information	38.89	22.22	11.11	11.11	16.67	18	2.44	1.51	25.00	21.43	25.00	14.29	14.29	28	2.71	1.37	0.057
Component 1: Site Quality		40.73	21.19	14.90	12.25	10.93	302	2.31	1.40	40.00	20.85	23.66	8.45	7.04	355	2.22	1.26	0.102
	Component 2: Mission Support	32.28	21.26	12.60	18.11	15.75	127	2.64	1.49	23.57	17.14	18.57	16.43	24.29	140	3.01	1.51	0.029
Overall Site Satisfaction		37.93	20.11	14.66	13.79	13.51	348	2.45	1.44	36.34	19.27	22.20	10.00	12.20	410	2.42	1.38	0.272

* Frequency of Responses

P is the chi-square test probability of association

Based on the comparison between these two surveys, there was no significant difference in overall satisfaction, but satisfaction with navigation appears to have declined. In addition, some improvements seem to have been made in perceptions of mission support. Group discourse led to the conclusion that the Clickable Top Nav and Product Information experiments offered no real improvements in decision-making costs, but that they may have improved the ability to complete the desired shopping mission. Given that the Clickable Top Nav experiment was only a minor navigational change, it was suggested that other factors

might be influencing the reduction in navigational satisfaction during the 2nd Survey time period. During an ELT group discussion it was suggested that customers could be responding negatively to the Fly-Out Menu experiments that were running concurrently with the 2nd Survey.

Moving on to consider the 3rd on-site survey, there is no change in overall site satisfaction over the period between 1st Survey and the 3rd Survey (Table 5-19). However, consistent with the 2nd Survey, perceptions of navigation appear to have declined ($P = 0.018$). In addition, product information perceptions have also declined ($P < 0.001$). In line with these changes, perceptions of site quality ($P = 0.010$) and mission support ($P = 0.002$) also appear to have declined. Perceptions of product information ($P < 0.001$), site quality ($P = 0.030$) and mission support ($P < 0.001$) all also declined relative to perceptions during the 2nd Survey time period.

Table 5-19 : 1st, 2nd and 3rd Survey X² Analysis – Decision Costs

Theme	Codes	1st Survey - Response Scale *					Scale Descriptives			3rd Survey - Response Scale *					Scale Descriptives			X ² Probability	
		Strongly Negative		Neutral Opinion		Strongly Positive	n	Mean	SD	Strongly Negative		Neutral Opinion		Strongly Positive	n	Mean	SD	P (1v3)	P (2v3)
		1	2	3	4	5				1	2	3	4	5					
Theme Decision Costs		31.03	18.97	10.34	24.14	15.52	58	2.74	1.50	30.23	32.56	16.28	16.28	4.65	43	2.33	1.20	0.140	0.134
	Navigation	27.50	17.50	10.00	30.00	15.00	40	2.88	1.48	24.00	32.00	20.00	16.00	8.00	25	2.52	1.24	0.018	0.753
	Product Information	38.89	22.22	11.11	11.11	16.67	18	2.44	1.51	38.89	33.33	11.11	16.67	0.00	18	2.06	1.08	0.000	0.000
	Component 1: Site Quality	40.73	21.19	14.90	12.25	10.93	302	2.31	1.40	33.84	32.32	19.19	11.11	3.54	198	2.18	1.13	0.010	0.030
	Component 2: Mission Support	32.28	21.26	12.60	18.11	15.75	127	2.64	1.49	25.37	35.82	16.42	14.93	7.46	67	2.43	1.23	0.002	0.000
	Overall Site Satisfaction	37.93	20.11	14.66	13.79	13.51	348	2.45	1.44	30.64	27.27	17.85	11.78	12.46	297	2.48	1.37	0.286	0.249

* Frequency of Responses
P is the chi-square test probability of association

Given low sample sizes for the navigation and product information codes, it is possible that the X² rejection of the null hypothesis is due to type I error. In support of this point, when considering the aggregation of the two decision cost codes, into the decision cost theme, the probability of a change in perceptions diminishes and does not meet the 0.05 threshold ($P = 0.140$). As a consequence, it is possible that the decline in the Navigation and Product Information mean perception scores, since the 1st survey, is as a result of random sampling rather than a real reduction in perceptions.

Qualitative Insights

Decision cost feedback from customers prior to the beginning of site experimentation often related to the lack of clarity provided by the product images, such as “some pictures are not very clear” and “I’d like to see different sized girls not a zero in every photo.” From a navigational perspective, there were many comments relating to site ease of use, such as “website is so confusing”, and “search gave 2186 of everything”. Likely because there were no experimental changes to the product image quality, commentary regarding these images remained consistent across the research timeframe. However, feedback increased significantly with regarded to products being shown on the website that were unavailable

for purchase. Even after making changes to the site, navigational perspectives continued to support the earlier viewpoints and difficulty using the site. Much of the navigational feedback related to aspects of the site that did not involve experimentation, such as “allow for multiple color selections” and “98% of the time your filters are pointless.”

What surfaced from co-generated reflection on the qualitative insights is that the experimentation did not impact upon all potential decision-cost driving facets of the website. As a result, even if some progress was made, consumers continued to provide feedback on areas of the site where they wanted to see improvements or were experiencing frustration. Therefore, the qualitative insights provided a potential treasure trove of valuable information for BGC. These insights also indicated a lack of meaningful progress from the changes that had been tested and implemented.

5.1.6 Hypothesis 3: Results

While it does appear that satisfaction with site quality and mission support declined over the research time period, in light of the qualitative feedback, these changes cannot be directly attributed to decision costs. There was also no statistically supported change in specific perceptions related to overall decision costs. Therefore, the null hypothesis (*consumer decision costs have no impact upon customer satisfaction*) must be accepted.

The insights generated from the qualitative data provided insight into a host of required future actions. Interestingly, the co-research group seemed overwhelmed by the amount of feedback generated through the survey, and tried to discount it, while the ELT seemed to more broadly acknowledge its importance. My reflection on this was that the feedback highlighted further performance shorting-comings rather than proving that the co-research team was making the progress demanded of them. Regardless of the ambivalent response of the co-research group, these findings seemed to fuel the ELT’s desire to gain additional customer insights, and they frequently asked for them during weekly discussions. Another definitive outcome was the creation of a Director of Consumer Insights position, aimed generating similar insights from the brick-and-mortar store locations and marrying those with the views collected online. This new role reported directly to the CMO.

5.1.7 Hypothesis 4: Analysis

For hypothesis 4, the null hypothesis to be tested was:

$$H_{4_0}: \mu = K ; \text{Purchase justification has no impact upon customer satisfaction}$$

The literature states that increasing purchase justification will increase purchase satisfaction (Sela, et al., 2009; Böhm and Pfister, 1996). Further, perceptions of product value are highly correlated with consumer satisfaction (Deveraj, et al., 2002). Therefore, it should be

possible to increase BGC’s customer satisfaction by improving the ability to justify purchases. A focus on perceptions of the value of products should be a viable means to test this perspective, per Simonson (1999) and Sela, et al. (2009).

In order to test H4, two different test treatments were implemented at different times: *Colorization Experiment*, and *Fly-Out Menu Experiment v.2*. Based on the results of these experimental treatments, only the *Colorization* experiment was implemented on the website for all visitors to experience (implemented in research week 8). In order to measure the impact on customer satisfaction, survey results were collected prior to the A/B experiments being conducted (1st Survey), after the first two experiments were implemented (2nd Survey), and after all four experiments were implemented (3rd Survey). The timing of the surveys compared to the experiments is illustrated earlier in this chapter, in Table 5-14.

From the qualitative feedback, only eleven responses and twenty-five responses were generated by the 1st and 2nd surveys, respectively. The 3rd Survey only generated three responses pertaining to purchase justification. In part, the low amount of feedback may be because product value has the highest overall satisfaction rating across all of the survey data at a mean of 4.00 across twenty-three pieces of total feedback. Product value and communication of prices are the two codes that form the purchase justification theme. The closed-ended pricing questions also scored relatively highly at a 3.25 mean satisfaction rating across two hundred and forty-six responses.

Table 5-20 shows the qualitative feedback response measures for the purchase justification theme and the communication of prices and product value codes. Given sparse data at the code level, it is suggested that the analysis of the responses be conducted only at the theme level. From a comparison of the two surveys, it can be seen that the mean rating score increased from 2.91 to 3.12 with a high degree of probability ($P < 0.001$). On this basis, it does appear that the colorization experiment led to an improved level of customer satisfaction.

Table 5-20 : 1st and 2nd Survey X² Analysis – Purchase Justification

Theme	Codes	1st Survey - Response Scale *					Scale Descriptives			2nd Survey - Response Scale *					Scale Descriptives			X ² Probability	
		Strongly Negative		Neutral Opinion		Strongly Positive	n	Mean	SD	Strongly Negative		Neutral Opinion		Strongly Positive	n	Mean	SD		P (1v2)
		1	2	3	4	5				1	2	3	4	5					
Purchase Justification		27.27	27.27	0.00	18.18	27.27	11	2.91	1.63	28.00	4.00	20.00	24.00	24.00	25	3.12	1.54	0.000	
	Communication of Prices	42.86	42.86	0.00	14.29	0.00	7	1.86	1.00	44.44	11.11	22.22	22.22	0.00	9	2.22	1.23	0.000	
	Product Value	0.00	0.00	0.00	25.00	75.00	4	4.75	0.44	18.75	0.00	18.75	25.00	37.50	16	3.63	1.46	0.000	

Qualitative Insights

Purchase justification feedback from customers prior to the beginning of site experimentation most often focused on the great product value on the site and issues with promotional redemptions. Examples of this feedback are: “*can’t use promo code*”, “*did not get the twenty dollars of the dress*”, “*great pricing*”, “*reasonable prices*”. Immediately following the implementation of the Colorization experiment to all website visitors the commentary became even more polarized. Many customers continued to report high levels of satisfaction with the value of products, but there was also an increase in the volume of commentary regarding product appearing to be overpriced. For example, “*very overpriced*” and “*overpriced and final sale???* *You are ridicules!!!*” (sic).

Group reflection on the qualitative feedback led to the conclusion that the changes to accentuate product pricing had some impact on the emotional reaction of customers to the assortment as a whole. We debated how customers seemed to have become more price-sensitive and concerned about maximizing their product utility through a focus on value. I supported this dialogue with similar findings in the literature (e.g. Diehl, et al., 2003). Therefore, we concluded that increasing purchase justification through product value accentuation fundamentally changed the perception of the assortment and therefore altered satisfaction scales and measures.

5.1.8 Hypothesis 4: Results

The quantitative analysis of the purchase justification data did appear to show an increase in satisfaction following the colorization test implementation. Some members of the co-research group concluded that this was proof that the null hypothesis can be rejected. However, myself and other members of the group believed the qualitative feedback did not support any perceptual improvements. In addition, given the small sample sizes, especially in the 1st Survey, any conclusion to reject the null hypothesis would be highly questionable. Consequently, despite a lack of group consensus, a conclusion was reached that the null hypothesis (purchase justification has no impact upon customer satisfaction) cannot be rejected, despite the X^2 analysis results.

On a practical basis, it appears likely that the colorization experiment did not harm satisfaction regarding purchase justification. So, even without rigorous scientific support, BGC may feel that this change was valid to implement as a potential source of improved customer perceptions. Yet, the qualitative feedback also suggested to several group members that these changes may have triggered an unexpected emotional response and fundamentally altered assortment perceptions. As a consequence, I recommended to the ELT that future research is conducted to better understand those perceptions and to develop change actions that clearly do not manifest in any potentially negative price perceptions. This caused discussion about the best ways to respond to this recommendation. The resulting actions are covered later in this this paper.

As I reflected on the results of discussions, and my role in the process, I came to believe that a lack of consensus amongst the co-research group on the hypothesis 4 result implications was an important outcome for the broader research effort. Greenwood and Levin (2007) state that ‘hardline’ consensus is not a requirement of action research, far more important is the organizational learning generated by the debate. As such, this point in the research effort seemed to signal a new level of organizational awareness, the ability to debate the research data and to disagree with my views. I viewed this lack of consensus as a sign that the business was heading toward a positive knowledge-building outcome from these experiments, regardless of whether the experiments proved fruitful or not.

5.1.9 Hypothesis 5: Analysis

For hypothesis 5, the null hypothesis to be tested was:

$$H5_0 : \mu = K ; \text{ During visit and post-visit customer satisfaction does not differ}$$

Mogilner, et al. (2014) found that the temporal outlook of a customer affects their satisfaction with a particular task. For example, a consumer shopping for an item that is associated with a future need will have a higher level of satisfaction than a customer shopping for an immediate need (Mogilner, et al., 2014). Following this logic, it was hypothesized that this phenomenon should also work in reverse, where satisfaction levels may be higher after a shopping event than during the shopping event.

Table 5-21 : During and Post-Visit Survey Overview

On-Site Survey Type	Research Timing	Type of Data Collected	Survey Responses	Surveys with Ratings	Surveys with Qual Feedback
3rd Survey	Immediately following website experimentation and full change implementation	Overall and sub-category site satisfaction, customer typology, open-ended feedback	318	297	215
Post-Visit Survey	Immediately following website experimentation and full change implementation	Overall and sub-category site satisfaction, customer typology, open-ended feedback	225	206	58

In order to test H5₀ two different surveys were distributed to visitors of the website. One survey was made available while customers shopped the website, the other survey was distributed to customers immediately upon exiting from the website (see Table 5-21). The on-site survey was available to any website visitor who clicked on the ‘Feedback’ link on any page on the website. The Post-Visit Survey was randomly distributed to two-percent of website visitors who also browsed three pages or more on the website.

To determine if and where satisfaction may differ between on-site and off-site surveys, each code was analyzed along with the total site satisfaction score provided on each completed survey. Table 5-22 provides a summary of the response frequencies across the rating scale of each survey, and by qualitative code. From this table, you will notice that there is a

separate line to track the satisfaction ratings provided by those that chose not to provide qualitative feedback (no comment). This code was captured so that their overall site perspectives could still be represented through this analysis. Pearson chi-square analysis was used to analyze the mean difference probabilities.

Table 5-22 : 3rd Survey and Post-Visit Survey X² Code Analysis – Time to Experience

CODES	3rd Survey - Response Scale					Scale Descriptives			Post-Visit Survey - Response Scale					Scale Descriptives			X ² Probability
	Strongly Negative		Neutral Opinion		Strongly Positive	n	Mean	SD	Strongly Negative		Neutral Opinion		Strongly Positive	n	Mean	SD	
	1	2	3	4	5				1	2	3	4	5				
No Comment	24.39	13.41	14.63	14.63	32.93	82	3.18	1.60	3.38	2.03	15.54	46.62	32.43	148	4.03	0.93	0.000
Security & Privacy	100.00	0.00	0.00	0.00	0.00	1	1.00	0.00						0	N/A	0.00	N/A
Fulfillment	0.00	0.00	100.00	0.00	0.00	1	3.00	0.00	0.00	0.00	50.00	0.00	50.00	2	4.00	1.01	N/A
Software Bugs	36.84	34.21	21.05	7.89	0.00	38	2.00	0.95	20.00	40.00	20.00	20.00	0.00	5	2.40	1.03	N/A
Site Performance	0.00	75.00	0.00	25.00	0.00	4	2.50	0.87	50.00	0.00	0.00	50.00	0.00	2	2.50	1.51	N/A
Task Support	25.00	37.50	18.75	6.25	12.50	16	2.44	1.28	33.33	22.22	0.00	0.00	44.44	9	3.00	1.84	0.000
Product Information	38.89	33.33	11.11	16.67	0.00	18	2.06	1.08	0.00	16.67	16.67	50.00	16.67	6	3.67	0.95	0.000
Service Quality	54.84	22.58	12.90	9.68	0.00	31	1.77	1.01	33.33	50.00	16.67	0.00	0.00	6	1.83	0.69	0.000
Desired Purchase Timeline	0.00	100.00	0.00	0.00	0.00	1	2.00	0.00						0	N/A	0.00	N/A
Availability of Desirable Product	28.00	34.67	22.67	8.00	6.67	75	2.31	1.16	14.29	4.76	19.05	28.57	33.33	21	3.62	1.37	0.000
Product Quality	50.00	0.00	0.00	50.00	0.00	2	2.50	1.51	0.00	0.00	100.00	0.00	0.00	1	3.00	0.00	N/A
Communication of Prices	0.00	0.00	50.00	50.00	0.00	2	3.50	0.50						0	N/A	0.00	N/A
Product Value	0.00	0.00	0.00	0.00	100.00	1	5.00	0.00	0.00	0.00	0.00	0.00	100.00	2	5.00	0.00	N/A
Navigation	24.00	32.00	20.00	16.00	8.00	25	2.52	1.24	0.00	25.00	25.00	25.00	25.00	4	3.50	1.12	0.000
TOTAL	30.64	27.27	17.85	11.78	12.46	297	2.48	1.37	7.28	6.31	16.02	39.32	31.07	206	3.81	1.16	0.000
w/o technology & fulfillment	30.31	25.59	17.32	12.20	14.57	254	2.55	1.41	6.60	5.58	15.74	40.10	31.98	197	3.85	1.13	0.000
w/o tech, fulfill & no comment	33.14	31.40	18.60	11.05	5.81	172	2.25	1.20	16.33	16.33	16.33	20.41	30.61	49	3.33	1.47	0.000
Component 1: Site Quality	33.84	32.32	19.19	11.11	3.54	198	2.18	1.13	14.89	17.02	21.28	25.53	21.28	47	3.21	1.36	0.000
Component 2: Mission Support	25.37	35.82	16.42	14.93	7.46	67	2.43	1.23	17.39	17.39	8.70	21.74	34.78	23	3.39	1.53	0.000

From a review of the coded survey data it can be concluded that site satisfaction perceptions are clearly different across these two surveys. Every X² probability, where sufficient data exists to conduct a chi-square test, has a p-value of < 0.001. On top of this, the mean rating of every code, including the two latent variables (site quality and mission support) is the same or higher in the Post-Visit Survey than it is in the 3rd on-site survey, which was distributed concurrently. These results also hold true when comparing and analyzing the qualitative theme data (see Table 5-23). Table 5-23 also shows the overall site satisfaction rating, excluding the ‘no comment’ code data. While exclusion of the ‘no comment’ data significantly reduces the overall sample size, it can be seen that this has no impact on the results. Again, the mean satisfaction score is shown to be different between the two surveys, supported by a p-value of <0.001. This total mean value is also shown to be consistent with each theme, where the value is higher post-visit than it is during the shopping visit.

Table 5-23 : 3rd Survey and Post-Visit Survey X² Theme Analysis – Time to Experience

THEMES	3rd Survey - Response Scale					Scale Descriptives			Post-Visit Survey - Response Scale					Scale Descriptives			X ² Probability
	Strongly Negative		Neutral Opinion		Strongly Positive	n	Mean	SD	Strongly Negative		Neutral Opinion		Strongly Positive	n	Mean	SD	
	1	2	3	4	5				1	2	3	4	5				
Decision Costs	30.23	32.56	16.28	16.28	4.65	43	2.33	1.20	0.00	20.00	20.00	40.00	20.00	10	3.60	1.03	0.000
Purchase Justification	0.00	0.00	33.33	33.33	33.33	3	4.00	0.82	0.00	0.00	0.00	0.00	100.00	2	5.00	0.00	N/A
Technology Support	33.33	38.10	19.05	9.52	0.00	42	2.05	0.96	28.57	28.57	14.29	28.57	0.00	7	2.43	1.18	0.009
Customer Centricity	44.68	27.66	14.89	8.51	4.26	47	2.00	1.15	33.33	33.33	6.67	0.00	26.67	15	2.53	1.59	0.000
Assortment Quality	28.57	33.77	22.08	9.09	6.49	77	2.31	1.17	13.64	4.55	22.73	27.27	31.82	22	3.59	1.34	0.000
Product Delivery	0.00	50.00	50.00	0.00	0.00	2	2.50	0.50	0.00	0.00	50.00	0.00	50.00	2	4.00	1.01	N/A
Security	100.00	0.00	0.00	0.00	0.00	1	1.00	0.00						0	N/A	N/A	N/A
TOTAL w/o No Comment	33.02	32.56	19.07	10.70	4.65	215	2.21	1.34	17.24	17.24	17.24	20.69	27.59	58	3.24	1.46	0.000

Qualitative Insights

Comparing qualitative feedback between the on-site and post-visit surveys provides insight into the quantitative results. Commentary provided in the on-site survey is far more direct and blunt, potentially reflecting a customer in the midst of trying to complete a task. For example, “can not pin images from the site,” “It’s also not that user-friendly in terms of how the favorites work and the overall navigation of the site,” and “social media site rejected the post as spam.” In contrast, the post-visit survey comments were altogether nicer “It would be nice to open up a new page from the picture,” “it would be nice to have all the information in one location instead of having to move from location to location!” Not all of the post-visit feedback was more softly delivered, but the co-research group was collectively struck by how different the message delivery was between the two survey methods.

5.1.10 Hypothesis 5: Results

Research into H5₀ found strong support for the theories of Mogilner, et al. (2014). This research may have even expanded Mogilner, et al’s (2014) theory by showing that shopping event recency is likely to have a significant impact on customer satisfaction. This is an important finding for any scholar or practitioner who is forming conclusions based on consumer satisfaction data. To wit, it appears to be important to collect satisfaction data with the same level of event recency in order for comparisons to be drawn across the data. These results also suggest that customer perceptions may soften, and even improve, as time passes. Based on these results, the null hypothesis (during visit and post-visit customer satisfaction does not differ) was rejected.

5.1.11 Hypothesis 6: Analysis

For hypothesis 6, the null hypothesis to be tested was:

$$H6_0 : \mu = K ; \text{Customer satisfaction is not related to sales revenue}$$

The literature posits that consumer satisfaction is the primary antecedent and influencer of consumer purchase behavior (Udo, Bagchi and Kirs, 2010; Kuo, et al., 2009; Rajic and Dado, 2013). In turn, higher levels of customer satisfaction will be associated with higher levels of sales revenue, all other things being equal (Naik and Srinivasan, 2015; Olsen, et al., 2014; Wicks and Roethlein, 2009; Udo, et al., 2010). In order to test this hypothesis data needed to be collected that measured site satisfaction and sales revenue continuously across the research time period.

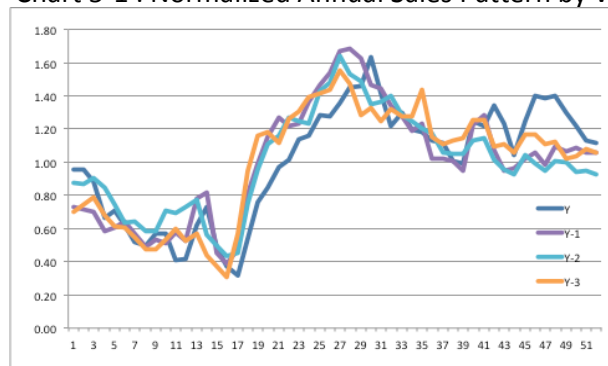
Sales revenue data by week was collected using the IBM *Coremetrics* application. In order to ensure that the sales revenue data was not influenced by seasonal demand variations, four

complete years of weekly sales revenue data was collected. This data included the sales revenue during the research year, and for the three prior years. The prior year data was used to perform a seasonal decomposition of the research year data. Seasonal decomposition is a common requirement when analyzing retail data, especially when comparing data across many weeks and months (e.g. Van Heerde and Bijmolt, 2005; Leeflang, et al., 2008). This is illustrated in chart 5-1, which shows normalized annual weekly sales demand by week. The data was normalized by dividing each week of sales data into each year's average weekly sales figure, allowing for a relative comparison of the changes in the weekly demand pattern.

For this analysis the Holt-Winters method of seasonal decomposition was used, as this is one of the most widely used and effective methods for accomplishing this task (Gardner, 1985; Gardner, 2006). The Holt-Winters method works by constructing a sales pattern from historical sales data, and then using it to decompose the data to extract a trend. For this analysis, the seasonal decomposition was accomplished using the prior three years of sales data in order to construct a

seasonal demand pattern. This pattern compared to the prior three years of data is shown in chart 5-2. The extracted seasonal demand pattern was then used to deseasonalize the research year sales data (e.g. Gardner, 1985). Once the data was successfully seasonally decomposed it was ready for comparison to the site satisfaction data.

Chart 5-1 : Normalized Annual Sales Pattern by Week



In order to generate a weekly site satisfaction score for comparison to the sales data, all on-site surveys with an overall site satisfaction rating were ordered by day and aggregated to weekly response counts and mean rating values. This data included data collected in weeks not previously reported in this research. Specifically, it included the previously omitted on-site survey responses from research weeks four to eleven and twenty-three to twenty-four. By including data across all weeks where overall site satisfaction and revenue data was generated it became possible to study the interrelationship of these two variables. The aggregated data is shown in Table 5-24.

Chart 5-2 : Seasonal Sales Pattern

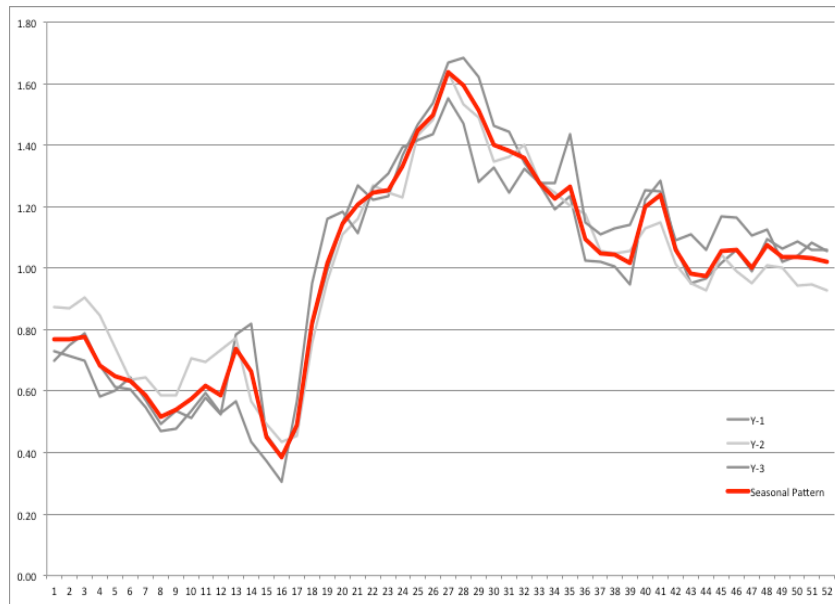


Table 5-24 : Combined On-Site Survey and Sales Revenue

Research Week	Combined On-Site Survey Descriptives						Scale Adjusted Sales (\$)	
	Weekly Response Count	Weekly Mean Satisfaction Value	SD	5-Week Rolling Average	Var to Rolling Average	Var to Total Mean SiteSat Value	Deseasonalized	Unadjusted
1	141	2.68	0.30			0.04	10.11	15.16
2	128	2.84	0.18			0.20	9.86	16.14
3	147	2.65	0.36			0.01	10.80	17.24
4	151	2.52	0.35			-0.12	11.42	17.30
5	105	2.69	0.65	2.67	0.01	0.04	13.84	19.39
6	87	2.61	0.74	2.66	-0.05	-0.04	12.18	16.81
7	35	2.46	0.80	2.60	-0.14	-0.19	10.66	14.46
8	30	2.70	0.28	2.59	0.11	0.05	12.10	15.45
9	23	2.39	0.95	2.61	-0.22	-0.25	11.63	14.25
10	21	2.43	1.04	2.55	-0.12	-0.22	11.13	14.09
11	26	2.85	0.36	2.57	0.28	0.20	12.30	13.46
12	26	2.81	1.08	2.65	0.16	0.16	12.66	13.28
13	30	2.40	0.64	2.58	-0.18	-0.25	11.52	12.01
14	25	2.68	0.98	2.63	0.05	0.03	11.58	11.80
15	17	3.06	0.80	2.73	0.33	0.41	12.37	14.84
16	44	1.95	0.61	2.46	-0.51	-0.69	11.73	14.51
17	21	2.19	1.10	2.36	-0.17	-0.45	15.07	15.93
18	24	3.00	1.42	2.47	0.53	0.35	14.92	14.65
19	21	2.76	1.35	2.47	0.29	0.12	12.77	12.43
20	27	2.74	0.88	2.45	0.29	0.10	14.04	14.81
21	31	3.00	0.83	2.77	0.23	0.35	15.66	16.61
22	162	2.51	0.50	2.65	-0.15	-0.14	16.46	16.47
23	181	2.65	0.36	2.63	0.02	0.01	15.50	16.66
24	203	2.81	0.26	2.69	0.12	0.16	14.94	15.47
25	88	2.93	0.50	2.72	0.21	0.28	13.99	14.47
26	71	2.74	0.57	2.71	0.03	0.09	13.04	13.45
27	32	2.45	0.42	2.75	-0.30	-0.19	12.99	13.24
28	39	2.26	0.58	2.74	-0.49	-0.39	17.90	13.72
29	30	2.47	0.36	2.66	-0.19	-0.17	17.90	13.72
30	29	2.37	0.92	2.51	-0.14	-0.28	17.98	13.98
31	30	2.57	0.54	2.42	0.15	-0.07	19.90	13.56
	2024	2.65	0.74			R² with Site Sat :	0.010	0.013

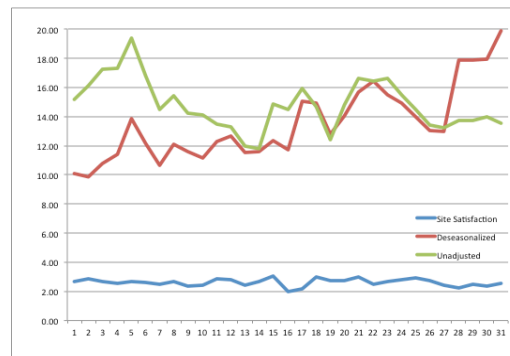
Several types of analysis were conducted on the data shown in Table 5-24. Regression analysis yielded low R² values, indicating little relationship between overall site satisfaction

and either deseasonalized sales ($R^2 = 0.010$) or actual sales revenue ($R^2 = 0.013$). To determine whether there was any lagged affect between sales revenue and site satisfaction, weekly mean satisfaction was cross-correlated with deseasonalized sales and actual sales. The results are shown in Table 5-25. This was to test whether overall site satisfaction ratings would translate to sales at a later timeframe, even if there were no immediate relationship to sales revenue. From the data shown in Table 5-25, it can be seen that there are no meaningful correlations between site satisfaction and either of the sales revenue variables at lags of up to five weeks. Cross-correlation was tested up to a difference of ten weeks, and still offered no signs of a statistically significant relationship.

Table 5-25 : Cross-Correlation Summary

Variable 1	Weekly Mean Satisfaction Value			
Variable 2	Deseasonalized Sales		Actual Sales	
Series Lag	Cross Correlation	Std. Error	Cross Correlation	Std. Error
-5	-0.156	0.196	-0.198	0.196
-4	-0.007	0.192	0.037	0.192
-3	0.225	0.189	0.385	0.189
-2	0.050	0.186	0.151	0.186
-1	-0.003	0.183	-0.007	0.183
0	-0.074	0.180	0.114	0.180
1	-0.327	0.183	-0.024	0.183
2	-0.119	0.186	0.186	0.186
3	0.129	0.189	0.347	0.189
4	0.073	0.192	0.021	0.192
5	0.026	0.196	-0.117	0.196

Chart 5-3 : Satisfaction and Revenue Data by Week



5.1.12 Hypothesis 6: Results

Despite considerable support in the literature for the relationship between consumer satisfaction and sales revenue, no relationship was found to exist between BGC’s customer and their performance. This is further illustrated in Chart 5-3, which actually shows a strong increase in deseasonalized sales relative to site satisfaction. The unadjusted actual sales in this chart highlight the volatility of the seasonal demand pattern. In contrast to the volatility in purchases, overall site satisfaction remained relatively consistent.

Individual and group reflection on these results called into question the efficacy of overall site satisfaction measurement. The analysis conducted into H6 was somewhat different from the research covered by the extant literature. The main difference being that this research focused on aggregate site satisfaction scores rather than tracking the behavior of individual consumers. It is likely that this is the main reason why this research does not validate the findings from the literature. Yet, in this research context, the null hypothesis H6₀ (customer satisfaction is not related to sales revenue) must be accepted.

While the findings from the literature were not supported, this analysis has brought to light an important finding. Because no relationship between overall site satisfaction and revenue

seems to exist, it calls into question the usage of an aggregated site satisfaction measure in the first place. Such a finding validates the points of Wilson (2002) and Olsen, et al., (2014) who contend that retailers have become too reliant on measures than deriving insights that can and should affect actions. As a consequence of the dialogue surrounding these results BGC's ELT recognized that they needed to exercise more caution when reviewing aggregate site satisfaction data. Until these results were discussed, the ELT had been demanding a weekly response from the digital team regarding the fluctuations in the daily site satisfactions scores. This new knowledge seemed to lead to the business to more productive site dialogue rather than knee-jerk reactions based on the score alone. In fact, this realization led BGC to eliminate the aggregated site satisfaction measure in favor of a weekly categorization of the open-ended site feedback responses into meaningful action opportunities.

5.1.13 Hypothesis 7: Analysis

For hypothesis 7, the null hypothesis to be tested was:

$$H7_0: \mu = K ; \text{Product information displays have no impact upon customer satisfaction}$$

Espoused theory in the literature suggests that improvements in the information provided to customers will also increase their level of satisfaction with the assortment and with their overall shopping experience (Mogilner, et al., 2008; Johnsen, 2009; Broniarczyk, et al., 1998; Hoch et al., 1999; Lim, et al., 2015). Therefore, it should be possible to observe changes in customer satisfaction based on adjustments to product information displays, including the availability of rich media (Demangeot and Broderic, 2010; Zhenui, et al., 2010).

In order to test $H7_0$ three different test treatments were implemented at different times: *Video Experiment, Product Description Experiment and the Size Display Experiment*. Based on the results of these experimental treatments, only the *Product Description* and the *Size Display* experiments were implemented on the website for all visitors to experience (implemented in research weeks 11 and 24, respectively). In order to measure the impact on customer satisfaction, survey results were collected prior to the A/B experiments being conducted (1st Survey), after the first experiment was implemented (2nd Survey), and after both experiments were implemented (3rd Survey). The timing of the surveys compared to the experiments is illustrated earlier in this section, in Table 5-14.

From the qualitative feedback, thirty-one responses, fifty-eight responses, and twenty-three relevant responses were generated by the 1st, 2nd and 3rd surveys, respectively. Product information analysis comprised of the summation of four specific feedback codes: communication of prices, product quality, product value, and product information. The mean response values were compared across the surveys using Pearson's chi-square, consistent with the methodology used elsewhere in this paper. Table 5-27 highlights the

results generated by the 1st and 2nd surveys, the latter of which immediately followed the implementation of the Product Description experiment.

It is interesting to note that perceptions of pricing appeared to improve ($P < 0.001$) from the 1st to the 2nd surveys. A member of the co-research group suggested that this result was generated by the *Colorization* experiment (implemented in week 11), which also impacted upon the communication of product pricing information. Indeed, overall satisfaction with product information appeared to improve between the 1st and 2nd surveys ($P = 0.006$). However, this trend appeared to reverse itself by the time the 3rd survey data was collected (see table 5-26).

Table 5-26 : 1st Survey and 3rd Survey χ^2 Analysis – Product Information

Theme	Codes	1st Survey - Response Scale *					Scale Descriptives			3rd Survey - Response Scale *					Scale Descriptives			X ² Probability	
		Strongly Negative		Neutral Opinion		Strongly Positive	n	Mean	SD	Strongly Negative		Neutral Opinion		Strongly Positive	n	Mean	SD	P (1v3)	P (2v3)
		1	2	3	4	5				1	2	3	4	5					
Total Product Information		35.48	22.58	6.45	12.90	22.58	31	2.65	1.61	34.78	26.09	13.04	21.74	4.35	23	2.35	1.278	0.001	0.003
Communication of Prices		42.86	42.86	0.00	14.29	0.00	7	1.86	1.00	0.00	0.00	50.00	50.00	0.00	2	3.50	0.50	N/A	N/A
Product Quality		50.00	0.00	0.00	0.00	50.00	2	3.00	2.01	50.00	0.00	0.00	50.00	0.00	2	2.50	1.51	N/A	N/A
Product Value		0.00	0.00	0.00	25.00	75.00	4	4.75	0.44	0.00	0.00	0.00	0.00	100.00	1	5.00	0.00	N/A	N/A
Product Information		38.89	22.22	11.11	11.11	16.67	18	2.44	1.51	38.89	33.33	11.11	16.67	0.00	18	2.06	1.08	0.000	0.000

* Frequency of Responses
P is the chi-square test probability of association

When assessing the mean value differences between the 1st and 3rd surveys, it becomes clear that levels of satisfaction declined markedly ($P = 0.001$). Sentiment also declined significantly from the 2nd survey to the 3rd survey ($P = 0.003$). From these quantitative results it appears that the experimentation did have a measureable impact on customer satisfaction. However, it was through a parallel review of the qualitative feedback that a deeper understanding was attained.

Qualitative Insights

As mentioned earlier, product information feedback in the first survey largely revolved around requests for improved imagery. For example, “*get some models that don’t look sickly*”, and “*some pictures are not very clear*”. Feedback in the 2nd survey was consistent with this, for example “*There could also be more pictures of the dresses from different angles rather than just from the model to the manican*” and “*the size chart is not helpful*” (sic). In contrast, the 3rd survey feedback appeared to pick up a new phenomena that was causing considerable frustration with the customer “*Waste of time clicking on items only to find out they are sold out*”, “*Please add a way to search only gowns available. I kept picking styles that were sold out*”, and “*If something is sold out you should remove it from the website or put a sold out sticker over it.*”

From several rounds of dialogue with the co-research group and the ELT, it was determined that what caused the feedback that was such a focus during the 3rd survey was an organizational change in inventory management practices. This change caused items that were previously available for sale on the website, to no longer be available online. Yet, because the sold out style was still available in some store locations, it remained an active part of the online product display. In essence, this inventory management change acted as an additional, unintended, test of customer perceptions in response to product information changes. In this case, not clearly highlighting products that were sold out had a significant negative impact on customer satisfaction levels.

Table 5-27 : 1st Survey and 2nd Survey X² Analysis – Product Information

Theme	Codes	1st Survey - Response Scale *					Scale Descriptives			2nd Survey - Response Scale *					Scale Descriptives			X ² Probability
		Strongly Negative		Neutral Opinion		Strongly Positive		Mean	SD	Strongly Negative		Neutral Opinion		Strongly Positive		Mean	SD	
		1	2	3	4	5	n			1	2	3	4	5	n			
Total Product Information		35.48	22.58	6.45	12.90	22.58	31	2.65	1.61	31.03	12.07	22.41	17.24	17.24	58	2.78	1.48	0.006
	Communication of Prices	42.86	42.86	0.00	14.29	0.00	7	1.86	1.00	44.44	11.11	22.22	22.22	0.00	9	2.22	1.23	0.000
	Product Quality	50.00	0.00	0.00	0.00	50.00	2	3.00	2.01	80.00	0.00	20.00	0.00	0.00	5	1.40	0.80	N/A
	Product Value	0.00	0.00	0.00	25.00	75.00	4	4.75	0.44	18.75	0.00	18.75	25.00	37.50	16	3.63	1.46	N/A
	Product Information	38.89	22.22	11.11	11.11	16.67	18	2.44	1.51	25.00	21.43	25.00	14.29	14.29	28	2.71	1.37	0.057

5.1.14 Hypothesis 7: Results

The general conclusion of the co-research group was that the quantitative analysis of customer feedback appeared to highlight clear differences in customer perceptions based on changes in the level of product information. This also appeared to receive validation through analysis of the qualitative feedback. Group reflection on these results suggested that the targeted changes in product information led to some perceptual benefits. One of these changes, the Colorization experiment, was initially mooted as a price justification change but the data seemed to suggest it was also perceived as a product information change. It was, however, an unexpected information change, related to product purchase availability that led to the largest influence on customer satisfaction.

Without the collection and analysis of qualitative data, the impact of the product information testing effort could easily have been misunderstood and misconstrued. This data also highlighted a sizable customer service problem to BGC, for which they were then able to develop a resolution plan. Whether the product information impacts were intended or not, the results seem to suggest that product information, especially product purchase availability, is a strong influencer of consumer satisfaction levels. On this basis, the null hypothesis (product information displays have no impact upon customer satisfaction) can be rejected.

5.1.15 Hypothesis 8: Analysis

For hypothesis 8, the null hypothesis to be tested was:

$$H8_0: \mu = K ; \text{Product information displays have no impact upon sales revenue}$$

There is evidence in the literature that improvements in product communication can reduce product utility uncertainty and increase purchase likelihood (Liang, 2010). As an example, the provision of rich media as a means of communicating product information has been shown to increase purchase intent (Zhenhui, et al., 2010). However, there is some contradictory information in the literature, where providing too much informational can contribute to purchase deferral (Urbany, et al., 1989). Therefore the motivation for testing $H8_0$ was to determine if and how revenue was impacted by product information displays, thus allowing BGC to tailor the website accordingly.

Three specific tests were constructed to test this hypothesis: *Product Description Experiment*, *Video Experiment*, and the *Size Display Experiment*. However, as noted in the analysis of H7, a number of other product information related factors also seemed to have an influence on consumer perceptions: the *Colorization Experiment* and a change in inventory management practices resulting in a lack of communication regarding items that were out of stock on the website. It is the three main site experiments that form the basis of analysis into H8, but the conclusions were also influenced by reflection on these other experiments.

Product Description Experiment

The *Product Description* experiment involved switching the order of the add-to-cart buttons with the product description information, on the product detail page (PDP). An example of this information display change is shown in appendix 7. Prior to implementing this test, and in the control group, the product description was provided alongside an image of the product, and the website user had to scroll down the page to find the add-to-cart buttons. The idea behind the product description change was to understand the impact of deemphasizing product information in favor of making it easier to find the add-to-cart button. One member of the co-research team positioned this as a test of whether an informational or an executorial focus was more important to the customer.

The Product Description experiment ran for approximately four full weeks, between research weeks 4 and 11. Using *Monetate's* A/B testing application, website traffic was randomly split 50/50 between the experiment and the control. In total 2,877,202 website sessions occurred during this period, with 1,437,229 experiencing the Product Description experimental treatment, and 1,439,973 experiencing the control treatment. The analysis

involved the same clickstream metrics as was used in the testing for each of the prior hypothesis analyses.

In order to test for differences between the treatment and control, paired sample *t*-tests were performed comparing each metric across the 24 days of the experiment. Any metric comparisons that resulted in a two-tailed *P* value of 0.05 or greater were discarded from the analysis. The only exception being revenue, given that it was the main target of this testing effort. A summary of the paired samples *t*-test results is shown in table 5-28.

Table 5-28 : Product Description Experiment *t*-test Significant Results Summary

Variable	Sessions	Mean	% Change in Mean	N	Std. Deviation	Std. Error Mean	t	df	P
NewVisitorConversionExperiment	1437229	.012	-2.3%	24	.002	.000	-2.427	23	.023
NewVisitorConversionControl	1439973	.012		24	.002	.000			
AddToCartExperiment	1437229	.052	1.8%	24	.004	.001	3.727	23	.001
AddToCartControl	1439973	.051		24	.004	.001			
CartAbandonmentExperiment	1437229	.731	0.8%	24	.035	.007	2.598	23	.016
CartAbandonmentControl	1439973	.725		24	.037	.008			
BounceRateExperiment	1437229	.184	-0.5%	24	.006	.001	-2.105	23	.046
BounceRateControl	1439973	.185		24	.007	.001			
RevenuePerSessionExperiment	1437229	1.822	0.9%	24	.388	.079	0.525	23	.604
RevenuePerSessionControl	1439973	1.806		24	.379	.077			

The *t*-test analysis does not show a statistically supported difference in revenue between the experiment and control groups. However, new visitor conversion is down on significant basis ($P = 0.023$), suggesting that product information is highly important to customers that may be unfamiliar with the brand and the website. Conversely, add-to-cart did increase ($P = 0.001$), a metric that in the other hypothesis analysis has often been associated with revenue. Apparently offsetting this increase in add-to-cart is an increase in cart abandonment ($P = 0.016$). The increase in cart abandonment, suggests that the emphasis on add to cart execution only generated a superficial change in behavior, and ultimately did not result in any impact on revenue. From discussion with the ELT, the CMO suggested an interest in understanding if the reduction in new visitor conversion would have any long-term impact on new customer acquisition and therefore future sales revenues. However, no immediate action recommendations were agreed, based on this specific dialogue.

Table 5-29 : Stepwise Regression Summary – Product Description Experiment

Variable	Partial regression coefficient (B)	Standard error	Standardised regression coefficient (Beta)		VIF	F-statistic / Significance of F		R2 / Adjusted R2
			t-statistic					
Intercept	-1.70	0.15				1514.81	0.996	
ConversionRateExperiment	137.93	2.65	0.93	51.95	1.452	0.000	0.995	
AverageOrderValueExperiment	0.01	0.00	0.14	8.27	1.382			
AverageTimeOnSiteExperiment	0.01	0.00	0.06	3.76	1.074			

Dependent Variable: RevenuePerSessionExperiment

A stepwise regression analysis was performed to better understand the variables most highly associated with changes in sales revenue in the product description experiment (see Table 5-29). This modeling exercise validated that conversion rate was the variable most highly associated with sales revenue, a consistent finding across each site experiment. Conversion rate contributed .975 out of the .996 model R². This finding further supports concerns about the impact of the product description change on new site visitors.

Table 5-30 : Correlations – Product Description Experiment

Pearson Correlations	RevenuePerSession Experiment
RevenuePerSession	1.000
ConversionRate	.987
NewVisitorConversion	.951
AddToCart	.754
CartAbandonment	-.903
BounceRate	-.223
AveragePageViews	.087
AverageTimeOnSite	-.145
AverageOrderValue	.618

To provide further insight into these test results an exploratory factor analysis (EFA) was performed on both the experimental and control group data (see Table 5-31). In both factorial analyses, revenue and

Table 5-31 : Exploratory Factor Analysis – Product Description Experiment

Exploratory Factor Analysis - Experiment

Factor Items	Factor Loadings by Component			Extracted Communi
	1	2		
ConversionRateExperiment	0.988			0.978
NewVisitorConversionExperiment	0.969			0.941
AddtoCartExperiment	0.727			0.566
CartAbandonmentExperiment	-0.918			0.851
BounceRateExperiment		-0.880		0.833
RevenuePerSessionExperiment	0.992			0.984
AveragePageViewsExperiment		0.826		0.686
AverageTimeOnSiteExperiment		0.547		0.354
AverageOrderValueExperiment	0.625			0.510

Eigenvalue 6.703
% of the variance explained 74.5

Exploratory Factor Analysis - Control

Factor Items	Factor Loadings by Component			Extracted Communi
	1	2	3	
ConversionRateControl	0.994			0.993
NewVisitorConversionControl	0.962			0.969
AddToCartControl	0.764		0.527	0.869
CartAbandonmentControl	-0.892			0.906
BounceRateControl		-0.765	0.540	0.890
RevenuePerSessionControl	0.961			0.990
AveragePageViewsControl		0.823		0.721
AverageTimeOnSiteControl		0.752		0.674
AverageOrderValueControl			-0.848	0.839

Eigenvalue 7.850
% of the variance explained 87.2

new visitor conversion were grouped together. The only change in factor groupings was the inclusion of average order value in the revenue component in the experiment, compared to it being represented in a separate component in the control group. Comparing the correlations of each variable with sales revenue between the two test groups also shows a change in average order value and new visitor conversion rate (Table 5-30). Where new visitor conversion had a lower correlation to revenue in the experiment than in the control, average order value and overall conversion gained in importance. The importance of add-to-cart also increased significantly in the experiment, according to the correlations.

Video Experiment

The *Video* experiment involved the removal of product videos on the product detail page (PDP). Prior to implementing this test, and in the control group, a video of a model demonstrating the product was provided alongside the still images of the product. An example of the Video control treatment is provided in appendix 8. The business idea behind removal of the product videos was to understand the impact of providing less information to the customer as a means of saving significant production costs for BGC. The CFO stated that production of product videos required a significant time and financial investment. Any analysis that validated the ability to stop video production, without any to impact sales revenue, would present a cost-saving opportunity to the business. However, the CMO expressed concern about the broader brand implications. In support of the CMO’s concerns, the literature posits that a reduction in product information should lead to a reduction in purchase intent and revenue.

Table 5-32 : Video Experiment *t*-test Significant Results Summary

Variable	Sessions	Mean	% Change in Mean	N	Std. Deviation	Std. Error Mean	t	df	P
ConversionRateExperiment	693284	.016	-3.1%	23	.002	.000	-2.342	22	.029
ConversionRateControl	1617774	.016		23	.002	.001			
NewVisitorConversionExperiment	693284	.013	-3.5%	23	.002	.000	-1.930	22	.067
NewVisitorConversionControl	1617774	.013		23	.002	.000			
AddToCartExperiment	693284	.058	-1.9%	23	.004	.001	-2.986	22	.007
AddToCartControl	1617774	.060		23	.004	.001			
RevenuePerSessionExperiment	693284	1.934	-3.8%	23	.282	.059	-2.420	22	.024
RevenuePerSessionControl	1617774	2.011		23	.318	.066			

The Video experiment ran for approximately three full weeks, between research weeks 6 and 9. Using *Monetate’s* A/B testing application, website traffic was randomly split 30/70 between the experiment and the control. In total 2,311,058 website sessions occurred during this period, with 693,284 experiencing the removal of videos experimental treatment, and

Table 5-33 : Correlations – Video Experiment

Pearson Correlations	RevenuePerSession Experiment	RevenuePerSession Control
RevenuePerSession	1.000	1.000
ConversionRate	.924	.983
NewVisitorConversion	.866	.953
AddToCart	.588	.690
CartAbandonment	-.887	-.920
BounceRate	-.141	-.237
AveragePageViews	.071	.089
AverageTimeOnSite	-.272	-.221
AverageOrderValue	.346	.358

1,617,774 experiencing the control treatment. The analysis involved the same clickstream metrics as was used in the testing for each of the prior hypothesis analyses.

In order to test for differences between the treatment and control, paired sample *t*-tests were performed comparing each metric across the 23 days of the experiment. A summary of the paired samples *t*-test results is shown in table 5-32. The *t*-test results highlight a significant impact from the removal of product videos. In particular, revenue declined almost four percent ($P = 0.024$), likely caused by a reduction in add-to-cart ($P = 0.007$) and conversion rate ($P = 0.029$) (included in Table 5-32).

To better understand these results a stepwise regression analysis was conducted. The results of this analysis are shown in Table 5-34. This analysis highlights the importance of conversion rate in both the experiment and control groups. However, a review of the regression model correlation coefficients (see Table 5-33) shows an impact on the relationship between revenue and conversion rate, especially new visitor conversion. The correlation coefficient of add-to-cart also declines from the control to the experimental treatment, further validating the *t*-test results.

Table 5-34 : Stepwise Regression Summary – Video Experiment

Regression Results - Experiment

Variable	Partial regression coefficient (B)	Standard error	Standardised regression coefficient (Beta)	t-statistic	VIF	F-statistic / Significance of F		R ² / Adjusted R ²
Intercept	-1.76	0.11		-16.29		1143.50	0.000	0.991
ConversionRateExperiment	125.59	2.80	0.93	44.85	1.001			0.990
AverageOrderValueExperiment	0.01	0.00	0.37	17.85	1.001			

Dependent Variable: RevenuePerSessionExperiment

Regression Results - Control

Variable	Partial regression coefficient (B)	Standard error	Standardised regression coefficient (Beta)	t-statistic	VIF	F-statistic / Significance of F		R ² / Adjusted R ²
Intercept	-1.82	0.13		-13.53		3000.31	0.000	0.997
ConversionRateControl	124.58	1.72	0.95	72.30	1.039			0.996
AverageOrderValueControl	0.02	0.00	0.18	13.36	1.039			

Dependent Variable: RevenuePerSessionControl

An exploratory factor analysis also supported the linkages between conversion rate, new visitor conversion rate, and add-to-cart. As Table 5-35 shows, EFA did not draw out any new relationships between the different treatments. However, the explanation of variance did decline slightly for component 1 in the experiment at 46.8%, compared to 49.6% in the

control. This finding is consistent with the reduced correlations found by the stepwise regression analysis.

The combined analysis conducted into the behavioral differences between the treatment and control groups does suggest a marked impact. As suggested by the literature, purchase intent clearly declined due to the removal of rich media from the website. The analysis also suggests that the reduction in the product information available to new visitors has a significant impact on their desire to purchase. This finding is also support by the analysis conducted into the Product Description experiment.

Table 5-35 : Exploratory Factor Analysis – Video Experiment

Exploratory Factor Analysis - Experiment

Factor Items	Factor Loadings by Component			Extracted Communalities
	1	2	3	
ConversionRateExperiment	0.985			0.992
NewVisitorConversionExperiment	0.951			0.921
AddtoCartExperiment	0.690		0.574	0.824
CartAbandonmentExperiment	-0.897			0.865
BounceRateExperiment		-0.927		0.893
RevenuePerSessionExperiment	0.967			0.981
AveragePageViewsExperiment		0.967		0.955
AverageTimeOnSiteExperiment		0.848		0.786
AverageOrderValueExperiment			-0.943	0.908

Eigenvalue 8.124
% of the variance explained 90.3

Exploratory Factor Analysis - Control

Factor Items	Factor Loadings by Component			Extracted Communalities
	1	2	3	
ConversionRateControl	0.997			0.996
NewVisitorConversionControl	0.975			0.961
AddToCartControl	0.740			0.761
CartAbandonmentControl	-0.905			0.899
BounceRateControl		-0.918		0.922
RevenuePerSessionControl	0.979			0.989
AveragePageViewsControl		0.956		0.961
AverageTimeOnSiteControl		0.919		0.927
AverageOrderValueControl			0.939	0.934

Eigenvalue 8.350
% of the variance explained 92.8

Size Display Experiment

The *Size Range* experiment, in contrast to the prior two experimental treatments, involved the addition of product information (size availability) to the product detail page (PDP). Prior

to implementing this test, and in the control group, the sizes available for purchase in any particular product were not immediately shown to the customer on the PDP. An example of the two Size Display treatments is provided in appendix 10. This experiment was added to the A/B testing effort based on initial reviews of the qualitative feedback provided by customers. There appeared to be significant frustration at the lack of product availability and the lack of product information regarding product out-of-stocks. Therefore, the Size Display treatment was developed in order to improve product information clarity regarding purchase availability. This experimental change was evidence that a robust participatory action research process was being undertaken, where initial plans and methodologies are adjusted and influenced by co-generated knowledge and discourse (Greenwood and Levin, 2007).

The Size Display experiment ran for approximately three weeks, between research weeks 22 and 24. Using *Monetate's* A/B testing application, website traffic was randomly split 50/50 between the experiment and the control. In total 2,392,286 website sessions occurred during this period, with 1,198,557 experiencing the increased display of size information in the experimental treatment, and 1,193,729 experiencing the control treatment. The analysis involved the same clickstream metrics that were used in the testing of each of the prior hypothesis analyses.

In order to test for differences between the treatment and control, paired sample *t*-tests were performed comparing each metric across the 25 days of the experiment. A summary of the paired samples *t*-test results is shown in table 5-36. The *t*-test results highlight a significant impact from the addition of product size availability information. In particular, revenue increased ($P = 0.013$), likely caused by an increase in add-to-cart ($P < 0.001$) and conversion rate ($P = 0.004$). Interestingly, the number of page views ($P < 0.001$) and the average time spent on the site also increased ($P < 0.001$), while the number of store appointments made declined ($P = 0.024$).

Table 5-36 : Size Display t-test Significant Results Summary

Variable	Sessions	Mean	% Change in Mean	N	Std. Deviation	Std. Error Mean	t	df	P
ConversionRateExperiment	1198557	.017	3.9%	25	.003	.001	3.222	24	.004
ConversionRateControl	1193729	.016		25	.003	.001			
NewVisitorConversionExperiment	1198557	.013	3.8%	25	.003	.001	1.981	24	.059
NewVisitorConversionControl	1193729	.012		25	.002	.000			
AddToCartExperiment	1198557	.065	3.9%	25	.006	.001	6.142	24	.000
AddToCartControl	1193729	.063		25	.006	.001			
BounceRateExperiment	1198557	.150	-0.7%	25	.010	.002	-2.266	24	.033
BounceRateControl	1193729	.151		25	.010	.002			
RevenuePerSessionExperiment	1198557	2.085	4.8%	25	.397	.079	2.673	24	.013
RevenuePerSessionControl	1193729	1.990		25	.357	.071			
AveragePageViewsExperiment	1198557	16.941	1.7%	25	.452	.090	8.225	24	.000
AveragePageViewsControl	1193729	16.655		25	.521	.104			
AverageTimeOnSiteExperiment	1198557	53.524	1.5%	25	4.066	.813	5.329	24	.000
AverageTimeOnSiteControl	1193729	52.755		25	4.145	.829			
PDPMakeanAppointmentClickExperiment	1198557	.003	-6.7%	25	.000	.000	-2.414	24	.024
PDPMakeanAppointmentClickControl	1193729	.003		25	.000	.000			
ClickSaleinNavigationMobileMenuExperiment	1198557	.012	3.8%	25	.002	.000	2.703	24	.012
ClickSaleinNavigationMobileMenuControl	1193729	.012		25	.002	.000			
ClickHeaderSignUpLoginExperiment	1198557	.023	2.3%	25	.003	.001	2.626	24	.015
ClickHeaderSignUpLoginControl	1193729	.022		25	.003	.001			

Table 5-37 : Stepwise Regression Summary – Size Display Experiment

Regression Results - Experiment

Variable	Partial regression coefficient (B)	Standard error	Standardised regression coefficient (Beta)	t-statistic	VIF	F-statistic / Significance of F		R ² / Adjusted R ²
Intercept	-2.32	0.10		-24.34		3537.33	0.000	0.997
ConversionRateExperiment	132.27	1.63	0.97	81.37	1.001			
AverageOrderValueExperiment	0.02	0.00	0.28	23.92	1.001			

Dependent Variable: RevenuePerSessionExperiment

Regression Results - Control

Variable	Partial regression coefficient (B)	Standard error	Standardised regression coefficient (Beta)	t-statistic	VIF	F-statistic / Significance of F		R ² / Adjusted R ²
Intercept	-1.95	0.09		-22.36		2613.09	0.000	0.996
ConversionRateControl	121.46	1.85	0.91	65.73	1.011			
AverageOrderValueControl	0.02	0.00	0.32	23.02	1.011			

Dependent Variable: RevenuePerSessionExperiment

Table 5-38 : Exploratory Factor Analysis – Size Display Experiment

Exploratory Factor Analysis - Experiment

Factor Items	Factor Loadings by			Extracted Communalities
	1	2	3	
ConversionRateExperiment	0.982			0.996
NewVisitorConversionExperiment	0.941			0.962
AddtoCartExperiment	0.899			0.865
CartAbandonmentExperiment	-0.819			0.851
BounceRateExperiment		-0.920		0.889
RevenuePerSessionExperiment	0.943			0.994
AveragePageViewsExperiment		0.863		0.792
AverageTimeOnSiteExperiment		0.783		0.659
AverageOrderValueExperiment			0.988	0.981

Eigenvalue 7.989
% of the variance explained 88.8

Exploratory Factor Analysis - Control

Factor Items	Factor Loadings by			Extracted Communalities
	1	2		
ConversionRateControl	0.992			0.994
NewVisitorConversionControl	0.966			0.967
AddToCartControl	0.907			0.827
CartAbandonmentControl	-0.871			0.861
BounceRateControl		-0.793		0.745
RevenuePerSessionControl	0.909			0.941
AveragePageViewsControl		0.837		0.751
AverageTimeOnSiteControl		0.847		0.721
AverageOrderValueControl		0.783		0.613

Eigenvalue 7.419
% of the variance explained 82.4

The *t*-test results are interesting because they again suggest that new visitor conversion is impacted by product information displays. In the Size Display experiment, it so happened that additional information seemed to cause an increase in new visitor and total visitor conversion rates. Another interesting phenomenon is the reduction in store appointment making and increase in time on the site. This suggested to the co-research group that providing this additional product information increased customer engagement online, but perhaps diminished their need to visit a store location. Therefore, while this experimental treatment may have benefited website sales, the overall impact to the business may not be as positive. Additional analysis in the form of stepwise regressions (Table 5-37) and EFA (Table 5-38) was also conducted, but they did not contribute any additional insights to the Size Display treatment comparisons.

5.1.16 Hypothesis 8: Results

The three different experiments that were developed to test H_{8_0} , proved to be highly effective. The results of each experiment also proved to be very consistent. In particular, new visitor conversion was highly correlated with product information changes in all of the experiments. Where increases in product information resulted in increased new visitor conversion, and reductions in product information resulted in reductions in new visitor conversion. Add-to-cart data also held the same relationship to product information in each of the experiments. The experimentation also resulted in an impact to revenue in the Video and Size Display experiments, but not in the Product Description experiment. The research group concluded that the addition or removal of product information has a more dramatic effect than the on page movement of product information.

Reflection upon the results generated by each of the three experiments does lead to the conclusion that the level of product information is correlated with purchase intent and revenue. Therefore, the null hypothesis H_{8_0} (*product information displays have no impact upon sales revenue*) can be rejected. While the rejection of the null hypothesis seems appropriate to the scope of this research, discussion of these results also raised questions about the overall revenue impact to the business. Specifically, the CEO articulated a concern that increased product information in one channel could lead to an increase in purchasing in that channel, but an offsetting decrease in another channel. While an answer to this concern was not within the scope of this research effort, the discussion point seemed to highlight an understanding of the value customer research and its implications to the business.

5.1.17 Hypothesis 9: Analysis

For hypothesis 9, the null hypothesis to be tested was:

$$H_{9_0} : \mu = K ; \text{There are no differences in satisfaction levels between customer types}$$

There is a view in the literature that customer perceptions are affected by their task orientation (typology) (Gupta and Kabadayi, 2010; Ha and Lennon, 2010; Moe, 2003;

Janiszewski, 1998). Janiszewski (1998) defined customers as fitting one of two broad types: goal-directed or exploratory. Goal-directed customers will have a near-term focus on performing a transaction, and will, therefore, be most satisfied by a website that is highly usable for that task (Gupta and Kabaday, 2010). Exploratory customers will be more focused on gathering information for a future purchase need, and will, therefore, be most satisfied with a website that provides strong levels of browsing and product information support (Gupta and Kabadayi, 2010). Moe (2003) further stratified these typologies into four distinct types. If satisfaction is influenced by these typologies, as suggested by the literature, then it should be possible to measure these differences and use them to identify where the website’s strengths and weaknesses.

In order to collect customer typology information a survey question was included in the 3rd on-site survey and the post-visit survey. The question asked, “*What was the primary reason for visiting the site today?*” Each response option was selected from a drop down menu with each possible response corresponding to a customer typology (see Table 5-39). In total 445 survey responses were generated across the two surveys (246 on-site, 199 post-visit). Each visit reason response was aggregated along with the corresponding site satisfaction score provided by the customer. Given the findings from H5, we agreed to analyze each survey independently due to the clear differences in satisfaction ratings between on-site and off-site customers.

Table 5-39 : Customer Typology Assignment by Survey Response Option

Primary reason for visiting the site today?	Janiszewski (1998) Typology	Moe (2003) Typology	Intent to Purchase (Moe, 2003)
To make a purchase online	Goal-directed	Directed Buying	Now
To save product favorites for a future store visit		Search/Deliberation	Soon
To find a store			Soon
To schedule a store appointment			Soon
Research for a future online purchase	Exploratory	Knowledge Building	In the future
Research for a future in-store purchase			In the future
Just to browse		Hedonistic	Just browsing
To look for event planning information			Just browsing
Other	Other	Other	Other

Chi-square analysis was conducted to identify the probability of rating distribution differences between each of the customer typologies. Table 5-40 shows the number of responses generated according to each of the Janiszewski (1998) and Moe (2003) customer types. Chi-square probabilities were calculated for each response combination and are displayed in a matrix format on the right side of the table. The mean rating values are quite distinct across each of the typologies, and this is supported by *p*-values that are less than 0.05 at all typology intersections. The only exception to the differences in typologies was with the 'Other' response category, which seems to be most similar to the 'Search/Deliberation' typology.

Table 5-40 : On-site Customer Typology Responses and X² Difference Probabilities

	3rd Survey - Response Scale *					Scale Descriptives			X ² Probability				
	Strongly Negative		Neutral Opinion		Strongly Positive								
Janiszewski (1998) Typology	1	2	3	4	5	N	Mean	SD	P (GD)	P (EX)	P (Oth)		
Goal-directed	39	22	12	11	8	92	2.21	1.34		0.007	0.690		
Exploratory	23	29	28	15	19	114	2.81	1.36	0.007		0.003		
Other	17	7	6	7	3	40	2.30	1.38	0.690	0.003			
Moe (2003) Typology	1	2	3	4	5	N	Mean	SD	P (DB)	P (S/D)	P (KB)	P (HE)	P (Oth)
Directed Buying	24	16	7	5	1	53	1.92	1.07		0.001	0.001	0.000	0.045
Search/Deliberation	15	6	5	6	7	39	2.59	1.57	0.001		0.036	0.000	0.309
Knowledge Building	21	24	20	13	12	90	2.68	1.34	0.001	0.036		0.001	0.029
Hedonistic	2	5	8	2	7	24	3.29	1.33	0.000	0.000	0.001		0.000
Other	17	7	6	7	3	40	2.30	1.38	0.045	0.309	0.029	0.000	
Total	79	58	46	33	30	246	2.50	1.38					

* Number of Responses

P is the chi-square test probability of association

Table 5-41 : Post-Visit Customer Typology Responses and X² Difference Probabilities

	Post-Visit Survey - Response Scale *					Scale Descriptives			X ² Probability				
	Strongly Negative		Neutral Opinion		Strongly Positive								
Janiszewski (1998) Typology	1	2	3	4	5	N	Mean	SD	P (GD)	P (EX)	P (Oth)		
Goal-directed	5	3	6	27	26	67	3.99	1.16		0.246	0.000		
Exploratory	6	5	23	45	36	115	3.87	1.07	0.246		0.000		
Other	4	3	3	5	2	17	2.88	1.41	0.000	0.000			
Moe (2003) Typology	1	2	3	4	5	N	Mean	SD	P (DB)	P (S/D)	P (KB)	P (HE)	P (Oth)
Directed Buying	4	1	5	12	14	36	3.86	1.29		0.004	0.482	0.000	0.000
Search/Deliberation	1	2	1	15	12	31	4.13	0.99	0.004		0.009	0.000	0.000
Knowledge Building	6	4	17	38	31	96	3.88	1.11	0.482	0.009		0.029	0.000
Hedonistic	0	1	6	7	5	19	3.84	0.90	0.000	0.000	0.029		0.000
Other	4	3	3	5	2	17	2.88	1.41	0.000	0.000	0.000	0.000	
Total	15	11	32	77	64	199	3.82	1.17					

* Number of Responses

P is the chi-square test probability of association

Table 5-41 shows the same data for the post-visit survey. A review of the mean rating values immediately draws a contrast to the more pronounced differences seen on the on-site survey. There is no statistically supported difference between the two Janiszewski (1998) typologies. Moe's (2003) typologies appear to be more clearly differentiated. There are significant probability of difference *p*-values (*P* < 0.05) between three of the four Moe (2003) typologies, with 'Knowledge Building' and 'Directed Buying' the only combination to have

very similar mean rating values and distributions ($P = 0.482$). All of the other possible typology comparisons meet the 0.05 p -value threshold.

Because there was a greater probability of differences across each of the typologies in the on-site survey, further analysis was conducted into this data. The goal was to identify if additional distinctions could be drawn to help inform additional website design and experimentation foci. In order to facilitate this, a k -means clustering approach was used to group typology, visit frequency and non-purchase reason survey responses with each of the site satisfaction scores. The satisfaction scores used were overall satisfaction, ease of use, site performance, product pricing and willingness to recommend to a friend. This latter satisfaction rating variable was collected on a 10-point Likert scale basis compared to the 5-point scale used for the balance of the variables. In order to reduce the possible impact on the k -means clustering calculations, this 10-point scale was converted to a 5-point scale. SPSS v22.0 was used to perform the statistical analysis.

Due to the use of the k -means clustering algorithm it was necessary to determine the number of clusters to group the data into. There is support in the literature for determining the number of clusters based on a visual

review of the data (e.g. Sedlmair, et al., 2012). As a visual review starting point, the data was ranked by mean rating score and put into a contour chart (see Chart 5-4). This chart allowed for a visual review of the data by response rating score. From the group's review of this chart it was determined that approximately four clusters of data were likely to exist: 1. Responses weighted heavily with low rating scores (left most side of the chart), 2. Responses with low to moderate scores, 3. Responses with moderate to high scores, and 4. Responses with high rating scores (right most side of the chart).

This approach was further validated through a hierarchical clustering. Review of the dendrogram

Chart 5-4 : Contour Chart of Satisfaction Ratings by Typology Response

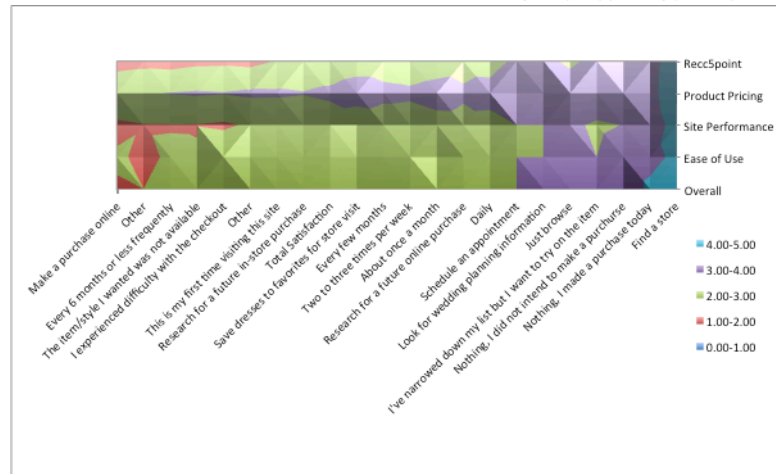
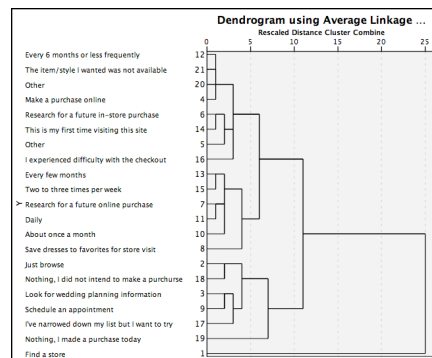


Chart 5-5 : Hierarchical Clustering Dendrogram



Source: SPSS v22.0

(chart 5-5) that resulted from the hierarchical clustering further supports four broad groups of response types. Using a four-group cluster target, a *k*-means clustering analysis was performed on the data. Table 5-42 provides a summary of the analysis results. Notable, is the significance of the *F* statistic of each variable, suggesting that good cluster distinctions were found.

Table 5-42 : *k*-means clustering results summary

Variable	Cluster Centers				ANOVA			
	1	2	3	4	Mean Square	df	F	Sig.
Overall	4.75	2.60	3.39	2.15	3	3	40.817	.000
EaseofUse	4.25	2.53	3.17	2.20	2	3	32.588	.000
SitePerformance	4.50	2.39	3.06	1.97	3	3	90.070	.000
ProductPricing	5.00	3.33	3.51	3.05	1	3	24.689	.000
Recc5point	4.75	2.39	3.21	1.75	4	3	98.342	.000

Table 5-43 shows how each survey response was clustered by each of the clustering methods. The *k*-means and hierarchical clustering methods generated virtually identical cluster selections. The only difference being that “*research for a future in-store purchase*” was grouped with the fourth cluster grouping instead of the second cluster grouping from the *k*-means analysis. From a typology perspective, inclusion of this response makes more intuitive sense based on where the *k*-means analysis placed it, as this grouping generally seems to be related to product research and the knowledge building typology.

In general, the cluster groupings did provide additional insights into the survey response data. Each *k*-means cluster grouping was reviewed and described, as shown at the bottom of Table 5-43. Customers seeking to find a store location were clearly the most satisfied, and were distinct from all other groupings. Next, on a mean rating basis, were customers who were just browsing the assortment or actually completed a purchase (cluster 3). The third ranked grouping (cluster 2), were customers who visited quite frequently in order to actively research an intended future purchase. Lastly, were customers who visit infrequently and experienced difficulty completing their desired task (cluster 4).

Reflection on the cluster analysis results suggests that task execution is a primary rating differentiator. Customers, who completed their tasks, seemed to be more satisfied with the site. Whereas, those that were unable to execute or were seeking to build their product knowledge seemed to rate the site more punitively. These insights suggested to some members of the co-research group that the site was not easy to navigate for those without prior site knowledge. Our group discourse on these results led us to conclude that the site needs to focus more on task orientation and the enablement of each customer typology and shopping mission type.

Table 5-43 : Survey Response Clusters

3rd Survey Responses	Satisfaction Ratings					Cluster Selections	
	Overall	Ease of Use	Site Performance	Product Pricing	Recc5point	k-Means Cluster	Hierarchical Cluster
Find a store	4.75	4.25	4.50	5.00	4.75	1	5
Research for a future in-store purchase	2.50	2.38	2.13	3.05	2.18	2	8
Research for a future online purchase	2.82	2.70	2.64	3.18	2.57	2	6
Save dresses to favorites for store visit	2.30	2.30	2.09	3.64	2.39	2	6
About once a month	2.68	2.39	2.61	3.46	2.43	2	6
Daily	2.75	2.75	2.58	3.46	2.52	2	6
Every few months	2.48	2.50	2.34	3.32	2.41	2	6
Two to three times per week	2.64	2.66	2.36	3.22	2.26	2	6
Just browse	3.30	3.00	3.15	3.85	2.73	3	7
Look for wedding planning information	3.25	3.00	3.00	3.25	3.38	3	7
Schedule an appointment	3.00	3.00	3.00	3.00	3.25	3	7
I've narrowed down my list but I want to try on the item first	3.29	3.00	2.88	3.65	3.41	3	7
Nothing, I did not intend to make a purchase	3.25	3.33	3.02	3.73	2.98	3	7
Nothing, I made a purchase today	4.22	3.67	3.33	3.56	3.50	3	7
Make a purchase online	1.92	2.09	1.81	2.92	1.55	4	8
Other	2.30	2.30	2.10	3.15	1.73	4	8
Every 6 months or less frequently	2.11	2.08	1.97	3.03	1.64	4	8
This is my first time visiting this site	2.41	2.33	2.10	3.14	2.03	4	8
I experienced difficulty with the checkout	2.17	2.67	1.88	3.04	1.81	4	8
Other	2.00	1.87	1.96	2.94	1.70	4	8
The item/style I wanted was not available	2.16	2.05	1.97	3.10	1.78	4	8
Total Satisfaction	2.50	2.46	2.30	3.24	2.19		

Cluster Descriptions	Mean Ratings by Cluster Description						
	Overall	Ease of Use	Site Performance	Product Pricing	Recc5point	k-Means Cluster	Hierarchical Cluster
Find a store	4.75	4.25	4.50	5.00	4.75	4.65	4.65
just browsing or did purchase/will purchase soon	3.39	3.17	3.06	3.50	3.21	3.27	3.27
researching future purchase, visiting daily to every few months	2.60	2.53	2.39	3.33	2.39	2.65	2.68
Visit infrequently, had difficulty purchasing the item I wanted	2.15	2.20	1.97	3.05	1.75	2.22	2.25

5.1.18 Hypothesis 9: Results

Both the chi-square typology and cluster analysis highlighted that satisfaction differences exist based on task orientation. Interestingly, these differences were less significant post-visit than they were during the website visit. To this point, Janiszewski’s (1998) and Moe’s (2003) customer typologies were all found have different levels of satisfaction to the $P < 0.04$ level. Only customers who chose not to divulge their shopping reasons were found to have similar site perceptions to the other survey responders. The cluster analysis helped to explain the key differences between each customer type.

According to the analysis, it was concluded that there was sufficient evidence to reject the null hypothesis (there are no differences in satisfaction levels between customer types). This is an important finding because it highlights where BGC has additional opportunity to engage with customers and improve overall site satisfaction levels. In particular, we agreed that improvements needed to be made to site navigation, especially for customers who are new to the site or conducting research into future purchases. The research also suggested that post-visit surveys offer less action oriented feedback as perceptions tend to improve, and strength of opinion diminishes, once the customer has left the website.

5.2 Principal Researcher Summary of the Findings

The combined results, generated by the research into each hypothesis, were intriguing in many respects. From a practice-based perspective the results challenged conventional thinking and long-held practices. The research process also challenged BGC to rethink its capability requirements, and to understand the benefits of an action research-based methodology. While the results were not intended to be generalizable, they do provide insights that could provide support for future academic research. For example, the results challenge common industry site satisfaction measurement and reporting practices and expand upon similar findings in the literature (e.g. Olsen, et al., 2014).

Table 5-44 : Hypothesis Results Summary

Hypotheses		Results
H1 ₀	Consumer decision costs have no impact upon sales revenue	Accept
H2 ₀	Purchase justification has no impact upon sales revenue	Reject
H3 ₀	Consumer decision costs have no impact upon customer satisfaction	Accept
H4 ₀	Purchase justification has no impact upon customer satisfaction	Accept
H5 ₀	During visit and post-visit customer satisfaction does not differ	Reject
H6 ₀	Customer satisfaction is not related to sales revenue	Accept
H7 ₀	Product information displays have no impact upon customer satisfaction	Reject
H8 ₀	Product information displays have no impact upon sales revenue	Reject
H9 ₀	There are no differences in satisfaction levels between customer types	Reject

5.2.1 Consumer Decision Costs

The literature suggests that there are a wide number of factors that contribute to consumer decision making costs, such as product pricing, assortment size, availability of comparable products, customer assortment knowledge, assortment filtering support, and website ease of use (Betancourt and Gautschi, 1990; Krishnan, et al., 2002; Mogilner et al., 2008; Urbany, et al., 1989; Green and Pearson, 2009; Belanche, et al., 2012). Many of these factors contribute to the amount of time a customer must spend to reach a purchase decision, where the larger the time commitment the higher the decision cost and the lower the intent to purchase (Simonson, 1999; Betancourt and Gautschi, 1990). More recent internet-based research posits that time spent visiting a website is often an indicator of loyalty (Gupta and Kabadayi, 2010; Green and Pearson, 2009; Belanche, et al., 2012).

Based on the extant literature, it was determined that some of BGC's website performance issues could be improved through a focus on reducing consumer decision costs. The potential conflict of opinions in the literature highlighted the importance of a nuanced approach to the analysis, supported by both quantitative and qualitative insights. As a consequence, a series of rigorous and carefully constructed experiments were deployed and

analyzed. The resultant findings were derived from three separate website experiments and measured through behavioral (clickdata) analysis and through closed and open-ended survey feedback. The outcome provides a comprehensive, robust mixed-methods supported, assessment of the consumer decision cost findings from the literature and their relevance to BGCs operating practices.

The decision cost experimentation started with a basic and intuitive change to improve the navigation of the website. The impact was a significant reduction in the website bounce rate and time spent on site, while page views increased. These changes indicated that visitor engagement was improved, while also reducing the decision costs required to accomplish tasks. Changes to product navigation menus, to make them more readable and to improve product search, resulted in fewer page views and less time on site. While this may indicate that customers were able to find products more efficiently, it did not result in any significant change in revenue. In fact, none of the experiments led to any meaningful impacts on revenue, indicating that the time and decision cost reductions did not increase purchase intent as posited by the literature.

Consistent with the lack of impact on revenues, the experimentation also had no aggregate impact on site satisfaction levels. There were actually statistically significant declines in website ease of use and product information ratings across the research time period ($P < 0.02$), both of which are contributors to decision costs according to the literature. It is possible that this was caused by a negative reaction to the experiments. What appears more likely is that due to the breadth of consumer decision cost influences, an incremental change to a small subset of these influencers was insufficient to counteract other site activities and customer experiences with the retail environment at large. Given the dynamic nature of the marketplace, this finding calls into question the ability to aggregate consumer opinions as a meaningful business and quantitative research measurement.

Despite being unable to reject the null hypothesis there were sufficient positive indicators to suggest that some of the experimental changes were beneficial to customers. In particular, the reduction in time costs from the *Clickable Top Nav* and *Fly-Out Menu v3* experiments may have longer-term customer perception benefits than was measureable in this research time frame. Also, improvements in bounce rate and the increasing importance of new visitor conversion suggests that the changes may be particularly benefitting visitors with less experience with the site. In contrast, the change implemented with the *Fly-Out Menu v1* experiment highlights the importance of consumer testing. The business was extremely confident about implementing the *v1* change, yet the results were not generally positive, and this change may have actually added some navigational confusion. Without the ability to conduct structured A/B testing, and the discourse fostered by the action research process, it is possible that this change would have been implemented without an adequate check and balance. The result of this action would have been organizational confusion as to the reasons for the resulting underperformance in the business.

While it does appear that satisfaction with site quality and mission support declined over the research time period, alternative causal factors were highlighted by the qualitative feedback. Therefore, satisfaction changes cannot be directly attributed to the changes in decision costs. There was also no statistically supported change in specific perceptions related to overall decision costs. These findings raise serious doubts about BGCs ability to use aggregated quantitative satisfaction measures to understand customer website sentiment. In a more generalizable context, this research supports similar concerns raised in the literature (e.g. Olsen, et al., 2014) regarding the ability to aggregate quantitative customer site satisfaction feedback for use in business decision-making.

5.2.2 Purchase Justification

According to the literature, purchase intent can be influenced by affecting a customer's justification to conduct a purchase (Betancourt and Gautschi, 1990; Meeker, et al, 2009; Simonson, 1999). A primary means of accomplishing a change in purchase justification is the communication of product value through promotional pricing, or comparative product offerings (Simonson, 1999). By providing greater purchase justification, academic research has shown that increased levels of purchasing and post purchase consumer satisfaction can be attained (Betancourt and Gautschi, 1990; Meeker, et al., 2009; Simonson, 1999; Deveraj, et al., 2002; Sela, et al., 2009; Böhm and Pfister, 1996). Based on these research inputs, website experiments were developed to better highlight product value and promotional pricing.

In order to influence purchase justification three different experiments were conducted. These experiments primarily focused on changes to the emphasis and visibility of promotional pricing and promotional product. From a revenue generation perspective, two important conclusions emerged. First, it was found that increased revenue was generated by increased levels of purchase justification ($P < 0.05$). Second, it was concluded by the business that using price promotions to increase purchase justification was leading to a reduction in revenues. This latter point was not supported to a high level of statistical significance due to the decision to end the experimentation due to its perceived negative impact. Nonetheless, the experimentation supported the claims from the literature, but with a potential warning that there may be a point of diminishing returns. As such, it highlighted the need for continued experimentation and analysis to find the revenue-maximizing method for increasing purchase justification.

In order to determine the impact of the purchase justification experiments on customer satisfaction, survey data was collected before and after the experiments were implemented to the total site visitor population. Data was collected using both closed and open-ended survey questions. Consistent with the findings from the research into customer satisfaction related to decision costs, there was insufficient evidence to confirm a relationship between purchase justification and satisfaction. While some quantitative support was found for a linkage between purchase justification and customer satisfaction, this finding was not

corroborated by qualitative analysis. In fact, there were some indications, from the qualitative data, that a focus on product value as a means of improving purchase justification was altering website perceptions and brand expectations.

While revenue improvements were ultimately attained from the experiments, the long-term implications from a brand perception perspective need to be monitored. Diehl and colleagues (2003) found that using pricing to increase purchase justification increases customer price sensitivity over time. Given an increase in price related qualitative feedback following the experimentation, it is possible that the revenue improvements are not sustainable and that constant testing, refinement and customer perception monitoring is required. Morgan, Anderson and Mittal (2005) posit that price sensitivity can be reduced through the use of customer feedback data to improve customer experience. Therefore, BGC should continue to study both the short and long-term impacts that are generated from changes in purchase justification. It seems that a positive revenue outcome can be generated, but that it requires a more cohesive strategy than just a focus on purchase justification alone.

5.2.3 Product Information

According to the literature, how product information is communicated to customers can also influence sales revenue and customer satisfaction. From a revenue standpoint, product information has the ability to increase purchase intent by reducing utility uncertainty (Liang, 2010). However, if too much information is provided, it can lead to declines in revenue due to purchase deferrals (Urbany, et al., 1989). Improving the method and quality of product information provided to customers has also been shown to improve customer satisfaction (Lim, et al., 2005; Mogilner, et al., 2008; Johnsen, 2009). From these research insights, experiments were constructed to try and improve both immediate sales revenue and customer satisfaction, in the belief that this would also lead to increases in future purchase likelihood.

Three separate experiments were implemented, each targeting different components of product information. These experiments measured the reaction to changes in the presentation of written product information, product videos, and stock availability. Each of these experiments produced comparable results with respect to the influence on new visitor conversion and add-to-cart rates. Specifically, there was a consistently significant ($P < 0.05$) and positive relationship between the quality and availability of product information and new visitor conversion and overall add-to-cart rates. This finding highlighted the business importance of making the website accessible to all customers, regardless of their familiarity with the brand.

Product information changes were also found to directly influence sales revenue. For example, removing product videos from the site resulted in a decline in revenue ($P < 0.03$). Also, improving information regarding the availability of products that were available for immediate purchase increased revenue ($P < 0.02$). The video experiment finding was

important, as the business had identified that ceasing production of the videos would save millions of dollars in annual expenses. However, this testing indicated that elimination of videos would have a more meaningful financial impact on the business, and would more than offset any expense savings. It was also found that changing the location of product information on a particular web page did not influence revenue generation; it was the addition or removal of information that directly influenced revenue.

In line with the significance of the revenue impacts, changes in product information also had an impact on customer sentiment. Interestingly, the influencers of customer satisfaction did not appear to come from the direct experimentation into product information. It was actually two unexpected sources of product information change that seemed to influence customer sentiment. First, customers seemed to respond positively to the change in the colorization of the product prices ($P < 0.001$), which was not initially intended to validate a response to changes in product information. Second, a change in inventory management practices caused an increase in product out-of-stocks on the website. The lack of product availability was not communicated to customers until they attempted to purchase the item. In response, there was a large amount of survey feedback regarding this issue. Based on the qualitative analysis that was conducted it was determined that this unexpected product information issue caused the decline in product information (and overall) satisfaction ($P < 0.001$).

This research finds that product information can have both positive and negative impacts on revenue and customer satisfaction. When the quality of information is improved, both revenue and satisfaction can be improved. However, the research also highlights the delicate nature of the retail ecosystem and the need for qualitative insights as an input into business decision-making. Had a mixed methods approach not been adopted for this research study, it is possible that the decline in customer satisfaction levels would have been attributed to the experimentation. While many of the findings from this research are difficult to generalize, the importance of a mixed methods research approach for future practice-based research is apparent.

5.2.4 Customer Satisfaction

Research into customer satisfaction, as the primary antecedent of business performance, pervaded every experimental treatment and piece of analysis. From the research into decision costs, purchase justification, and product information, only product information was shown to have a significant impact on satisfaction levels. While definitive associations have not been proven across each research subject, this lack of measurability is actually a key finding. In particular, the usage of quantitative satisfaction measures in the large-scale practice based environment, which comprised the target research environment, appears to be of questionable efficacy.

The usefulness of an aggregated quantitative satisfaction measure was further highlighted when considering the overall relationship of satisfaction and revenues across the research

time period. In fact, no relationship was found to exist between overall site satisfaction and sales revenue ($R^2 = 0.010$). This relationship was also tested on a lagged basis, and no relationship was found within a ten week time period of the change in revenue performance. As a consequence, no broad-based support was found for the literature that proposes satisfaction as the primary indicator of business performance (e.g. Diehl and Poynor, 2010; Newman and Werbel, 1973; Wicks and Roethlein, 2009).

An area where aggregated customer satisfaction measurement was viable was in assessing the impact of temporal outlook. This area of the research was based on the theories of Mogilner and colleagues (2014), where these authors posited that customers focused on immediate needs would have a lower relative level of satisfaction. The findings from this research actually expand on Mogilner, et al's (2014) position by showing that customer satisfaction is different, and higher, post-shopping visit than it is during shopping visit ($P < 0.001$). From the analysis, it seems that measuring attitudes during the shopping visit leads to more detailed and critical feedback. Whereas, perspectives following a shopping visit tend to be more general in nature and less harshly rated. In this specific business context, it highlights the usefulness of each type of customer survey. Where, onsite attitudes are most useful for identifying site problems, and post-visit attitudes are most useful for assessing general brand sentiment. What is also clear is that onsite and post-visit satisfaction cannot be directly compared and almost certainly should not be combined to produce an aggregated satisfaction measure.

Another key finding from the research into customer satisfaction is the relationship with customer typology. There is wide ranging support in the literature for the relationship between customer perceptions and customer task orientation (e.g. Gupta and Kabadayi, 2010; Ha and Lennon, 2010; Moe, 2003; Janiszewski, 1998). Therefore, measurement of satisfaction levels across different customer task types can highlight where a website has opportunities for improvement (Peng and Kim, 2014; Lee and Kozar, 2012). This research found strong support for the theories espoused in the literature, especially those of Moe (2003) and Janiszewski (1998). There was clear support for each of the typologies defined by these authors, to a p -value of < 0.01 . While no areas scored highly, goal-directed customers were clearly less satisfied than customers who were just visiting to more generally browse the assortment. This research highlighted the business need to prioritize improvements in site navigation and transactional execution. The research into customer typology also further validated the importance of during visit feedback collection as the primary method of identifying areas for website improvement.

Chapter 6: Action Research and Personal Reflections

The goal of action research is to create sustainable change in the organization where the research is conducted (Greenwood and Levin, 2007; Coghlan and Brannick, 2010). In accordance with this objective, it is important to discuss how new knowledge was created and leveraged as part of the research process. Reflexivity is an essential component of the knowledge generation process, especially with respect to the Principal Researcher and key stakeholders (Greenwood and Levin, 2007; Pedler, 2008; Coghlan and Brannick, 2010). In line with this, the purpose of this chapter is to provide a summary of the organizational and research benefits derived from the adoption of an action research methodology for this thesis project. An additional benefit of action research is the requirement for personal reflection as a means of identifying biases and personal developmental opportunities. Therefore, this chapter incorporates personal reflections into the action research process discussion.

6.1 Project Initiation Phase

The genesis of this project was the poor performance of the BGC website following a change in the software platform. At the time, it was the belief of BGC's executive leadership team (ELT) that the business lacked the processes and talent to identify and remedy the source of the issues. Due to this perceived void, the ELT was very open to new ideas and a change in approach. They also had a desire to understand the underlying problems and to facilitate a solution as quickly as possible. On this basis, the creation of a co-research team and the implementation of an action research methodology were accomplished with only a few middle managers showing signs of resistance to the project efforts. Interestingly, one of those seeming to resist change was actually part of the co-research team.

At the outset, there were some resource capacity concerns. For example, the Director of Digital Technology raised the following question in one of the first co-research team meetings:

"I believe the first tests will be basic in nature, and only limited IT resources will be required. Can you confirm that the business users will be able to accomplish these tests without support from us?"

In part, these concerns were likely due to the on-going effort to remedy the significant software bugs that existed at the outset of the project. From reflection upon this individual's general demeanor, it may be that there was also concern about the benefits that would be derived from the process. This person's demeanor seemed to be one of ambivalence toward the project. Ambivalence is characterized by a contradiction of attitudes and actions (Piderit, 2000). In this case, this individual provided slower than expected response times to questions, late delivery of agreed actions, and was short tempered with other members of the co-research team. Yet, he was also offering verbal support for the goals of the project.

The literature suggests that ambivalence can be caused by a belief that change is needed, but that the current solution is inadequate or will disrupt established routines (Piderit, 2000; Oreg and Sverdlik, 2011). A method for breaking ambivalent attitudes is to develop a shared vision for the project, and to develop an understanding of the underlying concerns that may exist (Dooley, 1997; Ford and Ford, 1994). As a result, I spent time with this individual trying to build the relationship and gain insight into any other concerns. I also made sure that this individual's opinions were included, and that he got credit for any contributions that were made by himself or his team. Over time, these actions seemed to reduce his ambivalence toward the project. However, shortly after the project concluded this individual decided to leave the organization to take another job opportunity. It is not believed that the project contributed to this decision, but it may provide further grounding for this person's demeanor during the project.

Beyond the initial ambivalence, the general response to the project, in its early stages, was one of curiosity. While the project was additive to their responsibilities, the majority of the co-research team was clearly excited to participate. However, there needed to be a period of knowledge building. For the first week or two this involved meeting on a daily basis to discuss the objectives and how we would measure success. We also spent considerable time discussing the process. In particular, as information about the experimentation became known around the organization, it generated a lot of suggestions for changes to the website. As a result, there was a considerable concern on the co-research team about capturing the requests and providing feedback, as noted by one of the co-researchers:

“I am concerned about the number of requests we are getting. We are getting asked to test changes on the site without being provided with the right details to setup the tests correctly, and we have no good way to provide feedback and update the business on our priorities.”

Based on these concerns we created a test request form that collected all of the necessary information. We also agreed to review each request as part of our weekly meetings and created a process for providing feedback to each requestor. While the input from the business was valuable for informing our experiment subjects and designs, we could not incorporate all of the requests into our immediate research efforts. Consequently, we had to inform some requestors that their test requests would be handled after the initial round of experiments were complete. The Chief Financial Officer further reinforced the need for communication, and the following conversation occurred between the CFO and me, during a group meeting:

CFO: *“How will we know when a test result is showing a beneficial result to the business, and how do we see what is happening?”*

Principal Researcher: *“It is our job as a [co-research] team to keep you informed and to provide you with any actions we would recommend. You are welcome to have access to the Monetate toolset to see our current testing data, but it would probably be better to wait until we have an opportunity to synthesize the data into findings and recommendations.”*

CFO: *“OK, but I would like to get regular updates, and I would like us to accelerate any potentially positive results so that we can receive the benefits as soon as possible.”*

This dialogue was important for establishing the ELTs involvement in the process and for understanding the desire to act on the results. In this meeting, we agreed to a weekly update on the performance of each experiment, and the insights we were deriving from the website testing and from the customer survey data. From a customer feedback perspective, the ELT, which includes the CEO, requested to see a weekly summary of the satisfaction scores being collected. Sharing this data raised awareness of common points of feedback to

the ELT, but it also meant that the leadership team became sensitive to week-to-week changes in the quantitative satisfaction scores. This sensitivity to the feedback data manifested in the desire to derive actions from the weekly fluctuations, rather than taking a more measured and broader time-period viewpoint on the required actions.

The discussions that occurred at the beginning of the research process were essential to finding a common ground for executing the experiments and taking action on the derived insights. Retrospectively, the bedding down of the action research methodology only occurred once an initial plan of action was developed. This initial plan became the catalyst for all of the organizational discussions about how to conduct, prioritize and communicate the actions to the key stakeholders. This dialogue also led to clear revisions in priorities and approach, but also fostered engagement and buy in from the stakeholders. The outcome was a final prioritized set of actions and meeting and decision points where reflection and plan revisions could be discussed and agreed. The working practices ultimately derived were highly consistent with the action research methodology proposed by Coghlan and Brannick (2010) but, consistent with action research objectives, the actual approach was developed democratically (Greenwood and Levin, 2007; Ivankova, 2015).

From a personal perspective, the project initiation phase was daunting. My knowledge of action research application, at this point in my research development, was primarily formed through reading and learning set interactions during the DBA coursework. I had formed learning sets within my prior business role, but I was in full control of the objectives and measures of success. So, this project was the first time I was attempting to utilize action research in an unfamiliar setting and in such a high profile manner. In addition, I had never faced this type of problem before. Yet, despite these uncertainties, I needed to appear to be confident in the approach, even though I was not. Had I appeared to waver in my belief that we could make business improvements, it is possible that the project would not have been approved. It is also possible that the ELT and co-research team would have more severely questioned the actions being taken, limiting our ability to take and learn from action.

Meta-learning

My confidence in the project stemmed from my belief that the combination of collaboration, discourse and theory would eventually uncover a path to improvement. However, had I not selected the right literature, constructed the right process, or worked with the right people, the results could have been very different. In fact, it must be stated that had I been more versed in the practice of action research, this project may have generated even better results for the business. Despite my personal shortcomings, the project initiation phase of the project was successful in gaining business alignment on the objectives, on the agreed actions, and on the formation of a co-research team and process.

It should also be noted that at this stage in the project, all participants were focused on short-term actions as a means of performance improvement. My focus was on setting up and reinforcing the action research process and to gaining alignment with the participants. I

was also trying to build relationships and trust, through one-on-one meetings, chatting in the break-room and hallways and finding opportunities to eat lunch or dinner together. The business participants were solely focused on immediate performance improvements. The reflection on actions, by the participants and I, had not yet really begun. We felt that we were in position to move to the next step and start taking action. However, I was unsure whether all of the foundational process and trust building work was ready to bear the weight of the project and produce any level of positive outcome.

6.2 Experimentation Phase

Following some initial trial and error regarding experiment setup, three experiments were launched within days of one another: the *clickable top nav*, *colorization*, and *product description* experiments. There was a lot of interest amongst the co-researchers as to whether the results would prove or disprove the hypotheses and show the potential for improvements in customer perceptions and revenues. This engagement was fostered through constant email and in-person dialogue about the results, which in turn encouraged each co-researcher to offer an interpretation of the results to date.

The large amount of dialogue also seemed to build ownership sentiment in each of the co-research team members. In fact, I was a little surprised at how quickly the members of the co-research team wanted to take over management of the experimentation process. For the first month of the project, I had been in the leadership position and had managed all points of the activity. However, within two weeks of the first experiments being implemented, other members of the co-research team started to take ownership of the process. At first, I saw this as a slight on my leadership, and I was a little offended. Through reflection and removing my personal emotions from this change in leadership, I gained a better perspective. In particular, I now see that this was an important step in the development of the organization and their confidence to take on a complex website testing process; an ability they did not possess before the project began.

Looking back on the change in ownership of the implementation of experiments, I believe two factors contributed to knowledge building in the organization. First, I presented a clear, but strong, point of view on the experiments that we needed to undertake. I also provided a strong perspective on the objectives for experimentation and their importance to project success. This point view was informed by my review of the literature and from my understanding of how BGC would measure success. Second, alongside the experimental reasons clarity, I fostered debate and involvement. Providing such a solid starting point seemed to give the co-research team a common focal point for discussions. Functional silos and internal politics had likely contributed to their previous inability to collaborate on a viable path to resolution. As an outsider, I did not bring any political baggage into our discussions, only debate about whether the experiments I was suggesting were executable and appropriate. This environment and the ability to debate my point of view fostered engagement and eventually built the knowledge and confidence of the business to enact site

changes. In my view, this was the practical effect of the implementation of action research cycles (e.g. Coghlan and Brannick, 2010) into the business environment.

During the weekly discussion process, the co-research team presented the preliminary results from each experiment to the ELT. One of the first tests, the colorization experiment, looked immediately positive and triggered an important discussion that would shape the decision-making process for the balance of the project:

CEO: *“Let’s roll it out to the whole site, we could use the extra business.”*

Principal Researcher: *“We have not reached a point where we can conclusively say that performance will improve, and we do not fully understand the ramifications of the change. I recommend that we wait another week or two until we are more confident in our assessment of the impact”.*

CEO: *“Very well. This effort looks like it is off to a positive start! Let’s discuss this again as soon as you have a recommendation.”*

In effect, the agreement became that the co-research team would provide guidance to the ELT on when actions should be taken. This was an important point for the business as a whole, as it created a dialogue about statistical confidence levels and when sufficient evidence exists to make a decision. Like with many businesses, there was pressure to generate improved business performance as rapidly as possible. In my experience, suggesting continued testing, when it appears that positive results are being generated, is often frowned upon. However, the problems that the project was founded upon were likely caused by a headlong rush to generate business benefits without a full understanding of the ramifications. While this was not explicitly stated by the ELT, the agreement to delay the full implementation of experiments, until the implications were better understood, seemed to be based on the tacit understanding of what had previously transpired.

Meta-learning

As I further reflect on this stage of the research effort, I realize that I spent much of my time building confidence in the co-research team. As previously noted, the capabilities of the co-research team were viewed skeptically by the ELT, and the co-research team seemed keenly aware of that. I believed that the co-research team was essential to the success of the overall effort. So, I dedicated my time to the empowerment of the group, toward building the acceptance that trying something new and failing was fine, provided that learning occurred in response. At the time, I don’t think I realized quite how much time I was dedicating to this confidence building effort, but I did know that it was critical to success.

As part of the confidence building effort, I also used the weekly forum with the ELT to give the co-research team credit for the work they were doing, to allow them to present positive outcomes and the new knowledge being generated. Over a period of weeks this appeared to build confidence with the ELT and improve relations between the two groups. I base my

assessment in this regard on improvements in the tone of dialogue in the weekly group meetings. The general meeting tone evolved from a questioning one, to a supportive one. In addition, informal conversations between the co-research group and I also appeared to move toward actions and positive outcomes and away from “us versus them” discourse.

For the ELT, the change in co-research group perceptions may have also started to highlight that the current website problems were not purely tactical issues due to a poorly considered website re-platform. This meta-development was probably important to the business being able to move into the recommendations phase. Yet, it was probably more important as a starting point for the meta-learning that would occur during that phase.

6.3 Recommendations Phase

Because the business was unwilling to wait for the entire project scope to be tested and validated before making progress, each experiment was managed somewhat independently. Each experiment was developed, debated, and implemented as previously discussed. As each single experiment generated conclusive results, the co-research team discussed and interpreted the results. These discussions centered on any tangible financial impacts and on the broader, less immediately tangible, customer and brand implications. With respect to the customer implications, the survey data was reviewed in parallel with the site experimentation, but robust analysis of the data did not occur until the latter stages of the project. This meant that the customer implications were factored in, but not comprehensively understood, as each site experiment decision was being made. As a consequence, the co-research team needed to interpret the customer feedback based on their previous working experiences rather than the utilization of the robust qualitative analysis that occurred later.

In hindsight, the desire of the business to proceed with an implementation decision in reaction to each individual experiment generated mixed results. While regular decision-making activity maintained interest and positive perceptions of the project effort, some actions were not broadly understood. For instance, the product information experiments were a good example of where the desire for an accelerated decision-making process impeded the ability of the project to generate learning and progress for the organization. With full consideration of the quantitative and qualitative results, the importance of product information can be quite clearly seen and understood, as detailed in the findings chapter of this thesis. However, before these broader findings could be presented, business decisions were required.

From a budgetary perspective, the production of videos was under scrutiny and became the genesis of one of the project experiments related to product information. This experiment demonstrated the importance of videos to the customer and to the performance of the business, meaning that the opportunity to reduce expense in this area was not feasible. So, a decision was taken to reduce the budget in another, and similar area, which may not have occurred had the full analysis of product information been presented prior to this decision. I

became aware of the potential impact of the budget decision when I was invited to attend a review of the marketing budget with the ELT team. In this meeting I felt compelled to interject my perspective into the budgetary review process:

CMO: *“In order to meet next year’s budget requirements we will need to stop taking product photos on a model, and we will begin shooting on a bust form.”*

Principal Researcher: *“I think that this would be a mistake. The website research conducted over the last several months shows that product information, like pictures and videos, are extremely important to the customer, especially new customers. For example, with the video experiment, we saw a 3.8% reduction in sales when we removed videos. There also appears to be strong evidence of an impact on customer sentiment related to the level of product information that is provided. It is, therefore, possible, even likely, that there will also be a direct sales ramification if you further reduce the ability of customers to understand the product offering through the use of bust forms versus models.”*

VP, Digital: *“I agree that this is a concern, and would likely have a negative impact on sales.”*

CEO: *“In light of this feedback, we need to schedule a smaller group meeting to discuss alternative areas for budget cuts and determine how best to meet our financial objectives.”*

This example related to the product information experimentation highlights the importance of being able to complete the action research process for each key area of the project. In this case, only an interpretation of how the individual video, colorization, and verbiage placement experiments impacted the business had been discussed. As such, the opportunity to present the results of these experiments and their combined implications to the overall concept of product information displays was clearly crucial for generating knowledge and a change in behavior. The need to promote research rigor and validity, through the reinforcement of the action research cycles, is well documented in the literature (e.g. Coghlan and Brannick, 2010; Kirwan and Conboy, 2009). This requirement was covered in both the literature review and methodology chapters of the thesis. Yet, business requirements, and the need for a democratic involvement, impinge on the ability to strictly adhere to the literature at every stage of the research process.

From a personal perspective I challenged myself to consider how I may have been able to present the product information findings earlier. My conclusion was that we needed to complete our co-research team action research cycles and that I needed to personally reflect on the results before I could form broader connections that were ready for discussion with others. My conclusion is that this may just be evidence of the natural push and pull of trying to conduct rigorous participatory research in a practice-based environment. It seems inevitable that methodological conflicts will occur wherever business decision-making occurs

on a daily basis, with or without the benefit of the knowledge generated from the research. With the benefit of this reflection, I believe that I should have recommended that any decisions, related to the areas being researched, be managed in a different process and deferred, where possible, until the study was complete.

A positive piece of personal feedback occurred during the latter stages of the project, where the VP of Digital stated:

"I appreciate your approach. You didn't come in here and just tell us what to do, or try to convince us that you knew best. You actually worked with us to solve the problems and you were careful not to overstep your role."

This feedback helped to validate that I had taken the right approach to fostering engagement from the business and the co-research team. His feedback was also commentary on the political nature of the project, and some points of contention that had arisen between the digital team and the ELT related to the problems the website was facing prior to project initiation. In the generation and presentation of results, I tried to steer all of the stakeholders away from a focus on past mistakes, and toward the results, findings and implications of the project work. This seemed to diffuse tensions and reduce the time being spent discussing the source of the problems and direct that time towards resolution; an approach the Head of ecommerce clearly appreciated.

The culminating act in the recommendations phase was to summarize and present a final set of findings. The first step was to synthesize all of the knowledge generated from the experiments and customer surveys into an initial set of findings. This was then presented to the co-research team for their review and feedback. This process generally occurred on an individual basis and resulted in adjustments to the business interpretations and implications of the findings. Once this validation and adjustment work had been completed, the definitive recommendations were presented to the ELT. The feedback at the conclusion of the project was that the project was a success and that business performance had not only stabilized but had improved markedly since the project began. The aggregate business improvements during the latter stages of the project effort were illustrated in chart 5-3 (page 94).

Meta-learning

By the end of the recommendations phase, organizational knowledge had developed considerably. While full trust between the ELT and co-research team had not been restored, significant improvements had been made. It also appeared to me that the ELT and the co-research team had a general recognition that the source of the initial website problems was probably due more to capability gaps than any other reason. This seemed to be acknowledged by the ELT through commentary about how the business had improved, but it was stated that it was the new test-and-learn capability that would be important to future success. This sentiment was also apparent in the co-research group, who continued to

generate new testing ideas and seemed to thrive on the ability to generate new knowledge for the benefit of the business.

6.4 Action Research Outcomes – Contribution to Practice

At the time of writing, it is now one year since the project effort began and the benefits of this participatory research effort are clearly in evidence. This research project has provided the catalyst for the creation of new functional departments, new business roles, new tools, new processes, new methods of analysis, and a sustained improvement in the web business. At the formal conclusion of the project effort, it was unclear what broader impacts were going to be derived. In hindsight, I think I expected the new knowledge we generated to spark an immediate change in business thinking. In reality, this new knowledge seems to have caused reflection amongst the Executive Leadership Team (ELT) and a more measured response than I initially anticipated.

In the months immediately following completion of the project, the co-research team stayed intact. Its core mission to identify, test and implement improvements to the website continued. I remained involved via phone and email as an “on-call” advisor to the co-research team should they want to discuss any of their proposed activities. At this point, the business had gone from being unable to develop and validate website problem remedies (project initiation) to having a cross-functional team that was skilled at doing so (project completion). The performance of the web business had also improved and was now growing compared to the prior year’s performance. However, in the first few months following project completion, it appeared that this might be the extent of the project’s impact.

When I did not see a revolutionary impact on broader business behaviors, I hoped that this was a sign of a more evolutionary impact. Concerned, I went back to the literature and was drawn to Stacey’s (2011) views on change through evolutionary social dynamics. Whether it was social dynamics or simply time to reflect that was needed, six months following the project the ELT seemed to be ready to take action. At this point, I was asked to return to the business on a full-time basis, with a focus on the brick-and-mortar business in particular. As part of this organizational change, the CEO had determined that the organization needed a stronger customer insights and analytics capability across all channels of business. The following organizational and capability changes were subsequently implemented:

- Hired a Director of Consumer Insights to collect brick-and-mortar customer insights
- Created a Business Analytics Group and Hired a Director of Business Analysis
- Hired additional analysts to focus on quantitative and qualitative data collection and analysis
- Implemented an A/B testing software application for usage in brick-and-mortar testing
- Expanded the role of *Monetate’s* software in daily website decision-making
- Hired new, more customer-focused, leaders for the Digital channel and for customer service that supported the digital channel

- Eliminated aggregated site satisfaction measures and reporting
- Hired a new Director of Digital Technology, with a background in technology change management

In addition to the organization and capability changes that were implemented, three additional project contribution proof points have emerged. Foremost is that these new capabilities have directly led to improvements in business performance. Website customer conversion has improved by more than thirty basis points, and it is acknowledged by the business that these improvements have come directly from the test and learn capabilities that were implemented. Over the last six months, this conversion improvement has helped the business grow more than forty percent to the prior year. The new test and learn capability in the brick-and-mortar business is also credited with halting a one-million dollar store fixture change expense that would have proceeded, and impacted the business negatively, without the benefit of this new capability.

The second major proof point of project contribution is related to a further redesign effort of BGC's website. During the project effort, BGC's Leadership team realized that a website redesign was required in order to improve its appeal to customers. In large part, sharing the open-ended customer feedback, which was collected during the project, sparked this realization. Discussion about this feedback caused further discussion and reflection about the feedback and the initiation of a website redesign effort. As part of this effort, the ELT also determined that more robust customer insights and testing activities were required than they had ever undertaken before. As a consequence, the new website design elements were tested in lab environments, focus group meetings were held, and A/B testing was undertaken on each major point of change. Adjustments to the designs were then made based on the feedback. I don't believe that this process would have been attempted at all, without this action research effort, and certainly not with the level of customer centrality that was used to assess the changes.

The third major proof point occurred recently, with the implementation of the new website design. Before agreeing that the website changes could be implemented, the CEO warned the Digital, Technology and Marketing teams that they must incorporate the new knowledge into their approach. As a result, rigorous systematic testing was completed to ensure that no customer-facing software 'bugs' would exist at the point of implementation. Customer service was trained on the new design, in advance, so that they could assist with any customer concerns. The change was also limited to a small portion of site visitors, and slowly expanded based on success. The result was a smooth implementation of website changes and no apparent negative impacts from the change.

Meta-learning

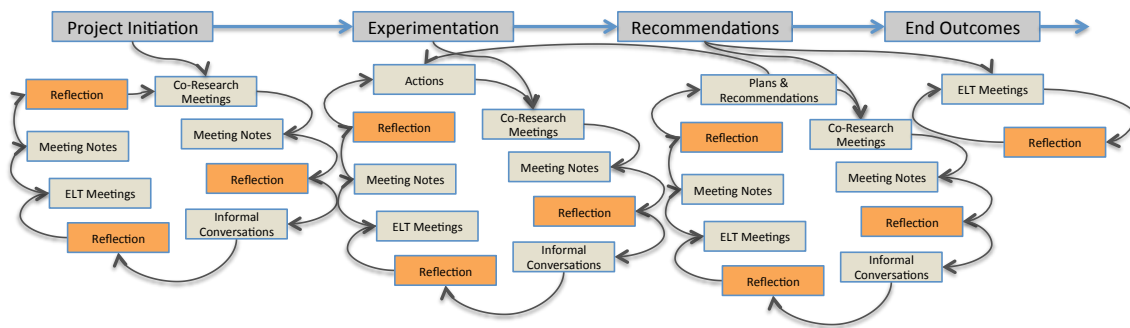
My expectation was for this project to spark an immediate management response. Perhaps this was naïve on my part. I held much stock in the ability of participatory action research to lead to significant and sustainable organizational change. What I did not appreciate was the

time required for the newly generated knowledge to manifest itself in the organization. With reflection, I better appreciate the need for individuals in an organization to derive meaning from the generation of new knowledge, and the time needed to consider how and when to take action. It also appears that an approval process, even for the actions proposed by the CEO, was required before any changes could be implemented.

Underpinning the ultimate changes there must have been a deep meta-learning cycle amongst the ELT. In effect, what was recognized through the actions that were taken is that spot treatments were only short-term solutions. Therefore, the improvements generated by the hypothesis testing were not the most important outcome for this research effort. It was the reflection on those actions, the findings, and the results, that encouraged deeper analysis and learning to occur in the organization. In fact, it was the recognition that capability gaps existed in the organization that was the real outcome from this action research project. Had we not had the forums for discourse, it would have been easy for these leadership learning's to be lost, and for the organization to have been solely focused on immediate performance improvements.

Table 6-1 illustrates the numerous points of discourse and reflection that occurred at each stage of the project. This table highlights how the co-research group and I experienced the project from a learning perspective. During each meeting we each took notes, and I specifically noted not only agreements, but also specific dialogue and tension points. After this, it was clear that reflection on the meetings was occurring based on numerous informal conversations that would take place discussing the outcomes of each meeting. This cycle was generally repeated throughout the project effort, with variations based on the specific activities being attempted. It was also clear that the ELT had undertaken their own reflective and sense-making processes, which were manifested through the questions and direction we received from them during the research effort. This was further evidenced after project conclusion, through the implementation of numerous organizational capability improvements aimed at filling the gaps identified through the project work.

Table 6-1 : Meta-Learning Cycles and Touch-Points



From a personal perspective, I also went through a similar, and intimate, reflection and meta-learning process, which continues to this day as I look back on this effort. Of particular note is the impact on my view of leadership. In the past, I approached leadership from a directive stance, where I encouraged others to follow my lead. What I realized through this project is that ceding control encourages participation and learning in others. I do not believe that the broader meta-learning outcome would have occurred at BGC had I taken a more domineering leadership position. I am sure that I could have led the co-research group to implement tests and that I could have led them to derive insights, which we would have presented to the ELT. Yet, the dialogue generated through open discourse forums, and the new thinking and insights that sprang from that, likely would not have occurred if I were setting the agenda of topics for each meeting. It is this realization that has challenged my notion of leadership and fundamentally changed my outlook. I now aim to manage democratically: to listen, to discuss and reflect, before trying to lead the way forward.

All in all, a year on from project inception, I could not have hoped for a better way for the organization to respond to the new knowledge that we co-generated. While I made mistakes and also had to compromise my own views and those substantiated by the literature, this process truly seems to have led to significant and sustainable improvements to the business. Consequently, I can confidently declare that the contribution to practice that I sought has, in fact, been made even if the outcomes were not exactly aligned with my expectations and hypotheses.

Chapter 7: Conclusion

The stated goal of this thesis project was to enable the business to identify actions that could be taken to improve customer perceptions and business performance. In addition, the research aimed to identify actions that should be avoided. Sustainable organizational change through process and new knowledge development was also sought and supported by an action research methodology. This final chapter of the thesis will summarize progress against these objects and discuss any potential academic implications from the research findings.

This research attempted to identify a relationship between specific business actions, revenue performance and customer satisfaction on an aggregated and large-scale basis across the population of a commercial website. The objective of which was to identify factors that did and did not influence customer sentiment and business performance on this commercial website. In parallel with the identification of performance and sentiment influencers, the research aimed to provide the catalyst for organizational learning and change. As part of the desired organizational change was the aspiration to develop an improved and sustainable set of business practices that would remain after the conclusion of the project.

In order to accomplish the research objectives, an action research methodology was employed. This method proved to be highly valuable for gaining business engagement and for fostering the discourse and reflection required to generate new knowledge. This process was successful in the creation of a co-research team that continued to operate after the thesis project work concluded. Approximately six months after project completion, the new knowledge generated by the project had also directly led to a series of organizational and business practice changes which included the creation of new functional departments, new business roles, changes in business practices, and the implementation and usage of improved analytical tools. In addition to these longer-term organizational and capability impacts, generated by cycles of meta-learning, this project was also credited with enhancements that directly led to immediate business performance improvements. Consequently, it can be concluded that this action research effort made a significant contribution to practice.

Upon reflection into the academic implications, the findings from the research into decision costs, purchase justification, and product information suggest that customer satisfaction is extremely difficult to interpret at scale. Where academics have successfully associated satisfaction with specific actions, the associations have invariably occurred in small-scale and controlled lab-type environments. In laboratory-based research it seems possible to build direct linkages between customer attitudes and business performance. Yet, at the scale that comes with real-world experimentation the ability to derive meaningful associations becomes far more problematic. As a consequence, this research has highlighted the essential need to generate customer attitude and behaviorally linked insights through disaggregated methods of analysis and data collection. In order to generate meaningful insights through practice-based research, the importance of mixed methods of analysis was clearly highlighted. Without the usage of both quantitative and qualitative data the research findings would only have been partially formed, and would likely have resulted in ineffective action recommendations.

The conclusion that the linkage between consumer attitudes and behaviors requires disaggregated data collection and analysis has implications for academics and practitioners.

From a practitioner perspective, it challenges the notion that aggregated site satisfaction scores provide a meaningful business measure; this research has shown that they do not. Therefore, as an outcome of this research, it is recommended that business managers immediately reconsider the usage of aggregated satisfaction measures. The findings also highlight the need to differentiate customer feedback based on visit recency and task orientation.

From an academic perspective, the findings provide further evidence of the need to develop business performance measures that enable meaningful actions. In this regard, this thesis research may prove to be a useful foundation for subsequent research in this area. In particular, there appears to be a need for further research into commercial site satisfaction measures and the connection to business actions, with special emphasis on the collection, interpretation, and incorporation of qualitative insights. The results of this research extend the theories espoused by Olsen, et al. (2014) by validating those theories but also demonstrating that open-ended customer feedback synthesis is a practical replacement for aggregated satisfaction measures.

Table 7-45 : Summary of Results: Tested Academic Theories

Tested Academic Theory	References	Results
Reductions in decision-making costs will lead to increased purchasing	Betancourt & Gautschi (1990); Peng & Kim (2014); Lee & Kozar (2012)	Unsupported
Product discount communication changes can impact revenue	Betancourt & Gautschi (1990); Meeker, et al. (2009); Simonson (1999)	Supported
Customer satisfaction is influenced by decision-making costs	Peng & Kim (2014); Lee & Kozar (2012); Zhenhui, et al. (2010)	Unsupported
Customer satisfaction can be improved by increasing purchase justification	Sela, et al. (2009); Böhm and Pfister (1996)	Unsupported
Customer satisfaction is influenced by temporal outlook	Mogilner, et al., 2014	Supported
Higher levels of customer satisfaction correspond to higher sales revenue	Diehl & Poynor (2010); Newman & Werbel (1973); Wicks & Roethlein (2009)	Unsupported
Customer satisfaction is influenced by the quality of information provided	Mogilner, et al. (2008); Johnsen (2009); Broniarczyk, et al., (1998)	Supported
Product information quality will increase customer purchasing	Mogilner, et al. (2008); Simonson (1999); Lim, et al. (2015)	Supported
Knowledge of each type of customer is essential to improvement development	Gupta & Kabadayi (2010); Ha & Lennon (2010); Moe(2003); Janiszewski(1998)	Supported

While the main findings and benefits of the research are not generalizable, the academy may also gain some methodological insights from the documented practice of an action research methodology. Of particular note, is that the main benefits from this action research project were derived many months after project conclusion. This means that it was the meta-learning generated from the action research process, not actual hypothesis testing that was the primary benefit and organizational change driver. As such, it was an emergent outcome and not the prescribed outcome that created the value. Consideration of these points should provide strong instruction for future action researchers and reinforce the importance of discourse, reflection, time and emergent outcomes as critical components of action research projects.

A number of other academic theories were utilized in the design of experiments. The results of this theoretical testing may be of academic interest related to how the theories were tested at real-world scale. The benefits of this review are likely to be as influential regarding

which testing methods did not produce meaningful results, as it is for those that actually did work. Table 7-45 highlights some of the theories that were tested as part of this research effort.

The linkage between decision-making costs and purchasing was unproven, but potential improvements in cross-channel purchasing and new visitor conversion suggest avenues for further research. Product discount communication testing supported the theories of Betancourt and Gautschi (1990), Meeker, et al. (2009) and Simonson (1999). However, from a practice-based perspective, this research highlighted that negative performance is also a potential outcome from these changes. Linkages between customer satisfaction and decision-making costs and purchase justification could not be demonstrated. Through this research, it became clear that linking any specific action and aggregate customer satisfaction measures was a futile task given the dynamic nature of the real-world environment. This was also true when trying to link aggregate customer satisfaction with aggregate business performance; both measures seemed unrelated at a company-wide level.

This research extended the work of Mogilner, et al. (2014) by demonstrating that shopping event recency has a significant impact on customer satisfaction. As a consequence, future research into consumer satisfaction should consider event recency. In particular, this research showed that customer satisfaction was lower during a task attempt than it is following a task attempt, regardless of the success of the task attempt. In addition, any collection of customer feedback intended to derive practice-based improvements should occur while the customer is trying to complete their tasks, not afterward.

Strong support was found for the theories of Mogilner, et al. (2008), Johnsen (2009) and Simonson (1999) related to product information quality. This research extended upon these theories, by showing that product information appears to impact customer satisfaction, especially where the quality of information is central to task completion. In addition, a further extension of these theories was found, where product information quality was found to impact new customers more than existing customers. It was also found that product information quality was more important than product information placement.

Strong support was also found for various customer typology theories (e.g., Gupta and Kabadayi (2010), Ha and Lennon (2010), Moe (2003), Janiszewski (1998)). The research found that it is essential to understand customer task orientation types when considering how the customer experience will be supported. Customer typology must be considered when designing how customers will interact with a brand, and website designs should cater to these different shopping missions. The research into customer typology differences also highlighted the importance of collecting customer feedback during their visit, rather than post visit.

In summary, the research did fulfill the objectives promised at the outset. Yet, the process of accomplishing these objectives was messy and less than perfect. My reflection is that the

nature of participatory action research requires that the Principal Researcher is comfortable to cede control and operate as part of a dynamic and democratic research effort. This inevitably leads to a less directly controllable research process, and to actions and results that were not anticipated at the outset. In fact, at times, the action research process can feel out of control and provides good instruction in Principal Researcher humility. However, through the utilization and reinforcement of the action research cycles, the process can also be extremely effective for driving positive and sustainable organizational change through cycles of reflection and meta-learning.

References and Appendices

References

Alexander, A 2010, 'Roles change as retailers rationalize assortments', *Drug Store News*, 32, 6, pp. 42-46.

Alfaro, J, and Corbett, C 2003, 'The Value of SKU Rationalization in Practice (The Pooling Effect Under Suboptimal Inventory Policies and Nonnormal Demand)', *Production and Operations Management*, 12, 1, pp. 12-29.

Ayanso, A, and Mokaya, B 2013, 'Efficiency Evaluation in Search Advertising', *Decision Sciences*, 44, 5, pp. 877-913.

Ayanso, A., and Yoogalingam, R. 2009, Profiling Retail Web Site Functionalities and Conversion Rates: A Cluster Analysis. *International Journal Of Electronic Commerce*, 14(1), 79-113.

Bart, Y, Shankar, V, Sultan, F, and Urban, G 2005, 'Are the Drivers and Role of Online Trust the Same for All Web Sites and Consumers? A Large-Scale Exploratory Empirical Study', *Journal Of Marketing*, 69, 4, pp. 133-152.

Baumol, W, and Ide, E 1956, 'Variety in Retailing', *Management Science*, 1, p. 93-101.

Belanche, D, Casaló, L, and Guinalú, M 2012, 'Website usability, consumer satisfaction and the intention to use a website: The moderating effect of perceived risk', *Journal Of Retailing And Consumer Services*, 19, pp. 124-132.

Betancourt, R, and Gautschi, D 1990, 'Demand Complementarities, Household Production, and Retail Assortments', *Marketing Science*, 2, p. 146-161.

Bloom, P, Gundlach, G, and Cannon, J 2000, 'Slotting Allowances and Fees: Schools of Thought and the Views of Practicing Managers', *Journal Of Marketing*, 64, 2, pp. 92-108.

Böhm, G, and Pfister, H 1996, 'Instrumental or emotional evaluations: What determines preferences?', *Acta Psychologica*, 93, 1-3, p. 135-148.

Bower, P 2006, 'How The SandOP Process Creates Value in the Supply Chain', *Journal Of Business Forecasting*, 25, 2, pp. 20-32.

Broniarczyk, S, and Griffin, J 2014, 'Decision Difficulty in the Age of Consumer Empowerment', *Journal Of Consumer Psychology (Elsevier Science)*, 24, 4, pp. 608-625.

Broniarczyk, S, Hoyer, W, and McAlister, L 1998, 'Consumers' Perceptions of the Assortment Offered in a Grocery Category: The Impact of Item Reduction', *Journal Of Marketing Research (JMR)*, 35, 2, pp. 166-176.

- Brownstone, D, and Train, K 1998, 'Forecasting new product penetration with flexible substitution patterns', *Journal Of Econometrics*, 89, 1-2, p. 109-129.
- Brynjolfsson, E, Yu (Jeffrey), H, and Smith, M 2003, 'Consumer Surplus in the Digital Economy: Estimating the Value of Increased Product Variety at Online Booksellers', *Management Science*, 11, p. 1580-1596.
- Bucklin, R, and Sismeiro, C 2003, 'A Model of Web Site Browsing Behavior Estimated on Clickstream Data', *Journal Of Marketing Research (JMR)*, 40, 3, pp. 249-267.
- Bucklin, R, and Sismeiro, C 2009, 'Click Here for Internet Insight: Advances in Clickstream Data Analysis in Marketing', *Journal Of Interactive Marketing*, 23, Anniversary Issue, pp. 35-48.
- Cetinã, I, Dumitrescu, L, and Vinerean, S 2014, 'Exploring Consumer Engagement In An E-Setting: A Qualitative Research Of Marketing Executives', *Economic Computation and Economic Cybernetics Studies and Research*, 48, 2, pp. 1-20.
- Chang, H, and Chen, S 2009, 'Consumer perception of interface quality, security, and loyalty in electronic commerce', *Information and Management*, 46, 7, pp. 411-417.
- Changsoo, S, and Tadisina, S 2008, 'Development of e-service quality measure for internet-based financial institutions', *Total Quality Management and Business Excellence*, 19, 9, pp. 903-918.
- Chernev, A ,2011, 'Product assortment and consumer choice: An interdisciplinary review', *Found Trends Mark*, 6, 1-61.
- Chernev, A, and Hamilton, R 2009, 'Assortment Size and Option Attractiveness in Consumer Choice Among Retailers', *Journal Of Marketing Research (JMR)*, 46, 3, pp. 410-420.
- Chiang, R, Chu, H, and Wang, Y 2013, 'Prediction of members' return visit rates using a time factor', *Electronic Commerce Research And Applications*, 12, 5, p. 362-371.
- Chongwon, P, and Hye-Won, L 2010, 'What Makes A Case Study Really Qualitative?: Show Me Your Evidence, Please!', *English Teaching*, 65, 4, pp. 79-101.
- Coelho, P, and Esteves, S 2007, 'The choice between a five-point and a ten-point scale in the framework of customer satisfaction measurement', *International Journal Of Market Research*, 49, 3, pp. 313-339.
- Cole, S 2005, 'Comparing mail and web-based survey distribution methods: Results of surveys to leisure travel retailers', *Journal Of Travel Research*, 43, 4, p. 422-430.
- Corbett, C, Blackburn, J, and van Wassenhove, L 1999, 'Partnerships to Improve Supply Chains. (cover story)', *Sloan Management Review*, 40, 4, pp. 71-82.

- Creswell, J W, 2014, *Research Design : Qualitative, Quantitative, and Mixed Methods Approaches* / John W. Creswell – 4th ed.: Los Angeles ; SAGE Publications, 2014.
- Creswell, J, and Creswell, J 2013, *Qualitative Inquiry and Research Design : Choosing Among Five Approaches* / John W. Creswell, n.p.: Los Angeles ; SAGE Publications, 2013.
- Cronin Jr., J, Brady, M, and Hult, G 2000, 'Assessing the Effects of Quality, Value, and Customer Satisfaction on Consumer Behavioral Intentions in Service Environments', *Journal Of Retailing*, 76, 2, p. 193-218.
- Dabholkar, P, Thorpe, D, and Rentz, J 1996, 'A Measure of Service Quality for Retail Stores: Scale Development and Validation', *Journal Of The Academy Of Marketing Science*, 24, 1, p. 3-16.
- Darley, W, Blankson, C, and Luethge, D 2010, 'Toward an integrated framework for online consumer behavior and decision making process: A review', *Psychology and Marketing*, 27, 2, pp. 94-116.
- Dass, M, Kumar, P, and Peev, P 2013, 'Brand vulnerability to product assortments and prices', *Journal Of Marketing Management*, 29, 7/8, pp. 735-754.
- Demangeot, C, and Broderick, A 2010, 'Consumer perceptions of online shopping environments: A gestalt approach', *Psychology and Marketing*, 27, 2, pp. 117-140.
- Devaraj, S, Fan, M, and Kohli, R 2002, 'Antecedents of B2C Channel Satisfaction and Preference: Validating e-Commerce Metrics', *Information Systems Research*, 3, p. 316
- Diehl, K, Kornish, L, and Lynch, J 2003, 'Smart Agents: When Lower Search Costs for Quality Information Increase Price Sensitivity', *Journal Of Consumer Research*, 30, 1, pp. 56-71.
- Diehl, K, and Poynor, C 2010, 'Great Expectations?! Assortment Size, Expectations, and Satisfaction', *Journal Of Marketing Research (JMR)*, 47, 2, pp. 312-322.
- Dommeyer, C, and Moriarty, E 1999, 'Comparing two forms of an e-mail survey: embedded vs. attached', *International Journal Of Market Research*, 42, 1, pp. 39-50.
- Donaldson, L. (2008) 'Vita contemplativa—Following the scientific method: how I became a committed functionalist and positivist', *Organization Studies*, 26 (7), pp.1071-1088.
- Dooley K.J. (1997) 'A complex adaptive systems model of organizational change', *Nonlinear Dynamics, Psychology, and Life Science*, 1 (1), pp.69–97.
- Dukes, A, Geylani, T, and Srinivasan, K 2009, 'Strategic Assortment Reduction by a Dominant Retailer', *Marketing Science*, 28, 2, pp. 309-319.
- Easterby-Smith, M, Thorpe, R, and Jackson, P 2012, *Management Research* / Mark Easterby-Smith, Richard Thorpe And Paul Jackson, n.p.: Los Angeles ; SAGE, 2012.

- Feintzeig, R, 2012, 'Lessons From the Death of Circuit City', retrieved from: <http://blogs.wsj.com/bankruptcy/2012/10/25/lessons-from-the-death-of-circuit-city/>
- Finn, A 2011, 'Investigating the non-linear effects of e-service quality dimensions on customer satisfaction', *Journal Of Retailing And Consumer Services*, 18, pp. 27-37.
- Ford, J, and Ford, L 1994, 'Logics Of Identity ,Contradiction, and Attraction in Change', *Academy Of Management Review*, 19, 4, pp. 756-785.
- Fuentes-Blasco, M, Saura, I, Berenguer-Contrí, G, and Moliner-Velazquez, B 2010, 'Measuring the antecedents of e-loyalty and the effect of switching costs on website', *Service Industries Journal*, 30, 11, pp. 1837-1852.
- Ganesh, J, Reynolds, K, Lockett, M, and Pomirleanu, N 2010, 'Online Shopper Motivations, and e-Store Attributes: An Examination of Online Patronage Behavior and Shopper Typologies', *Journal Of Retailing*, 86, pp. 106-115.
- Gardner Jr., ES 1985, 'Exponential Smoothing: The State of the Art', *Journal Of Forecasting*, 4, 1, pp. 1-28.
- Gardner, JS 2006, 'Exponential smoothing: The state of the art—Part II', *International Journal Of Forecasting*, 22, pp. 637-666.
- Gatautis, R, and Vaiciukynaite, E 2013, 'Website Atmosphere: Towards Revisited Taxonomy of Website Elements', *Economics and Management*, vol. 18, no. 3, pp. 537-544.
- Gaur, S, and Agrawal, R 2006, 'Service Quality Measurement in Retail Store Context: A Review of Advances Made Using SERVQUAL and RSQS', *Marketing Review*, 6, 4, pp. 317-330.
- Gilliland, M 2011, 'SKU Rationalization: Pruning Your Way to Better Performance', *Journal Of Business Forecasting*, 30, 3, pp. 28-31.
- Google Analytics, 2015, <http://www.google.com/analytics/>
- Grandcolas, U, Rettie, R, and Marusenko, K 2003, 'Web Survey Bias: Sample or Mode Effect?', *Journal Of Marketing Management*, 19, 5/6, pp. 541-561.
- Greaves, F, Lavery, A, Ramirez Cano, D, Moilanen, K, Pulman, S, Darzi, A, and Millett, C 2014, 'Tweets about hospital quality: a mixed methods study', *BMJ Quality and Safety*, 23, 10, p. 838-846.
- Green, D, and Pearson, J 2009, 'The Examination Of Two Web Site Usability Instruments For Use In B2C E-Commerce Organizations', *Journal Of Computer Information Systems*, 49, 4, pp. 19-32.
- Greene, J, Speizer, H, and Wiitala, W 2008, 'Telephone and Web: Mixed-Mode Challenge', *Health Services Research*, 43, 1p1, pp. 230-248.

- Greiner, L.E. (1972) 'Evolution and Revolution as Organizations Grow', *Harvard Business Review*, July-August, pp. 37-46.
- Gupta, R, and Kabadayi, S 2010, 'The relationship between trusting beliefs and Web site loyalty: The moderating role of consumer motives and flow', *Psychology and Marketing*, 27, 2, pp. 166-185.
- Gwo-Guang, L, and Hsiu-Fen, L 2005, 'Customer perceptions of e-service quality in online shopping', *International Journal Of Retail and Distribution Management*, 33, 2, pp. 161-176.
- Ha, Y, and Lennon, S 2010, 'Online visual merchandising (VMD) cues and consumer pleasure and arousal: Purchasing versus browsing situation', *Psychology and Marketing*, 27, 2, pp. 141-165.
- Hamilton, A, 2008, 'Why Circuit City Busted, While Best Buy Boomed', retrieved from: <http://content.time.com/time/business/article/0,8599,1858079,00.html>
- Handelsman, M, and Munson, J 1985, 'On Integrating Consumer Needs for Variety with Retailer Assortment Decisions', *Advances In Consumer Research*, 12, 1, pp. 108-112.
- Hansen, C. L., Mortensen, N. H., Hvam, L., and Harlou, U. 2012, 'Calculation of Complexity Costs—An Approach for Rationalizing a Product Program', In *Proceedings of NordDesign Conference*.
- Harper, D, 2014, 'retail', *Online Etymology Dictionary*, available at: <http://www.etymonline.com/index.php?term=retail> (accessed 20 January 2015).
- Hauser, J, Urban, G, Liberali, G, and Braun, M 2009, 'Website Morphing', *Marketing Science*, 28, 2, pp. 202-223.
- Hoch, S, Bradlow, E, and Wansink, B 1999, 'The Variety of an Assortment', *Marketing Science*, 4, p. 527-546.
- Hoonakker, P, and Carayon, P 2009, 'Questionnaire Survey Nonresponse: A Comparison of Postal Mail and Internet Surveys', *International Journal Of Human-Computer Interaction*, 25, 5, pp. 348-373.
- Howe, KR 1988, 'Against the Quantitative-Qualitative Incompatibility Thesis or Dogmas Die Hard', *Educational Researcher*, 8, p. 10.
- Huang, P, Lurie, N, and Mitra, S 2009, 'Searching for Experience on the Web: An Empirical Examination of Consumer Behavior for Search and Experience Goods', *Journal Of Marketing*, 73, 2, pp. 55-69.
- Huang, T, and Van Mieghem, J 2014, 'Clickstream Data and Inventory Management: Model and Empirical Analysis', *Production and Operations Management*, 23, 3, pp. 333-347.

- Huber, J, Payne, J, and Puto, C 1982, 'Adding Asymmetrically Dominated Alternatives: Violations of Regularity and the Similarity Hypothesis', *Journal Of Consumer Research*, 9, 1, pp. 90-98.
- Ivankova, N 2015, *Mixed Methods Applications in Action Research: From Methods to Community Action*, Los Angeles ; SAGE, 2015.
- Iyengar, S, and Lepper, M, 2000, 'When choice is demotivating: Can one desire too much of a good thing?', *Journal of Personality and Social Psychology*, 79, 995–1006.
- Janiszewski, C 1998, 'The Influence of Display Characteristics on Visual Exploratory Search Behavior', *Journal Of Consumer Research*, 25, 3, pp. 290-301.
- Jiang, P, and Rosenbloom, B 2005, 'Customer intention to return online: Price perception, attribute-level performance, and satisfaction unfolding over time', *European Journal Of Marketing*, 39, 1-2, p. 150-174.
- Jifeng, L, Sulin, B, and Han, Z 2012, 'The Effectiveness of Online Shopping Characteristics and Well-Designed Websites on Satisfaction', *MIS Quarterly*, 36, 4, pp. 1131-A9.
- Johnsen, M 2009, 'Just in the niche of time', *Drug Store News*, 31, 6, pp. 19-22.
- Johnson, E, Moe, W, Fader, P, Bellman, S, and Lohse, G 2004, 'On the Depth and Dynamics of Online Search Behavior', *Management Science*, 50, 3, pp. 299-308.
- Johnson, R, and Onwuegbuzie, A 2004, 'Mixed Methods Research: A Research Paradigm Whose Time Has Come', *Educational Researcher*, 7, p. 14.
- Katz, J 2009, 'Buried Under Inventory', *Industry Week/IW*, 258, 10, pp. 16-18.
- Kegelart, T., and Van Vyve, M. (2014). *A conic optimization approach for SKU rationalization* (No. CORE Discussion Paper (2014/34)). UCL.
- Kernstock, J, and Brexendorf, T 2012, 'Corporate Brand Integration in Mergers and Acquisitions - An Action Research-Based Approach', *Corporate Reputation Review*, 15, 3, p. 169.
- Kim, D, Ferrin, D, and Rao, H 2009, 'Trust and Satisfaction, Two Stepping Stones for Successful E-Commerce Relationships: A Longitudinal Exploration', *Information Systems Research*, 2, p. 237-257.
- Kim, E, and Jackson, V 2009, 'The Effect of E-SERVQUAL on e-Loyalty for Apparel Online Shopping', *Journal Of Global Scholars Of Marketing Science*, 19, 4, p. 57-65.
- Kim, S, and Stoel, L 2004, 'Apparel retailers: website quality dimensions and satisfaction', *Journal Of Retailing And Consumer Services*, 11, pp. 109-117.

- Kirwan, O, and Conboy, K 2009, 'An Action Research Case Study of the Facilitators and Inhibitors of E-Commerce Adoption', *International Business Research*, 2, 2, pp. 48-56.
- Kocher, J 2009, "Redesign with SEO in mind", *Multichannel Merchant*, 5, 9, pp. 19-21.
- Kohavi, R, Longbotham, R, Sommerfield, D, and Henne, R 2009, 'Controlled experiments on the web: survey and practical guide', *Data Mining and Knowledge Discovery*, 18, 1, pp. 140-181.
- Kök, A, and Xu, Y 2011, 'Optimal and Competitive Assortments with Endogenous Pricing Under Hierarchical Consumer Choice Models', *Management Science*, 57, 9, pp. 1546-1563.
- Krishnan, T, Koelemeijer, K, and Rao, R 2002, 'Consistent Assortment Provision and Service Provision in a Retail Environment', *Marketing Science*, 21, 1, pp. 54-73.
- Kuo, Y, Wu, C, and Deng, W 2009, 'The relationships among service quality, perceived value, customer satisfaction, and post-purchase intention in mobile value-added services', *Computers In Human Behavior*, 25, Including the Special Issue: The Use of Support Devices in Electronic Learning Environments, pp. 887-896.
- Kwan, I, Fong, J, and Wong, H 2005, 'An e-customer behavior model with online analytical mining for internet marketing planning', *Decision Support Systems*, 41, 1, pp. 189-204.
- Ladhari, R 2010, 'Developing e-service quality scales: A literature review', *Journal Of Retailing And Consumer Services*, 17, pp. 464-477.
- Lai, AW 1994, 'Consumption Schemata: Their Effects on Consumer Decision Making', *Advances In Consumer Research*, 21, 1, pp. 489-494.
- Lau, J, and Wong, W 2000, 'Behavioural surveillance of sexually-related risk behaviours for the cross-border traveller population in Hong Kong: the evaluation of the overall effectiveness of relevant prevention programmes by comparing the results of two surveillance surveys', *International Journal Of STD and AIDS*, 11, 11, pp. 719-727.
- Lee, S, and Koubek, R 2010, 'The effects of usability and web design attributes on user preference for e-commerce web sites', *Computers In Industry*, 61, Human-Centered Computing Systems in Industry - A Special Issue in Honor of Professor G. Salvendy, pp. 329-341.
- Lee, Y, and Kozar, K 2012, 'Understanding of website usability: Specifying and measuring constructs and their relationships', *Decision Support Systems*, 52, pp. 450-463.
- Leeflang, P, Parreño Selva, J, Van Dijk, A, and Wittink, D 2008, 'Decomposing the sales promotion bump accounting for cross-category effects', *International Journal Of Research In Marketing*, 25, pp. 201-214.

Lewin, K, and Lewin, G 1948, *Resolving Social Conflicts: Selected Papers On Group Dynamics*, n.p.: New York: Harper, 1948.

Lewis, I, Watson, B, and White, K 2009, 'Internet versus paper-and-pencil survey methods in psychological experiments: Equivalence testing of participant responses to health-related messages', *Australian Journal Of Psychology*, 61, 2, pp. 107-116.

Li, X, Cheng, C, Kim, H, and Petrick, J 2008, 'A systematic comparison of first-time and repeat visitors via a two-phase online survey', *Tourism Management*, 29, pp. 278-293.

Liang, G 2010, 'Capturing Consumption Flexibility in Assortment Choice from Scanner Panel Data', *Management Science*, 56, 10, pp. 1815-1832.

Lim, J, Al-Aali, A, and Heinrichs, J 2015, 'Impact of satisfaction with e-retailers' touch points on purchase behavior: The moderating effect of search and experience product type', *Marketing Letters*, 26, 2, pp. 225-235.

Long, M, and McMellon, C 2004, 'Exploring the determinants of retail service quality on the Internet', *Journal Of Services Marketing*, 18, 1, pp. 78-90.

Mahler, D, and Bahulkar, A 2009, 'Smart complexity', *Strategy And Leadership*, 37, 5, p. 5-11.

Mantrala, M, Levy, M, Kahn, B, Fox, E, Gaidarev, P, Dankworth, B, and Shah, D 2009, 'Why is Assortment Planning so Difficult for Retailers? A Framework and Research Agenda', *Journal Of Retailing*, 85, 1, p. 71-83.

MAXQDA, software for qualitative data analysis, 1989-2015, VERBI Software – Consult – Sozialforschung GmbH, Berlin, Germany.

Meeker, B, Parikh, D, and Jhaveri, M 2009, 'The complexity conundrum', *Marketing Management*, 18, 1, p. 54-59.

Mehrabian, A, and Russell, J 1976, *An Approach To Environmental Psychology*, n.p.: Cambridge (Mass.): M.I.T. Press, 1976.

Moe, WW 2003, 'Buying, Searching, or Browsing: Differentiating between Online Shoppers Using In-Store Navigational Clickstream', *Journal of Consumer Psychology*, 1/2, p. 29-39.

Moe, W, and Fader, P 2004, 'Dynamic Conversion Behavior at E-Commerce Sites', *Management Science*, 50, 3, pp. 326-335.

Mogilner, C, Aaker, J, and Kamvar, S 2014, 'How Happiness Affects Choice', *Journal Of Consumer Research*, pp. S312-S326.

Mogilner, C, Rudnick, T, and Iyengar, S 2008, 'The Mere Categorization Effect: How the Presence of Categories Increases Choosers' Perceptions of Assortment Variety and Outcome Satisfaction', *Journal Of Consumer Research*, 35, 2, pp. 202-215.

- Montgomery, A, Shibo, L, Srinivasan, K, and Liechty, J 2004, 'Modeling Online Browsing and Path Analysis Using Clickstream Data', *Marketing Science*, 23, 4, pp. 579-595.
- Morgan, N, Anderson, E, and Mittal, V 2005, 'Understanding Firms' Customer Satisfaction Information Usage', *Journal Of Marketing*, 69, 3, pp. 131-151.
- Morison, T, Gibson, A, Wigginton, B, and Crabb, S 2015, 'Online research methods in psychology: Methodological opportunities for critical qualitative research', *Qualitative Research In Psychology*, 12, 3, pp. 223-232.
- Mullin, T 2005, 'Reinventing Supermarkets', *Chain Store Age*, 81, 7, p. 28.
- Murray, C, Talukdar, D, and Gosavi, A 2010, 'Joint Optimization of Product Price, Display Orientation and Shelf-Space Allocation in Retail Category Management', *Journal Of Retailing*, 86, 2, pp. 125-136.
- Naik, K, and Srinivasan, S 2015, 'An assessment of departmental store service effectiveness using a modified SERVQUAL approach', *Journal Of Business and Retail Management Research*, 9, 2, pp. 27-43.
- Newman, J, and Werbel, R 1973, 'Multivariate Analysis of Brand Loyalty for Major Household Appliances', *Journal Of Marketing Research (JMR)*, 10, 4, pp. 404-409.
- Nowotny, H, Scott, P, and Gibbons, M 2003, 'Introduction: `Mode 2' Revisited: The New Production of Knowledge', *Minerva: A Review Of Science, Learning and Policy*, 41, 3, p. 179.
- O'Connell, V 2008, 'Reversing Field, Macy's Goes Local', *Wall Street Journal - Eastern Edition*, 21 April.
- Olsen, L, Witell, L, and Gustafsson, A 2014, 'Turning customer satisfaction measurements into action', *Journal Of Service Management*, 25, 4, pp. 556-571.
- Oreg, S, and Sverdluk, N 2011, 'Ambivalence toward imposed change: The conflict between dispositional resistance to change and the orientation toward the change agent', *Journal Of Applied Psychology*, 96, 2, pp. 337-349.
- Oyeyemi, G, Adewara, A, and Adeyemi, R 2010, 'Complex Survey Data Analysis: A Comparison of SAS, SPSS and STATA', *Asian Journal Of Mathematics and Statistics*, 3, 1, pp. 33-39.
- Parasuraman, A 2013, 'Finding Service Gaps in the Age of e-Commerce', *IESE Insight*, 17, pp. 30-37.
- Parasuraman, A, and Zeithaml, V 1994, 'Alternative Scales for Measuring Service Quality: A Comparative Assessment Based on Psychometric and Diagnostic Criteria', *Journal Of Retailing*, 70, 3, pp. 201-230.

Parasuraman, A, Zeithaml, V, and Berry, L 1985, 'A Conceptual Model of Service Quality and Its Implications for Future Research', *Journal Of Marketing*, 49, 4, pp. 41-50.

Parasuraman, A, Zeithaml, V, and Berry, L 1988, 'SERVQUAL: A Multiple-Item Scale for Measuring Consumer Perceptions of Service Quality', *Journal Of Retailing*, 64, 1, pp. 12-40.

Parasuraman, A, Zeithaml, V, and Malhotra, A 2005, 'E-S-QUAL a multiple-item scale for assessing electronic service quality', *Journal Of Service Research*, 7, 3, p. 213-233.

Pellegrini, M 2010, 'Rethink Assortment', *Progressive Grocer*, 1, 2, pp. 16-18.

Peng, C, and Kim, Y 2014, 'Application of the Stimuli-Organism-Response (S-O-R) Framework to Online Shopping Behavior', *Journal Of Internet Commerce*, 13, 3/4, pp. 159-176.

Petnji Yaya, L, Marimon, F, and Casadesus Fa, M 2012, 'Measuring E-Service Quality: Reviewing E-S-QUAL', *Economic Quality Control*, 27, 1, pp. 19-41.

Piderit, S 2000, 'Rethinking Resistance and Recognizing Ambivalence: A Multidimensional View of Attitudes Toward an Organization Change', *Academy Of Management Review*, 25, 4, pp. 783-794.

Rajaram, K 2001, 'Assortment planning in fashion retailing: methodology, application and analysis', *European Journal Of Operational Research*, 129, 1, pp. 186-208.

Rajic, T, and Dado, J 2013, 'Modelling the relationships among retail atmospherics, service quality, satisfaction and customer behavioural intentions in an emerging economy context', *Total Quality Management and Business Excellence*, 24, 9-10, pp. 1096-1110.

Romand Jr., N, Donovan, C, Hsinchun, C, and Nunamaker Jr., J 2003, 'A Methodology for Analyzing Web-Based Qualitative Data', *Journal Of Management Information Systems*, 19, 4, pp. 213-246.

Rutz, O, and Bucklin, R 2012, 'Does banner advertising affect browsing for brands? clickstream choice model says yes, for some', *Quantitative Marketing and Economics*, 10, 2, pp. 231-257.

Salanova, M, Agut, S, and Peiró, J 2005, 'Linking Organizational Resources and Work Engagement to Employee Performance and Customer Loyalty: The Mediation of Service Climate', *Journal Of Applied Psychology*, 90, 6, pp. 1217-1227.

Schmitz, C, You-Cheong, L, and Lilien, G 2014, 'Cross-Selling Performance in Complex Selling Contexts: An Examination of Supervisory- and Compensation-Based Controls', *Journal Of Marketing*, 78, 3, pp. 1-19.

Sedlmair, M, Tatu, A, Munzner, T, and Tory, M 2012, 'A Taxonomy of Visual Cluster Separation Factors', *Computer Graphics Forum*, 31, 3pt4, pp. 1335-1344.

Sela, A, Berger, J, and Liu, W 2009, 'Variety, Vice, and Virtue: How Assortment Size Influences Option Choice', *Journal Of Consumer Research*, 35, 6, pp. 941-951.

Senecal, S, Kalczynski, P, and Nantel, J 2005, 'Consumers' decision-making process and their online shopping behavior: a clickstream analysis', *Journal Of Business Research*, 58, Retailing Reseach :things change, things stay the same, pp. 1599-1608.

Shah, S, and Corley, K 2006, 'Building Better Theory by Bridging the Quantitative–Qualitative Divide', *Journal Of Management Studies*, 43, 8, pp. 1821-1835.

Simonson, I 1999, 'The Effect of Product Assortment on Buyer Preferences', *Journal Of Retailing*, 75, 3, p. 347.

Singh, V, Hansen, K, and Blattberg, R 2006, 'Market Entry and Consumer Behavior: An Investigation of a Wal-Mart Supercenter', *Marketing Science*, 25, 5, pp. 457-476.

Sinha, A, Inman, J, Yantao, W, and Park, J 2005, 'Attribute Drivers: A Factor Analytic Choice Map Approach for Understanding Choices Among SKUs', *Marketing Science*, 24, 3, pp. 351-359.

Stacey, R.D. 2011, *Strategic management and organisational dynamics: the challenge of complexity*. 6th ed. Harlow: Pearson.

Studenmund, AH 1992, *Using Econometrics : A Practical Guide*, n.p.: Boston, Mass. [u.a.].

Su, Q, and Chen, L 2015, 'A method for discovering clusters of e-commerce interest patterns using click-stream data', *Electronic Commerce Research And Applications*, 14, 1, pp. 1-13.

Szymanski, D, and Hise, R 2000, 'e-Satisfaction: An Initial Examination', *Journal Of Retailing*, 76, 3, pp. 309-322.

Tarnanidis, T, Owusu-Frimpong, N, and Marciniak, R 2010, 'Consumer choice: between explicit and implicit reference points', *Marketing Review*, 10, 3, pp. 269-286.

Teddlie, C, and Yu, F 2007, 'Mixed Methods Sampling: A Typology with Examples', *Journal Of Mixed Methods Research*, 1, 1, pp. 77-100.

Tezza, R, Bornia, A, and Andrade, D 2011, 'Measuring web usability using item response theory: Principles, features and opportunities', *Interacting With Computers*, 23, pp. 167-175.

Thomke, S, and Manzi, J 2014, 'The Discipline of Business Experimentation', *Harvard Business Review*, 92, 12, pp. 70-79.

Treiblmaier, H, and Pinterits, A 2010, 'Developing Metrics For Web Sites', *Journal Of Computer Information Systems*, 50, 3, pp. 1-10.

Udo, G, Bagchi, K, and Kirs, P 2010, 'An assessment of customers' e-service quality perception, satisfaction and intention', *International Journal Of Information Management*, 30, 6, p. 481-492.

Urbany, J, Dickson, P, and Wilkie, W 1989, 'Buyer Uncertainty and Information Search', *Journal Of Consumer Research*, 16, 2, pp. 208-215.

Van Heerde, H, and Bijmolt, T 2005, 'Decomposing the Promotional Revenue Bump for Loyalty Program Members Versus Nonmembers', *Journal Of Marketing Research (JMR)*, 42, 4, pp. 443-457.

Verhagen, T, and van Dolen, W 2009, 'Online purchase intentions: A multi-channel store image perspective', *Information and Management*, 46, 2, pp. 77-82.

Wang, Y, Hernandez, M, and Minor, M 2010, 'Web aesthetics effects on perceived online service quality and satisfaction in an e-tail environment: The moderating role of purchase task', *Journal Of Business Research*, 63, Advances in Internet Consumer Behavior and Marketing Strategy, pp. 935-942.

Wang, Y, Minor, M, and Wei, J 2011, 'Aesthetics and the online shopping environment: Understanding consumer responses', *Journal Of Retailing*, 87, pp. 46-58.

Wicks, A, and Roethlein, C 2009, 'A Satisfaction-Based Definition of Quality', *Journal Of Business and Economic Studies*, 15, 1, pp. 82-97.

Wilkerson, J, Iantaffi, A, Grey, J, Bockting, W, and Rosser, B 2014, 'Recommendations for internet-based qualitative health research with hard-to-reach populations', *Qualitative Health Research*, 24, 4, pp. 561-574.

Wilson, A 2002, 'Attitudes towards customer satisfaction measurement in the retail sector', *International Journal Of Market Research*, 44, 2, pp. 213-222.

Woodruff, R, Cadotte, E, and Jenkins, R 1983, 'Modeling Consumer Satisfaction Processes Using Experience-Based Norms', *Journal Of Marketing Research (JMR)*, 20, 3, pp. 296-304.

Wu, W, Lee, C, Fu, C, and Wang, H 2014, 'How can online store layout design and atmosphere influence consumer shopping intention on a website?', *International Journal Of Retail and Distribution Management*, 42, 1, pp. 4-24.

Wulf, T 2011, 'Good to Great to Gone', *Training*, 48, 3, pp. 20-22.

Ya Lan, W, Tainyi, L, Pin, L, and Hsi Peng, L 2015, 'Contribution and Trend to Quality Research--a literature review of SERVQUAL model from 1998 to 2013', *Informatica Economica*, 19, 1, pp. 34-45.

Zeithaml, V, Parasuraman, A, and Malhotra, A 2002, 'Service Quality Delivery Through Web Sites: A Critical Review of Extant Knowledge', *Journal Of The Academy Of Marketing Science*, 30, 4, pp. 362-375.

Zhenhui, J, Chan, J, Tan, B, and Wei Siong, C 2010, 'Effects of Interactivity on Website Involvement and Purchase Intention', *Journal Of The Association For Information Systems*, 11, 1, pp. 34-59.

Zikmund, W 1997, *Business Research Methods* / William G. Zikmund, Fifth Edition, n.p.: Orlando; The Dryden Press 1997.

Appendix

Appendix 1: Initial Literature Supported Web Survey Design

(only used for the Thesis Proposal; this was not implemented due to discussions with the co-research group)

Shopping frequency : Lim, et al. (2015), Verhagen and Van Dolen (2009),

1. How frequently do you visit this site?
 - a. Daily
 - b. Two to three times per week
 - c. Every two to three weeks
 - d. About once a month
 - e. Every few months
 - f. Every 6 months or less frequently
 - g. This is my first time visiting this site

Consumer Typology Determinant Questions : Janiszewski (1998), Moe (2003)

2. Which of the following best describes the type of event you were shopping for today?
 - a. Wedding
 - b. Prom / Homecoming
 - c. First Communion
 - d. Quinceanera
 - e. Graduation
 - f. Formal party
 - g. Other
3. What type of product were you looking for today?
 - a. Dresses only
 - b. Dresses and Accessories
 - c. Accessories only
 - d. Shoes only
 - e. Tuxedos
 - f. Other products
 - g. Other services
4. What was your primary reason for visiting this site today?
 - a. Research for a future online purchase
 - b. Research for a future in-store purchase
 - c. To make a purchase online
 - d. To save product favorites for future store visit
 - e. Just to browse
 - f. To find a store
 - g. To schedule a store appointment
 - h. To look for event planning information
 - i. Other, please specify
5. Were you able to find what you were looking for today?
 - a. Yes
 - b. No
 - c. Didn't have a specific style or item in mind yet
6. (IF NO IN QUES 5) Which of the following describe why you couldn't find what you were looking for?
 - a. I couldn't find the specific item I was looking for
 - b. I couldn't find a style that appealed to me
 - c. My size was not available
 - d. I had difficulty using the site (navigation, searching, etc.)
 - e. Other, please specify
7. (IF NO IN QUES 5) Which best describes what you plan to do next.
 - a. I plan to visit the website again later
 - b. I plan to visit the retail store
 - c. I plan to shop other online retailers
 - d. I plan to shop other retail stores
 - e. Don't Know / Unsure
 - f. Other, please specify
8. Did you make a purchase a _____ m today?
 - a. Yes
 - b. No
9. (IF NO IN QUES 8) Which of the following best describes why you did not make a purchase today?
 - a. I've narrowed down my list but I want to try them on item first
 - b. I'm still looking
 - c. The item/style I wanted was not available to order
 - d. I experienced difficulty with the checkout process
 - e. Other, please specify

Customer Satisfaction (E-S-QUAL) : Parasuraman, Zeithaml and Malhotra (2005)

- Efficiency
10. This site makes it easy to find what I need
 11. It makes it easy to get anywhere on the site.
 12. It enables me to complete a transaction quickly.
 13. Information at this site is well organized.
 14. It loads its pages fast.
 15. This site is simple to use.
 16. This site enables me to get on to it quickly.
 17. This site is well organized.
- System Availability
18. This site is always available for business.
 19. This site launches and runs right away.
 20. This site does not crash.
 21. Pages at this site do not freeze after I enter my order information
- Fulfillment
22. It delivers orders when promised.
 23. This site makes items available for delivery within a suitable timeframe
 24. It quickly delivers what I order.
 25. It sends out the items ordered.
 26. It has in stock the items the company claims to have.
 27. It is truthful about its offerings.
 28. It makes accurate promises about delivery of products.
- Privacy
29. It protects information about my Web-shopping behavior.
 30. It does not share my personal information with other sites.
 31. This site protects information about my credit card.
- Price Perceptions : Jiang and Rosenbloom (2005)
32. The price of products on this site met my expectations
 33. The price of products on this site were lower than I expected
 34. The price of products on this site are consistent with other similar sites

Appendix 2: First On-Site Questionnaire

1. Overall satisfaction rating of the site
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
2. Product satisfaction
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
3. Look and feel of the site
5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
4. Navigation of the site
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
5. How Likely are you to return to this website?
 - a. 10-point Likert (0 – 10)
6. How Likely are you to recommend this website to a friend?
 - a. 10-point Likert (0 – 10)
7. Please leave your feedback!
 - a. Open-ended dialogue entry box was provided

Appendix 3: Second On-Site Questionnaire

1. Overall satisfaction rating of the site
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
2. The site makes it easy to find what I need
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
3. The site performs to my expectations
5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
4. The price of the products meets my expectations
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
5. How frequently do you visit _____.com? (Please choose one...)
 - a. Daily
 - b. Two to three times per week
 - c. About once a month
 - d. Every few months
 - e. About every 6 months or less frequently
 - f. This is my first time visiting this website
6. What was the primary reason for visiting the site today? (Please choose one...)
 - a. Research for a future online purchase
 - b. Research for a future in-store purchase
 - c. To make a purchase online
 - d. To save product favorites for a future store visit
 - e. Just to browse
 - f. To find a store
 - g. To schedule a store appointment
 - h. To look for event planning information
 - i. Other
7. Were you able to find what you were looking for today? (Please choose one...)
 - a. Yes, I was able to complete my task
 - b. No, I couldn't find the specific item I was looking for
 - c. No, I couldn't find any products that appealed to me
 - d. No, the item I wanted was out of stock
 - e. No, I had difficulty using the site
 - f. No, Other
8. What, if anything, prevented you from making a purchase today? (Please choose one...)
 - a. Nothing, I made a purchase today
 - b. Nothing, I did not intend to make a purchase today
 - c. I've narrowed down my list but I want to try the item on first
 - d. The item/style I wanted was not available to order
 - e. I experienced difficulty with the checkout process
 - f. Other
9. How likely would you be to recommend _____.com to a friend?
 - a. 10-point Likert (0 – 10)
10. Which of the following best describes the type of event you were shopping for today?
 - a. Wedding
 - b. Prom / Homecoming
 - c. First Communion
 - d. Quinceanera
 - e. Graduation
 - f. Formal party
 - g. Other
11. Please leave additional comments here:
 - a. Open-ended dialogue entry box was provided

Appendix 4: Post-Visit Questionnaire

1. Overall satisfaction rating of the site
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
2. Willingness to recommend
 - a. 10-point Likert (0 – 10)
3. What was the primary reason for visiting the site today? (Please choose one...)
 - a. Research for a future online purchase
 - b. Research for a future in-store purchase
 - c. To make a purchase online
 - d. To save product favorites for a future store visit
 - e. Just to browse
 - f. To find a store
 - g. To schedule a store appointment
 - h. To look for event planning information
 - i. Other
4. Were you able to complete your task? (Please choose one...)
 - a. Yes, I was able to complete my task
 - b. No, I couldn't find the specific item I was looking for
 - c. No, I couldn't find any products that appealed to me
 - d. No, the item I wanted was out of stock
 - e. No, I had difficulty using the site
 - f. No, I had difficulty during checkout
 - g. No, I had difficulty making an appointment
 - h. No, Other
5. The site makes it easy to find what I need
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
6. The look and feel of the site is appealing
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
7. The site loads pages fast
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
8. The site does not freeze/crash/give errors.
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
9. The site allows me to sort/narrow products by attributes that are important to me
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
10. The product descriptions provide the necessary information I need to evaluate items
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
11. The product images give me a clear depiction of the product
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
12. The price of the products meets my expectations
 - a. 5-point Likert measure (Strongly Negative – Neutral – Strongly Positive)
13. Please leave additional comments here:

Open-ended dialogue entry box was provided

Appendix 5: Side-by-Side Survey Comparison and Academic Literature Support

Survey Questions by Survey

Literature Category	1st On-Site	2nd On-Site	Post-Visit	Literature Support	Scale	Scale Support	Data Type
e-satisfaction	Overall satisfaction rating of the site	Overall satisfaction rating of the site	Overall satisfaction rating of the site	Ha & Lennon (2010); Belanche, et al. (2012); Diehl & Poynor (2010)	5-point Likert	Ha & Lennon (2010); Diehl & Poynor (2010); Szymanski & Hise (2000)	Quantitative
Assortment Perceptions	Product satisfaction			Mantrala, et al. (2009); Hoch, et al. (1999); Chernev & Hamilton (2009)	5-point Likert	Petnji Yaya, et al. (2012)	Quantitative
Aesthetics	Look and feel of the site		The look and feel of the site is appealing	Wang, et al. (2010); Wang, et al. (2011)	5-point Likert	Petnji Yaya, et al. (2012)	Quantitative
Navigation / Ease of Use	Navigation of the site	The site makes it easy to find what I need	The site makes it easy to find what I need	Lee & Kozar (2012); Peng & Kim (2014)	5-point Likert	Parasuraman, et al. (2005)	Quantitative
Navigation / Ease of Use			The site allows me to sort/narrow products by attributes that are important to me	Diehl, et al., (2003); Hoch, et al., (1999); Mogilner, et al. (2008)	5-point Likert	Parasuraman, et al. (2005)	Quantitative
Loyalty Intentions	How Likely are you to recommend this website to a friend?	How likely would you be to recommend _____.com to a friend?	Willingness to recommend	Diehl & Poynor (2010); Betancourt & Gautschi (1990); Belanche, et al. (2012)	10-point Likert	Coelho & Esteves (2007)	Quantitative
Loyalty Intentions	How Likely are you to return to this website?			Parasuraman, et al. (2005); Belanche, et al. (2012); Long & McMellon (2004)	10-point Likert	Coelho & Esteves (2007)	Quantitative
Site Performance		The site performs to my expectations	The site does not freeze/crash/give errors.	Parasuraman, et al. (2005)	5-point Likert	Parasuraman, et al. (2005)	Quantitative
Site Performance			The site loads pages fast	Parasuraman, et al. (2005)	5-point Likert	Parasuraman, et al. (2005)	Quantitative
Value Perceptions		The price of the products meets my expectations	The price of the products meets my expectations	Jiang & Rosenbloom (2005); Betancourt & Gautschi (1990)	5-point Likert	Petnji Yaya, et al. (2012)	Quantitative
Shopping Frequency		How frequently do you visit _____.com? (Please choose one...)		Lim, et al. (2015); Verhagen & van Dolen (2009)	Multiple Choice	Verhagen & van Dolen (2009)	Quantitative
Customer Typology		What was the primary reason for visiting the site today? (Please choose one...)	What was the primary reason for visiting the site today? (Please choose one...)	Moe (2003); Janiszewski (1998); Ganesh, et al. (2010)	Multiple Choice	Verhagen & van Dolen (2009)	Quantitative
Customer Typology		Which of the following best describes the type of event you were shopping for today?		Moe (2003); Janiszewski (1998); Mogilner, et al. (2014)	Multiple Choice	Verhagen & van Dolen (2009)	Quantitative
Task Support		Were you able to find what you were looking for today? (Please choose one...)	Were you able to complete your task? (Please choose one...)	Broniarczyk & Griffin (2014); Chernev & Hamilton (2009); Treiblmaier & Pinterits (2010)	Multiple Choice	Verhagen & van Dolen (2009)	Quantitative
Purchase Intent		What, if anything, prevented you from making a purchase today? (Please choose one...)		Liang (2010); Urbany, et al. (1989); Belanche, et al. (2012)	Multiple Choice	Verhagen & van Dolen (2009)	Quantitative
Product Information			The product descriptions provide the necessary information I need to evaluate items	Diehl, et al., (2003); Urbany, et al. (1989); Janiszewski (1998)	5-point Likert	Parasuraman, et al. (2005)	Quantitative
Product Information			The product images give me a clear depiction of the product	Belanche, et al., (2012); Zhenhui, et al. (2010); Urbany, et al. (1989); Janiszewski (1998)	5-point Likert	Parasuraman, et al. (2005)	Quantitative
Open-Ended	Please leave your feedback!	Please leave additional comments here:	Please leave additional comments here:	Ivankova (2015); Cresswell (2014)	Open text entry	Ivankova (2015); Cresswell (2014)	Qualitative

Appendix 6: Colorization A/B Test Treatments

Control Treatment

[ADD TO CART](#) [SAVE TO MY FAVORITES](#) [FIND A STORE](#) [MAKE AN APPOINTMENT](#)

Experimental Treatment

[ADD TO CART](#) [SAVE TO MY FAVORITES](#) [FIND A STORE](#) [MAKE AN APPOINTMENT](#)

Experimental Treatment

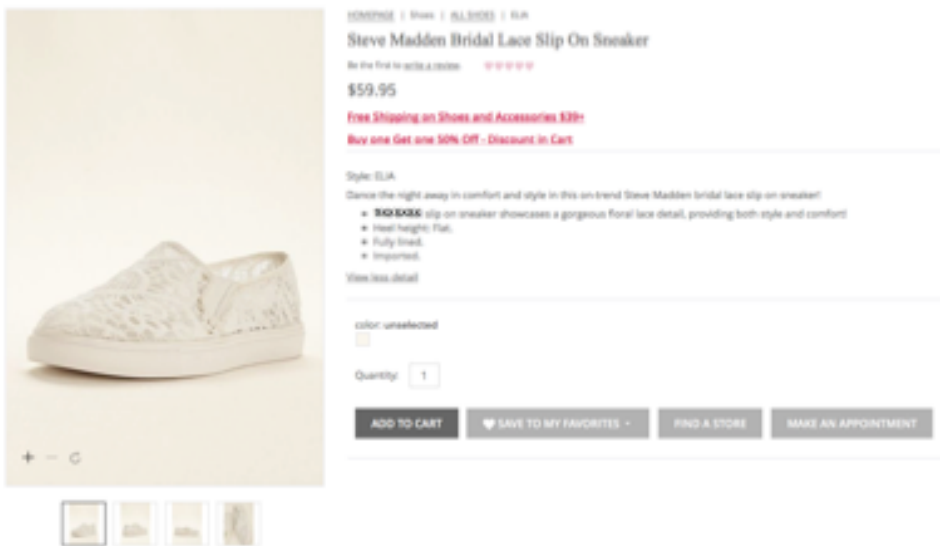
~~\$699.00~~ \$649.00

Control Treatment

~~\$699.00~~ \$649.00

Appendix 7: Product Description A/B Test Treatments

Control Treatment



CONTROL | SHOES | SLIP-ONS | SIA

Steve Madden Bridal Lace Slip On Sneaker

Be the first to act! A. 2020 ★★★★★

\$59.95

Free Shipping on Shoes and Accessories \$39+

Buy one Get one 50% OFF - Discount in Cart

Style: SIA

Dance the night away in comfort and style in this on-trend Steve Madden bridal lace slip on sneaker!

- **XXXXXXXX** slip on sneaker showcases a gorgeous floral lace detail, providing both style and comfort
- Heel height: Flat.
- Fully lined.
- Imported.


[View less detail](#)

color: unselected

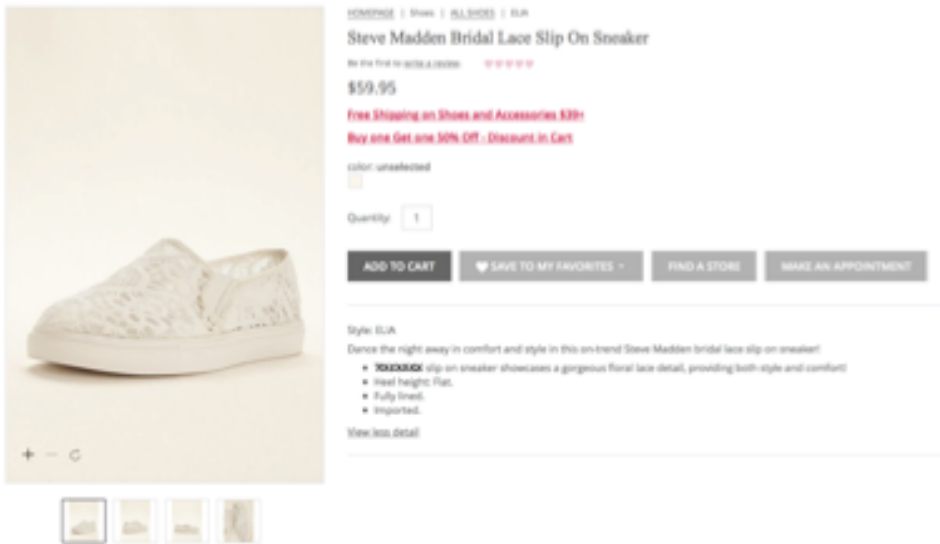
Quantity:

[ADD TO CART](#) [SAVE TO MY FAVORITES](#) [FIND A STORE](#) [MAKE AN APPOINTMENT](#)

+ - C



Experimental Treatment



CONTROL | SHOES | SLIP-ONS | SIA

Steve Madden Bridal Lace Slip On Sneaker

Be the first to act! A. 2020 ★★★★★

\$59.95

Free Shipping on Shoes and Accessories \$39+

Buy one Get one 50% OFF - Discount in Cart

color: unselected

Quantity:

[ADD TO CART](#) [SAVE TO MY FAVORITES](#) [FIND A STORE](#) [MAKE AN APPOINTMENT](#)


Style: SIA

Dance the night away in comfort and style in this on-trend Steve Madden bridal lace slip on sneaker!

- **XXXXXXXX** slip on sneaker showcases a gorgeous floral lace detail, providing both style and comfort
- Heel height: Flat.
- Fully lined.
- Imported.

[View less detail](#)

+ - C

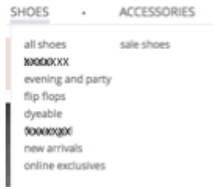


Appendix 8: Video A/B Test - Control Treatment

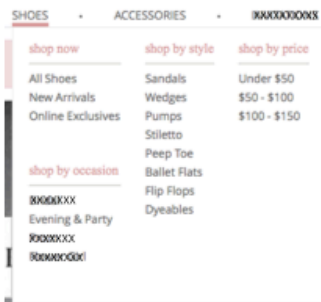


Appendix 9: Fly-Out Nav A/B Test Treatments

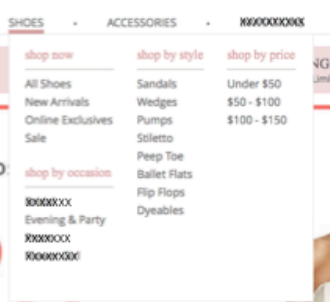
Control Treatment



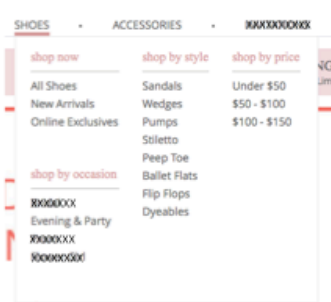
Experimental Treatment v.1



Experimental Treatment v.2



Experimental Treatment v.3



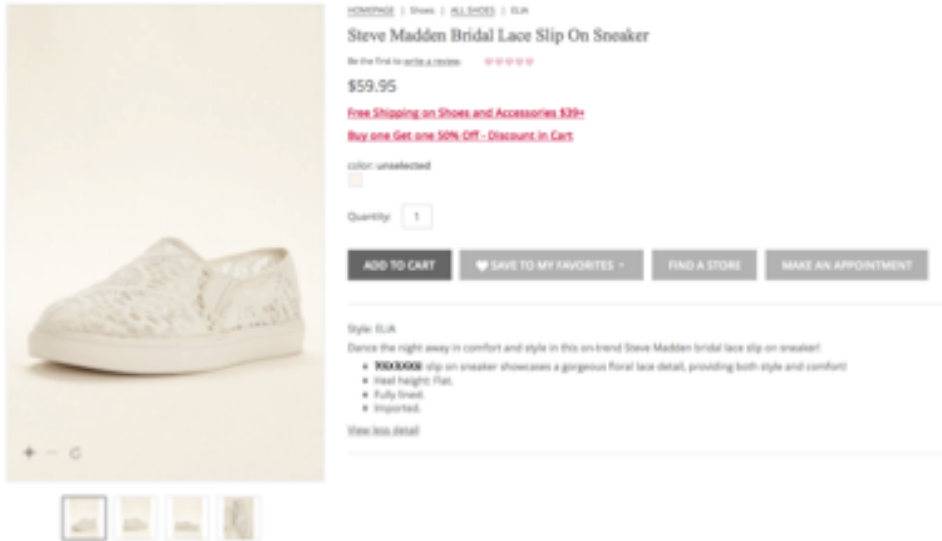
Experimental Treatment v.2

(Top Nav Header 'Sale' Colorization)



Appendix 10: Size Display A/B Test Treatments

Control Treatment



SHOES > Sneakers > ALL SHOES > Slip On

Steve Madden Bridal Lace Slip On Sneaker

Be the first to write a review ★★★★★

\$59.95

Free Shipping on Shoes and Accessories \$39+
Buy one Get one 50% Off - Discount in Cart

color: unselected

Quantity:

[ADD TO CART](#) [SAVE TO MY FAVORITES](#) [FIND A STORE](#) [MAKE AN APPOINTMENT](#)

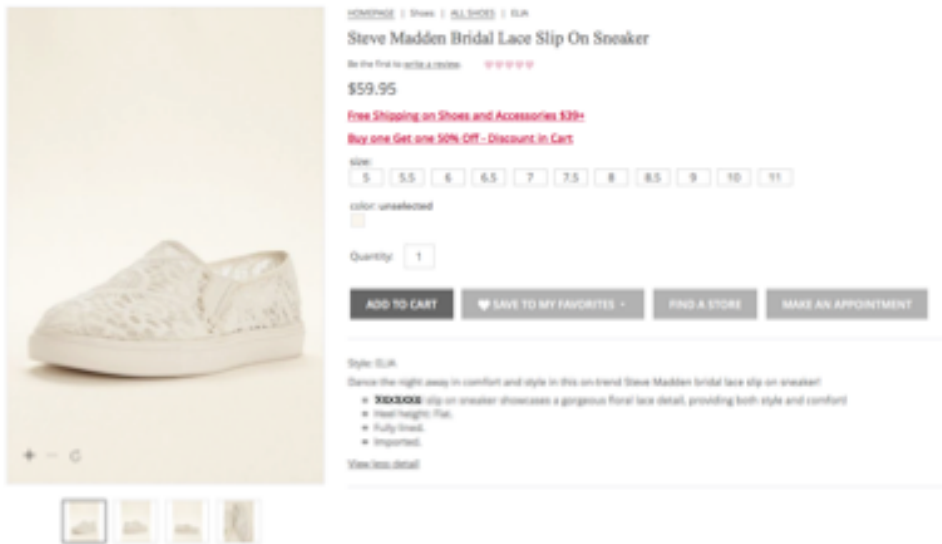
Style: EJA

Dance the night away in comfort and style in this on-trend Steve Madden bridal lace slip on sneaker!

- **XXXXXX** slip on sneaker showcases a gorgeous floral lace detail, providing both style and comfort
- Inset height: Flat.
- Fully lined.
- Imported.

[View less detail](#)

Experimental Treatment



SHOES > Sneakers > ALL SHOES > Slip On

Steve Madden Bridal Lace Slip On Sneaker

Be the first to write a review ★★★★★

\$59.95

Free Shipping on Shoes and Accessories \$39+
Buy one Get one 50% Off - Discount in Cart

size:

color: unselected

Quantity:

[ADD TO CART](#) [SAVE TO MY FAVORITES](#) [FIND A STORE](#) [MAKE AN APPOINTMENT](#)

Style: EJA

Dance the night away in comfort and style in this on-trend Steve Madden bridal lace slip on sneaker!

- **XXXXXX** slip on sneaker showcases a gorgeous floral lace detail, providing both style and comfort
- Inset height: Flat.
- Fully lined.
- Imported.

[View less detail](#)

Appendix 11: Clickstream Data Definitions

- ⟨ Average page views
 - **Definition:** The average number of pages viewed per session.
Average Page Views might go down if site navigation or search is improved. An increase in Page Views might be due to enhanced content, relevant content targeting, or a generally heightened interest of users visiting the site.
 - **Example:** If users collectively visit a total of 150 pages over 30 sessions, the average number of pages viewed is 5.
 - **Formula:** Total Page Views / Total Sessions
- ⟨ Cart Abandonment
 - **Definition:** The percentage of visitors who view their cart with at least 1 item in it during their session, but do not proceed to make a purchase.
Provides visibility into customers who select products but ultimately defer their purchase. Such incidences may indicate a desire to purchase in a physical store, to return on a different device, to return once the customer has reviewed their other purchase options (via other retail locations/sites), to purchase elsewhere, or not to purchase at all.
 - **Example:** If the site has 500 total sessions where a customer views product in the cart, and 10 of those sessions did not complete a purchase, this represents a 2% Cart Abandonment rate.
 - **Formula:** Sessions with Cart Item Viewed and No Purchase / Total Sessions with Cart Item Viewed
- ⟨ New Visitor Conversion
 - **Definition:** Percentage of first-time visitors (anyone that entered the site without a site-specific tracking cookie or pixel) who makes a purchase.
Providing relevant content, and providing an experience that establishes trust are potential methods for improving new visitor conversion.
 - **Example:** If the site has 100 total sessions by “new visitors” and 25 of those sessions are “new visitors” who made a purchase, then the “New Visitor Conversion” rate is 25%.
 - **Formula:** New Visitor Sessions with Purchase / Total New Visitor Sessions
- ⟨ Add to Cart
 - **Definition:** The percentage of visitors who view their cart with at least 1 item in it during their session.
This is helpful for evaluating customer engagement and success in the navigation and search process. Customers who “Add to Cart” have generally found items they are interested in purchasing, so increases in this measure likely mean improvements in the ability to find products or improved perceptions of the products found.
 - **Example:** If the site has 100 total sessions, and 7 of those sessions viewed the cart which contains an item, then you have a 7% Add to Cart rate.
 - **Formula:** Sessions with Cart Item Viewed / Total Sessions
- ⟨ Average Time on Site
 - **Definition:** The average time spent browsing per session.
Time spent might go down if navigation or search is improved. An increase in time might be due to improved content, relevant content targeting, or a generally heightened interest of visitors on the site. This metric measures the entire span of time, and does not guarantee that visitors were engaged for the full duration (such as if they left their browser window open).
 - **Example:** If the site has 200 sessions, and those sessions have a combined browsing time of 500 minutes, then the Average Time on Site is 2.5 minutes.
 - **Formula:** Total Session Time (seconds) / Total Sessions
- ⟨ Average Order Value
 - **Definition:** The average amount spent on an order.
Increased order value can be the result of relevant content targeting, successful upselling, higher cart quantities (such as add-ons and cross-sells), promotional spending thresholds, and attracting the right type of visitors through off-site marketing efforts.
 - **Example:** If the site had 500 orders with a combined purchase total of \$10,000, the average order value is \$20.
 - **Formula:** Total Revenue for Orders / Total Number of Orders
- ⟨ Bounce Rate
 - **Definition:** The percentage of visitors who view only one page on the site and then leave.
A high bounce rate may indicate irrelevant content, improperly targeted content or ineffective inbound-marketing experiences.
 - **Example:** If the site had 1,000 total sessions, and 80 of those sessions viewed one page and then left the site, then the site has an 8% Bounce Rate.
 - **Formula:** Sessions with Only One Page View / Total Sessions
- ⟨ Conversion Rate
 - **Definition:** Percentage of visitors who make a purchase during their session.
Assuming a constant number of site visits, an increase in conversion rate represents additional purchases, regardless of the number of items or value of those purchases
 - **Example:** If the site has 1,000 total sessions and 60 of those sessions resulted in a purchase, the site has a 6% conversion rate.
 - **Formula:** Sessions with Purchase / Total Sessions
- ⟨ Revenue Per Session
 - **Definition:** The average spent by each visitor per site session.
Increased revenue can be the result of relevant content targeting, successful upselling, higher cart quantities (such as add-ons and cross-sells), promotional spending thresholds, and attracting the right type of visitors through off-site marketing efforts.
 - **Example:** If the site has 20 total sessions, with a combined purchase total of \$1,000, then each session is worth \$50.
 - **Formula:** Total Revenue / Total Sessions

Appendix 12: Survey Response Summary

CODDS	1st Survey - Response Scale					2nd Survey - Response Scale					3rd Survey - Response Scale					X ² Probability							
	Strongly Negative		Neutral Opinion		Strongly Positive		Strongly Negative		Neutral Opinion		Strongly Positive		Strongly Negative		Neutral Opinion		Strongly Positive						
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5		1	2	3	4	5		
No Comment	0.00	0.00	12.50	50.00	37.50	0.00	50.00	0.00	0.00	50.00	0.00	50.00	0.00	0.00	50.00	0.00	50.00	0.00	0.00	50.00	0.00	0.00	N/A
Security & Privacy	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Fulfillment	60.00	13.33	6.67	6.67	13.33	40.00	40.00	0.00	20.00	0.00	40.00	0.00	20.00	0.00	40.00	0.00	20.00	0.00	20.00	0.00	0.00	0.00	N/A
Software Bugs	49.59	27.64	16.26	4.88	1.63	47.37	19.30	25.44	5.26	2.63	11.4	1.96	1.09	0.076	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Site Performance	45.45	31.82	18.18	4.55	0.00	28.57	28.57	0.00	14.29	0.00	14.29	7	2.43	1.30	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Task Support	26.47	17.65	14.71	17.65	23.53	10.81	10.81	10.81	18.92	48.65	37	3.84	1.41	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Product Information	38.89	22.22	11.11	11.11	16.67	25.00	14.29	14.29	14.29	28	2.71	1.37	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Service Quality	40.00	20.00	13.33	16.67	10.00	52.11	22.54	16.90	5.63	2.82	71	1.85	1.08	0.002	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Desired Purchase Timeline	50.00	0.00	50.00	0.00	0.00	33.33	16.67	16.67	25.00	8.33	12	2.58	1.39	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Availability of Desirable Product	14.63	2.44	21.95	21.95	39.02	26.03	19.18	31.51	9.59	13.70	73	2.66	1.33	0.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Product Quality	50.00	0.00	0.00	0.00	50.00	80.00	0.00	20.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Communication of Prices	42.86	42.86	0.00	14.29	0.00	44.44	11.11	22.22	22.22	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Product Value	0.00	0.00	0.00	25.00	75.00	18.75	0.00	18.75	25.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Navigation	27.50	17.50	10.00	10.00	15.00	29.03	29.03	19.35	9.68	12.90	31	2.48	1.35	0.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
TOTAL	37.93	20.11	14.66	13.79	13.51	36.34	19.27	22.20	10.00	12.20	410	2.42	1.38	0.272									
w/o technology & fulfillment	27.66	14.36	13.83	21.28	22.87	32.04	18.66	21.13	11.97	16.20	284	2.62	1.45	0.019									
w/o tech, fulfill & no comment	28.89	15.00	13.89	20.00	22.22	32.27	18.44	21.28	12.06	15.96	282	2.61	1.45	0.040									
Component 1: Site Quality	40.73	21.19	14.90	12.25	10.93	40.00	20.85	23.66	8.45	7.04	355	2.22	1.26	0.102									
Component 2: Mission Support	32.28	21.26	12.60	18.11	15.75	17.14	18.57	16.43	24.29	140	3.01	1.51	0.029										

CODDS	1st Survey - Response Scale					2nd Survey - Response Scale					3rd Survey - Response Scale					X ² Probability							
	Strongly Negative		Neutral Opinion		Strongly Positive		Strongly Negative		Neutral Opinion		Strongly Positive		Strongly Negative		Neutral Opinion		Strongly Positive						
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5		1	2	3	4	5		
No Comment	24.39	13.41	14.63	14.63	32.93	3.38	2.03	15.54	46.62	32.43	148	4.03	0.93	0.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Security & Privacy	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	4.00	1.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Fulfillment	0.00	0.00	100.00	0.00	0.00	0.00	0.00	50.00	20.00	50.00	2	2.40	1.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Software Bugs	36.84	34.21	21.05	7.89	21.05	0.00	40.00	0.00	20.00	0.00	5	2.50	1.51	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Site Performance	0.00	75.00	0.00	25.00	0.00	0.00	0.00	0.00	50.00	0.00	2	3.00	1.84	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Task Support	25.00	37.50	18.75	6.25	12.50	0.00	22.22	16.67	16.67	16.67	6	3.67	1.84	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Product Information	38.89	32.32	11.11	11.11	12.50	0.00	22.22	16.67	16.67	16.67	6	3.67	1.84	0.05	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Service Quality	54.84	22.58	12.90	9.68	0.00	33.33	50.00	16.67	0.00	0.00	6	1.83	0.99	0.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Desired Purchase Timeline	0.00	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0	1.83	0.99	0.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Availability of Desirable Product	28.00	34.67	22.67	8.00	6.67	14.29	4.76	19.05	28.57	33.33	21	3.62	1.37	0.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Product Quality	50.00	0.00	0.00	50.00	0.00	0.00	0.00	100.00	0.00	0.00	0	3.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Communication of Prices	0.00	0.00	50.00	50.00	0.00	0.00	0.00	0.00	0.00	0.00	0	5.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Product Value	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	25.00	25.00	4	3.50	1.12	0.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
Navigation	24.00	32.00	20.00	16.00	8.00	0.00	25.00	25.00	25.00	25.00	4	3.50	1.12	0.000	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	N/A
TOTAL	30.64	21.27	17.85	11.78	12.46	7.28	6.31	16.02	39.32	31.07	206	3.81	1.16	0.000									
w/o technology & fulfillment	30.31	25.59	17.32	12.20	14.57	6.60	5.58	15.74	40.10	31.98	197	3.85	1.13	0.000									
w/o tech, fulfill & no comment	33.14	31.40	18.60	11.05	5.81	16.33	16.33	16.33	20.41	30.61	49	3.33	1.47	0.000									
Component 1: Site Quality	33.84	32.32	19.19	11.11	3.54	14.89	17.02	21.28	25.53	21.28	47	3.21	1.36	0.000									
Component 2: Mission Support	25.37	35.82	16.42	14.93	7.46	17.39	17.39	8.70	21.74	34.78	23	3.39	1.53	0.000									

THEMES	1st Survey - Response Scale					2nd Survey - Response Scale					3rd Survey - Response Scale					X ² Probability											
	Strongly Negative		Neutral Opinion		Strongly Positive		Strongly Negative		Neutral Opinion		Strongly Positive		Strongly Negative		Neutral Opinion		Strongly Positive										
	1	2	3	4	5	n	Mean	SD	1	2	3	4	5	n	Mean		SD	1	2	3	4	5	n	Mean	SD	P	
Decision Costs	31.03	18.97	10.34	24.14	15.52	58	2.74	1.50	27.12	25.42	22.03	11.86	13.56	59	2.59	1.36	0.000	27.27	27.27	27.27	27.27	27.27	27.27	27.27	27.27	0.000	
Purchase Justification	27.27	27.27	0.00	18.18	27.27	11	2.91	1.53	28.00	4.00	20.00	24.00	24.00	25	3.12	1.54	0.000	48.97	28.78	16.55	4.83	17.19	39.53	17.19	3.51	1.11	0.035
Technology Support	48.97	28.78	16.55	4.83	17.19	64	2.67	1.51	46.28	19.83	25.62	4.96	3.31	121	1.99	1.11	0.035	32.81	18.75	14.06	2.33	20.93	20.93	17.19	2.53	1.53	0.433
Customer Centricity	16.28	2.33	20.93	5.88	11.76	43	3.65	1.44	37.96	18.52	14.81	10.19	18.52	108	2.53	1.53	0.433	58.82	11.76	11.76	0.00	5.88	11.76	17	2.58	1.34	0.000
Assessment Quality	58.82	11.76	11.76	0.00	0.00	17	2.00	1.42	29.49	17.95	30.77	8.97	12.82	78	2.41	1.34	0.000	100.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.000
Product Delivery	100.00	0.00	0.00	0.00	0.00	2	1.00	0.00	35.29	23.53	11.76	23.53	5.88	17	2.41	1.34	0.000										#DIV/0!
Security														0													
TOTAL	38.82	20.59	14.71	12.94	12.94	340	2.41	1.44	36.52	19.12	22.30	10.05	12.01	408	2.42	1.38	0.754										

THEMES	1st Survey - Response Scale					2nd Survey - Response Scale					3rd Survey - Response Scale					X ² Probability											
	Strongly Negative		Neutral Opinion		Strongly Positive		Strongly Negative		Neutral Opinion		Strongly Positive		Strongly Negative		Neutral Opinion		Strongly Positive										
	1	2	3	4	5	n	Mean	SD	1	2	3	4	5	n	Mean		SD	1	2	3	4	5	n	Mean	SD	P	
Decision Costs	30.23	32.56	16.28	16.28	4.65	43	2.33	1.20	0.00	20.00	20.00	40.00	20.00	10	3.60	1.03	0.000	30.23	32.56	16.28	16.28	4.65	43	2.33	1.20	0.000	
Purchase Justification	0.00	0.00	33.33	33.33	33.33	3	4.00	0.82	0.00	0.00	0.00	0.00	100.00	2	5.00	0.00	N/A	0.00	0.00	33.33	33.33	33.33	3	4.00	0.82	0.000	
Technology Support	33.33	38.10	19.05	9.52	0.00	42	2.05	0.96	28.57	28.57	14.29	28.57	0.00	7	2.43	1.18	0.009	44.68	27.66	14.89	8.51	4.26	47	2.00	1.15	0.000	
Customer Centricity	28.57	33.27	22.08	9.09	0.00	77	2.31	1.17	33.33	33.33	6.67	0.00	26.67	15	2.53	1.59	0.000	28.57	33.27	22.08	9.09	0.00	77	2.31	1.17	0.000	
Assessment Quality	0.00	50.00	50.00	0.00	0.00	2	2.50	0.50	0.00	0.00	0.00	0.00	50.00	2	4.00	1.01	N/A	0.00	50.00	50.00	0.00	0.00	2	2.50	0.50	0.000	
Product Delivery	100.00	0.00	0.00	0.00	0.00	1	1.00	0.00	0.00	0.00	0.00	0.00	50.00	0	N/A	N/A	N/A	0.00	0.00	0.00	0.00	0.00	0	N/A	N/A	N/A	
Security														0													
TOTAL	33.02	32.56	19.07	10.70	4.65	215	2.21	1.34	0.118	0.129	17.24	17.24	20.69	27.59	58	3.24	1.46	0.000									

Appendix 13: Charts, Tables and Figures Summary

Charts

Chart 1-1 : Customer Site Satisfaction (Feb '13 – Jan '15)	Pg 7
Chart 4-1 : Normalized Annual Sales Pattern by Week	Pg 47
Chart 5-1 : Research Groups and Primary Direction of Power	Pg 94
Chart 5-2 : Seasonal Sales Pattern	Pg 95
Chart 5-3 : Satisfaction and Revenue Data by Week	Pg 96
Chart 5-4 : Contour Chart of Satisfaction Ratings by Typology Response	Pg 112
Chart 5-5 : Hierarchical Clustering Dendrogram	Pg 112

Tables

Table 2-1 : Consumer Costs	Pg 16
Table 2-2 : Survey Approaches in the Literature	Pg 23
Table 2-3 : Theoretical Research Summary	Pg 27
Table 3-1 : Research Hypothesis Summary	Pg 42
Table 4-1 : Surveys Used to Gather Customer Feedback	Pg 53
Table 4-2 : Qualitative Analysis Themes and Codes	Pg 54
Table 4-3 : A/B Test Construct	Pg 60
Table 4-4 : Behavioral Measurement Variables	Pg 63
Table 4-5 : Consumer Typology Measures	Pg 64
Table 4-6 : Research Activity by Research Week	Pg 67
Table 5-1 : Clickable Top Nav <i>t</i> -test Significant Results Summary	Pg 71
Table 5-2 : Stepwise Regression Summary – Clickable Top Nav	Pg 71
Table 5-3 : Fly-Out Menu v1 <i>t</i> -test Significant Results Summary	Pg 72
Table 5-4 : Exploratory Factor Analysis – Fly-Out Menu v1	Pg 73
Table 5-5 : Stepwise Regression Summary – Fly-Out Menu v1	Pg 74
Table 5-6 : Fly-Out Menu v3 <i>t</i> -test Significant Results Summary	Pg 75
Table 5-7 : Exploratory Factor Analysis – Fly-Out Menu v3	Pg 76
Table 5-8 : Stepwise Regression Summary – Fly-Out Menu v3	Pg 77
Table 5-9 : Colorization <i>t</i> -test Significant Results Summary	Pg 79
Table 5-10 : Exploratory Factor Analysis – Colorization	Pg 79
Table 5-11 : Stepwise Regression Summary – Fly-Out Menu v3	Pg 80
Table 5-12 : Fly-Out Menu v2 <i>t</i> -test Significant Results Summary	Pg 81
Table 5-13 : Stepwise Regression Summary – Fly-Out Menu v2	Pg 82
Table 5-14 : Survey Data Collection and Site Change Implementation by Research Week	Pg 83
Table 5-15 : On-Site Survey Overview	Pg 84
Table 5-16 : Exploratory Factor Analysis – Qualitative Feedback Codes	Pg 85
Table 5-17 : 1 st Survey Summary – Decision Cost Measures	Pg 86
Table 5-18 : 1 st Survey and 2 nd Survey X^2 Analysis – Decision Costs	Pg 86
Table 5-19 : 1 st , 2 nd and 3 rd Survey X^2 Analysis – Decision Costs	Pg 87
Table 5-20 : 1 st and 2 nd Survey X^2 Analysis – Purchase Justification	Pg 89
Table 5-21 : During and Post-Visit Survey Overview	Pg 91
Table 5-22 : 3 rd Survey and Post-Visit Survey X^2 Code Analysis – Time to Experience	Pg 92
Table 5-23 : 3 rd Survey and Post-Visit Survey X^2 Theme Analysis – Time to Experience	Pg 92
Table 5-24 : Combined On-Site Survey and Sales Revenue	Pg 95
Table 5-25 : Cross-Correlation Summary	Pg 96
Table 5-26 : 1 st Survey and 3 rd Survey X^2 Analysis – Product Information	Pg 98
Table 5-27 : 1 st Survey and 2 nd Survey X^2 Analysis – Product Information	Pg 99
Table 5-28 : Product Description Experiment <i>t</i> -test Significant Results	Pg 101
Table 5-29 : Stepwise Regression Summary – Product Description	Pg 101
Table 5-30 : Correlations – Product Description Experiment	Pg 102
Table 5-31 : Exploratory Factor Analysis – Product Description	Pg 102
Table 5-32 : Video Experiment <i>t</i> -test Significant Results Summary	Pg 103
Table 5-33 : Correlations – Video Experiment	Pg 103
Table 5-34 : Stepwise Regression Summary – Video Experiment	Pg 104

Tables (continued)

Table 5-35 : Exploratory Factor Analysis – Video Experiment	Pg 105
Table 5-36 : Size Display <i>t</i> -test Significant Results Summary	Pg 107
Table 5-37 : Stepwise Regression Summary – Size Display Experiment	Pg 108
Table 5-38 : Exploratory Factor Analysis – Size Display Experiment	Pg 108
Table 5-39 : Customer Typology Assignment by Survey Response Option	Pg 110
Table 5-40 : On-site Customer Typology Responses and X^2 Difference Probabilities	Pg 111
Table 5-41 : Post-Visit Customer Typology Responses and X^2 Difference Probabilities	Pg 109
Table 5-42 : <i>k</i> -means clustering results summary	Pg 113
Table 5-43 : Survey Response Clusters	Pg 114
Table 5-44 : Hypothesis Results Summary	Pg 115
Table 7-45 : Summary of Results: Tested Academic Theories	Pg 136

Figures

Figure 1-1 : Action Research Process in Practice	Pg 9
Figure 3-1 : Conceptual Intervention Model	Pg 35
Figure 3-2 : Conceptual Research Model	Pg 37
Figure 3-3 : Research Conceptualization Using the S-O-R Framework	Pg 37
Figure 3-4 : Conceptual Hypothesis Model	Pg 38
Figure 4-1 : Action Research Methodology	Pg 43
Figure 4-2 : Weekly Action Research Cycles	Pg 48
Figure 4-3 : Mixed Method Analysis Type Usage	Pg 51
Figure 4-4 : MMAR Research Approach Overview	Pg 52
Figure 6-1 : Meta-Learning Cycles and Touch-Points	Pg 53

Appendix 14: Ethical Approval

Dear Mark Crystal

I am pleased to inform you that the DBA Ethics Committee has approved your application for ethical approval for your study. Details and conditions of the approval can be found below:

Committee Name: DBA Ethics Committee

Title of Study: Linking consumer satisfaction to consumer behavior and retailer revenues: an empirical analysis of e-commerce performance influencers

Student Investigator: Mark Crystal

School/Institute: School of Management

Approval Date: 19 June 2015

The application was APPROVED subject to the following conditions:

- The researchers must obtain ethical approval from a local research ethics committee if this is an international study
- University of Liverpool approval is subject to compliance with all relevant national legislative requirements if this is an international study.
- All serious adverse events must be reported to the Sub-Committee within 24 hours of their occurrence, via the Research Integrity and Governance Officer (ethics@liv.ac.uk)
- If it is proposed to make an amendment to the research, you should notify the Committee of the amendment.

This approval applies to the duration of the research. If it is proposed to extend the duration of the study as specified in the application form, the Committee should be notified.

Kind regards

DBA Ethics Committee

University of Liverpool Management School in Partnership with Laureate Online Education