Examining the effectiveness of activation techniques on consumer behavior in temporary loyalty programs

Bies, Suzanne

DOI:
10.26116/kzfe-y651

Publication date:
2022

Document Version
Publisher's PDF, also known as Version of record

CentER

Examining the Effectiveness of Activation Techniques on Consumer Behavior in Temporary Loyalty Programs

SUZANNE M.T.A. BIES

Tilburg University

Examining the Effectiveness of Activation Techniques on Consumer Behavior in Temporary Loyalty Programs

Proefschrift ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. W.B.H.J. van de Donk, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in de Aula van de Universiteit op vrijdag 1 juli 2022 om 10.00 uur door

Suzanne Maria Theresia Apolonia Bies

geboren op 24 augustus 1991 te Eindhoven

Promotores:	prof. dr. ir. Bart J.J.A.M. Bronnenberg prof. dr. Els Gijsbrechts
Leden promotiecommissie:	prof. dr. Tammo H.A. Bijmolt prof. dr. Els Breugelmans prof. dr. Marnik G. Dekimpe dr. Sarah E.C. Gelper

Made possible by BrandLoyalty.

Acknowledgements

I am very pleased to be able to present you with this dissertation, which undoubtedly would not have been completed without the help, guidance and support of several important people. As such, it is only fitting I thank them and attempt to do them justice in this next section.

First and foremost, I would like to thank my supervisors Bart and Els, without who's guidance this dissertation would quite simply not exist. I realize just how lucky I have been to have been supervised by not one, but two such outstanding scholars, who on top of everything also complement each other so well and work together with such respect. I have learned a lot from you both, but you also individually had a great influence on my development as a researcher. Bart, your ability to come at a problem with a different angle to consider, has been inspiring and a great contribution during meetings, often providing renewed momentum. You are not only a very skilled researcher, but also a very kind person. I want to thank you for making time for me, even though you were extremely busy or even an ocean and several time zones away. You also always kept a close eye on me and made sure to prevent me from becoming too hard on myself, and passed your passion for research onto me. For all this and more, thank you! Els, you helped me in more ways than you know. You are not only a truly impressive and extraordinarily modest researcher, but also a really wonderful supervisor! You always make time, no matter how busy you are (your extremely fast replies still astound me to this day), and I could always come to you when I got stuck. The level of efficiency and structure in your work is something that I still aspire to and that has helped me tremendously. I would also like to thank you for always being so patient, honest, and overall committed to making me a better researcher. I owe you so much, I cannot thank you enough!

Els, Marnik, Sarah and Tammo, I am very honored that you were willing to be a part of my doctoral committee. Thank you for all your comments and suggestions that helped to improve this dissertation, and for taking time out of your busy schedules. I would like to give special thanks to Marnik. Throughout my years at Tilburg, you have always given very valuable feedback whenever I presented earlier versions of my work at PhD Camps, for which I am very grateful.

Next, I would like to thank BrandLoyalty for supporting my PhD financially, and providing me with such unique data. I would especially like to thank Arnoud, Christian, Koen, Lina, Ruud and Steven for all their help throughout the years. Thank you for answering all my questions and pointing me in the right direction. Koen, your infectious enthusiasm for loyalty programs inspires me to this day, and I cannot wait to see what you come up with next.

I would like to thank the entire marketing group at Tilburg for their continued support over the years. Thank you for all your feedback during seminars and for the nice conversations during lunches or Friday afternoon drinks. Special thanks to Scarlett, Heidi and Nancy, you are the backbone of the department. Thank you for all your administrative support but also for all our nice conversations. Heidi, thank you for the book! Ana, Anouk, Astrid, Bernadette, Constant, Elke, Esther, Georgi, Joep, Nazli, and Yan, thank you for all our interesting discussions and for our annual Sinterklaas festivities.

I am also very grateful for the amazing support I have received from my new colleagues in Groningen. Thank you for making me feel so welcome and at home so quickly. Starting a new job in these circumstances was quite challenging, but your instant support has been heartwarming and unvaluable!

My dissertation would not have been completed without the support of a wonderful group of friends. Georgi and Elke, thank you for not only being PhD colleagues, but for your friendship and encouragement! Vilma, thank you for agreeing to be my paranymph, and for becoming part of my family. Phil and Kim, you are my oldest friends for a reason, thank you

for always being so supportive! Christian, Dominique, Erik, Harrie, Iris, Laura, Marie, Nicoline, Peter, and Tijm, you are not only Nick's friends, but very quickly also became mine. Thank you for making me feel so welcome in your group and always being supportive and offering some much-needed relaxation.

I could also not have finished this dissertation without the amazing support of my family. Mom and dad, thank you for believing in me so fully and doing everything within your power to help me achieve my goals. I could not have done this without you. Though I am my own (very stubborn) person, you have shaped me into the woman I am today. Know that I love you very much and will always be grateful! Anton and Frank, thank you for always being protective of your little sister! I would also like to thank you both and Mariëtte, for always being genuinely interested in what I was doing and asking the right questions. Frank and Monique, thank you for welcoming me so wholeheartedly into your family. I realize how lucky I am with such wonderful in-laws. Thank you for your support and all your help with James, as I attempted to finish my PhD. Joyce and Mardien, thank you for always being there to offer some time away from work and for helping me relax!

Finally, to the two most important men in my life. James, you are the light of my life! Seeing your smile and happy demeanor instantly fills me with pride and joy. I love you more than I can express. Last, but in no way least, Nick, my love. There are simply not enough words in the English language, or any other for that matter, that can express the gratitude I feel towards you. You are without fault the kindest and most caring person I know. Your support, both workwise as well as personally, has gone above and beyond. Thank you for always making me smile. I proudly dedicate this dissertation to you. I owe you everything and I love you more than I can say!

Table of Contents

1. General Introduction

Loyalty programs (LPs) are commonly used in the retail-, travel-, and financial services industry. Their goal is to increase customer loyalty through offering additional (often monetary) benefits to consumers. Generally speaking, we can make a distinction between two types of programs. First, there are permanent programs which run without end, where consumers continuously save points based on their spending which they can redeem at any point in time to obtain a reward. Second, there are the temporary programs, which typically only run for several weeks, after which rewards can no longer be redeemed and saved points lose their value. While these temporary programs have received far less attention in prior literature than their permanent counterparts, they often result in larger increases in consumer expenditures (Bijmolt & Verhoef, 2017). They are also becoming increasingly popular (Bombaij, Gelper, & Dekimpe, 2022). Due to their brief duration and the required (logistical) expertise in running them, temporary programs are often managed by a third party, the *program operator*, who is, among other things, responsible for the stock and quality of rewards, and promotion materials.

When it comes to ensuring the success of LPs, however, a key component is ‘customer engagement’, i.e., the extent to which consumers participate in LPs and actively save for and collect rewards (and in doing so alter their purchasing behavior). However, extant studies show that consumers’ attention to, and interest in, the LP decreases over time (Dorotic, Verhoef, Fok, & Bijmolt, 2014) – as evidenced by declining (purchase and) redemption activity, and failure to redeem collected stamps or points even when the threshold for redemption is exceeded (Drèze & Hoch, 1998; Kwong, Soman, & Ho, 2011; Stourm, Bradlow, & Fader, 2015). This loss of engagement is not limited to the traditional permanent loyalty programs: an industry report from GfK (2015) finds that less than half of the participants that start saving in temporary programs actually redeem a reward. It follows that due to the limited time span of TLPs, it is even more crucial to ensure effective means of increasing customer engagement for these programs, as there is only a small window of opportunity to effectively activate consumers. Therefore, awareness needs to be built immediately and program salience needs to be properly maintained, before the program comes to an end and retailers, program operators and consumers all lose their chance to benefit.

There are different activation techniques that retailers and program operators can use to engage customers in the program, that can help increase consumers' purchasing activities and redemption behavior. Offline they can for example use direct mailing, leaflets, tv advertising, and in-store activation, whereas online they can utilize social media advertising and email communication. In addition, the existence of mobile apps has paved the road for program operators to add mobile push messaging to their marketing instruments.

Though it is crucial for companies to determine how to maintain LP salience and increase consumer participation, previous literature on these techniques is scarce. While there has been research on some of the activation techniques in the context of LPs (e.g., direct (e)-mailing), there are still many forms that remain unexplored in an LP context or otherwise (e.g., in-store activation, mobile push messaging, etc.). In addition, it is unclear when (e.g., at the start or end of the program), or for who (what type of consumers or stores), such activation techniques work better, and on which metric (e.g., purchasing activities or redemption behavior) they might do so. To illustrate, program activation techniques might not stimulate more spending for heavier buyers at the chain who can more easily achieve their saving goals to begin with, but it might be critical to remind light buyers and keep them motivated. As such, to date, no clear-cut strategy has been developed yet regarding how to maintain program salience. This has led to prior calls from the literature to offer new insights in this domain (Ailawadi, Beauchamp, Donthu, Gauri, & Shankar, 2009; Dorotic et al., 2014). This dissertation entitled "Examining the Effectiveness of Activation Techniques on Consumer Behavior in Temporary Loyalty Programs" aims to fill this gap.

In Chapter 2 – "How Push Messaging Impacts Consumer Spending and Reward Redemption in Store-Loyalty Programs" – we examine the causal effect of in-app mobile push messages on consumer participation and reward collection in temporary loyalty programs. Existing LP studies have found little impact of traditional forms of communication, such as direct mail or e-mail (Lewis, 2004; Dorotic, 2010; Dorotic et al., 2014), and while the advancement of mobile apps provides a promising new way to directly reach the consumer, apps too often show a rapid drop in usage (van Heerde, Dinner, & Neslin, 2019). To date, however, it has not been empirically determined whether (in-app) mobile push messaging can help increase continued customer engagement throughout the program. Therefore, this study uses a panel dataset from a large-scale field experiment with a randomized control group, covering consumer spending before and during the program. With this dataset, we estimate heterogeneous treatment effects of push messaging on a variety of program engagement measures. In addition, we determine the impact of saving dynamics on the effect of push messaging and look into how the timing of sending messages influences their effectiveness.

In Chapter 3 – "Drivers and Consequences of In-Store Promotion-Execution Quality: An Analysis for Temporary Loyalty Programs" – we focus on the effect of the quality of execution of in-store support plans on sales, including factors such as reward replenishment, training of store personnel, and management of in-store signage and displays. While industry reports have noted the importance of and lack of proper retail execution (Gomez & Sides, 2015; POPAI, 2015), little prior empirical research exists that determines the extent to which deviations from planned in-store support take place, nor has it been determined what the sales impact of these deviations is. To this end, this study uses a unique dataset that contains weekly information on in-store support factors and execution scores, from stores running a temporary loyalty program. We combine the in-store execution data with sales data to create a panel dataset that allows us to empirically determine the effect of each factor. Using pre- and post-program sales, we separate the impact of deviations from planned execution from the effect of the campaign itself and other (temporal or cross-sectional) factors.

In Chapter 4 – "Determining the Difference in Effectiveness of Different Message Types in Store-Loyalty Programs" – we explore whether differences exist in the effectiveness of different types of mobile push messages. In particular, while in Chapter 2 we focus on the causal impact of mobile push messages, it is unclear if this effect is the same for different types of messages. To explore this in more detail, we use panel data covering the same program studied in Chapter 2 and focus on two distinct message types commonly used in loyalty programs – engagement- vs. promotion-oriented messages. In addition, we focus on the extent to which these differences depend on consumer types, and how they vary for different outcome variables. We find that promotion messages are more effective than engagement messages in general, but that the increase in impact is largest for spending and for heavier buyers at the chain.

In Chapter 5, we summarize the main findings of the dissertation and reflect on the key implications for retailers and program operators. In addition, we discuss some of the limitations and explore possible avenues for future research.

2. How Push Messaging Impacts Consumer Spending and Reward Redemption in Store-Loyalty Programs

2.1 Introduction

Loyalty programs (LPs) are ubiquitous and are in widespread use in the retail-, travel-, and financial services industry, to name a few. These include long-term/permanent LPs, in which points/miles/stamps are saved continuously to be converted into rewards (money or goods), as well as shorter term/temporary LPs that operate for several weeks, after which saved points or stamps lose their value. In 2015, 52% of consumers in the Netherlands participated in such short-term LPs (GfK, 2015). Furthermore, between 2006 and 2015, the use of these LPs by retailers has more than quadrupled in Asia and North America, and even increased six-fold in Europe. The 'market' for such LPs was estimated at about two billion Euros worth of redemption value in 2012 and has been growing since.[1]

In many LPs, consumers receive stamps at the checkout register based on their purchasing amount (e.g., one stamp per 10 euro spent). Consumers can then save these stamps in order to redeem rewards for a pre-specified number at a later point in time, but before the program ends. For short-term LPs, popular reward categories are crystal-, cooking-, and cutting ware, and the rewards can be obtained at a discount or are free. A key component of the success of these programs is 'customer engagement', i.e., whether the program can activate consumers to save stamps and collect rewards which, in turn, encourages future purchases. Yet, extant studies show that after initial excitement, the salience of, and interest in, LPs typically wears off – as evidenced by declining (purchase and) redemption activity (Dorotic, Verhoef, Fok, & Bijmolt, 2014). Especially light-to-moderate customers of the retailer are found to accumulate points far beyond the redemption thresholds and fail to collect rewards (Kwong, Soman, & Ho, 2011; Stourm, Bradlow, &

[1] Based on an internal report of BrandLoyalty, conducted by McKinsey in 2012.

Fader, 2015; Lal & Bell, 2003; Liu, 2007). Such loss of customer engagement is observed in permanent loyalty programs (PLPs), but also in temporary ones (TLPs), where redemption rates are additionally found to be low (GfK, 2015; Drèze & Hoch, 1998). The critical question is then: How can companies maintain LP salience and consumer participation?

Extant LP studies find little impact of traditional (direct mail or e-mail) communication on program sales or redemption (Lewis, 2004; Dorotic, 2010; Dorotic, et al., 2014). Mobile apps have been advanced as a promising way to increase consumer engagement, in general (van Heerde, Dinner, & Neslin, 2019) and in a LP context in particular (Reinartz & Linzbach, 2018). At the same time, app adoption is not enough: continued app use (stickiness) is important (Wang, Krishnamurthi, & Malthouse, 2018), yet apps, too, "are notorious for a rapid drop in usage" (van Heerde et al., 2019). These authors suggest that repeated push notifications could play a role here, but do not document this empirically (in general or in the context of LPs).

Previous studies on mobile marketing indicate that mobile coupons and advertising text messages can be more effective than direct mail or email (Reichhart, Pescher & Spann, 2013), and can influence consumers' immediate and planned behavior (Fang, Gu, Luo, & Xu, 2015). However, they also show that the impact of mobile push is context dependent (Bart, Stephen, & Sarvary, 2014) and can, in some instances, be negative (Tsang, Ho, & Liang, 2004). Moreover, also with mobile messages, incrementality is an issue – for instance, with mobile coupons, the question is whether redemption will lead to *additional* sales (Fong, Fang, & Luo, 2015). These issues become particularly relevant for mobile-push reminders in a loyalty-program context, because the repeated messages could come with decreasing returns, and the reminders may trigger consumers to redeem piled-up stamps rather than collect additional stamps (spend more). Therefore, whether and to what extent push notifications increase redemption and sales in the context of a LP (app) remains an unsettled issue.

Our paper intends to fill this gap. Our core objective is to document the impact of mobile push (within apps) on purchases and redemption in a TLP. In so doing, we not only evaluate the effect of individual messages but also consider the impact of the entire push program, including the number and timing of messages.

Additionally, we shed light on the relative impact of push notifications on spending vs. redemption. This is particularly relevant in a TLP setting for two reasons. First, if push messages enhance redemption, this creates a caveat for retailers. On the one hand, redemption may be beneficial because redeemers, feeling more rewarded, are more likely to develop a positive attitude towards the retailer, and to subsequently engage in extra stamp collection (and, thus, spending, see also Dorotic et al., 2014, p. 351). On the other hand, retailers typically incur a cost on each redeemed reward. For retailers, mobile messaging is successful only if it enhances customer engagement in the program while maintaining a sound balance between stamp collection and redemption. Second, short-term LPs are often planned and operated by a third party, the 'program operator,' for profit. The business model of the program operator (and the supplier of reward products) is to make money on each reward redeemed by consumers. So, for these stakeholders, redemption is the primary – if not the only – performance metric.

To determine the causal impact of push messaging, we use a unique panel data set covering a large controlled 18-week field experiment in the Fall of 2016. The data involve 46,504 program-participating consumers shopping at 11,895 stores. Our data track consumers' expenditure and redemption during the short-term LP. In the field experiment, a random subset of panelists (the control group) receives no push messaging after week 3 of the program. The program operator also selects small independent random samples of households who do not receive particular single messages within the program. All other panelists receive the same push messages each week, i.e., without targeting. The program operator varies the number of push messages across weeks from 1 to 3 messages per week.

The random assignment in the communication treatment and the panel structure of our data allow us to use a regression framework that avoids some of the endogeneity concerns in LP- and mobile communication research. For instance, rather than making a contrast between self-selected program participants and non-participants, we compare similar participants exposed to different push messaging regimes assigned at random. In addition, with ample data at the individual level to account for household fixed effects, we can address any unintended selection in the communication experiment (e.g., messages may fail for technological reasons related to household income). For this reason, and being conservative, our preferred specification features two-way (household and time) fixed effects.

Our results show that push messaging strongly enhances the *number of stamps redeemed*, especially among heavy customers at the chain (whose stamp balance quickly builds up within the app). Importantly, though, mobile push also has a strong positive impact on consumer *spending* in the course of the program, and this spending lift, too, increases with high levels of pre-program expenditures. Overall, relative to a random control condition of app users who do not receive push messages after the third week of the program, consumers treated with the push plan redeem about twice as many stamps, and spend about 14% more on average during the program. For the retailer, this results in an 11% margin increase on treated customers throughout the program period – a sizable gain. Hence, unlike traditional communication devices, mobile push messaging clearly enhances program engagement among customers who install the app, and improves performance for the different stakeholders (i.e., the retailer, the program operator, and the manufacturer of rewards). Using our estimates as a basis for simulating the dynamic impact of additional push messages, we uncover managerial guidelines for the targeting and timing of mobile push in short-term LPs.

Messages targeted at heavy customers are particularly effective at increasing both stamps saving and redemption. The timing effect is more ambivalent: back-end loading (i.e., scheduling push late rather than early in the program) increases redemption, whereas messages sent in mid-program weeks have the largest impact on expenditure and stamp collection. Therefore, stamps saving and redemption goals should be carefully balanced when scheduling the push plan.

The remainder of this paper is organized as follows. Section 2 discusses the literature and Section 3 provides the conceptual framework. Section 4 presents the panel data set that tracks program participation and reward collection. Section 5 contains the empirical model, while Section 6 gives the estimation results. Section 7 discusses the implications, and Section 8 concludes.

2.2 Background Literature

Our paper bridges two streams of literature: that on LPs, and that on mobile marketing. Table 2.1 summarizes key empirical studies, which we discuss below.

2.2.1 Loyalty Programs

A large body of literature has looked into the effectiveness of LPs (see, e.g., Bijmolt, Dorotic, & Verhoef, 2011; Breugelmans et al., 2015; and Kim, Steinhoff, & Palmatier, 2021 for reviews and discussion). Table 2.1, Panel A gives an overview of the studies most relevant for our purpose. These field studies shed light on the performance impact of LPs, in consumer packaged goods (CPG) as well as Non-CPG settings. While they predominantly focus on permanent LPs, TLPs have received attention as well (Drèze & Hoch, 1998; Lal & Bell, 2003; Minnema, Bijmolt, & Non, 2017; Taylor & Neslin, 2005; Bombaij & Dekimpe, 2020). Overall, these studies indicate that LPs can enhance sales, in a dynamic fashion, and differently so depending on consumers' prior purchase rate at the chain. By awarding stamps or points in proportion to consumers' purchases amounts, LPs – and short-term LPs in particular – may stimulate consumers to spend more at the retailer (e.g., Bijmolt et al., 2011; Breugelmans et al., 2015; Dorotic et al., 2014; Meyer-Waarden, 2007; Lewis, 2004; Taylor & Neslin, 2005; Chauduri, Voorhees, & Beck, 2009). These spending effects evolve endogenously. When consumers' stamps balance is still low, the need to accumulate enough stamps to obtain a future reward may motivate them to spend more at the retailer (the 'points-pressure' mechanism, Bijmolt & Verhoef, 2017). Conversely, consumers may also accelerate their purchases as they move closer to their goal of obtaining the reward (the 'goal-gradient hypothesis,' Kivetz, Urminsky, & Zheng, 2006).[2] In all, though, the observed spending increases are often quite modest, especially among consumers with already high pre-program purchase levels at the chain (Liu, 2007; Bijmolt et al., 2011).

Moreover, consumers' attention to, and interest in, the LP is found to decline over time (Dorotic et al., 2014). As the program progresses, consumers – in particular, light-to-moderate buyers – often fail to redeem the collected points (Drèze & Hoch, 1998; Kwong et al., 2011; Stourm et al., 2015), even if the threshold for redemption is exceeded. In some instances, between 50% and 70% of potential LP rewards are never collected (Bijmolt et al., 2011). This loss of engagement is not limited to PLPs: a recent industry report finds that less than half of the participants in TLPs actually redeem a reward (GfK, 2015). To the extent that receiving a reward may revive excitement about the LP (Dorotic et al., 2014) and trigger consumers to step up their expenditures (because of behavioral reinforcement or positive feelings toward the retailer, 'the rewarded-behavior mechanism,' Blattberg, Kim, & Neslin,

[2] These *positive* points-pressure and goal-gradient effects on spending materialize if consumers have not reached the reward threshold yet (that is, when their 'distance' is still positive, and their 'balance' below the minimum number of stamps to collect a reward). In that region, points-pressure increases spending at lower levels of balance (higher levels of distance); and the goal-gradient mechanism lifts sales at higher levels of balance (lower levels of distance). We contend that, as consumers' stamps balance exceeds the reward threshold, (i) the goal-gradient effect is no longer operative (the goal is reached), and (ii) following the points-pressure mechanism, consumers become less inclined to step up their spending as their number of already accumulated stamps grows farther above the redemption threshold.

2008; Kopalle, Sun, Neslin, Sun, & Swaminathan, 2012), this failure to redeem may be problematic. Continued program participation thus becomes a critical issue.

Even so, extant studies shed little light on how to maintain or increase program salience and engagement. Surprisingly few papers have investigated the impact of retailer communication in the course of the program. Notable exceptions are Lewis (2004), Dorotic (2010) and Dorotic et al. (2014), who study the impact of emails, and Danaher, Laszlo and Danaher (2020), who include both emails and direct mails.[3] Overall, these studies suggest that retailer communication has only a limited effect on LP sales or redemption.[4] We contribute to this literature by studying the impact of new and possibly more effective forms of direct to consumer communication, in particular (in app) mobile messaging.

2.2.2 Mobile Marketing

A second stream of literature to which we aim to contribute is mobile marketing, e.g., the use of mobile apps, and mobile communication (see e.g., Andrews, Goehring, Hui, Pancras, & Thornswood, 2016; Grewal, Bart, Spann, & Zubcsek, 2016; Lamberton & Stephen, 2016; and Shankar et al., 2016 for reviews and discussion). Table 2.1, Panel B summarizes key field studies on the performance impact of mobile marketing.

Mobile Apps. Mobile applications ('apps') have been advanced as a potential way to engage consumers and stimulate spending, in general (Grewal et al., 2016, Shankar et al., 2016, van Heerde, et al., 2019), and in the context of LPs: "Mobile becomes a highly viable channel for customized application-based […] reward programs" (Reinartz & Linzbach,

[3] Wang et al. (2018) include email and direct mail as controls but do not report on the results.
[4] An exception is Danaher et al. (2020), who find that regularly sending direct mails to inform LP members about their point statements encourages program participation. However, their objective and setting is very different from ours. First, they consider a long-term coalition program – with no points expiration, and very little incentive for consumers to step up their purchases at one retailer. Second, every LP member in their setting receives such statements on a regular basis (i.e., there is no 'control group', and no clear time variation in message reception). Though this does not constitute a problem for identifying which types of rewards should be promoted to consumers in different states of LP engagement (their objective), it makes it hard to infer the causal impact of messages. Third, even these authors find that sending emails has little or no influence on program engagement, leaving the question whether other forms of communication (like in-app push) would be effective.

2018, p. 316). Several studies empirically report on the impact of retailer apps in general, and show that they can increase online and offline sales (e.g., Liu et al., 2019; Narang & Shankar, 2019; van Heerde et al., 2019; Zubcsek, Katona, & Sarvary, 2017). Still, Gu and Kannan (2021) find that in a highly competitive setting (i.c. hotels), app adoption can reduce spending, especially among consumers with low app engagement. In a loyalty-program setting, Wang et al. (2018) find that linking a retailer's LP to a mobile app may increase program sales and redemption. However, Kim, Wang and Malthouse (2015) show that when consumers discontinue using the app, their spending levels decrease again. Sustained sales lift thus only comes with continued app usage ('sticky apps', Kim et al., 2015), yet consumers are found to quickly lose interest in apps over time (Appel, Libai, Muller, & Shachar, 2020; van Heerde et al., 2019). Repeated notifications sent from within the app might act as reminders and maintain interest, but this is not documented yet empirically.

Mobile communication. Besides apps, mobile communication, that is, "content sent by or on behalf of advertisers and marketers to a mobile device at a time other than when the subscriber requests it" (Unni & Harmon, 2007, p.30) has also received extensive attention in the academic literature (see Table 2.1, Panel B). This form of direct communication has gained appeal with the growth of smartphone use (Grewal et al., 2016). Extant studies on mobile communication have mainly focused on 'stand-alone' mobile coupons (Danaher, Smith, Ranasinghe, & Danaher, 2015; Dubé, Fang, Fong, & Luo, 2017; Park, Park, & Schweidel, 2018; Zubcsek et al., 2017; Mills & Zamudio, 2018; Fong et al., 2015) or mobile ads (Andrews et al., 2016; Bart et al., 2014; Luo, Andrews, Fang, & Phang, 2014; Fang et al., 2015; Osinga, Zevenbergen, & van Zuijlen, 2019; Li, Luo, Zhang, & Wang, 2017), mostly delivered as text messages. Apart from reaching consumers in real-time and at low cost, these mobile (push) messages are found to grab attention, trigger recall, provide contextually-relevant content, and foster intimacy/engagement (Bellman, Potter, Treleaven-Hassard, Robinson, & Varan, 2011; Grewal et al., 2016; Bart et al., 2014; Shankar et al., 2016; Reinartz & Linzbach, 2018) – thereby increasing sales of the promoted items in the immediate period and beyond (Fang et al., 2015). However, push messaging also has downsides. (Too many) messages may lose effect and even lead to irritation (see, e.g., Tsang et al., 2004). Moreover, being sent to personal devices, mobile push messages can be perceived as intrusive and cause privacy concerns (Andrews et al., 2016), which, in turn, may negatively affect purchases behavior (Phelps, D'Souza, & Nowak, 2001). Finally, for push messages that offer a benefit to the consumer, incrementality can be an issue – the question being whether the additional sales make up for the cost of granting the benefit (Fong et al., 2015).

While extant papers (as shown in Table 2.1, Panel B) have documented the impact of (i) 'stand-alone' mobile messages (coupons or ads) (ii) for individual products (iii) typically distributed through SMS, no study to date has studied the impact of (i) push-notification schedules, (ii) in the context of LPs, (iii) sent within a mobile application.[5] Yet, the effect of the latter messages may be quite different. On the one hand, push messages sent within a mobile application may receive more attention and suffer less from privacy concerns, because consumers opted in for the app. Moreover, LP-related notifications are typically not limited to a particular promoted item but pertain to the entire spending amount at the retailer, and, to the extent that these messages alert consumers to already 'earned' rewards, they may generate a more positive response. On the other hand, the use of *repeated* messages within the LP may lead to a loss in effectiveness, and it remains unclear whether the incremental sales for the retailer – if any – make up for the cost of increased redemption.

To summarize this section, extant studies on LPs highlight the need to enhance and maintain program engagement (salience and redemption), but fail to identify forms of

[5] Li et al. (2017) study in-app push messages, but focus on individual messages outside of a LP context.

communication that clearly serve this purpose. In turn, the literature on mobile marketing emphasizes the potential of mobile apps and messages to trigger attention and foster sales, but does not document to what extent these effects apply to in-app push-notification schedules for LPs. Our paper intends to fill this gap.

Table 2.1: Summary of relevant literature

Panel A: Loyalty programs												
Study	LP		App	Communication		Prior customer purchase rate	Outcome		Dynamics	Setting		Findings
	TLP	PLP		Type	#		Sales related	Redemption		Real Life	CPG	
Bolton, Kannan & Bramlett (2000)	×	✓	×	×	×	✓	✓[v]	×	✓	✓	×	LP members overlook or discount negative evaluations of company vis-a-vis competition in their repatronage.
Bombaij & Dekimpe (2020)	×	✓	×	×	×	×	✓	×	×	✓	✓	Identify how program design, retailer and country characteristics moderate PLP sales
Bombaij, Gelper & Dekimpe (2022)	✓	×	×	×	×	×	×	✓	×	✓	✓	Identify how program design, retailer and country characteristics moderate TLP redemptions. Point to difference between PLPs and TLPs.
Danaher, Laszlo & Danaher (2020)	×	✓	×	DM, email	×	✓	✓	✓	✓	✓	✓	LP members transition between 'activity' states. Promoting specific types of rewards at different states may improve LP effects. DM generally more effective than email
Dorotic, Verhoef, Fok & Bijmolt (2014)	×	✓	×	DM	×	(✓)	✓	✓	✓	✓	✓	Increasing redemption enhances salience and purchases. Mailings have small impact at best. Impact depends on customers' prior purchases.
Dreze and Hoch (1998)	✓	×	×	×	×	×	✓	✓	✓	✓	✓	TLP increases sales. Low redemption makes it profitable.
Dreze & Nunes (2011)	×	✓	×	×	×	×	✓	×	✓	✓	×	Successive successes (status attainment) enhance effort (extra sales).
Chauduri, Voorhees & Beck (2019)	×	✓	×	×	×	×	✓	×	✓	✓	✓	LPs significantly increase sales and (with a delay) profit in the short and the long run.
Kim, Wang & Malthouse (2015)	×	✓	✓	×	×	✓	✓	×	✓	✓	×	App adoption and continued use increase future spending. When customers discontinue using the app, their spending levels decrease
Kopalle, Sun, Neslin, Sun & Swaminathan al. (2012)	×	✓	×	×	×	✓	✓	✓	✓	✓	×	Both the frequency reward and customer tier components of LPs generate incremental sales.
Lal & Bell (2003)	✓	×	×	×	×	✓	✓	✓	✓	✓	✓	Programs are profitable because incremental sales to casual shoppers offset subsidies to loyal customers.
Leenheer, van Heerde, Bijmolt & Smidts (2007)	×	✓	×	×	×	×	✓	×	×	✓	✓	Small positive effect of PLP membership on share-of-wallet. Correcting for self-selection = important.
Lewis (2004)	(✓)*	(✓)*	×	email**	×	✓	✓	×	✓	✓	✓	Accounts for long term effects. LP successful for increasing annual sales.
Liu (2007)	×	✓	×	×	×	✓	✓	✓	✓	✓	✓	Heavy customers before the LP redeem more but do not buy more, light customers buy more and become more loyal.
Liu & Yang (2009)	×	✓	×	×	×	✓	✓	×	×	✓	×	With competing LPs. larger firms benefit more from their LPs. In non-expandable cats, LP effect erodes with # of programs.
Mägi (2003)	×	✓	×	×	×	×	✓	×	×	✓	✓	Impact of loyalty cards on consumer share of wallet is mixed. Competing cards may cancel each other out.
Meyer-Waarden (2007)	×	✓	×	×	×	✓	✓	×	✓	✓	✓	LPs increase customer lifetimes. The effect decreases with multiple card memberships of nearby stores and is higher for high-share customers.

Meyer-Waarden & Benavent (2009)	✗	✓	✗	✗	✗	✓	✓	✗	✓	✓	✓	Heavier (lighter) buyers adopt loyalty cards earlier (later). Late card adopters do not buy more at the chain in the longer run.
Minnema, Bijmolt & Non (2017)	✓	✗	✗	✗	✗	✓	✓	✗	✓	✓	✓	IRP results in incremental shopping trips. Bonus premiums enhance the impact of price cuts on brand sales. IRPs and bonus premium are especially (but not exclusively) effective for households that collect the premiums.
Rossi (2018)	✗	✓	✗	✗	✗	✓	✓	✗	✓	✓	✗	LP effects are driven by a segment of frequent customers who value rewards more than money (price-insensitive reward-seeking = PIRS consumers).
Sharp & Sharp (1997)	✗	✓	✗	✗	✗	✓	✓	✗	✓	✓	✓	LPs do not systematically increase sales at participating stores.
Stourm, Bradlow & Fader (2015)	✗	✓	✗	✗	✗	✗	✗	✓	✗	✓	✗	Redemption is low. Customers differ in their motivation to redeem. Redemption behavior is driven by cognitive and psychological incentives.
Taylor & Neslin (2005)	✓	✗	✗	✗	✗	✗	✓	✓	✓	✓	✓	TLP = profitable. Find both a points-pressure and rewarded-behavior impact. TLP can appeal to customers not interested in the PLP.
Wang, Krishnamurthi & Malthouse (2018)	✗	✓	✓	DM, email	✗	✓	✓	✓	✓	✓	✓	Mobile app and a flexible reward structure increase sales and redemption from LPs (especially from occasional customers), and spending at partners that sell high-penetration categories.
Wang, Lewis, Cryder & Sprigg (2016)	✓	✗	✗	✗	✗	✗	✓	✗	✓	✓	✗	TLP increased purchasing in the post-promotion period. Marketers should set reachable goals, and be cautious about goal failures among loyal customers.
Zhang & Breugelmans (2012)	✗	✓	✗[***]	✗	✗	✗[iv]	✓	✗	✗	✓	✓	IBLPs increase (decrease) sales among previous non-LP (LP) members. reduced attrition, and attracted more new customers.
Panel B: Mobile Marketing: Apps and Mobile Communication												
Study	LP		App	Communication		Heterogeneity:	Outcome		Consumer	Setting		Communication
	TLP	PLP		Type	#	Prior Purchases	Sales	Redemption	Response Dynamics	Real Life	CPG	
Mobile Apps												
Gu & Kannan (2021)	✗	✗	✓	email	✗	✓	✓	✗	✓	✓	✗	In competitive setting (hotels), app adoption decreases spending, especially when used for search and with low app engagement.
Kim, Wang & Malthouse (2015)	✗	✓	✓	✗	✗	✓	✓	✗	✓	✓	✗	Sticky apps can lift brand sales poorly designed apps can hurt the brand.
Li, Luo, Zhang & Wang (2017)	✗	✗	✓	m-ad, SMS & in-app	✗	✗	✓	✗	✓	✓	✗	Consumer response to mobile promos depends on weather. Prevention ads lower (improve) the promo impact in sunny (rainy) weather.
Narang & Shankar (2019)	✗	✗	✓	✗	✗	✓	✓	✗	✓	✓	✗	App adopters spend more, return more, and yield higher net sales, omline and offline. Sales depend on app usage rather than availability.
Van Heerde, Dinner & Neslin (2019)	✗	✗	✓	✗	✗	✓	✓	✗	✓	✓	✗	Sales increases from retailer apps are higher for distant and for offline-only customers, due to higher app access (use).

Zubcsek, Katona & Sarvary (2017)	×	×	✓	m-coupon, in app	×	✓	×	✓	×	✓	✓	Co-located consumers respond to coupons in the same product category. Location history can be used to target next-day coupons.
Mobile Communication												
Andrews, Luo, Fang & Ghose (2016)	×	×	×	m-ad, SMS	×	×	✓	×	×	✓	×	Physical crowdedness increases consumer response to mobile ads.
Bart et al. (2014)	×	×	×	m-ad, webpage	×	×	(✓)[vi]	×	✓	✓	✓	Mobile ads are often ineffective, but can enhance recall and purchase intentions for high-involvement and utilitarian products.
Danaher et al. (2015)	×	×	×	m-coupon, SMS	×	×	×	✓	✓	✓	✓	Impact of m-coupons depends on face value, delivery time and location. Work best for short expiry times.
Dubé, Fang, Fong & Luo (2017)	×	×	×	m-coupon, SMS	×	✓	×	✓	×	✓	×	Competitive responses raise (lower) the profitability of behavioral (geographic) targeting; (a)symmetric pricing incentives soften (toughen) price competition.
Fang et al. (2015)	×	×	×	m-ad, SMS	×	×	✓	×	✓	✓	×	Location-based mobile promotions (LMP) induce impulsive, same-day purchases, but also create product awareness and sales for subsequent days (12). Location matters.
Fong, Fang & Luo (2015)	×	×	×	m-coupon, SMS	×	×	✓	×	×	✓	×	Promoting to consumers near a competitor's location generates extra sales without cannibalizing profits.
Ghose, Li & Liu (2019)	×	×	×	m-coupon, SMS	×	×	✓	✓	✓	✓	✓	Trajectory-based mobile targeting within a shopping mall increases redemption and (current and future) spending, especially for high-income consumers, during weekdays and for exploratory shopping.
Liu, Lobschat, Verhoef & Zhang (2019)	×	×	✓	×	×	✓	✓	×	✓	✓	×	Adding an app to a retailer's mobile website can increase customers' purchase incidence, frequency and order size, and enhance loyalty.
Luo, Andrews, Fang & Phang (2014)	×	×	×	m-ad, SMS	×	×	✓	×	✓	✓	×	Promoting at close (far) distance is more effective with same-day (one-day prior) targeting.
Mills & Zamudio (2018)	×	×	×	m-coupon, self-scan	✓	✓	×	✓	✓	✓	✓	Sheds light on drivers of m-coupon redemption under competition (coupon value, price range, and brand loyalty).
Osinga, Zevenbergen & van Zuijlen (2019)	×	×	×	m-banner ad	×	×	✓	×	✓	✓	×	Mobile banner ads increase offline sales, but do not affect online sales.
Park, Park & Schweidel (2018)	×	×	×	m-coupon, SMS	✓	✓	✓	×	✓	✓	×	Price discount and free sample mobile coupons increase consumers' purchase likelihood and spending during the redemption period. For free sample coupons, the effect lasts beyond that period.
Wang, Krishnamurthi & Malthouse (2018)	×	✓	✓	DM, email	×	✓	✓	✓	✓	✓	✓	Mobile app and a flexible reward structure increase sales and redemption from LPs (especially from occasional customers), and spending at partners that sell high-penetration categories.
Our Study	✓		✓	✓	✓	✓	✓	✓	✓	✓	✓	

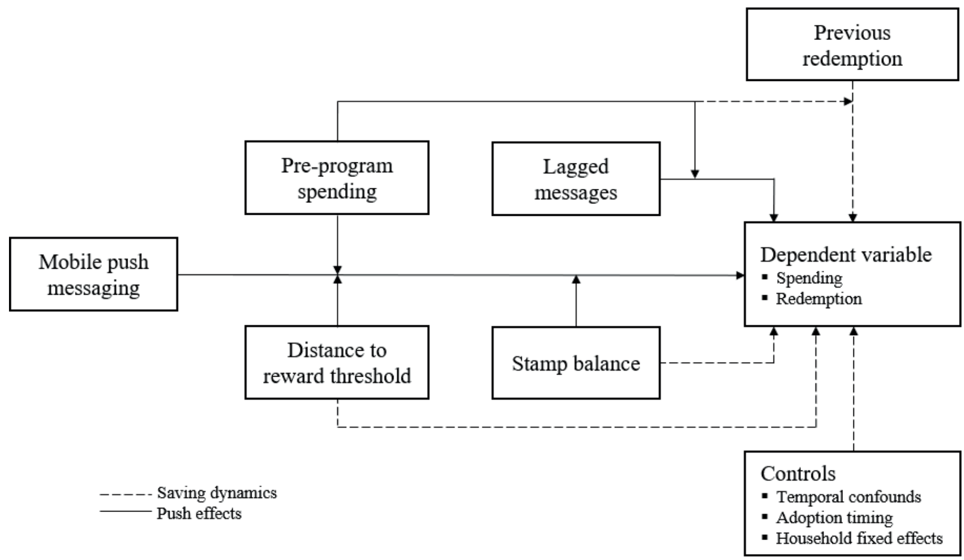
DM = direct mail, # = number of messages. * The program has an annual time constraint, ** email coupon independent of LP, *** the retailer used emails and store flyer announcements, but these could not be distinguished from the program change itself, [iv] Prior LP membership accounted for, not sales, [v] repatronage, [vi] purchase intention. Note: Table1 includes empirical (field) studies, as published in JM, JMR, Mkt Sc, Man Sc, JAMS, IJRM and JR.

2.3 Conceptual Framework

2.3.1 Overall Framework and Setting

Figure 2.1 presents our conceptual framework. Our two outcomes of interest are consumer spending (and, hence, *number of stamps earned*), and redemption (the *number of stamps traded in for rewards)* during the program. Our framework accounts for several shifters of savings and redemption (dotted lines) that operate independently from push messaging. These are consumers' *stamps balance*, *distance to the reward threshold*, and *previous redemption*. Our focus, though, is on how mobile push messages (left side of Figure 2.1) intervene in these processes and ultimately affect in-store spending and stamp redemption (solid lines in Figure 2.1). We discuss our expectations about these effects below.

Figure 2.1: Conceptual framework



2.3.2 Immediate Effects

Main message effects. As indicated above, consumer engagement in LPs typically declines over time, because the program becomes less salient and/or consumers lose interest

(Dorotic et al., 2014). Push messages can help to remedy this. First, they can serve as a simple reminder of the program and the associated rewards. Second, mobile messages can strengthen latent consumer goals, including reward-seeking (Shankar et al., 2016) and thus help increase consumers' stamp saving efforts and in-store spending. Third, push messages may remind consumers of the efforts already put in thus far (e.g., extra trips or purchases at the retailer), which will become sunk should the consumer not keep up those efforts or fail to see them through by redeeming a reward ('switching costs', e.g., Kim, Shi, & Srinivasan, 2001; Taylor & Neslin, 2005; Kivetz, 2003). On the downside, as more messages are sent, these messages may come to lose effect (and, possibly, even cause irritation). Moreover, being sent to personal devices, mobile messages may create privacy concerns. However, to the extent that push notifications are only sent to consumers who voluntarily opted in for the app, we do not expect the latter to be too much of a problem. So, on the whole, we expect a positive but marginally diminishing main effect of messages on spending and redemption.

Interactions with distance and balance. Building on Kivetz et al.'s (2006) observation that goal proximity increases promotion sensitivity, we anticipate higher push effects when distance is low, as the consumer is reminded that he put in a lot of effort already and is almost there. We expect a negative interaction effect on spending between push-messages and balance. On the one hand, when balance is still (well) below the reward threshold, the message may remind consumers of what they still have to accomplish, and provide a trigger to save for a reward (Koo & Fischbach, 2008). At the same time, as consumers' stamps balance grows beyond the threshold, messages may make consumers more aware not only of the program, but also of the already collected stamp surplus, and become less effective in stimulating extra spending. As for redemption, we propose a larger message effect when balance is high and consumers can collect their reward without (much) further effort.

Conversely, when the distance is too high, the consumer still cannot redeem any stamps even if pushed, so we expect a negative interaction on redemption.

2.3.3 Dynamics and Heterogeneity

Lagged effects. Besides an immediate impact, we also expect push messages to have a 'lagged' effect, i.e., affect consumers' spending and redemption behavior in the next period(s). For spending, this effect can go both ways. On the one hand, it may be positive due to a delayed message-response (e.g., Braun & Moe, 2013): consumers may only (have the opportunity to) visit the chain in the period(s) following the message and, being reminded of the program, may increase their expenditures at that time. On the other hand, similar to a post-promotion dip (see, e.g., Neslin & van Heerde, 2008), the lagged effect may be negative due to purchase acceleration: consumers who increased their purchases at the time of the message and built up inventory may subsequently buy less. Because it is not clear upfront which of these two forces dominate, we have no directional expectation for the lagged-message effect on spending. As for redemption: ceteris paribus (that is: controlling for stamps balance), we do not expect previous-period messages to drive down the current number of stamps redeemed. Rather, these push messages may have made the reward program more salient and therefore serve as a reminder to trade in stamps for rewards.

Pre-program spending. As discussed above, consumers' pre-program purchase levels at the retailer affect their program participation (spending and redemption). An important question, then, is how it shapes their response to mobile push messages. For spending, the answer could go either way. On the one hand, lighter buyers may find the redemption-threshold more difficult to attain, and be less prone to spend more in response to a push message. (Too many) push messages targeted at this segment could even produce a reactance effect and lead consumers to spend less (Stauss, Schmidt, & Schoeler, 2005). Moreover, these buyers may not be in the habit of visiting the chain on a frequent basis, and thus have less opportunity to (immediately) step up their expenditures unless the message convinces them to incur an extra visit. On the other hand, while heavier buyers who regularly visit the chain may have more *opportunity*, they may also have less *room* for further spending increases – the so-called ceiling effect (Bijmolt et al., 2011). We leave the net outcome of these forces on consumers' immediate and lagged message response as an empirical issue. When it comes to redemption, we expect higher pre-program purchase levels to enhance the push-messaging effects: heavier spenders can more easily surpass the redemption threshold whenever they receive a message and trade in their stamps for a reward in response, while low-spenders may become frustrated by these messages (Stauss et al., 2005).

Given that redemption not only incites future spending but also comes at a cost, a key question for the retailer is how the spending and redemption impact of push messaging net out, over time, for each customer group – something our estimation results will shed light on.

2.4 Data

2.4.1 Program Details and Data Sources

Our data cover an 18-week LP, operated by BrandLoyalty, and running from August 1, 2016, until November 30, 2016, across 12,135 stores in a major retail chain in Indonesia. Consumers save for rewards by collecting stamps, earned at a fixed rate of one stamp per 40,000 IDR (Indonesian Rupiah) – equivalent to approximately US$2.8 – spent in one of the chain's stores. Stamps can be collected physically or digitally. Digital collection requires installing an app, available from the start of the program onwards. Throughout the program, push messages can be sent to the shoppers with this app. While installing the app is not required to participate in the program, consumers only start receiving push messages after they install the app. To use the app, consumers need to link the installed app to their customer card. When a customer card is scanned at the check-out, the consumer's stamp balance is automatically updated by the number of stamps earned or redeemed.

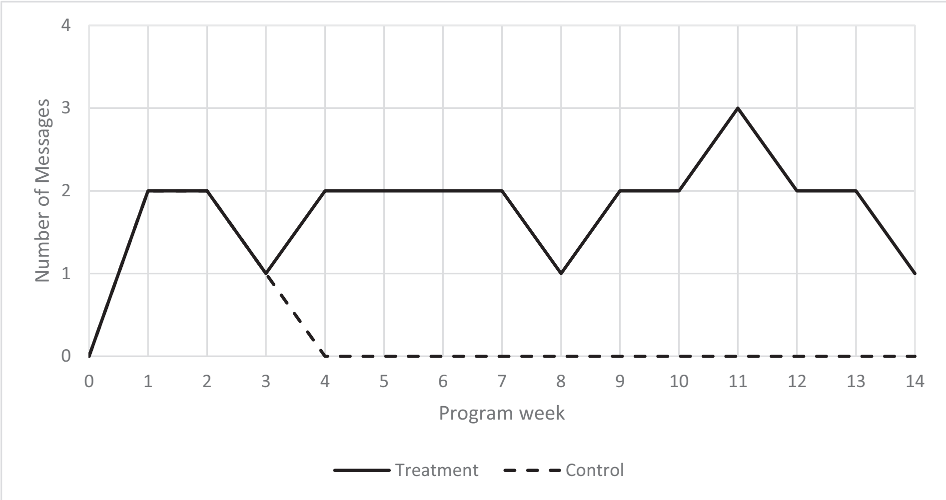
The rewards for this program are plastic containers to store food, offered at a large discount. To obtain a reward, consumers hand in a pre-specified number of stamps, and pay a strongly reduced price (below the wholesale price of the reward for the retailer). The containers vary in size and in the number of stamps required to redeem each item. Appendix A1 (Table A1.1) provides details on the reward items and associated stamps requirements. Consumers can collect and redeem stamps during the first 14 weeks of the program but can only redeem stamps during the final 4 weeks (clean-up weeks). However, during these final weeks, the program operator only pushes messages to consumers who have a balance high enough to still redeem a reward. To avoid potential bias, we estimate the effects based on the 14 program weeks only, and we exclude these final clean-up weeks. Note that all treated consumers receive the same number and type of push messages during program weeks, i.e., the program operator does not target specific groups of consumers during the program weeks.

Because our objective is to assess the impact of push messages sent within the program app, our population of interest consists of consumers who adopt the app (and hence, by necessity, are also card holders of the chain). As part of the measurement system to evaluate the push campaign, the program operator designed exogenous variation in push messages in two different ways. First, a randomly selected subgroup of 2,305 app-installers only receive push messages in the first 3 weeks of the program, and no messages thereafter. These customers serve as the 'control group.' The construction of this group selects on app installation in the first 2 weeks of the program. The remaining customers, who also opted in for the app, do receive push messages in later program weeks, and make up the 'treatment group.' We focus on those households in the treatment group who install the app in the exact same period as the control households, i.e., the first 2 weeks of the program. This ensures that the control and treatment group are comparable on unobservables that correlate with the app installation timing, and leaves us with 44.2K treated shoppers who received push messages throughout the program.

In addition, the program operator withheld individual push messages from a subsample of the consumers in the treatment condition, throughout the program. This resulted in randomly selected subsamples that received one message less in a given week. Depending on the week, the size of such a subsample was 3-9% of the 44.2K shoppers. Each subsample was drawn independently across messages from the pool of treated consumers. Membership of these additional control groups is not correlated with pre-program spending levels ($\rho = .001, N = 651{,}056, p = 0.358$). The resulting variation in push messages is exogenous. Finally, the delivery of messages through the mobile app may fail for additional reasons including canceled phone service, outages, and the like. The fraction of failed messages is not trivial: of the 651K shopper-weeks with messages, 26% have at least one failed message, and message delivery to a given customer can be persistently poor. However, these failures are not the result of targeting by the retailer or program operator. Also, the correlation between missing messages and spending levels prior to the program is small (albeit significant given the large number of observations; $\rho = -.038, N = 651{,}056, p < 0.001$).[6] As such, the failed messages provide an additional source of variation in the number of messages received. The push message schedule as used by the program operator is shown in Figure 2.2.[7]

[6] We note that even if there is a link between number of failed messages and customer spending prior to the program, this does not jeopardize our identification of the message effects because our model includes household fixed effects and interactions with pre-program spending.
[7] We are not aware of the options to block push messages, nor do our data offer information in this regard. However, even if consumers can block push messages and some did, the results from our field experiment still show that push messaging has a positive impact on spending and redemption.

Figure 2.2: Number of push messages sent per week of the program



With these program details in mind, our study merges two datasets. The first dataset contains app information at the household-day level, and captures when consumers go the store, how many stamps they redeem on each transaction, and when they receive push messages. The app data from this first dataset only provides information on the number of stamps redeemed from the point a consumer adopts the app until the end of the program.[8] The second, complementary, dataset contains the exact spending amounts for the same households, obtained from the chain's database, for 18 weeks before the start of the program and for an additional 18 weeks covering the program. To combine these two sources, we first aggregate messages, transactions, and stamp redemption information from the app data to the household-week level.[9] The merged dataset then contains information for households that

[8] Though we do not observe redemption behavior before app adoption, we do not believe this renders our study invalid for two reasons. First, given the observed redemption behavior and pace of accumulating stamps, there will generally be low incidence of redemption in the first two weeks. Second, such redemption cannot be the result of push messages because the consumer receives none. In addition, we account for household specific fixed effects, so heavy users' baseline redemption is taken into account in our analysis.

[9] To the best of our knowledge, messages are sent to everyone who is a recipient at the same time. There is likely variation with respect to when households read the message even when it is sent at the same time but we have no information on this. We only know when messages are received, not whether or when they are read. If a message is received in the evening, it may not be read until the next morning. Aggregation to the week level helps overcome this issue, and we assume that this variation does not impact our weekly results.

subscribe to the app at some point during the first 2 weeks of the program. The data also cover households' purchase history with the retailer prior to the program. We will later use this information to distinguish between more-vs.-less heavy buyers at the chain. Because we are interested in the effect of push messages and not the program itself, the analysis of push message effectiveness uses the periods covering the program (during which push messages could be sent). In total, our data cover 651K household-weeks for estimation.

2.4.2 Sample Statistics

Table 2.2 provides sample statistics on the average weekly expenditures and number of visits to the chain running the LP. It does so for the sample as a whole, and also separately for the treatment and control group. In addition, it shows the split between subgroups with different levels of pre-program spending: the 'low-quartile' (lowest-25% customers), 'median' (25%-75%), and 'high-quartile' group (top-25% spenders).

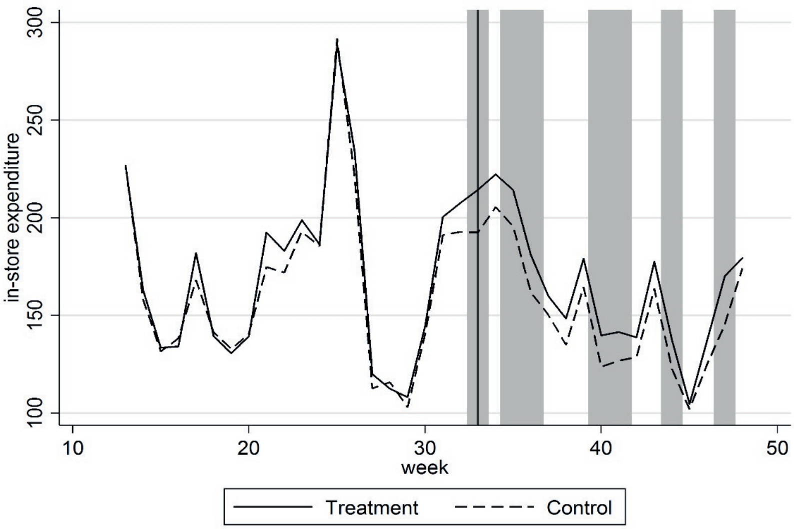
During pre-program weeks, consumers spend 167,400 IDR per week, and undertake 2.33 weekly trips to the chain, on average. At the same time, we observe large variation within the customer base: mean spending amounts and visits prior to the program range between 20,300 IDR (.57 visits) for the low-quartile group, and 463,500 IDR (5.57 visits) for the high-quartile customers. This will make the analysis of differences in push-message responsiveness between lighter and heavier buyers particularly relevant. Comparing consumers' behavior before and during the program, we observe a slight increase in weekly spending (on average: from 167,400 to 169,200 IDR) and visit frequency (on average: from 2.33 to 2.47 visits per week). These increases cannot be interpreted as a main effect of the loyalty program, because they are confounded with a time trend among program participants.

Table 2.2: Household shopping activity by treatment group, program status and household quartiles

	Mean	SD	N	0-25% HH Mean	25-75% HH Mean	75-100% HH Mean
Weekly expenditures (1000 IDR)						
Pre-program						
Control	163.4	218.1	2,305	20.3	91.6	455.9
Treatment	167.6	225.6	44,199	20.4	93.1	463.5
Total	167.4	225.3	46,504	20.4	93.1	463.2
During program						
Control	155.5	184.3	2,305	59.9	117.5	331.4
Treatment	169.9	199.5	44,199	62.6	129.4	358.0
Total	169.2	198.8	46,504	62.5	128.8	356.7
Number of trips						
Pre-program						
Control	2.28	2.64	2,305	0.57	1.57	5.49
Treatment	2.33	2.64	44,199	0.57	1.59	5.57
Total	2.33	2.64	46,504	0.57	1.58	5.57
During program						
Control	2.31	2.31	2,305	1.04	1.90	4.47
Treatment	2.48	2.47	44,199	1.11	2.03	4.74
Total	2.47	2.47	46,504	1.10	2.03	4.73

Our interest, though, is not with the main effect of the LP, but with the impact of push messaging on consumer saving and redemption behavior, which we will assess by monitoring consumers in the control versus the treatment group. Comparing the pre-program behavior of these two groups, we find that pre-program spending levels are comparable (163,400 IDR for the control group vs. 167,600 IDR for the treatment group), as are the store visits (2.28 vs. 2.33 times per week). We conducted a parallel-trends test in the pre-program periods by testing for the equality of time trends across the treatment vs. control groups. Applying a linear trend model to the pre-program expenditure data, the difference in trends between the treated and control group is very small at 100 IDR per week (s.e. = 394.87) and we cannot reject the null hypothesis that both groups exhibit similar purchase dynamics prior to the program. Figure 2.3 shows the average weekly household spending for the treated and control households across different periods in time. It confirms that the expenditures of the treated and the control are highly similar in the pre-program weeks.

Figure 2.3: In-store expenditure by treatment and time



Note: The shaded time-periods denote weeks with significant differences in the average in-store expenditure between households in the treated and the control groups. The vertical line denotes week 33 after which differences in treatment start. None of the differences between the treated and the control group are significant before week 33.

Zooming in on the shifts from pre-program to within-program weeks in Table 2.2, we observe that the control group spends about 155,500 IDR during program weeks (against 163,400 IDR pre-program). As above, this difference cannot be interpreted as a program effect. The effect of push messaging is informed by comparing this drop to the spending shift for the treatment group. For the treated households who keep receiving push messages, expenditure is almost the same during the program (169,900 IDR) as before (167,600 IDR). This is also visible in Figure 2.3. Although the pre-program trends are similar, from week 34 onward, as the control households no longer obtain messages, the participants who obtain push messages spend consistently more on average than those who do not obtain push messages. Similarly, whereas households in the control group display almost the same shopping frequency during the program as they did before (2.31 vs 2.28 trips per week), households in the treated group show an increase (from 2.33 to 2.48 trips per week). These statistics give a first indication that push messaging enhances the outcomes of the program – an effect that we will separate out more carefully with our formal modelling approach.

Table 2.3: Heterogeneity in program participation and rewards redemption

	Weekly stamp collection		**Fraction of redemption incidence**		**Stamps redeemed per week**		**N**
Control	Mean	SD	Mean	SD	Mean	SD	
Average household	3.73	4.80	0.13	0.16	2.91	4.50	2,305
0-25% households	1.33	2.89	0.06	0.11	1.04	2.81	598
25-75% households	2.72	2.89	0.12	0.13	2.09	2.97	1,136
75-100% households	8.25	6.25	0.26	0.19	6.52	6.15	571
Treatment							
Average household	4.10	5.18	0.15	0.17	3.25	4.85	44,199
0-25% households	1.42	2.69	0.06	0.10	1.09	2.61	11,028
25-75% households	3.01	3.34	0.12	0.13	2.31	3.28	22,116
75-100% households	8.93	6.70	0.28	0.20	7.27	6.56	11,055
Total							
Average household	4.08	5.16	0.15	0.17	3.23	4.83	46,504
0-25% households	1.42	2.70	0.06	0.10	1.09	2.62	11,626
25-75% households	3.00	3.32	0.12	0.13	2.30	3.26	23,252
75-100% households	8.90	6.68	0.28	0.20	7.23	6.54	11,626

Table 2.3 provides more detailed statistics on households' saving behavior during program weeks. It reports sample statistics of stamp collection, fraction of weeks with redemption, and number of stamps redeemed per week, in the treatment and control group and for households with different pre-program spending levels. Several points are worth noting. First, stamp collection and redemption monotonically increase with consumers' spending prior to the program. Second, with a mean of 4.08 stamps collected per week, the average consumer should not find it too hard to exceed the modal redemption threshold of 10 stamps (which we will use as the threshold in our main models).[10] However, as we already

[10] We also considered a threshold of five stamps (the lowest requirement, which applied to one of the rewards). As can be seen from Appendix A5, the pattern of results remained the same. Because the 5-stamps reward was not a collectable (unlike the other rewards), and because the majority of rewards required (at least) 10 stamps (see Appendix A1), we retained the 10-stamps threshold in our main model.

saw for expenditures, the distribution of weekly stamp collection across consumers is heavily skewed, with a value of only 3.01 for the median group, and only 1.42 for the low-quartile group – indicating that some consumers have to save for weeks before they can collect a reward. Third, redemption is low. For instance, with 3.00 stamps collected per week, a 'median' household accumulates 42 stamps in the course of the 14 program weeks. However, the number of stamps actually traded in for a reward is much lower, at about 32 stamps for (= 14 × 2.30). The same holds for heavy buyers, who accumulate some 125 stamps in the course of the program, only 101 of which are redeemed. So, even consumers who expressed an interest in the program by installing the app often fail to redeem their rewards. This underscores the importance of mobile push as a potential tool to increase redemption.

2.5. Model

2.5.1 Empirical Issues

Recall from section 4.1 that our estimation sample includes households who, like the control households, installed the app in the first two weeks. This allows us to study the effect of push messaging holding installation timing of the app constant and rely solely on variation from random assignment into treatment and control.

In our empirical modeling, we further need to account for a number of confounds. First, in addition to a program effect, other temporal factors (e.g., trends, holidays) can influence households' stamp saving or redemption behavior. We control for these factors through time fixed effects. Second, the adoption of the app in itself may lead to increases in stamp saving or redemption behavior. To separate this app effect from the message effect, we include a step dummy that indicates when households adopt the app. Third, as suggested by the model-free evidence, household differences may exist that affect their stamp saving or redemption behavior. Though our control households are picked randomly and should not be different from the treated households, we use household-fixed effects to rule out any biases.

While the above considerations all pertain to measuring a common treatment effect of push messaging, we are additionally interested in *heterogeneous treatment effects.* As indicated in the conceptual part and Figure 2.1, we seek to determine the *moderating* impact of two sets of observable consumer behavioral characteristics. First, we allow the treatment effect to vary by different levels of household spending prior to the program. Second, we allow it to be heterogeneous with respect to different levels of observed past participation during the program, i.e., stamp balance (or the cumulative number of stamps collected minus redeemed), and distance from the redemption threshold (or how many stamps a consumer is away from a possible redemption). These two moderators relate to each other, as the distance to the reward threshold is equal to the balance minus the number of stamps required to get a reward. But, while distance can at most be equal to the reward threshold, balance can be much larger and, together, the two variables allow to flexibly capture the impact of available stamps below and above the threshold. The variation in the three moderator variables stems from the between-household difference in the level of spending (collection) and redemption.

To determine these treatment effects, we require exogenous variation in our messaging variables. First, as mentioned earlier, a random control group receives no messages after week 3 of the program. Second, there is variation in the number of messages sent over time, and several individual messages were withheld from a random subsample of the treated consumers in different weeks. Finally, a fraction of the messages failed to be sent, but such failure is unintentional and thought to be related to technical issues. In sum, our push messages vary across households and time without being selectively targeted.

2.5.2 Model Specification

To determine the effect of push messaging on consumers' spending (stamp collection), we use the following specification:

$$\begin{aligned} \text{SPEND}_{i,t} = {} & \alpha_i + \gamma_t + \beta_0 + \beta_1\text{DIST}_{i,t} + \beta_2\text{BALANCE}_{i,t} + \beta_3\text{PREVREDEMP}_{i,t} + \\ & \beta_4\text{ADOPT}_{i,t} + \beta_5\text{LOG\_MESSAGES}_{i,t} + \beta_6\text{LOG\_MESSAGES}_{i,t-1} + \\ & \beta_7\text{LOG\_MLESSAGES}_{i,t-2} + \beta_8\text{DIST}_{i,t} \times \text{PREVSPEND}_i + \\ & \beta_9\text{BALANCE}_{i,t} \times \text{PREVSPEND}_i + \beta_{10}\text{PREVREDEMP}_{i,t} \times \text{PREVSPEND}_i + \\ & \beta_{11}\text{ADOPT}_{i,t} \times \text{PREVSPEND}_i + \beta_{12}\text{LOG\_MESSAGES}_{i,t} \times \text{PREVSPEND}_i + \\ & \beta_{13}\text{LOG\_MESSAGES}_{i,t-1} \times \text{PREVSPEND}_i + \beta_{14}\text{LOG\_MESSAGES}_{i,t-2} \times \text{PREVSPEND}_i + \\ & \beta_{15}\text{LOG\_MESSAGES}_{i,t} \times \text{DIST}_{i,t} + \beta_{16}\text{LOG\_MESSAGES}_{i,t} \times \text{BALANCE}_{i,t} + \epsilon_{i,t}, \end{aligned} \tag{2.1}$$

where the dependent variable $\text{SPEND}_{i,t}$ is the total amount of money spent at the retailer by household i in week t. Equation 2.1 comprises three sets of explanatory variables. First, we include regressors that capture the dynamics of LP savings behavior (discussed in the conceptual part). The variables previous stamp balance ($\text{BALANCE}_{i,t}$), and distance to the reward-threshold ($\text{DIST}_{i,t}$) accommodate the points-pressure and goal-gradient mechanisms, respectively; while previous redemption ($\text{PREVREDEMP}_{i,t}$) captures possible rewarded-behavior effects. (We note that the correlation between balance and distance is not overly high (ρ= -.45), such that both variables can be included). We allow these saving dynamics to differ between more-or-less-heavy buyers at the chain ($\text{BALANCE}_{i,t} \times \text{PREVSPEND}_i$, $\text{DIST}_{i,t} \times \text{PREVSPEND}_i$, and $\text{PREVREDEMP}_{i,t} \times \text{PREVSPEND}_i$). Second, as indicated above, to cleanly assess the push-message effects we also need to rule out several confounds. To control for a possible app effect, we include a dummy variable for app-adoption timing ($\text{ADOPT}_{i,t}$), and also allow its impact to vary with household's prior spending at the chain ($\text{ADOPT}_{i,t} \times \text{PREVSPEND}_i$). Other household differences unrelated to the push messaging (e.g., the fact that larger households may need larger baskets and thus 'collect' more stamps) are controlled for through household fixed effects (α_i). To control for 'message-unrelated' changes in spending in the course of the program (e.g., consumers collecting more stamps at the start) or other temporal confounds, we also use time fixed effects (γ_t).

The third set of regressors relate to the push message effects – our key variables of interest. The variable $\text{LOG\_MESSAGES}_{i,t}$ captures the immediate effect of the number of messages received by the household in a given week, where we use a log transform to accommodate decreasing returns of multiple messages in the same week. We also include the lags of this variable to capture the post-message effect due to purchase acceleration or carryover. In addition to these main treatment effects, we bring in several moderators. To determine whether messaging is more or less effective depending on consumers' stamp balance or distance, we include their interactions with the same-period message variable ($\text{BALANCE}_{i,t} \times \text{LOG\_MESSAGES}_{i,t}$, $\text{DIST}_{i,t} \times \text{LOG\_MESSAGES}_{i,t}$). (In these interactions, we mean-center distance and balance). To assess whether message response differs between more vs. less heavy buyers at the chain, we allow the immediate and lagged message variables to interact with the household's (mean-centered) average spending before the program (PREVSPEND_i). For an exact specification of each of these variables, see Table 2.4.

Table 2.4: Variable descriptions

Variable	Description
<u>*Spending/Redemption variables*</u>	
$\text{SPEND}_{i,t}$	*Total spend (in 1000 Rupiah)*. Measured as the total amount of money spent at the retailer by household i in week t. Equal to zero if no spending takes place that week.
$\text{REDEEM}_{i,t}$	*Stamps redeemed*. Measured as the number of stamps that are redeemed by household i in week t. Equal to zero if no redemption takes place that week.
$\text{PREVREDEMP}_{i,t}$	*Previous redemption.* Dummy variable (0/1) that indicates whether household i has made a redemption in the previous week.
<u>*Message variables*</u>	
$\text{LOG\_MESSAGES}_{i,t}$	*Ln_messages.* This variable is calculated as $\ln(\text{MESSAGES}_{i,t} \times 100 + 1)$, where $\text{MESSAGES}_{i,t}$ equals the number of messages received by household i in week t.
$\text{LOG\_MESSAGES}_{i,t-x}$	*Lagged messages.* This variable is calculated as $\ln(\text{MESSAGES}_{i,t-x} \times 100 + 1)$, for x=1, 2; where $\text{MESSAGES}_{i,t-x}$ equals the number of messages received by household i in the previous week $(t-1)$ or the week before that $(t-2)$.
<u>*Moderators*</u>	
$\text{BALANCE}_{i,t}$	*Previous stamp balance.* Measured as the stamp balance of household i, prior to (before the start of) week t. It is calculated as

	$\text{BALANCE}_{i,t-1} + \text{COLLECT}_{i,t-1} - \text{REDEEM}_{i,t-1}$. Stamp collection during a given week is reflected in the balance of week $t + 1$.
$\text{DIST}_{i,t}$	*Distance to the reward threshold.* The number of stamps household i requires to reach the redemption threshold, prior to (before the start of) week t. In this case we consider the target/threshold to be ten stamps, as this is the modal number of stamps required, for which they can redeem a reward. (See Appendix A5 for a robustness check). If at least ten stamps have been collected, or in other words, if the previous stamp balance is equal to at least ten, distance to the target is equal to zero.
PREVSPEND_i	*Average spending prior to the program (in 1000 Rupiah).* Equal to the average (grand mean-centered) spending of household i in the weeks prior to the start of the program.
<u>*Controls*</u>	
$\text{ADOPT}_{i,t}$	*Adoption timing.* Step dummy that is equal to 1 from the app adoption time of household i onwards, 0 otherwise.
γ_t	*Week fixed effects*.
α_i	*Household fixed effects*.
$\text{CF}_{i,t}$	*Correction factor.* Included in the 'number of stamps redeemed' layer of the hurdle model, to capture the correlation with the redemption incidence layer (see McFadden & Dubin 1984); it is measured as $\left(1 - \hat{P}_{i,t}\right) * \frac{\ln\left(1-\hat{P}_{i,t}\right)}{\hat{P}_{i,t}} + \ln\left(\hat{P}_{i,t}\right)$, where $\hat{P}_{i,t}$ is the predicted redemption incidence probability for household *i* in week *t*.

As mentioned before, in-store spending directly results in stamp collection, for app users. Moreover, because our interest is in the impact of push messaging within the program (and not in the program effect as such), we estimate the spending model on observations during the program only - excluding pre-program weeks. This way, we get a cleaner assessment of household fixed effects within the program and, thus, of the additional influence of push messaging.

To measure the effect of push messaging on stamps-redemption behavior, we use a hurdle specification with the same regressors as in Equation 2.1, in which the first layer captures the probability of redemption incidence for household *i* in week *t* ($\text{REDEMP\_INC}_{i,t}$) through a binary-logit specification, and the second layer captures the number of stamps redeemed given incidence (REDEEM$_{i,t}$).[11,12] This hurdle model naturally accommodates the (many) weeks in which a household does not redeem and allows us to predict the 'previous redemption incidence' regressor for our dynamic simulations (see also below). We estimate the two layers sequentially but, following McFadden and Dubin (1984), capture their interdependence through a correction factor in the second layer (see Table 2.4 for details).

We use the sandwich estimator of variance (White-Huber standard errors) to accommodate heteroscedasticity.

2.6. Results

To build up the results, we consider four nested specifications (which all include fixed effects and the app-adoption dummy): (1) a basic model with the main effects of distance, balance, and previous redemption, but no impact of push messaging, (2) a model with current and lagged push messages added, (3) a model that also adds households' pre-program spending, and (4) the 'full' model with the message-distance and message-balance interactions added to the previous specification. Appendix A2 gives the results for all models. Below, we discuss the estimates of the full model, reported in Table 2.5.

2.6.1 Spending

Column 1 of Table 2.5 lists the results for store expenditure/stamp collection. Considering the control variables, we find that program-app adoption comes with an increase in spending and stamp collection (β = 111.2, p < .001) – as expected. Our results also corroborate the presence of pre- and post-reward saving dynamics. First, we find that a smaller stamps balance increases spending (β = -2.432, p < .001). This is in line with the

[11] We note that the main effect of PREVSPEND$_i$ is subsumed in the household fixed effects. We include household fixed effects in the spending model, and in the 'number of stamps' layer of the hurdle model but not in the redemption incidence (logit) layer (where we include the PREVSPEND variable instead), because this would create a dynamic panel-regression problem.

[12] Because the number of stamps for different rewards are not multiples of one another, and consumers can (thus) redeem almost any number of stamps on a given occasion, a count model is not feasible here, and we can use a continuous (linear) specification for the number of stamps redeemed given redemption incidence.

points-pressure mechanism: at levels of balance well below the savings threshold, the prospect of having to reach that (still remote) threshold within a given time frame leads consumers to make an extra effort, but this pressure wears off as the number of available stamps grows. Second, the negative coefficient of distance (β = -2.531, $p < .001$) supports the goal-gradient hypothesis that closeness to the redemption target motivates consumers to save more stamps (Kivetz et al., 2006). Third, previous redemption positively affects spending (β = 30.74, $p < .001$), in evidence of the rewarded-behavior mechanism (Drèze & Nunes, 2011).

Our focus, though, is on the impact of mobile push messages. Column 1 of Table 2.5 shows that push messages positively affect households' in-store spending and, thus, stamp collection. Considering the main effect, we obtain a positive immediate effect of push messaging on expenditures (β = 2.679, $p < .001$). Turning to the lagged-message variables, the effect is negative for the one-week lag (β = -1.800, $p < .001$), but insignificant for the two-week lag ($p > .10$). This suggests that, similar to a post-promotion dip, consumers compensate for the spending lift at the time of receiving the message by buying less in the subsequent week (but not after).

The interaction effects with pre-program spending are all significant, suggesting there is considerable heterogeneity in the effects of messaging along different levels of prior spending at the chain. Our implications section provides further detail.

Next, we report on how saving dynamics interact with messaging. As Table 2.5 shows, the available stamp balance slightly reduces the effect of push messages on spending (β = -.0685, $p <.10$). As for distance, we expected that a smaller distance to the reward threshold would strengthen the immediate effect of messaging, and our results confirm such a negative interaction (β = -.578, $p < .001$). To show how these effects play out in combination, Figure A3.1 in Appendix A3 plots the spending impact of push messages as a function of the consumer's available stamps.

Table 2.5: Estimation results

	Spending	Redemption Incidence	# of Stamps Redeemed
$\text{BALANCE}_{i,t}$	-2.432***	0.0137***	0.319***
	(0.201)	(0.000441)	(0.0292)
$\text{DIST}_{i,t}$	-2.531***	-0.185***	-0.686*
	(0.433)	(0.00364)	(0.350)
$\text{PREVREDEMP}_{i,t}$	30.74***	1.256***	4.777*
	(1.783)	(0.0108)	(2.163)
PREVSPEND_{i}		0.00150***	
		(0.0000423)	
$\text{ADOPT}_{i,t}$	111.2***		
	(3.123)		
$\text{LOG\_MESSAGES}_{i,t}$	2.679***	0.125***	0.515*
	(0.437)	(0.00356)	(0.241)
$\text{LOG\_MESSAGES}_{i,t-1}$	-1.800***	0.0391***	0.0241
	(0.257)	(0.00340)	(0.0945)
$\text{LOG\_MESSAGES}_{i,t-2}$	-0.192	0.0106***	0.0731
	(0.252)	(0.00292)	(0.0485)
$\text{BALANCE}_{i,t} \times$ PREVSPEND_{i}	-0.00227***	-0.0000152***	-0.0000734*
	(0.000326)	(0.000000506)	(0.0000358)
$\text{DIST}_{i,t} \times$ PREVSPEND_{i}	-0.00449***	0.000137***	0.000259
	(0.00128)	(0.00000472)	(0.000284)
$\text{PREVREDEMP}_{i,t} \times$ PREVSPEND_{i}	0.0446***	-0.000333***	0.0000332
	(0.00844)	(0.0000323)	(0.00104)
$\text{ADOPT}_{i,t} \times$ PREVSPEND_{i}	0.269***		
	(0.0255)		
$\text{LOG\_MESSAGES}_{i,t} \times$ PREVSPEND_{i}	0.0228***	0.00000740	0.0000505
	(0.00186)	(0.00000811)	(0.000156)
$\text{LOG\_MESSAGES}_{i,t-1} \times$ PREVSPEND_{i}	-0.00803***	-0.0000200*	0.000132
	(0.00157)	(0.00000839)	(0.000147)
$\text{LOG\_MESSAGES}_{i,t-2} \times$ PREVSPEND_{i}	-0.0155***	-0.0000360***	-0.000248*
	(0.00157)	(0.00000716)	(0.000122)
$\text{LOG\_MESSAGES}_{i,t} \times$ $\text{BALANCE}_{i,t}$	-0.0685*	-0.000468***	0.00335
	(0.0304)	(0.0000871)	(0.00281)
$\text{LOG\_MESSAGES}_{i,t} \times$ $\text{DIST}_{i,t}$	-0.578***	0.00497***	0.0272
	(0.0778)	(0.000744)	(0.0171)
$\text{CorrectionFactor}_{i,t}$			-2.467
			(2.020)
Constant	144.4***	-3.146***	5.038
	(4.988)	(0.0526)	(7.819)
Observations	651056	636981	93211
R^2 (pseudo-R^2)	0.042	(0.169)	0.168

Note: All models include time fixed effects, the (linear) models for spending and # stamps redeemed also include household fixed effects, the (logit) model for redemption incidence does not. Standard errors in parentheses. Reported R-squares net of fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Finally, we consider how the effectiveness of immediate and lagged messages depend on a household's pre-program spending with the retailer. The results show that the positive effect of immediate messages on spending increases with pre-program spending (β = .0228, p < .001). Immediate message effects are thus smaller for less-heavy buyers, who visit the store less regularly or progress towards the reward threshold more slowly and may be discouraged when they receive messages. As for lagged messages, the negative interactions with pre-program spending (1 lag: β = -.00803, 2 lags: β = -.0155, p < .001) implies that heavy buyers at the chain exhibit larger post-message spending dips. This is consistent with a 'ceiling' effect: because these customers already allocate (most of) their grocery budget to the chain, they have less room for extra outlay, and may partly compensate extra expenditures due to push by spending less subsequently.

2.6.2 Redemption Incidence and Number of Stamps Redeemed

Columns 2 and 3 of Table 2.5 report on the results for redemption incidence and stamps redeemed. Like before, the main effects of saving dynamics are as expected. Both redemption incidence, and the number of stamps redeemed given incidence, increase as the distance to the reward threshold goes down (incidence: β = -.185, #stamps: β = -.686, p < .001) and stamp balance goes up (incidence: β = .014, #stamps: β = .319, p < .001). Furthermore, we obtain a positive effect of previous redemption on current redemption incidence (β = 1.256, p < .001) and redeemed stamps (β = 4.777, p < .001).

The immediate effect of push messaging is positive and significant, for both incidence (β = .125, p < .001) and number of stamps (β = .515, p < .10). So, as anticipated, current messages contribute to program awareness and remind consumers about redemption possibilities. Whereas previous-week messages affected in-store expenditure negatively, we find that they enhance the enhance the propensity to redeem (1 lag: β = .0391, 2 lags: β = .0106, p < .001), but do not affect the number of stamps traded in (p > .10).

As for the moderating effects, we find a positive interaction coefficient between messages and distance (β = .00497, p < .001) and a negative interaction coefficient between messages and balance (β = -.000468, p < .001). Though surprising at first, these coefficients must be interpreted as adjustments to the already built-in interactions in the logit model, which are governed by the main effects of distance and balance.[13] Together, these coefficients *do* imply that, as expected, the message impact on redemption incidence increases with higher balance and decreases with higher distance – be it less so than would be dictated by the built-in interactions. Figure A3.2 in Appendix A3 plots the message effects on redemption as a function of available stamps.

2.7 Implications

2.7.1 Setup

As the model results show, the impact of push messages on consumers' shopping activity and program participation is statistically significant. We also find that it comes about in several ways, making it difficult to gauge the magnitude of the total (direct and indirect) effects of sending push messages. First, there is a direct effect of messaging on in-store spending and on redemption. Second, lagged messages influence current spending and redemption. Third, the impact of push messages depends on consumers' stamps balance and distance to the reward, quantities that endogenously evolve over time. Finally, more vs. less heavy buyers differ in their message response. Therefore, to assess the effects of push messaging over time for customers with different levels of (pre-program) spending with the retailer, we use our estimates as inputs for dynamic simulations.

Specifically, starting from the first week of the program, we predict consumers' spending and number of stamps redeemed in the course of the program for alternative push plans, and compare the predicted spending and redemption trajectories. We do so for

[13] The marginal impact of a message on the redemption probability is P*(1-P)*(B0+B1*Distance+B2*Balance), where B0, B1 and B2 are the main- and interaction- coefficients of message, and P is the 'Baseline' redemption probability. Because P increases in balance and decreases in distance, and because it is typically below .5, P*(1-P) will also increase (decrease) in balance (distance). Hence, as long as (B0+B1*Distance+B2*Balance) is positive, the message impact will be higher for higher balance and lower distance.

different customer profiles – i.e., lighter vs. heavier buyers at the chain. For each profile, we simulate 1000 trajectories (each corresponding to a set of draws for redemption incidence) and average the outcomes across those trajectories. Moreover, to account for uncertainty in the estimated parameters, we repeat these simulations for 1000 draws from the parameter sampling distributions, to assess the significance of the differences between alternative push plans. Appendix A4 gives more details on the simulation procedure.

2.7.2 To Push or Not to Push?

We first assess the impact of the retailer's current push plan, i.e., with the 'actual' number of push messages sent out during the program as depicted in Figure 2.2. We do this by contrasting the predicted spending and redemption levels of consumers 'treated' with that plan, with those of consumers in a 'control' scenario (who, as also shown in Figure 2.2, receive no more messages after the third program week).

We start by considering the average effects, displayed in the top row of Panel A of Table 2.6. As the table shows, the push plan has a significant positive impact on spending. Starting from week 4 (which is the week in which the 'actual' and 'control' plan begin to diverge), treated consumers spend 270,216 IDR more (or, +24,565 IDR more per week on average) during the remainder of the program. So, the push plan produces a substantial (14.02%)[14] lift in consumer expenditures.

We also obtain a positive impact on the number of stamps redeemed. Consumers treated with the push plan redeem 18.12 more stamps (or, 1.65 more stamps per week) throughout the remaining 11 program weeks than they would in the control scenario. This comes down to a doubling of stamp redemption (+105.8%)[15]. It follows that on average,

[14] Obtained as the spending lift among treated consumers during the treatment weeks (i.e., week 4 to 14), divided by the (simulated) baseline spending by control consumers during those same weeks, or: 270,216 IDR/ 1,927,595 IDR=.1402.
[15] Obtained as the redemption lift among treated consumers in weeks 4 through 14, divided by control consumers' (simulated) baseline redemption in those weeks, or: 18.12/17.12=1.058.

messaging 'closes' the redemption gap: the additional number of stamps redeemed (18.12) by far exceeding the extra number of stamps collected (i.e., the 'stamps' equivalent of 270,216 IDR, or 6.76 stamps). As such, by the end of the program, the push plan reduces the average consumers' stamp balance by about one fourth (-10.73 stamps, a 23.4% drop).

These numbers, however, pertain to the 'average' customer. Because the message effects depend on households' spending with the chain prior to the program, and because the pre-program spending distribution across households is right skewed,[16] we also consider the impact of the push plan for other customer types (i.e., with pre-program spending levels at the lower-quartile, median, and upper-quartile of the pre-program spending distribution). The results are again summarized in Panel A of Table 2.6.

Zooming in on spending first, we note that less-heavy customers of the chain (the lower quartile of the distribution) already react positively to the push plan. Being sent more messages (from week 4 onwards), increases their expenditures at the chain by 219,642 IDR (a 12.8% lift) throughout the remaining program weeks. More 'heavy' buyers at the chain show larger expenditure increases in response to push messaging (+244,480 IDR and a 13.4% lift for median, +377,954 IDR and a 16.2% lift for upper-quartile customers). A similar pattern is observed for redemption: the current push plan leads upper-quartile customers to redeem more additional stamps in absolute terms (+22.45 stamps) than lower-quartile consumers (+15.42 stamps) and customers with median pre-program spending levels (+16.8 stamps).[17] The large difference between effects on spending versus redemption confirms that especially for customers who regularly visit the store and have saved many stamps, push messages can serve as a trigger to redeem.

[16] The average pre-spending level is approximately equal to the 75th percentile of the spending distribution. Hence, our average household is considered a high spending household.
[17] All pairwise differences in spending and redemption lift between customer groups are significant at $p<.01$.

Summarizing, we find that the current push plan has a strong positive impact on expenditures, and especially entices heavy customers to spend more at the chain. First and foremost, though, it stimulates program participants to trade in collected stamps for rewards – a phenomenon that is also particularly strong among heavy buyers at the chain.

Table 2.6: Impact of adding extra messages

		Change in		
Consumer Type	**Baseline spending (in IDR, across program weeks*)**	**Spending (in IDR, across program weeks)**	**Number of Stamps Redeemed (across program weeks)**	**Stamp Balance (by the end of the program)**
Panel A: Actual push plan vs. control group				
Average	1,927,595	270,216	18.12	-10.73
Low Quartile	1,717,076	219,642	15.42	-9.30
Median	1,822,081	244,480	16.82	-10.11
Upper Quartile	2,336,295	377,954	22.45	-12.44
Panel B: One additional push message in week 4 of the program vs. actual push plan				
Average	2,197,821	2,710	0.085	-0.021
Low Quartile	1,936,718	2,298	0.079	-0.024
Median	2,066,561	2,475	0.081	-0.024
Upper Quartile	2,714,249	3,891	0.130	-0.039
Panel C: One additional push message in week 7 of the program vs. actual push plan				
Average	1,527,929	2,958	0.119	-0.062
Low Quartile	1,349,536	2,214	0.084	-0.043
Median	1,438,291	2,322	0.092	-0.042
Upper Quartile	1,881,481	4,122	0.174	-0.088
Panel D: One additional push message in week 10 of the program vs. actual push plan				
Average	928,264	2,502	0.187	-0.150
Low Quartile	818,568	2,182	0.159	-0.128
Median	873,202	2,183	0.162	-0.131
Upper Quartile	1,144,139	2,619	0.189	-0.149

* Baseline spending equals the total amount spent from the week of the change onward if consumers were not to receive the extra messages. Hence, in Panel A, baseline spending is the total amount spent from week 4 onward (in which the actual and control plan start to diverge) until the end of the program, by consumers who would not receive any messages during those weeks. In Panel B (C,D), baseline spending is the amount spent from week 4 (7, 10) onward, by consumers who would not receive the extra message in week 4 (7,10). Note: 1000 IDR is approximately equal to 0.07 USD. These numbers are considerable for Indonesian standards.

2.7.3 How Much Push, to Whom and When?

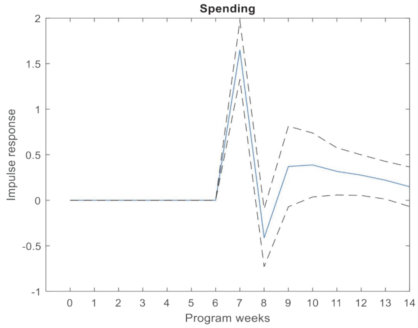
So far, we compared the actual push plan with the control scenario. However, managers can deviate from the current plan, and increase or reduce the number (and timing) of the messages as they see fit. To get a better feel for the impact of individual messages within the push plan, we consider the effect of adding one push message to the push plan actually used during the program. Because the message effects are dynamic, the return from an extra message may well depend on its timing. Hence, we report the results for three scenarios: one with an extra push in week 4 of the fourteen-week program, one with and extra message in week 7, and one with and extra message in week 10.

Figure 2.4 shows the impulse response (i.e., the over-time impact of an additional message in week 7) for the average consumer, for spending (Panel A) and number of stamps redeemed (Panel B), across program weeks. The solid lines indicate the mean effects, and the dashed lines represent 95% confidence bounds (+/- two standard deviations around the mean). To further clarify the dynamics, it also plots the evolution of the stamps balance – i.e., the number of stamps collected but not yet redeemed in a certain week (Panel C).

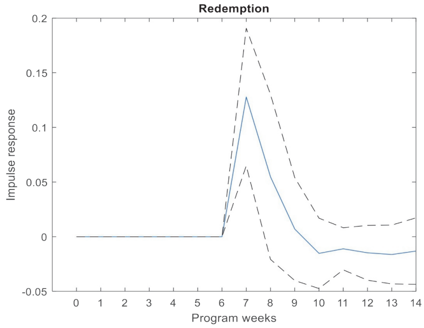
For spending, we see an initial spike in week 7 caused by the immediate effect of the extra push message in that week, a dip in the week after due to the negative lagged-message effects, and a return to positive values from week 9 onward. The plot for redemption also shows a spike at the time of the extra message (week 7), with a gradual return to the base level in subsequent weeks. The spending and redemption effects after week 9 are the result of the 'savings dynamics,' i.e., the impact of previous redemption, distance and balance. Indeed, the extra spending in week 9 follows from the 'rewarded behavior' created by the higher previous-week redemption, and from the ensuing reduction in stamps balance (see Figure 2.4, Panel C), which provides a trigger to again step up collection. Yet, this same reduction in available stamps dampens the number of stamps redeemed in the course of subsequent weeks – only to gradually return to the base levels. As the plots show, positive redemption effects materialize especially in the short run, whereas increases in spending partly follow from lower balance in later periods.

Figure 2.4: Impact of an extra message in week 7 to the current push plan (average hh)

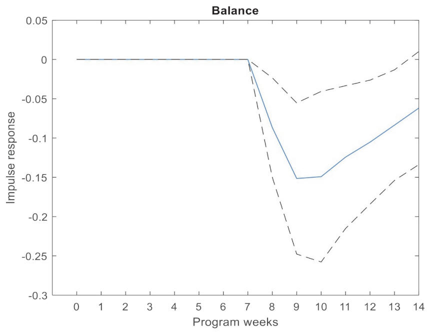
Panel A



Panel B



Panel C



Accumulating the impulse-responses in Figure 2.4 across program weeks, we obtain the effects reported in the top row of Table 2.6, Panel B – the total change in spending and number of stamps redeemed from adding the extra message in week 7 for the average customer. The table also indicates the change in stamps balance by the end of the program, and breaks down each of these effects for lighter, medium, and heavier buyers at the chain. Table 2.6, Panel C provides the corresponding figures for a message added at a later time in the program (i.e., week 10 instead of week 7). Two additional insights emerge.

First, distinguishing lighter from heavier buyer segments, we observe a similar pattern as for the total push-plan effect: the effect of an extra message on spending and redemption is larger for consumers with higher levels of pre-program expenditures.

Second, the impact of an individual message depends on its timing within the program. Later messages, which reach consumers by the time they have built up a larger stamps balance, yield stronger redemption effects. In contrast, spending lifts are highest for mid-program messages. As can be seen from Figure 2.4, this follows from three countervailing forces. First, push messaging increases spending within the same week (and entails a (smaller) spending dip in the week after). This produces a positive net effect especially early on in the program, when balance is still low. Second, push messaging also triggers redemption and thus reduces stamps balance, which stimulates customers to start collecting extra stamps (read: spending more) again. This redemption effect, and hence the associated upward spending shift, is bigger for messages sent later on. Third, however, for push messages near the end of the program, this positive 'uptake' effect is truncated in time. In all, this implies that from a spending perspective, messages around the middle of the program are most effective.

To rule out that this finding is idiosyncratic to the specific weeks in which we implemented the extra message, we consider alternative push plans with an additional message in still earlier or later weeks. The conclusion from these simulations comports with the findings above: adding more messages to the currently implemented plan enhances spending and redemption, especially for heavy buyers. However, the impact at the margin decreases for everyone. As for timing, the expenditure effects are highest for messages sent about halfway during the program. In contrast, later messages generate a stronger redemption lift.

2.7.4 Robustness Checks

To validate the robustness of our results, we run several alternative models. (Appendix A5 gives detailed results). First, because households may collect a reward that requires fewer stamps than the 'modal' threshold (10), we re-estimate the model with the threshold set to five (the lowest reward value). This model yields a similar pattern of coefficients, but somewhat lower message effects (see Appendix A5, Table A5.1). Because there is only one reward with a 5-stamp requirement, and because this type of reward is not likely to be selected multiple times (unlike the other rewards, it is not a collectable), we retain the 10-stamps threshold in our main model.

Second, instead of using the log of messages, we consider an alternative specification, in which we allow for separate main-effect parameters for one, two and three messages per week (the range observed in our dataset). This model not only allows for decreasing returns but also for possibly negative effects as the number of messages goes up. We find that this model yields the same overall shape of effects as that with log messages: the impact of an extra message in the same week is still positive but becomes smaller and ultimately insignificant (see Appendix A5, Table A5.2). This confirms the presence of diminishing returns and suggests that for the message frequencies observed in our data, irritation is not an issue. Because the log-messages model is more parsimonious, we retain it as our main model.

Third, we check whether the dynamic message effects prevail if we include lagged spending as an extra explanatory variable. Though we find this variable to be collinear with previous redemption (leading to a coefficient shift for that variable), the dynamic message effects show the same pattern as before (see Appendix A5, Table A5.3, Panel B).[18]

2.7.5 Profitability

Push messages enhance consumer spending at the retailer. However, they have an even stronger effect on number of stamps redeemed. In many LPs – including the one studied here – such redemptions come at a cost to the retailer. Even if consumers, in addition to handing in stamps, have to pay a certain amount to obtain a reward, this may not cover the

[18] We did not retain the model with lagged spending as our final specification, because the use of household fixed effects together with a lagged dependent variable might bias the individual coefficients of the spending model, and the use of Arellano-Bond estimators leads to very unstable outcomes (see also Appendix A5).

wholesale price of the reward charged by the program operator to the retailer – implying that the retailer leaves money on the table for each redemption. The critical question thus remains whether the margin on extra sales from push messaging compensates for the additional loss on redeemed rewards, i.e., whether the push plan is actually profitable for the retailer.

We use average wholesale prices and consumer reward payments to the retailer[19] obtained directly from the program operator to explore this. Let STAMPS_REWARD be the average number of stamps redeemed per collected reward (in our case: 11.938 stamps), and LOSS_REWARD be the average 'net loss' for the retailer per reward, that is: the difference between the reward payments made by the consumer to the retailer, and the wholesale-price paid by the retailer to the program operator (i.e.: 9,550.32 IDR). Let ΔREDEEM (ΔSPEND) be the extra number of stamps redeemed (amount spent) due to push messaging (as displayed in Table 2.6), and MARGIN the retailer's average gross margin as a fraction of the sales price (which we set at .25, a typical retailer margin for groceries). The profitability of a push plan can then be approximated as:

$$\Delta\mathrm{PROFIT} = \mathrm{MARGIN} * \Delta\mathrm{SPEND} - \mathrm{LOSS\_REWARD} * (\Delta\mathrm{REDEEM}/\mathrm{STAMPS\_REWARD})$$

Table 2.7: Profitability of push messaging (actual push plan vs. control group)

		Change in		
Consumer Type	**Baseline spending (in IDR, across program weeks*)**	**Margin on amount spent (in IDR, across program weeks)**	**Cost on redeemed stamps (in IDR, across program weeks)**	**Profit (in IDR, across program weeks)**
Average	1,927,595	67,554.0	14,492.4	53,061.5
Low Quartile	1,717,076	54,910.4	12,337.1	42,573.3
Median	1,822,081	61,119.9	13,458.9	47,661.0
Upper Quartile	2,336,295	94,488.5	17,963.6	76,524.9

* Baseline spending operationalized as in Table 1.6. Note: 1000 IDR is approximately equal to 0.07 USD. Margin obtained as change in spending (see Table 1.6) times .25, a typical retailer margin for groceries. Cost obtained as change in redeemed stamps (see Table 1.6) divided by average number of stamps per reward (11.938) times average retailer loss per reward (9,550.32 IDR) (Figures obtained from the company).

[19] These are "approximate" wholesale prices because not every redemption requires the same number of stamps or additional payment by the consumer. Though we do not have details on the exact reward(s) redeemed by individual households by redemption occasion, we know the relative occurrence of the different redemption types in the entire program. Our wholesale prices are weighted averages with these occurrences as weights.

Table 2.7 presents the results of these calculations, for the actual push plan vs. the control. Though push messaging comes at a higher redemption cost, this is more than offset by the margin on additional consumer spending, in particular for heavy buyers. Counterfactuals reveal that this also holds for extra messages added to the actual plan, e.g., in weeks 4 and 10, or even more so in week 7. Hence, mobile push leads to incremental profit – especially among heavy customers, and for messages sent about halfway the program.

2.8 Conclusions, Limitations and Future Research

Retailer LPs are pervasive in grocery retailing. However, even if they last only for a limited number of weeks, interest among program participants typically wears off, as evidenced by declining sales effects and failure to redeem stamps. Whereas traditional communication has not revealed very effective at maintaining program engagement, mobile push notifications sent to adopters of a loyalty-program app have been advanced as a potentially promising approach. Still, their effects are not clear-cut upfront. Though they may successfully act as reminders and maintain program salience, repeated messages to consumers' personalized mobile devices may lose effect or even become counterproductive. Moreover, if such messages primarily stimulate consumers to trade in already accumulated stamps rather than collect additional stamps, the net outcome for the retailer may not be profitable. Our study empirically documents these issues, using a very large-scale field experiment. In so doing, it contributes to the literature on LPs, but also on the, still young, body of work on the effects of mobile marketing.

2.8.1 Findings

Redemption. Mobile push messaging appears to be a very effective tool to stimulate redemption among program participants. This is not surprising, in view of previous evidence that LP participants often pile up stamps far beyond the redemption threshold. Push messages sent to consumers' personal mobile devices then remind them of 'earned' rewards and provide a trigger to redeem those rewards. Our findings attest to this. Push messages produce a large immediate lift in the number of stamps redeemed, especially among heavy buyers (who can easily redeem stamps on their regular store visits), and more so when distance to the threshold is lower and stamps balance higher. This effect carries through to the next week because of rewarded behavior, to be followed by weeks in which consumers rebuild their stamp balance. The net result is a strong increase in number of stamps redeemed, especially from notifications sent later on in the program, and among the chain's large customers.

Spending. Push messaging also substantially enhances customers' total stamp collection and spending at the chain. This effect materializes along several countervailing, routes. First, push messages have a positive immediate effect on spending – confirming that mobile messages can serve as a reminder of the program and of the gain from continued participation (Kivetz, 2003). This immediate effect is larger for customers with high pre-program spending – the group typically found to yield lower LP returns (Bijmolt et al., 2011) – who regularly visit the chain anyway and can thus easily 'act' upon the message. For lighter customers, this effect is smaller, possibly because these households find it harder to reach the reward thresholds. While we find evidence of decreasing returns (sending more messages in a given week leads to lower additional sales) we do not observe negative effects from extra push notifications, suggesting that among consumers who opted in for a program app these notifications do not easily trigger irritation or reactance – at least within the data range.

Second, we find a negative post-message effect, especially for the chain's largest customers. These customers may have less leeway to further increase their expenditures, and partly compensate for extra spending at the time of receiving the message by reducing their outlay subsequently. Third, beyond that period, the message effect turns positive again for some time. This follows from savings-dynamics mechanisms: rewarded behavior, and the lower stamps balance, give consumers an incentive to again step up their stamp collection efforts and make up for the spending dip. The outcome of these countervailing forces is a positive impact of push on spending, especially for messages sent halfway in the program. Hence, unlike (e)mailing programs for which modest results have been reported at best (Lewis, 2004; Dorotic, 2010), mobile push messaging reveals to be a powerful instrument to enhance stamps-saving behavior in LPs – and the expenditures that go with it.

Relative impact on spending and redemption. A notable observation is that push messaging has a stronger effect on redemption than spending, and thus closes the 'redemption gap.' Reminding consumers that they can earn a reward makes them more prone to redeem already collected stamps than to save extra stamps. For program operators, this is good news: it implies less money left on the table. For retailers, it increases the redemption cost associated with the program. Even so, our results reveal that this extra cost is more than compensated by the margin gain from additional spending, implying that push messaging is profitable for the retailer as well. For the program on hand, the net effect of the push plan amounts to 11% of the baseline margin on treated customers[20] – a non-negligible figure.

2.8.2 Management Implications

Our results have important implications for managers involved in short-term LPs. They show that ongoing push notifications are a highly effective tool to ensure continued program engagement among participants enrolled in the program app, i.e., to make these consumers spend more and, especially, redeem more stamps. Given that push messaging significantly enhances the number of stamps redeemed, it can be used to tackle the phenomenon of un-redeemed stamps piling up – answering Dorotic et al. (2014)'s earlier call. Indeed, in the program under study we find that the current push plan reduces the number of unredeemed stamps by about one fourth – a very large effect. The redemption lift can

[20] Approximated as the marginal gain from the push plan (53,062 IDR, Table 1.6), divided by gross profits (25% of baseline spending in Table 1.6 or 481,899 IDR).

substantially benefit the program operator and reward manufacturer – stakeholders that hardly received any attention in previous literature. For the retailer, the additional sales generated by the push program (in our setting: about a 14% lift) more than make up for the higher costs from increased redemption. Such a positive effect corroborates retailers' current tendency – as observed based on anecdotal evidence – to favor large numbers of messages.

Our results also provide retailers with guidelines on how to further improve the sales impact of their mobile push plans. Because the expenditure lift is particularly strong among heavy buyers (who typically account for a large portion of total turnover: in our setting: the top-25 customers are good for about 75% of retailer revenue), retailers have an interest in targeting those customers. As for timing: scheduling push halfway (rather than early or late in the program) may improve the sales-effectiveness of messages sent. Targeting and timing also shape the redemption effect of the program, but not in the same way. Messages sent in *later* program weeks (especially to heavy buyers), lead to the largest redemption increases. It follows that in setting up a push plan, the interests of retailers may not be aligned with those of the reward manufacturer or program operator – an issue especially relevant if the latter acts as the 'mobile-push captain'. Much will also depend on the program design in terms of consumer requirements to obtain rewards, and retailer payments to the program operator per redeemed reward. Typically, a balance is struck such that retailers can earn back the redemption costs through extra sales, while program operators also make a profit. Our results show that even if both parties benefit, push messaging shifts this balance in favor of the program operator and may become a relevant aspect in the total program design.

2.8.3 Limitations and future research.

While this study provides new insights, it also has limitations that entail opportunities for future research. First, our results pertain to one (temporary) program in one country. We considered the more typical setting of an 'investment' program, in which the retailer (at least partly) pays for the rewards. If the retailer earns money on redeemed rewards (as is the case for 'for profit' programs), the increased redemption caused by push messaging will translate into even higher gains. Also, while the processes that we uncover may generalize to most LPs, the magnitude of the specific effects may well be context-dependent. For instance, cultural differences may affect how many push messages are deemed acceptable by consumers. Likewise, the exact nature of the reward or the redemption threshold may alter the timing of specific effects (e.g., the speed with which balance is restored). Last but not least, the absence of irritation effects may follow from the fact that in the program under study, consumers never received more than three push messages per week. Venkatesan and Kumar (2004) report that irritation effects may exist when communication frequencies are higher than those observed in our data.

Second, a broader question is whether our results generalize to permanent LPs where consumers collect points for expenditure that can be converted into rewards, e.g., frequent flier miles for free flights (Airlines), preferred customer programs for free nights or extra amenities (Hotels), customer cards for free coffee (Starbucks), etc.. While we expect the mechanisms to remain similar, the size of the message effects may change in a way that is not clear upfront. On the one hand, PLPs are bound to suffer more strongly from loss of salience and interest (Bijmolt et al., 2011), which could make for a stronger reminder effect of push messaging. On the other hand, push messages create a lower 'sense of urgency' in PLPs than TLPs. Even if consumers are reminded of the PLP and its rewards, there is less time pressure to redeem, and even less so to step up spending to qualify for redemption before saved points expire. Future research should document which of these forces prevail.

Third, our objective was to assess the impact of in-app push notifications on consumer engagement in a TLP. Hence, our population was restricted to card holders enrolled in the store's permanent LP (a prerequisite for TLP app adoption). Within that population, to ensure a clean comparison of test and control households, we estimated our effects on the early app adopters. Still, these constitute an important group, representing about 50% of all adopters, and spending 70% more per capita than late adopters.[21]

Fourth, we measured the impact of messages sent to consumers following their app adoption (and our test and control groups were random samples of app adopters), however, we do not know whether or when consumers read these messages, or possibly even disadopted the app (i.e., remove it from their mobile device). By the end of the program, we *do* observe lack of activity inside the app among certain households, but this may well be due to infrequent household purchasing rather than app disadoption. Indeed, such inactivity in final program weeks occurs mainly for households with low pre-program spending. Also, we do not find any link between such inactivity and the number of messages received by the household,[22] suggesting that app disadoption – if any – is not driven by push notifications.

Fifth, though we discussed possible causes of positive and negative message effects – including, for instance, reminder, discouragement and stockpiling effects – we could not directly quantify these intermediate process measures. Moreover, the push-message impact (over time) may differ depending on whether messages primarily remind consumers about available rewards, encourage them to save for those rewards, or stimulate them to redeem already collected stamps. For lack of information on the message content, we could not disentangle those effects, but leave that as a relevant topic for future study.

Finally, for lack of data, we only considered the impact of push messages during program weeks. To the extent that increased redemption also creates extra customer goodwill

[21] A caveat is that our spending data are recorded using consumers' loyalty card scans, while consumers can also purchase at the store without using their card. If push messages encourage card usage, we cannot disentangle that from spending increases and our spending impact may be somewhat overestimated. However, because much of our spending lift comes from an increase in quantity per visit, not visits, we do not expect this effect to be strong.

[22] We find a positive correlation between households' still being 'active' in the last program week and their pre-program spending (.233). Also, 'inactivity' is equally prevalent among households in the test and control group (21.29% vs. 19.43%; correlation between being active in week 44 and being in the control group: -.0102).

and commitment in the longer run – an aspect that we cannot measure here – our estimates of the push-message effects are conservative. We leave this as an area for future research.

2.9 Appendix A

Appendix A1: Program Reward Structure

Table A1.1: Program Reward Structure

Product Description	Spend Requirement in stamps	Promotional Price	Recommended Retail Price	Discount consumer %
Manual pump	10	900		98%
	5	19900		67%
			59900	
Vacuum container Rec 600ml	15	4900		95%
	10	24900		74%
			94900	
Vacuum container Rec/tal 1L	15	9900		90%
	10	29900		70%
			99900	
Vacuum container Rec/tal 1,4L	15	14900		87%
	10	34900		70%
			114900	
Vacuum container Rec 1,2L	15	19900		83%
	10	39900		67%
			119900	
Container Medium Square 1,5L	15	29900		76%
	10	49900		60%
			124900	

Appendix A2: Full Estimation Results

Table A2.1: Regression results for stamp collection / store expenditure

	(1)	(2)	(3)	(4)
$\text{BALANCE}_{i,t}$	-3.415***	-3.417***	-2.720***	-2.432***
	(0.108)	(0.108)	(0.155)	(0.201)
$\text{DIST}_{i,t}$	-6.331***	-6.292***	-4.915***	-2.531***
	(0.248)	(0.248)	(0.288)	(0.433)
$\text{PREVREDEMP}_{i,t}$	34.13***	34.29***	30.47***	30.74***
	(1.896)	(1.897)	(1.782)	(1.783)
$\text{ADOPT}_{i,t}$	101.4***	97.65***	106.7***	111.2***
	(2.668)	(2.703)	(3.097)	(3.123)
$\text{LOG\_MESSAGES}_{i,t}$		1.328***	1.256***	2.679***
		(0.245)	(0.247)	(0.437)
$\text{LOG\_MESSAGES}_{i,t-1}$		-1.453***	-1.596***	-1.800***
		(0.257)	(0.255)	(0.257)
$\text{LOG\_MESSAGES}_{i,t-2}$		-0.0338	-0.0138	-0.192
		(0.253)	(0.251)	(0.252)
$\text{BALANCE}_{i,t} \times \text{PREVSPEND}_{i}$			-0.00227***	-0.00227***
			(0.000326)	(0.000326)
$\text{DIST}_{i,t} \times \text{PREVSPEND}_{i}$			-0.00459***	-0.00449***
			(0.00127)	(0.00128)
$\text{PREVREDEMP}_{i,t} \times \text{PREVSPEND}_{i}$			0.0452***	0.0446***
			(0.00844)	(0.00844)
$\text{ADOPT}_{i,t} \times \text{PREVSPEND}_{i}$			0.268***	0.269***
			(0.0254)	(0.0255)
$\text{LOG\_MESSAGES}_{i,t} \times \text{PREVSPEND}_{i}$			0.0228***	0.0228***
			(0.00164)	(0.00186)
$\text{LOG\_MESSAGES}_{i,t-1} \times \text{PREVSPEND}_{i}$			-0.00788***	-0.00803***
			(0.00156)	(0.00157)
$\text{LOG\_MESSAGES}_{i,t-2} \times \text{PREVSPEND}_{i}$			-0.0156***	-0.0155***
			(0.00157)	(0.00157)
$\text{LOG\_MESSAGES}_{i,t} \times \text{BALANCE}_{i,t}$				-0.0685*
				(0.0304)
$\text{LOG\_MESSAGES}_{i,t} \times \text{DIST}_{i,t}$				-0.578***
				(0.0778)
Constant	192.6***	192.5***	168.1***	144.4***
	(3.136)	(3.136)	(3.851)	(4.988)
Observations	651056	651056	651056	651056
R^2	0.038	0.038	0.042	0.042

All models include household and time fixed effects. Standard errors in parentheses. Reported R-squares net of fixed effects.

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2.2: Logit results for redemption (Y/N).

	(1)	(2)	(3)	(4)
$\text{BALANCE}_{i,t}$	0.00615***	0.00616***	0.0117***	0.0137***
	(0.000178)	(0.000180)	(0.000246)	(0.000441)
$\text{DIST}_{i,t}$	-0.174***	-0.171***	-0.163***	-0.185***
	(0.00135)	(0.00136)	(0.00145)	(0.00364)
$\text{PREVREDEMP}_{i,t}$	1.291***	1.257***	1.260***	1.256***
	(0.00978)	(0.00988)	(0.0108)	(0.0108)
PREVSPEND_{i}	0.00103***	0.00105***	0.00158***	0.00150***
	(0.0000161)	(0.0000162)	(0.0000411)	(0.0000423)
$\text{LOG\_MESSAGES}_{i,t}$		0.107***	0.106***	0.125***
		(0.00297)	(0.00309)	(0.00356)
$\text{LOG\_MESSAGES}_{i,t-1}$		0.0375***	0.0377***	0.0391***
		(0.00326)	(0.00339)	(0.00340)
$\text{LOG\_MESSAGES}_{i,t-2}$		0.00674*	0.00910**	0.0106***
		(0.00284)	(0.00291)	(0.00292)
$\text{BALANCE}_{i,t} \times \text{PREVSPEND}_{i}$			-0.0000152***	-0.0000152***
			(0.000000510)	(0.000000506)
$\text{DIST}_{i,t} \times \text{PREVSPEND}_{i}$			0.000136***	0.000137***
			(0.00000469)	(0.00000472)
$\text{PREVREDEMP}_{i,t} \times \text{PREVSPEND}_{i}$			-0.000337***	-0.000333***
			(0.0000323)	(0.0000323)
$\text{LOG\_MESSAGES}_{i,t} \times \text{PREVSPEND}_{i}$			-0.0000108	0.00000740
			(0.00000772)	(0.00000811)
$\text{LOG\_MESSAGES}_{i,t-1} \times \text{PREVSPEND}_{i}$			-0.0000208*	-0.0000200*
			(0.00000838)	(0.00000839)
$\text{LOG\_MESSAGES}_{i,t-2} \times \text{PREVSPEND}_{i}$			-0.0000364***	-0.0000360***
			(0.00000716)	(0.00000716)
$\text{LOG\_MESSAGES}_{i,t} \times \text{BALANCE}_{i,t}$				-0.000468***
				(0.0000871)
$\text{LOG\_MESSAGES}_{i,t} \times \text{DIST}_{i,t}$				0.00497***
				(0.000744)
Constant	-2.331***	-2.685***	-3.225***	-3.146***
	(0.0411)	(0.0415)	(0.0487)	(0.0526)
Observations	636981	636981	636981	636981
Pseudo-R^2	0.154	0.161	0.169	0.169

All models include time fixed effects. Standard errors in parentheses. The number of observations is exclusive of consumers who never redeem.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table A2.3: Stamps redeemed given redemption.

	(1)	(2)	(3)	(4)
BALANCE$_{i,t}$	0.329***	0.329***	0.340***	0.319***
	(0.0103)	(0.00986)	(0.0235)	(0.0292)
DIST$_{i,t}$	-1.265***	-1.266***	-0.675*	-0.686*
	(0.195)	(0.181)	(0.288)	(0.350)
PREVREDEMP$_{i,t}$	9.470***	9.429***	5.557**	4.777*
	(1.384)	(1.274)	(2.143)	(2.163)
LOG_MESSAGES$_{i,t}$		0.987***	0.610**	0.515*
		(0.129)	(0.200)	(0.241)
LOG_MESSAGES$_{i,t-1}$		0.205**	0.0472	0.0241
		(0.0673)	(0.0924)	(0.0945)
LOG_MESSAGES$_{i,t-2}$		0.0607	0.0762	0.0731
		(0.0456)	(0.0477)	(0.0485)
BALANCE$_{i,t}$ × PREVSPEND$_{i}$			-0.0000818*	-0.0000734*
			(0.0000357)	(0.0000358)
DIST$_{i,t}$ × PREVSPEND$_{i}$			0.000357	0.000259
			(0.000278)	(0.000284)
PREVREDEMP$_{i,t}$ × PREVSPEND$_{i}$			-0.000295	0.0000332
			(0.00104)	(0.00104)
LOG_MESSAGES$_{i,t}$ × PREVSPEND$_{i}$			0.000111	0.0000505
			(0.000153)	(0.000156)
LOG_MESSAGES$_{i,t-1}$ × PREVSPEND$_{i}$			0.000105	0.000132
			(0.000147)	(0.000147)
LOG_MESSAGES$_{i,t-2}$ × PREVSPEND$_{i}$			-0.000268*	-0.000248*
			(0.000122)	(0.000122)
LOG_MESSAGES$_{i,t}$ × BALANCE$_{i,t}$				0.00335
				(0.00281)
LOG_MESSAGES$_{i,t}$ × DIST$_{i,t}$				0.0272
				(0.0171)
Constant	-6.858	-10.93*	1.317	5.038
	(4.233)	(4.349)	(7.858)	(7.819)
CF$_{i,t}$	-6.845***	-7.017***	-3.219	-2.467
	(1.274)	(1.207)	(1.995)	(2.020)
Observations	93211	93211	93211	93211
R^2	0.167	0.168	0.167	0.168

All models include household and time fixed effects. Standard errors in parentheses. Reported R-squares net of fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Appendix A3: Marginal Impact of Push Messages

Figure A3.1: Impact of a push message on spending, redemption incidence and number of stamps redeemed in the same week, as a function of available stamps (for consumer with average pre-program spending)

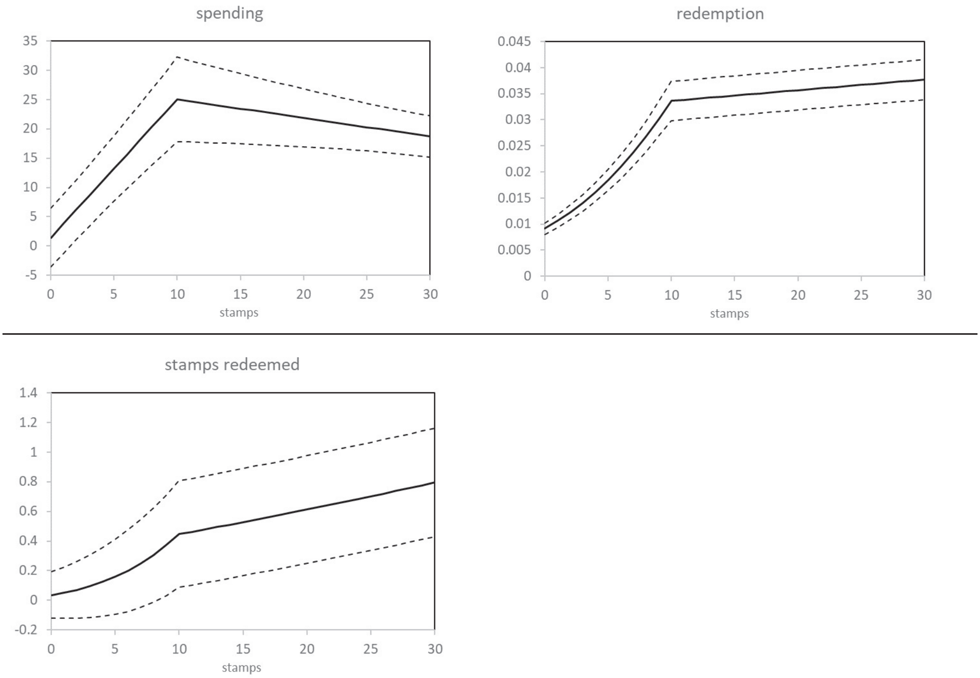
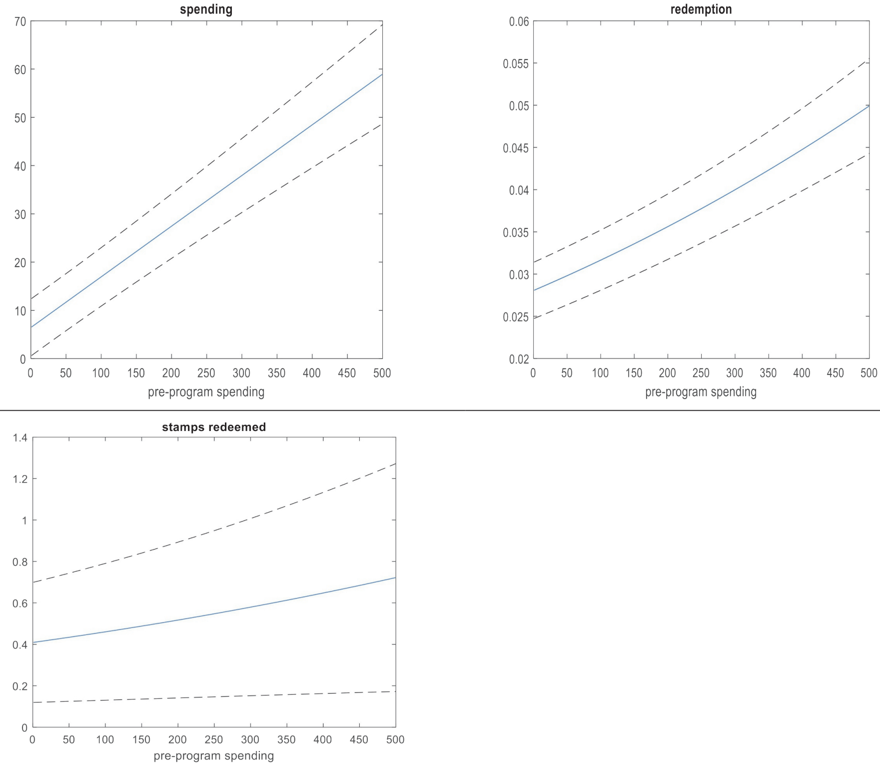


Figure A3.2: Impact of a push message on spending, redemption incidence and number of stamps redeemed in the same week, as a function of consumers' pre-program spending (in 1,000 IDR), for an average number of available stamps.



Appendix A4: Simulation Procedure

To assess the net effects of push messaging over time for customers with different levels of (pre-program) spending with the retailer, we use our estimates as inputs for dynamic simulations (see, e.g., Ataman, van Heerde, & Mela, 2010, and van Heerde, Gijsbrechts, & Pauwels, 2015, for a similar approach). Starting from the first week of the program, we predict consumers' spending and number of stamps redeemed in the course of the program for alternative push plans, and compare the predicted spending and redemption trajectories.[23] We do so for four different 'customer profiles' in terms of prior spending at the chain – i.e. for average, lower-quartile, median, and upper-quartile pre-program spending levels.

Specifically, we proceed as follows. In each week, we predict consumers' spending levels based on the estimates of Equation 2.1. Next, to dynamically forecast whether or not consumers will redeem, we use the logit-model estimates (i.e., the estimates from the first layer of the hurdle model) to predict the redemption incidence probability for the considered week. We then use a draw from a uniform distribution between zero and one, to convert this probability into a zero-one variable for that week (Specifically, we set redemption incidence to one if the draw is larger than the predicted probability, and to zero otherwise). If redemption incidence is zero, so is the number of stamps redeemed. If it is one, we predict the number of stamps redeemed using the estimates of the second layer of our hurdle model. We then move on to the next week, in which we set the lagged redemption-incidence regressor for spending, update the levels of balance and distance, and repeat the previous steps till the last week of the program, to obtain a complete trajectory (from the first till the last program week). We repeat this process for a thousand trajectories (sets of incidence draws) and then aggregate across these trajectories. We report the results of these averages in Table 2.6 of the

[23] We also consider *predicted* values for the base case, to rule out the impact of noise in the actual spending and redemption levels, and obtain a clean basis for comparison with the other scenarios.

main text. To account for uncertainty in the estimated parameters, we repeat these simulations for 1000 draws from the parameter sampling distributions, to assess the significance of the differences between alternative push plans, and obtain the confidence intervals in Figure 2.4 in the main text.

In the main simulations, we use logical restrictions: we use a zero lower bound for spending and number of stamps redeemed, and an upper bound for the number of stamps redeemed equal to the number of available stamps (that is, the stamps balance at the beginning of the week plus the number of collected stamps in the considered week). As a check, we also re-ran the simulations without such restrictions, and found the results to remain highly similar.

References Appendix A4:

Ataman, B., H. J. van Heerde, & C. F. Mela (2010). The long-term effect of marketing strategy on brand sales. *Journal of Marketing Research*, *47* (December), 866–82.

Van Heerde, H. J., E. Gijsbrechts & K. Pauwels (2015). Fanning the flames? How media coverage of a price war affects retailers, consumers, and investors. *Journal of Marketing Research*, *52* (October), 674-693.

Appendix A5: Robustness Checks

Table A5.1: Robustness against Alternative Distance Threshold of 5 – Panel A: Estimation Results

	Spending	Redemption Incidence	# of Stamps Redeemed
$\mathrm{BALANCE}_{i,t}$	-2.275***	0.0197***	0.239***
	(0.184)	(0.000458)	(0.0349)
$\mathrm{DIST5}_{i,t}$	-3.642***	-0.309***	0.844
	(0.655)	(0.00788)	(0.575)
$\mathrm{PREVREDEMP}_{i,t}$	29.49***	1.250***	-2.277
	(1.823)	(0.0107)	(2.069)
$\mathrm{PREVSPEND}_{i}$		0.00198***	
		(0.0000412)	
$\mathrm{ADOPT}_{i,t}$	111.1***		
	(3.119)		
$\mathrm{LOG\_MESSAGES}_{i,t}$	3.912***	0.125***	-0.262
	(0.544)	(0.00351)	(0.231)
$\mathrm{LOG\_MESSAGES}_{i,t-1}$	-1.800***	0.0391***	-0.222*
	(0.257)	(0.00339)	(0.0925)
$\mathrm{LOG\_MESSAGES}_{i,t-2}$	-0.155	0.0101***	0.0123
	(0.252)	(0.00291)	(0.0482)
$\mathrm{BALANCE}_{i,t} \times \mathrm{PREVSPEND}_{i}$	-0.00253***	-0.0000242***	0.0000260
	(0.000304)	(0.000000523)	(0.0000447)
$\mathrm{DIST5}_{i,t} \times \mathrm{PREVSPEND}_{i}$	-0.0104***	0.000194***	-0.00108*
	(0.00234)	(0.00000977)	(0.000425)
$\mathrm{PREVREDEMP}_{i,t} \times \mathrm{PREVSPEND}_{i}$	0.0453***	-0.000357***	0.00302**
	(0.00846)	(0.0000326)	(0.00104)
$\mathrm{ADOPT}_{i,t} \times \mathrm{PREVSPEND}_{i}$	0.270***		
	(0.0256)		
$\mathrm{LOG\_MESSAGES}_{i,t} \times \mathrm{PREVSPEND}_{i}$	0.0233***	0.00000607	0.000120
	(0.00187)	(0.00000825)	(0.000158)
$\mathrm{LOG\_MESSAGES}_{i,t-1} \times \mathrm{PREVSPEND}_{i}$	-0.00789***	-0.0000183*	0.000331*
	(0.00157)	(0.00000847)	(0.000145)
$\mathrm{LOG\_MESSAGES}_{i,t-2} \times \mathrm{PREVSPEND}_{i}$	-0.0154***	-0.0000413***	-0.0000475
	(0.00157)	(0.00000722)	(0.000126)
$\mathrm{LOG\_MESSAGES}_{i,t} \times \mathrm{BALANCE}_{i,t}$	-0.0492	-0.000504***	0.00598*
	(0.0280)	(0.0000906)	(0.00264)
$\mathrm{LOG\_MESSAGES}_{i,t} \times \mathrm{DIST5}_{i,t}$	-0.890***	0.00851***	-0.0178
	(0.122)	(0.00161)	(0.0309)
$\mathrm{CF}_{i,t}$			4.327*
			(1.952)
Constant	137.2***	-3.560***	31.47***
	(4.082)	(0.0529)	(8.014)
Observations	651056	636981	93211
R^2 (pseudo-R^2)	0.041	(0.160)	0.167

Standard errors in parentheses. Reported R-squares net of fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Panel B: Simulation Outcomes: Actual push plan vs. control group

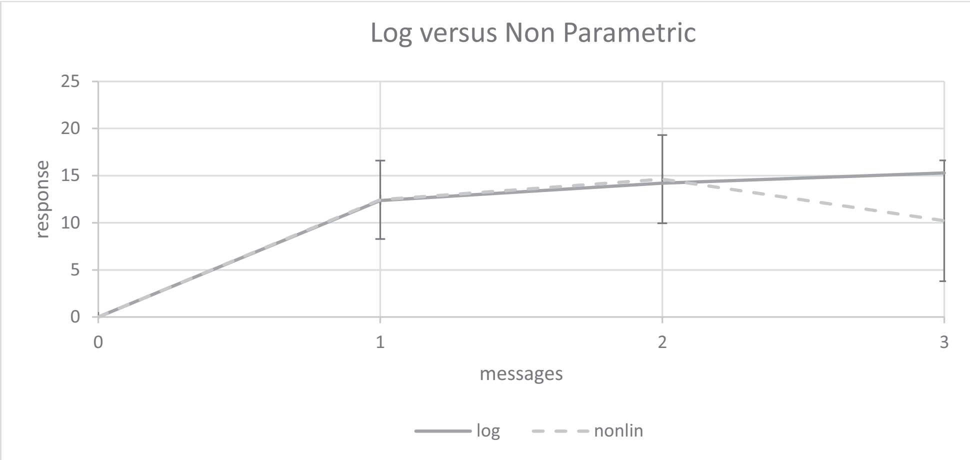
	Change in					
	Spending (in IDR, across program weeks)		**Number of Stamps Redeemed (across program weeks)**		**Stamp Balance (by the end of the program)**	
Model:	**Main**	**Threshold=5**	**Main**	**Threshold=5**	**Main**	**Threshold=5**
Consumer Type						
Average	270,216	218,739	18.12	16.39	-10.73	-10.87
Low Quartile	219,642	161,900	15.42	14.08	-9.30	-9.78
Median	244,480	190,994	16.82	15.28	-10.11	-10.40
Upper Quartile	377,954	337,653	22.45	20.60	-12.44	-12.28

Table A5.2: Robustness against Non-parametric Message Impact – Panel A: Estimation Results

	Spending	Redemption Incidence	# of Stamps Redeemed
$\text{BALANCE}_{i,t}$	-2.433***	0.0136***	0.322***
	(0.201)	(0.000440)	(0.0290)
$\text{DIST}_{i,t}$	-2.513***	-0.186***	-0.734*
	(0.433)	(0.00361)	(0.347)
$\text{PREVREDEMP}_{i,t}$	30.76***	1.252***	5.039*
	(1.784)	(0.0108)	(2.138)
PREVSPEND_{i}		0.00149***	
		(0.0000422)	
$\text{ADOPT}_{i,t}$	111.2***		
	(3.128)		
$\text{1 MESSAGE}_{i,t}$	12.44***	0.478***	2.131*
	(2.123)	(0.0185)	(0.948)
$\text{2 MESSAGES}_{i,t}$	14.63***	0.712***	3.107*
	(2.388)	(0.0196)	(1.349)
$\text{3 MESSAGES}_{i,t}$	10.21**	0.810***	3.144*
	(3.271)	(0.0339)	(1.560)
$\text{LOG\_MESSAGES}_{i,t-1}$	-1.793***	0.0360***	0.0251
	(0.256)	(0.00340)	(0.0898)
$\text{LOG\_MESSAGES}_{i,t-2}$	-0.169	0.0105***	0.0775
	(0.252)	(0.00292)	(0.0484)
$\text{BALANCE}_{i,t} \times$ PREVSPEND_{i}	-0.00227***	-0.0000152***	-0.0000763*
	(0.000326)	(0.000000506)	(0.0000356)
$\text{DIST5}_{i,t} \times$ PREVSPEND_{i}	-0.00449***	0.000137***	0.000295
	(0.00128)	(0.00000472)	(0.000281)
$\text{PREVREDEMP}_{i,t} \times$ PREVSPEND_{i}	0.0446***	-0.000330***	-0.0000694
	(0.00844)	(0.0000323)	(0.00103)
$\text{ADOPT}_{i,t} \times$ PREVSPEND_{i}	0.269***		
	(0.0255)		
$\text{LOG\_MESSAGES}_{i,t} \times$ PREVSPEND_{i}	0.0228***	0.00000843	0.0000504
	(0.00186)	(0.00000807)	(0.000156)
$\text{LOG\_MESSAGES}_{i,t-1} \times$ PREVSPEND_{i}	-0.00802***	-0.0000205*	0.000123
	(0.00157)	(0.00000838)	(0.000147)
$\text{LOG\_MESSAGES}_{i,t-2} \times$ PREVSPEND_{i}	-0.0155***	-0.0000363***	-0.000256*
	(0.00157)	(0.00000716)	(0.000122)
$\text{LOG\_MESSAGES}_{i,t} \times$ $\text{BALANCE}_{i,t}$	-0.0682*	-0.000459***	0.00328
	(0.0304)	(0.0000869)	(0.00281)
$\text{LOG\_MESSAGES}_{i,t} \times$ $\text{DIST}_{i,t}$	-0.581***	0.00505***	0.0292
	(0.0778)	(0.000737)	(0.0172)
$\text{CF}_{i,t}$			-2.731
			(2.002)
Constant	144.2***	-3.134***	4.061
	(4.992)	(0.0527)	(7.731)
Observations	651056	636981	93211
R^2 (pseudo-R^2)	0.042	(0.169)	0.168

Standard errors in parentheses. Reported R-squares net of fixed effects. * p < 0.05, ** p < 0.01, *** p < 0.001.

Note on Table A5.2: The pattern of main-message effects in the non-parametric robustness check also shows decreasing returns to messages, and it is statistically equivalent to that of the proposed specification – as shown in the graph below, which displays the (same-week) spending impact of one, two or three messages within a week, for an 'average' consumer. We note, though, that the coefficient of 'three messages' in the nonparametric model must be treated with caution, given the few instances in our data.



Panel B: Simulation Outcomes: Actual push plan vs. control group

	Change in					
	Spending (in IDR, across program weeks)		**Number of Stamps Redeemed (across program weeks)**		**Stamp Balance (by the end of the program)**	
Model:	**Main**	**Nonparametric Message Impact**	**Main**	**Nonparametric Message Impact**	**Main**	**Nonparametric Message Impact**
Consumer Type						
Average	270,216	278,485	18.12	18.01	-10.73	-10.67
Low Quartile	219,642	226,247	15.42	15.33	-9.30	-9.39
Median	244,480	251,653	16.82	16.64	-10.11	-10.05
Upper Quartile	377,954	393,831	22.45	22.99	-12.44	-12.86

Table A5.3: Robustness against Including Lagged Spending – Panel A: Estimation Results

	Spending	Redemption Incidence	# of Stamps Redeemed
$\text{BALANCE}_{i,t}$	-3.227***	0.00693***	0.306***
	(0.189)	(0.000471)	(0.0177)
$\text{DIST}_{i,t}$	0.0107	-0.173***	-0.735**
	(0.416)	(0.00367)	(0.230)
$\text{PREVREDEMP}_{i,t}$	-6.404***	0.879***	4.350***
	(1.737)	(0.0119)	(0.988)
$\text{SPENDING}_{i,t-1}$	0.157***	0.00113***	0.00217
	(0.00454)	(0.0000150)	(0.00114)
PREVSPEND_{i}		0.00117***	
		(0.0000443)	
$\text{ADOPT}_{i,t}$	108.0***		
	(3.074)		
$\text{LOG\_MESSAGES}_{i,t}$	2.772***	0.127***	0.564**
	(0.419)	(0.00364)	(0.176)
$\text{LOG\_MESSAGES}_{i,t-1}$	-1.663***	0.0390***	0.0403
	(0.262)	(0.00347)	(0.0728)
$\text{LOG\_MESSAGES}_{i,t-2}$	0.237	0.00962**	0.0767
	(0.247)	(0.00298)	(0.0469)
$\text{BALANCE}_{i,t} \times \text{PREVSPEND}_{i}$	-0.00175***	-0.0000122***	-0.0000702**
	(0.000306)	(0.000000552)	(0.0000255)
$\text{DIST}_{i,t} \times \text{PREVSPEND}_{i}$	-0.00317**	0.000145***	0.000331
	(0.00122)	(0.00000486)	(0.000214)
$\text{PREVREDEMP}_{i,t} \times \text{PREVSPEND}_{i}$	0.0304***	-0.000422***	-0.000348
	(0.00794)	(0.0000336)	(0.000838)
$\text{ADOPT}_{i,t} \times \text{PREVSPEND}_{i}$	0.321***		
	(0.0253)		
$\text{LOG\_MESSAGES}_{i,t} \times \text{PREVSPEND}_{i}$	0.0220***	-0.0000103	0.000000914
	(0.00180)	(0.00000833)	(0.000159)
$\text{LOG\_MESSAGES}_{i,t-1} \times \text{PREVSPEND}_{i}$	-0.00617***	-0.0000125	0.000130
	(0.00161)	(0.00000874)	(0.000137)
$\text{LOG\_MESSAGES}_{i,t-2} \times \text{PREVSPEND}_{i}$	-0.00855***	-0.00000401	-0.000200
	(0.00152)	(0.00000740)	(0.000111)
$\text{LOG\_MESSAGES}_{i,t} \times \text{BALANCE}_{i,t}$	-0.0921**	-0.000497***	0.00318
	(0.0284)	(0.0000927)	(0.00278)
$\text{LOG\_MESSAGES}_{i,t} \times \text{DIST}_{i,t}$	-0.492***	0.00583***	0.0316*
	(0.0738)	(0.000751)	(0.0152)
$\text{CF}_{i,t}$			-2.894*
			(1.348)
Constant	96.90***	-3.667***	2.154
	(4.960)	(0.0591)	(5.861)
Observations	651056	636981	93211
R^2 (pseudo-R^2)	0.063	(0.187)	0.168

Standard errors in parentheses. Reported R-squares net of fixed effects. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Panel B: Simulation Outcomes: Actual push plan vs. control group						
	Change in					
	Spending (in IDR, across program weeks)		**Number of Stamps Redeemed (across program weeks)**		**Stamp Balance (by the end of the program)**	
Model:	**Main**	**With lagged Spending**	**Main**	**With lagged Spending**	**Main**	**With lagged Spending**
Consumer Type						
Average	270,216	199,033	18.12	9.39	-10.73	-4.17
Low Quartile	219,642	130,379	15.42	6.89	-9.30	-3.38
Median	244,480	163,703	16.82	8.15	-10.11	-3.81
Upper Quartile	377,954	336,625	22.45	13.89	-12.44	-5.07

Note: We did not retain the model with lagged spending as our final specification. The reason is that the use of household fixed effects together with a lagged dependent variable may bias the individual coefficients of the spending model. While it is technically feasible to tackle this issue using Arellano-Bond (AB) estimators, the problem with these estimators is that the outcomes are very sensitive to choices for instruments (for which there is little guidance). We also found that to be the case here. The results in Table A5.3 are the 'regular' regression results, without AB correction. Although it is very easy to obtain much stronger effects of push messaging with AB than the ones we are reporting (and we did find much larger push message effects with Arellano-Bond estimators!), we are mindful of the fact that this is commonplace with cases where instruments such as lagged DVs (as used in AB) are weak. Such estimators often suffer from finite-sample bias and large sampling errors, which could make the 'cure' worse than the disease (Rossi, 2014). So, we consciously did not take this route.

3. Drivers and Consequences of In-Store Promotion-Execution Quality: An Analysis for Temporary Loyalty Programs

3.1 Introduction

Despite the increase in online grocery shopping in the wake of the Covid-pandemic, the bulk of grocery purchases continues to occur offline. As such, the physical store environment remains a critical customer touchpoint, and retail execution an important performance driver. Retail execution "*describes activities performed at the store-level […] aimed at increasing brick-and-mortar sales*" (Brogie, 2019) – including replenishment, training of store personnel, and management of in-store signage and displays (Nordfält, Grewal, Roggeveen, & Hill, 2014; Roggeveen & Grewal, 2018). These activities are essential not only for retailers themselves, but also for their external partners. Consumer packaged goods (CPG) brands rely on their promotional initiatives being implemented according to plan (Krishna, 2019). Likewise, the success of (temporary) loyalty programs (LPs), which can be seen as promotional campaigns for retailers, hinges on retailer commitment and proper implementation of in-store support[24]. According to an industry report by Deloitte (Gomez & Sides, 2015), in a highly competitive environment the execution of store strategy is crucial in gaining a competitive advantage.

While important, execution – and retail execution in particular – is also notoriously difficult. As indicated by Gomez and Sides (2015) "*Although the challenge of effective execution transcends industries, the retail industry is a particularly challenging environment in which to execute effectively. […] It is difficult to achieve consistency across thousands of stores, not to mention train, retain, and motivate thousands of hourly associates of varying ages with different levels of skill. Ninety percent of organizations fail to effectively execute*

[24] Based on discussions with managers at BrandLoyalty, a global loyalty program operator.

their strategies and realize the full benefits of their efforts." Especially for strategies laid out by external partners, commitment to execution is often poor (Sull, Homkes, & Sull, 2015). For instance, based on a large-scale audit among UK retailers, POPAI (2015) reports that whereas over 80% of interviewed CPG brand managers expected their promotional displays to be correctly implemented, only 41% of retail stores had the planned display, as defined by the brand manager, properly executed. Increasing promotion intensity and macro shifts are bound to worsen the situation: *"Every store is trying its best, but with fewer staff on the rota and more tasks to complete, store teams have a heavier workload than ever post-COVID-19. As a result, [...] standards [....] inevitably slip"* (Yoobic, 2020). It follows that poor retail execution and lack of compliance with promotional plans are an underestimated problem, and a growing cause for concern.

To date, however, little is known about the magnitude and drivers of execution deviations, or about their impact on the sales outcomes of in-store campaigns.[25] One reason is lack of data. Specific number on deviations from planned execution reported in the literature often pertain to strategies in general, and are highly variable and hard to compare (Cândido & Santos, 2015). Information on in-store compliance for *promotional* campaigns is seldom collected[26]. If so, it is typically obtained only at the start of the campaign. This is a problem because execution quality may vary in the course of the campaign (POPAI, 2015). For instance, it may deteriorate if replenishment falls short, displays become disorderly, or signs are taken down. It could also go up again if stores are replenished, or if store managers properly re-arrange promotion materials. Moreover, the collected information is often partial, in that it pertains only to a specific instrument (e.g., whether or not a display is placed, see Hacker, Floerkemeier, Sarma, & Schuh, 2010) and to a non-representative subset of stores.

[25] Insights into the consequences of poor retail execution are vague and imprecise – reported losses ranging from "3.7% of sales" (Skorupa, 2018) to "40% of the campaign's potential" (Mankins & Steele, 2017).
[26] According to POPAI (2015), as much as 30% of all P-O-P activity is still never or rarely measured.

Yet, stores may well differ in the quality of their execution for different instruments (Raman, DeHoratius, & Ton, 2001). Which stores are more subject to promotional execution deviations and why is an important managerial issue that, for lack of data, has remained largely unexplored.

Even if execution quality is measured, assessing the consequences of execution deviations is particularly challenging. It requires separating the impact of non-compliance from the effect of the campaign 'as such', and from other temporal or cross-sectional factors. Such rigorous analysis seems currently lacking, prompting industry analysts and scholars to call for more research on the topic (Ailawadi, Beauchamp, Donthu, Gauri, & Shankar, 2009; POPAI, 2015).

We intend to address this call. To this end, we consider the setting of temporary retailer loyalty programs (TLPs). These are campaigns at a retail chain operated by a third party (the 'program operator'), with the aim to enhance retailer sales (and reward redemption). During TLPs, which typically run for several weeks, consumers can collect stamps at any store of the retailer at a fixed spending rate per stamp (e.g., 1 stamp for every 10 dollars spent), which can then be used to redeem a reward at a large discount for a pre-specified number of stamps. Common examples of reward categories of TLPs include crystal-, cooking-, and cutting ware. Like other promotional campaigns, TLPs depend on in-store support and execution. In a typical scenario, the program operator sets up a plan for in-store support and provides materials and directions to retailers. However, individual stores may not comply with the plan, which, if unwarranted, could diminish (or jeopardize) the success of the TLP.

TLPs constitute a particularly interesting setting for our purposes. First, they represent an important and growing market: in recent years, the use of TLPs has become pervasive globally (Bombaij, Gelper, & Dekimpe, 2022). Second, TLPs involve multiple in-store instruments, including *signage* (e.g., posters and other marketing materials), *displays* (promotional racks containing rewards), *replenishment activities* (ensuring reward availability), and *staff* (drawing attention to/providing info about the program). Third, the planned support and timing/duration of the TLP are typically 'standardized' across retailer stores – leading to a common baseline against which the impact of store-specific non-compliance can be assessed. With this in mind, we aim to address the following research questions:

First, what is the quality of in-store execution of TLPs, for different in-store instruments, over time? Does this execution quality differ between stores, and what underlies these differences? Second, to what extent does imperfect compliance for these in-store instruments affect store sales? Third, what are the implications for retailers and program operators? How should they prioritize their efforts to improve retail execution?

To address these questions, we leverage a unique data set. In this data set, the program operator collects information on in-store execution throughout the program, based on a longitudinal survey across a large (representative) sample of participating stores. For these individual stores, we have data on their size, as well as on characteristics of the region in which they operate. We combine this information with sales data from these stores before, during, and after the program. This panel data set allows us to gauge and explain the magnitude of store-specific execution divergence. It also allows us to separate the sales effect of these deviations from the impact of the program (and associated 'planned' in-store support) as such, as well as from other temporal confounds.

Our contribution is twofold. First, we show that poor in-store execution of planned support is a non-negligible phenomenon. We measure the size of execution problems and show that deviations from planned support are often the rule and not the exception. Focusing on heterogeneity of program execution, we also study whether and to what extent the execution deviations of TLP's are concentrated in certain stores and uncover the link with store size and region characteristics. For instance, we report that the execution of in-store signage, display and reward availability is lower in smaller stores and densely populated markets, while staff knowledge about the program is worse in urban markets.

Second, the pervasiveness of execution problems motivates estimating the wedge between the effect on retailer performance of existing TLPs and the counterfactual effect of well-executed TLPs. To this end, we document the impact of deviations in TLP execution on sales outcomes. Rather than just assessing the effect of whether a support instrument is used (e.g., whether a display is placed – the typical measure in previous literature), we assess to what extent the instrument is managed according to standard (i.e., does the display look orderly, is it placed in the correct spot, etc.), and how that influences the sales outcome of the TLP. We establish the relative magnitude of these effects for the different types of in-store support, and for stores with different characteristics. We obtain an interesting pattern of *influence*s that, combined with the observed *levels* of execution deviations, can guide retailers and external parties in optimizing the sales effectiveness of promotional execution. For instance, investing in the training of store staff reveals to be particularly beneficial in large stores and areas with low socio-economic development. It is less effective in smaller stores, where ensuring display quality is of primary importance, and urban areas, where signage plays a larger role.

The study is organized as follows. Section 2 discusses relevant background literature and builds the conceptual framework. Section 3 presents the unique data used in this research and lays out the variable operationalizations. Section 4 describes the methodology, and section 5 provides the empirical results. Finally, section 6 concludes and discusses the implications.

3.2 Background Literature & Conceptual Framework

3.2.1 In-Store Marketing and In-Store Execution

Roggeveen and Grewal (2018) define in-store marketing as "*all activities that a retailer undertakes within the store to engage with customers and get customers to engage with the retailer's goods and services*" (p.1). These include visual signage (marketing materials), displays, merchandise availability, and service personnel (see, e.g., Nordfält, et al., 2014; Roggeveen & Grewal, 2018). Previous studies have documented the impact of in-store instruments on sales, in general and in the context of promotional campaigns. In-store marketing can increase consumers' time spent in the store (Baker, Levy, & Grewal, 1992; Donovan & Rossiter, 1982; Helmefalk & Hultén, 2017; Mehrabian & Russell, 1974) and shape their decisions at the point-of-purchase (Nordfält, et al., 2014) – which is important as most decisions are made on the spot (Stilley, Inman, & Wakefield, 2010). More specifically, signage and displays have been found to direct consumers' attention to the focal brand and to signal the presence of a promotion even if no actual price cut is offered – thereby strongly lifting brand sales (e.g., Bemmaor & Mouchoux, 1991; e.g., Dhar, Hoch, & Kumar, 2001; East, Eftichiadou, & Williamson, 2003; Inman, McAlister, & Hoyer, 1990). Merchandise availability, too, is shown to play a critical role: while less-than-complete stocks can signal scarcity and increase product appeal (Castro, Morales, & Nowlis, 2013; van Herpen, Pieters, & Zeelenberg, 2005), out-of-stocks typically lead to consumer dissatisfaction with the retailer and to foregone sales (Breugelmans, Gijsbrechts, & Campo, 2018; Byun & Sternquist, 2012; Corsten & Gruen, 2003). Finally, Baker, Parasuraman, Grewal, and Voss (2002) and Grewal, Roggeveen, Sisodia, and Nordfält (2017) argue that interactions with the store's staff contribute to the overall in-store experience and can influence customer engagement – and thus constitute a crucial component of in-store marketing (Roggeveen & Grewal, 2018). In brief, in-store instruments are found to be a determining factor to maintain and increase customer engagement and spending.

Yet, the impact of these in-store instruments hinges on proper execution thereof – which turns out to be particularly challenging. Promotions cause a significant burden on the store personnel (Chain Store Age, 2007) and execution on the store floor is typically far from perfect (Skorupa, 2018; Yoobic, 2020). At the same time, monitoring and increasing execution quality is costly (Krishna, 2019). Several industry reports therefore point to in-store execution failures and insufficient compliance with promotional plans as a critical problem (POPAI, 2015; Sull et al., 2015) – even if the actual size of the problem remains unclear due to a lack of consistent and uniform reporting of execution quality (Cândido & Santos, 2015).

Extant knowledge on in-store execution deviations is scant. Available evidence mostly stems from industry reports (POPAI, 2015; Skorupa 2018). While academic marketing literature on the quality of in-store execution is virtually non-existent, studies in supply-chain management do shed some light on the topic. Notable examples are Hacker et al. (2010), who analyze the promotion execution in 10 stores of a US retailer – tracking the timely placement of promotional displays with RFID and showing that only 28 % arrive on the shop floor within 3 days of the campaign start, and Raman et al. (2001), who study inventory record inaccuracies and misplaced SKUs, and the underlying causes, across multiple outlets of two leading retail chains. Together, these studies suggest that execution deviations (i) can be important, (ii) differ between stores, even within a given retail chain, (iii) are partly attributable to factors out of the store's control (e.g., the store's distribution center), but (iv) also originate from local store managers' lack of awareness and failure to see the potentially detrimental consequences. Still, extant studies on promotional execution[27]

[27] There is a stream of papers on execution problems in the supply chain – see, e.g., Raman et al. (2001) and references therein. However, most of these papers focus on the prevalence and consequences of stockouts due to poor inventory planning and have little to do with promotional execution.

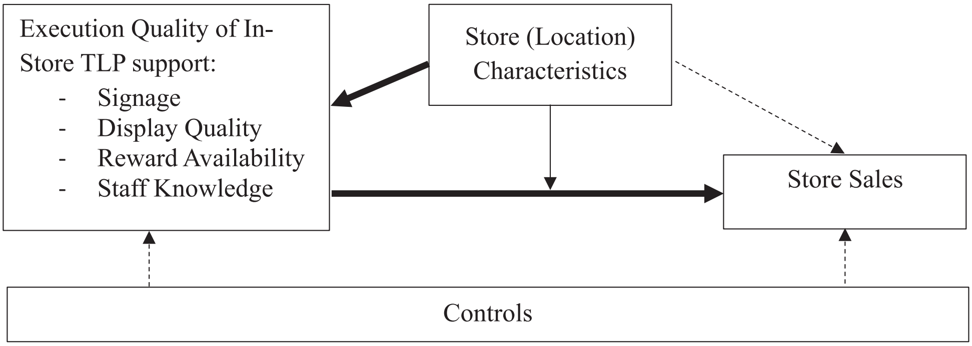
focus on only one or a few instruments, for a non-representative set of stores, often at the start of the program. To the best of our knowledge, no systematic evidence is available to date on the magnitude of non-compliance for different in-store promotion instruments, let alone on its antecedents or its implications.

In sum: in-store promotional execution is challenging, but potentially deficient, as it may seriously hamper the promotion outcomes. Still, the marketing literature to date sheds little light on the magnitude of execution deviations and the store characteristics that underlie them, nor does it document their performance implications. This gap is important as such. From a broader perspective, it even casts doubt on the previously reported impact of in-store actions. On the one hand, studies that assessed this impact in a controlled setting may well reveal the potential impact of in-store marketing instruments but overestimate their effect absent full control of the execution quality. On the other hand, in studies involving secondary field data, the instrument measures are often based on planned support or extrapolated from their use in a subset of sampled stores. In such settings, the outcome variables do reflect consumers' true reactions to the actual implementation, but the instrument measures do not capture the actual execution quality at the considered store. In either setting, the question remains to what extent deviations in in-store execution dampen the store outcomes – a question our research will address in the context of TLPs.

3.2.2 Conceptual Framework

Figure 3.1 presents our conceptual framework. First, we zoom in on the relevant in-store instruments in the context of TLPs – program signage, displays, staff knowledge and reward availability. We then reflect on the size of execution deviations in these instruments across stores, and on the store (location) characteristics that may underlie them. Finally, we assess how the implementation quality of four in-store support instruments affects TLP sales at the store.

Figure 3.1: Conceptual framework



In-store instruments in support of TLPs. Ample research has aimed to determine the effectiveness of retailer loyalty programs, and what makes these programs successful (see, e.g., Bijmolt, Dorotic, & Verhoef, 2010; Breugelmans, et al., 2015; Kim, Steinhoff, & Palmatier, 2021, for an overview). Though most literature to date has focused on permanent loyalty programs, TLPs are rapidly becoming increasingly popular (Bombaij et al., 2022). While TLPs have been found to increase sales (Bijmolt et al., 2010; Lewis, 2004; Taylor & Neslin, 2005), their success critically depends on the ability to draw attention and maintain program salience (Dorotic, Verhoef, Fok, & Bijmolt, 2014; GfK, 2015), thereby activating consumers to increase spending at the store to earn future rewards. To achieve this goal, program operators allocate considerable budgets to in-store program support. In-store instruments used to support TLPs include signage – e.g., posters, ceiling hangers; displays – racks containing the rewards and featuring information about the program; and the involvement of staff – who can alert customers to, and provide information on, the TLP. Moreover, the store typically functions as a collection point where *rewards* are physically held in stock, and where consumers can exchange saved stamps 'on the spot' for a reward at any time during program weeks. Of course, these in-store TLP instruments may only work well if they are properly implemented. Signs and displays must be put up in places where consumers come across them, displays must be 'kept' in an orderly fashion[28], and staff must be able to correctly inform customers. Still, as indicated above, in-store execution of promotional support is challenging, and this is likely no different in the context of TLP campaigns.

Execution deviations: magnitude and drivers. To develop expectations on what drives store differences in execution quality, we use a MOA (Motivation-Opportunity-Ability) framework. First, some store managers and personnel may be more *motivated* to comply and invest in proper execution than others. As indicated by Raman et al. (2001), this depends on the perceived stakes, i.e., the expected economic impact of execution deviations which, in turn, may vary with the market potential and competitive pressure in the market where the store operates. Stores with higher commercial stakes are bound to strive for higher execution quality for each of the four in-store instruments. Second, the *opportunity* to execute in-store support may differ between stores. This could have to do with the store's selling area – more space facilitating proper placement of signage and displays; its accessibility – which enables timely replenishment of rewards; and its social embeddedness – which fosters the development and exchange of staff knowledge. Third, in-store execution quality may depend on the professionalism of store managers and personnel, i.e., their *ability* to properly organize and maintain in-store activities. Again, professionalism will enhance the quality for each in-store instrument.

Though important, it is not feasible for program managers or retailer headquarters to assess the MOA of each store's personnel directly. However, the three constructs can be tied to readily observable characteristics of the store and the market in which it operates. Specifically, we contend that the size of the store, and the degree of socio-economic status,

[28] Castro et al. (2013) argue that for 'ingestible' products, unorderly displays (with a limited number of items) may signal popularity and even generate higher sales than orderly displays (that are fully stocked).

urbanization, and population density of the region, tap into the MOA of store personnel, and can thus serve as antecedents to in-store execution quality. We summarize these expectations in Table 3.1.

As for s*tore size,* we expect (i) a positive impact on motivation (larger stores having higher commercial stakes), (ii) a positive link with ability (large-store managers typically being more professional and better educated), and (iii) a positive impact on the opportunity to create a sensory atmosphere (Kotler, 1973), and properly place and maintain signage and displays. As such, we anticipate that larger stores have higher execution quality for all four instruments, and especially for signage and display.

Next, we anticipate that higher *socio-economic status* of the market (i) enhances motivation (upscale socio-economic areas have higher income and more sales potential), and (ii) ability (higher HDI areas have a more educated/professional workforce to draw from). However, (iii) the high cost of reduced worker productivity due to training (Siebert & Zubanov, 2009) is especially harmful in higher income areas where personnel is more expensive, due to which retailers may economize on the time to let personnel familiarize themselves with or communicate about the TLP. Hence, we expect regions with high socio-economic development to have higher execution quality for signage, display and reward availability, but not for staff knowledge.

Turning to *urbanization*, we expect (i) a positive impact on motivation (urban areas typically exhibit more intense competition), and (ii) a positive impact on ability (to the extent that urban workforce tends to have more resource management skills; Abel, Gabe & Stolarick 2012). However, (iii) urban areas may have lower accessibility because of heavier traffic, which may hamper reward replenishment. Together, this would imply higher execution quality for signage, display and staff, but lower reward availability, in urban regions.

Table 3.1: Expected impact of store (location) characteristics on execution quality

	Motivation	**Ability**	**Opportunity**			
	Commercial stakes	Profession-alism	Accessibility	Space	Social interaction	Net effect
Panel A: Link between (i) store characteristics and (ii) opportunity, motivation and ability						
Size (PPS)	+	+		+		+
HDI	+	+			-	+/-
Urban	+	+	-			+/-
Density	+	+	-		+	+
Panel B: Link between (i) opportunity, motivation and ability, and (ii) execution instruments						
Signage	+	+		+		+
Display	+	+		+		+
Staff	+	+			+	+
Reward	+	+	+			+

Note: The table should be read as follows. For instance, as indicated in Panel A, stores with higher pre-program sales (PPS) tend to have higher commercial stakes, are managed more professionally, and have more space available. In turn, as indicated in Panel B, commercial interest and professionalism make managers more motivated and able to execute signage and display and train staff, while more space gives the opportunity to properly manage signage and display. The impact of other location characteristics can be read from the table in a similar way. A "+/-" indicates that the location characteristic can have a different impact on the distinct execution instruments.

Finally, we postulate that *population density* has (i) a positive impact on motivation (more sales potential and thus higher commercial stakes), and (ii) a positive impact on ability (because densely populated areas have a broader workforce to draw from and hence better selection of personnel). This may foster proper implementation of in-store instruments, including signage and display. Moreover, people in those areas tend to have more (weak) social ties and interact with more people (Sato & Zenou 2014), which provides an extra incentive to invest in staff knowledge. Conversely, we expect (iii) a negative link between population density and store accessibility (more congestion), which is detrimental for reward availability.

Impact of execution quality on store sales. How does execution quality affect the TLP's sales implications? For signage, displays and staff involvement, we expect a clear positive impact. In-store *signage* reaches consumers when they are about to make purchase decisions (Nordfält, 2011; Stilley, et al., 2010), making it an effective program reminder.

Displays can provide powerful sensory stimuli: seeing the rewards, and being able to physically inspect them, may increase consumers' desire to acquire the rewards and increase their spending accordingly (Spence, Puccinelli, Grewal, & Roggeveen, 2014). Because the program typically lasts for several weeks – such that consumers have time to act but may also *need* time to accumulate stamps for rewards – TLP signage and displays may work like advertising messages, with delayed and accumulating effects across program weeks. S*taff* can alert consumers to the TLP and answer any questions regarding the program rules, rewards, and redemption modalities. As such, the more knowledgeable the staff with regard to the program, the more likely that staff interactions will further customers' engagement and up their spending to save for rewards.

While the direction of these effects may seem straightforward, their magnitude is not. Moreover, for reward availability, even the direction of the impact is not obvious. On the one hand, reward availability is important: if consumers see that the item they are saving for is not available and are unsure about future replenishment, they may be less motivated to keep up their saving (purchasing) efforts. On the other hand, limited availability can create a scarcity effect and render program rewards more desirable (Byun & Sternquist, 2012; Castro, et al., 2013; van Herpen, et al., 2005). It follows that incomplete availability of rewards can actually be beneficial. In sum, our empirical analysis will document how, and how strongly, deviations from planned support of in-store instruments affect sales – in absolute terms, and relative to one another.

To complicate matters, the impact of execution deviations may further depend on store (location) characteristics, in a way that is not clear upfront. For instance, consumers may more easily notice (temporary) reward unavailability in smaller stores, but also be more forgiving in case they occur (given the lower predictability of redemptions), and more willing to wait for replenishment. Or, as another example, consumers in upper-class areas may be less inclined to interact with store personnel but – being typically time-pressed – may also depend more strongly on store staff to be alerted to, and properly informed about, the program. Because the direction of these moderating effects is unclear a priori, we leave this as an empirical issue (see Datta, van Heerde, Dekimpe, & Steenkamp, 2022, for a similar procedure).

3.3 Data

3.3.1 Setting

The TLP that we study is set up by global loyalty program operator BrandLoyalty, and is typical of most TLPs. It is run at a large grocery retailer in Indonesia, over a period of 20 weeks. Consumers can collect stamps by spending money at the retailer, and trade in these stamps for rewards in the retailer's physical stores. One stamp is earned for every 30,000 IDR (Indonesian Rupiah) spent. The reward range for this program consists of 5 different outdoor dinnerware items that are offered at a large discount (65-87%). As is usually the case for TLPs, the program setup involves various types of in-store support. The program operator sets up a support plan that is common to all stores of the retailer where the TLP is run and involves all main in-store instruments (i.e., signage in the form of marketing posters to signal the program, racks to display the rewards, training materials to inform store personnel, and delivery/merchandising schedules to make the rewards available in the stores). The retailer then commits to implementing these instruments (plans) in the different stores.

3.3.2 Data sources

To answer our research questions, we make use of two main data sources. The first dataset, which is truly unique, systematically reports on the quality of the in-store execution of the TLP, at each of the retailer's individual outlets, during program weeks. The second dataset consists of transactional (sales) data, before, during, and after the program, for each of these outlets, along with store location data. Together, these data will allow us to (i) assess the size of execution deviations and explore some of the underlying drivers, and (ii) gauge their impact on the sales outcomes of the TLP, thereby controlling for cross-sectional and temporal confounds and addressing endogeneity concerns. A detailed description of the data is given below.

In-Store Execution. These data come from a systematic data collection effort by the program operator, in the form of a longitudinal survey. Each store of the retailer is contacted every week to fill out an online survey via a designated application created by the program operator that is accessible via smart phone or computer. The surveys are designed to determine how well these stores execute different aspects of in-store support, through a set of multiple-choice questions or 'select all that apply' questions. The data are unique in several respects. First, they cover *multiple in-store instruments,* i.e., activities related to (i) the quality of signage, (ii) displays, (iii) reward availability, and (iv) staff knowledge. Based on the survey responses, this dataset contains scores between 0-100 for each of the four instruments at the store-week level, where the maximum score ('100') is awarded in case of perfect execution of/full compliance with the planned support. For signage and display, we note that the scores capture the proper use of these instruments rather than the mere presence or absence. To ensure the store managers answer the survey questions truthfully, they are asked to add pictures that show the actual situation in their store. In addition, the program operator sends mystery shoppers to the store, without the store's knowledge, to verify their answers. Details on the survey questions and scores are provided in Appendix B1.

Second, the survey is administered across *all stores* (and not just a selection of stores deemed problematic or important). In addition, the response rate is very high (82.5% on average), with only 15% of stores having missing survey data on six weeks or more. Given that for these stores extrapolation of scores based on non-missing weeks is potentially problematic, these stores are dropped.[29] For the remaining stores, any missing scores are imputed by either taking the score from the previous or the subsequent week. In the end, weekly scores for 11.5K stores who regularly filled out the survey (85% of surveyed stores), remain.

Third, the survey is longitudinal. i.e., it monitors the execution quality of each in-store instrument from week to week. Only for 'staff knowledge', the questions always differ from one week to the next, to avoid the bias that would occur when stores learn the answers and then never answer incorrectly again. We explain how we deal with this in the next section.

Sales data. The second dataset contains transactional data and covers the spending of customer card owners[30] at the individual outlets of the retail chain. To enable merging of this dataset with the first one, we aggregate these data to the store-week level to create a panel dataset containing a 45-week period, from August 29, 2017 until July 9, 2018. For some stores, sales data are missing in certain weeks[31]. To err on the side of caution, we remove these stores from the data altogether. The sales data thus include 10,966 stores and cover the 20-week loyalty program, as well as 20 weeks prior to the program and 5 weeks post-program.

For 7,849 stores[32], we also have information on the store's geographic location, i.e., the island and the province in which the store is located. In total, the considered country includes 35 provinces that have very different socio-demographic and economic

[29] Closer analysis reveals that the stores with missing survey data are typically smaller outlets, which is not surprising, given that these stores may have less staff for back-end activities. However, because our analysis explicitly accommodates store-size differences (as explained below), we do not expect non-response bias to be a problem.
[30] We include only those card owners who shop at least twice during the data span, at least once during the program, and who spend at least 90% of their total spending at the retailer at those stores who fill out the survey frequently enough that we can reasonably impute missing scores.
[31] For the majority of these 589 stores, sales data are missing for pre-program weeks, which means proper baselines cannot be determined for these stores. Because these stores represent only 5% of the total store set, we do not expect their omission to jeopardize the representativeness of the data.
[32] There does not seem to be any pattern in the availability of store location data, as these stores cover the full range of store sizes, and are spread across the country.

characteristics (as further documented below). For each of these stores, we also know the distribution center that replenishes it (each store being replenished by one of the 29 distribution centers). The final dataset merges the sales and execution datasets and contains weekly store spending for a 45-week period, as well as execution scores of the four instruments at the store-week level, for the 20-weeks of the program, giving us approximately 322K observations.

3.3.3 Variable Operationalizations

In-Store execution variables. Our key variables consist of the retailer execution of the four in-store instruments. To operationalize these variables, we start from the raw survey scores, $\text{Exec\_Score}_{i,t,l}$, for the different stores (i), program weeks (t) and in-store instruments (l, where l = signage, display, reward availability, and staff knowledge). These raw scores will serve as our dependent variables in a first step – as further explained below. To properly capture the impact of execution quality on sales, we use transformations to construct the corresponding independent variables in the sales equations. For signage and display, to capture dynamics and allow the impact of these instruments to carry over to later weeks, we transform their scores into stock variables $\text{Signage}_{i,t}$ and $\text{DisplayQuality}_{i,t}$ (see Table 3.2 for details). For reward availability, keeping in mind the literature on scarcity effects, we include two variables: the score itself ($\text{RewardAvailability}_{i,t-1}$), which reflects the number and types of rewards in store, and its square ($\text{RewardAvailability\_Sq}_{i,t-1}$). Combined, these two allow for an inverted-U shape consistent with a scarcity effect: more rewards leading to increased saving (and spending) at first but taking away the incentive to speed up sales beyond a certain point. Recall that for signage, display and reward availability, stores receive approximately the same questions each week, which allows us to meaningfully monitor over-time changes. For staff knowledge though, to avoid response bias and to get a broader picture, the questions vary from week to week. Unfortunately, this makes a comparison between scores from one week to the next impossible. For this instrument, we thus settle for a cross-sectional indicator and measure to what extent the staff of a store is more or less knowledgeable compared to other stores. To this end, we mean-center the weekly (rescaled) scores across stores, by subtracting the average for that particular week. For each store, we then take the average of the store's mean-centered values across weeks. Hence, in our final dataset, a store's value for $\text{StaffKnowledge}_{i,t}$ has the same value for each of the program weeks.

Store and market characteristics. As a proxy for store size, we use the stores' average pre-program sales obtained from an initialization period (i.e., the first four weeks of our pre-program sales data).[33] Location characteristics are obtained from the World Bank and pertain to the province in which the store is located – the country covering 35 provinces. Urbanization is measured as the percentage of the provinces' population living is an urban area. Population density is obtained as the number of inhabitants per square kilometer. Socio-economic status is measured by the human development index (HDI) – a summary statistic of the population's education level, income, and health.

*Store sales***.** We will measure the impact of the instruments' execution quality on $\text{Sales}_{i,t}$, which is the total log sales (in Indonesian Rupiah) of store i in week t[34].

Details on the operationalizations can be found in Table 3.2.

Table 3.2: Variable descriptions

Variable	**Description**
<u>*Dependent variable*</u>	
$\text{SALES}_{i,t}$	*Total sales*. Measured as the log of value sales at store i in week t. Zero sales do not occur.
<u>*Execution of in-store instruments*</u>	
$\text{Signage}_{i,t}$	*Signage.* This stock variable contains the stock value for the signage (marketing materials) score of store i in week t, such that $\text{Signage}_{i,t} = \lambda * \text{Signage}_{i,t-1} + (1-\lambda) * \text{Signage\_Score}_{i,t}/100$. Variable equals zero for pre- and post-program weeks.

[33] We exclude the data used to create this variable from our estimation sample.
[34] We had no instances of zero sales.

$\text{DisplayQuality}_{i,t}$	*Display quality.* This stock variable contains the stock value for the display quality score of store i in week t, such that $\text{DisplayQuality}_{i,t} = \lambda * \text{DisplayQuality}_{i,t-1} + (1-\lambda) * \text{DisplayQuality\_Score}_{i,t}/100$. Variable equals zero for pre- and post-program weeks.
$\text{RewardAvailability}_{i,t-1}$	*Reward availability.* Variable equals reward availability score of store i in week $t-1$, divided by 100. Score equals zero for pre- and post-program weeks.
$\text{RewardAvailability\_Sq}_{i,t-1}$	*Square of the above Reward Availability Score.*
$\text{StaffKnowledge}_{i,t}$	*Staff knowledge.* This variable contains the average mean-centered score for staff knowledge of store i in week t, divided by 100. Score equals zero for pre- and post-program weeks and has the same positive value for all program weeks.
<u>*Antecedents/Moderators*</u>	
StoreSize_i	*Store Size.* This variable contains the mean-centered pre-program sales of store i using a 4-week initialization period prior to program start (these weeks are not included in the analyses). Used only as an interaction term as the main effect of this variable is subsumed in the store fixed effects.
Urbanization_i	*Degree of Urbanization.* Measured as the % of the local (province) population that lives in an urban area (World Bank).
Density_i	*Population density in the store's province.* Measured as the number of inhabitants per square kilometer (World Bank).
HDI_i	*Human Development Index.* Summary measure reflecting the population health, education and income of the store's province.
<u>*Controls*</u>	
$\text{ReligiousHoliday}_{i,t}$	*Religious Holiday*. Seasonal pulse dummy equal to 1 if a religious holiday took place in week t, weighted with the relative predominance of the corresponding religion in the geographic area of store i, and equal to 0 otherwise.
Program_Pulse_t	Dummy variable equal to 1 during program weeks and zero elsewhere.
Trend_t	Trend variable (i.e., equal to 1 in first week, 2 in second week, etc.).

Note: The Human Development Index (HDI) measures the achievements in three basic dimensions of human development: a long and healthy life, access to knowledge and a decent standard of living. The HDI is the geometric mean of normalized indices measuring achievements in each dimension. Data sources: World Bank.

3.3.4 Descriptives

Table 3.3, Panel A provides descriptives of the execution quality for the four in-store instruments. Two findings stand out. First, keeping in mind that ‘perfect execution’ (based on strategy of the program operator) would correspond with values of 100, we find that the actual scores leave room for improvement. This holds especially for signage and staff knowledge, with average scores below 70. In comparison, with an average score of about 89, the execution quality for displays is markedly higher. In all, this corroborates that in-store implementation of promotional support is challenging and that execution deviations warrant attention – be it more so for some in-store instruments than others. Second, the standard deviations point to substantial heterogeneity between individual outlets – underscoring the importance of exploring the underlying drivers.

Table 3.3: Descriptives

Panel A: Execution scores[a]

Variable	Mean	SD	Median	Lower quartile	Upper quartile
Signage	69.30	27.42	75	50	95
Display Quality	88.73	24.76	100	100	100
Reward Availability	74.69	31.83	100	60	100
Staff Knowledge	69.88	23.13	75	55	89

Panel B: Store (location) characteristics[b]

Variable	Mean	SD	Median	Lower quartile	Upper quartile
Store Size	.025	.65	.097	-.85	.71
HDI	74.27	2.00	73.58	72.60	77.08
Urbanization	56.91	19.67	47.58	36.40	81.80
Population Density	2518.23	4559.29	1022.01	346	7636

[a] Based on 156,980 observations (7849 stores * 20 program weeks)
[b] Based on 7849 stores. The last two columns represent the average for stores in the lower and upper quartile.

Table 3.3, Panel B provides summary statistics on store size, as well as urbanization, population density and socio-economic status of the province to which the store belongs. As the table shows, there is substantial variation in these indicators – which may well underlie the differences in execution quality.

Table 3.4: Correlation matrix

	StoreSize	HDI	Urban.	Density	Signage	Display	Reward Av.	Staff
StoreSize	1	-.045**	-.162**	-.071**	.234**	.256**	.058**	.146**
HDI		1	.579**	.778**	-.088**	-.081**	-.014	-.003
Urbanization			1	.829**	-.070**	-.102**	-.107**	-.095**
Density				1	-.120**	-.127**	-.084**	-.035**
Signage					1	.595**	.410**	.283**
Display						1	.486**	.220**
Reward Av.							1	.350**
Staff								1

Note: Reward Av.: reward availability, Urban.: urbanization. ** $p < .01$

Table 3.4 displays the raw correlations between the key variables. As the table shows, the overlap between the in-store execution variables remains limited. This shows that they are – indeed – separate constructs that may have different antecedents, and a potentially different impact on store performance. Next, several correlations between the execution quality scores and the store location features are statistically significant – indicating that the latter may be relevant antecedents. However, the store (location) features also show significant correlations with one another: not surprisingly, degree of urbanization and population density are positively related, and show an – albeit weaker – positive correlation with the level of socio-economic development. Close inspection reveals, however, that even among urbanized areas, our data show substantial variation in population density. Accordingly, including these variables jointly as drivers in our regressions (as discussed below) does not entail high variance inflation factors (highest vif equals 5.63), with separate variables contributing significantly to model fit. We conclude that accounting for each of these location characteristics is thus important, keeping in mind that our interpretation will be 'ceteris paribus' – after controlling for the other factors.

Turning to the changes over time, Figure 3.2 shows the evolution of the signage stock, display quality stock, and reward availability in the course of the program for the average store.[35] The execution quality of signage and display is higher initially, but then seems to slightly deteriorate. Reward availability, in comparison, shows a strongly declining trend with surrounding replenishment bumps. This underscores the importance of monitoring in-store execution throughout the campaign: assessing in-store implementation of the promotion plans only at the start of the campaign may not reveal the full picture.

Figure 3.2: Evolution of execution quality over time

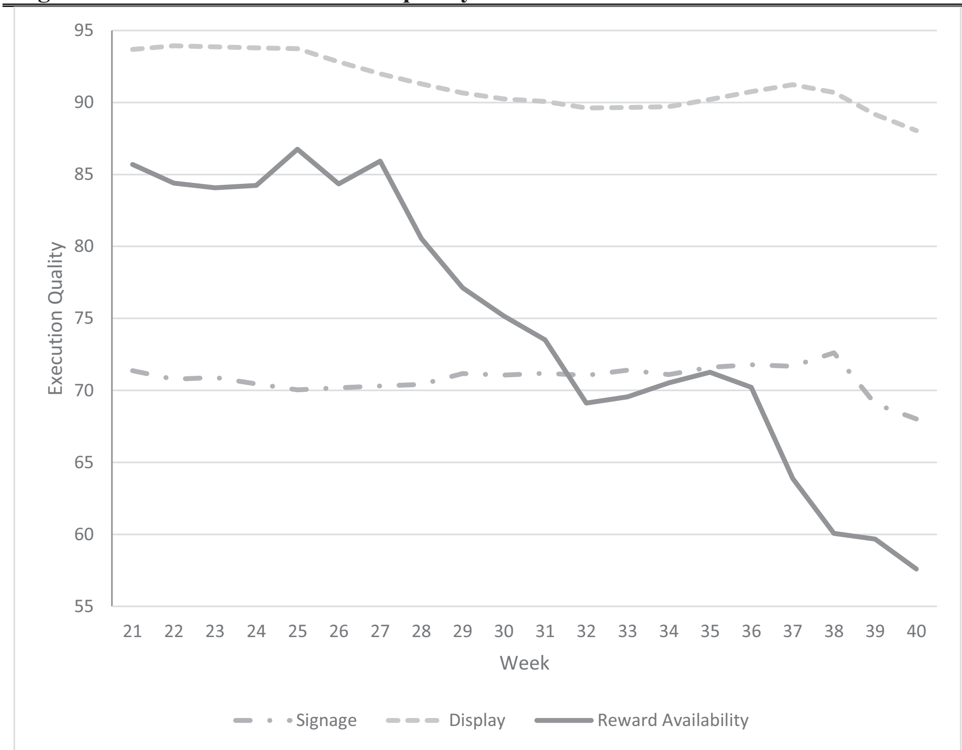
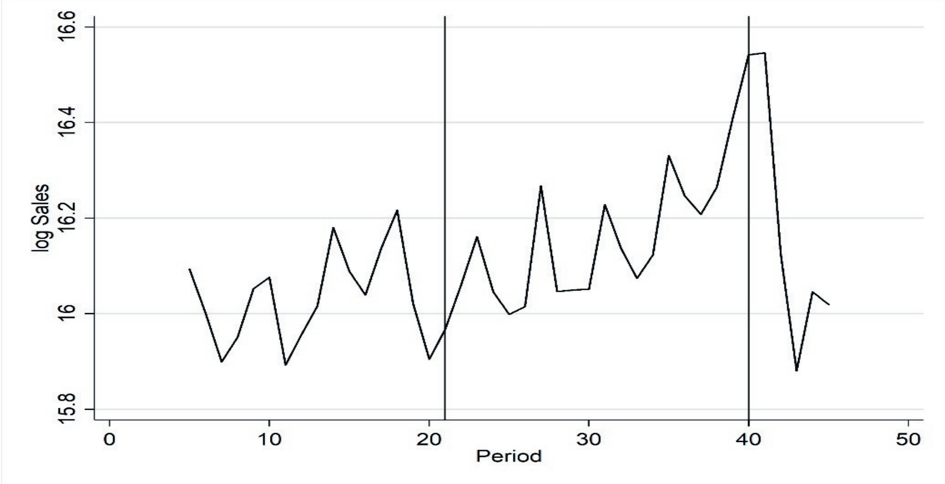


Figure 3.3 displays the weekly evolution of log sales, on average across stores, along with the program timing. The first (second) vertical line indicates the start (end) of the program. Overall, we see an increase in sales during the program weeks, with a dip post-program. At the same time, we also see a trend in sales prior to the program, and a huge peak

[35] Note that staff knowledge is not included in Figure 3.2, because the staff knowledge scores does not vary over time during the program weeks.

around the end of the program. The latter peak turns out to coincide with a religious holiday. This underscores the importance of using a formal model to separate the program effect (and the impact of execution deviations therein) from other temporal factors.

Figure 3.3: Average (log of) store sales per week



3.4 Methodology

Our analysis proceeds in two steps. We first examine the antecedents of execution quality. Next, we assess the impact of in-store execution on store sales, thereby accounting for possible endogeneity from step 1. We discuss both steps below.

3.4.1 Assessing the Antecedents of In-Store Execution Quality

To investigate the differences in execution quality between stores, we use a panel regression model, in which we stack the data for individual stores (i), program weeks (t), and in-store instruments (l). The dependent variable in this model is the (raw) execution quality score (as obtained from the surveys), for a given instrument and store in a specific week ($\text{Exec\_Score}_{i,t,l}$). As explanatory variables, we include, next to instrument-specific dummy variables ($\text{Dum\_Display}_{i,t,l}$, $\text{Dum\_Reward}_{i,t,l}$, $\text{Dum\_Staff}_{i,t,l}$, with signage as the reference instrument) the store (location) characteristics, for which we use instrument-specific coefficients to reflect the different impact laid out in our conceptualization. To capture evolutions in execution quality during program weeks, we add a trend variable for each instrument except staff knowledge, for which the survey questions change across weeks. We also include dummies for the separate distribution centers (from where rewards, displays and signage are shipped to the individual stores), which we mean-center within a province to avoid overlap with the province-specific antecedents. Finally, we add a normally distributed store-specific random component, ϵ_i, to accommodate any remaining (unobserved) differences in the stores' execution quality. The model thus becomes:

$$\begin{aligned}
\text{Exec_Score}_{i,t,l} = {} & \alpha_0 + \alpha_1 \text{Dum_Display}_{i,t,l} + \alpha_2 \text{Dum_Reward}_{i,t,l} \\
& + \alpha_3 \text{Dum_Staff}_{i,t,l} + \sum_{k=1}^{4} \alpha_{4,k} \text{Dum_Signage}_{i,t,l} * \text{StoreChar}_{i,k} \\
& + \sum_{k=1}^{4} \alpha_{5,k} \text{Dum_Display}_{i,t,l} * \text{StoreChar}_{i,k} \\
& + \sum_{k=1}^{4} \alpha_{6,k} \text{Dum_Reward}_{i,t,l} * \text{StoreChar}_{i,k} \\
& + \sum_{k=1}^{4} \alpha_{7,k} \text{Dum_Staff}_{i,t,l} * \text{StoreChar}_{i,k} + \alpha_8 \text{Dum_Signage}_{i,t,l} \\
& * \text{Trend}_t + \alpha_9 \text{Dum_Display}_{i,t,l} * \text{Trend}_t + \alpha_{10} \text{Dum_Reward}_{i,t,l} \\
& * \text{Trend}_t + \sum_{m=1}^{29} \alpha_{11,m} \text{DC}_{i,m} + \epsilon_i + \omega_{i,t,l}
\end{aligned} \tag{3.1}$$

where $\text{StoreChar}_{i,k}$ are the store-specific location characteristics, i.e., store size ($k = 1$), socio-economic status ($k = 2$), urbanization ($k = 3$) and population density ($k = 4$), and $\text{DC}_{i,m}$ are dummy variables equal to 1 if store i is replenished by distribution center m, which we mean-center across distribution centers in the same province. To accommodate heteroskedasticity and error dependencies, we estimate the model with robust standard errors.

3.4.2 Modelling the impact of In-Store Execution on Store Sales

In the next step, we assess the impact of the in-store execution quality on store sales. This comes with important challenges related to confounding factors and endogeneity.

3.4.2.1 Empirical challenges

Confounding factors. To properly assess the impact of execution quality, we need to control for three types of confounds. First, we need to separate the impact of the execution as such from other store characteristics that drive sales. We do so by including observations prior to and following the program and by including store fixed effects to capture differences in the stores' base sales.[36] Second, we need to distinguish the in-store execution effect from temporal confounds. We control for the latter through time fixed effects. Moreover, some seasonal effects – in particular: a religious-holiday impact causing a huge spike in period 39 and 40 – may be specific to the store's geographic setting. Because our retailer has stores all over Indonesia, with different islands having different main religions, it is important to control for this. We therefore include a seasonal pulse dummy for that holiday (which lasts several weeks), weighted with the relative predominance of the corresponding religion in the store's geographic area. Third, we must isolate the in-store execution effect from the 'mere program' effect (that is, the fact that the program is run). We note that the latter, too, is largely controlled for through the time fixed effects. Specifically, the time-fixed effects capture the fact that sales may go up and down during program weeks in a flexible pattern, thus avoiding spurious correlation between sales and in-store execution. For instance, if sales go up in the course of the program but execution deteriorates over time, we might overstate the execution effect (which could then be confounded with the 'mere program' effect). By separately controlling for any temporal changes in sales, the time-fixed effects take care of this issue. A complicating factor is that the impact of the program over time may further

[36] We use 'absorbing' regressions, such that we do not have to estimate store constants for each of the 8K stores.

depend on the store (location) characteristics – e.g., the TLP yielding higher sales lifts in urban areas, or stronger sales increases early on in larger stores. To cleanly separate such program effects from the impact of in-store execution, we add interactions between the store (location) characteristics on the one hand, and a program pulse dummy and a trend line on the other hand.

Endogeneity. An important (and related) concern is that the quality of in-store execution may be endogenous, which may lead to biased estimates. Again, such endogeneity can stem from different sources. First, we may face reversed causality if, rather than driving store sales, the execution score in a given week is the *result* of store sales in that same week. Because the survey responses are systematically collected on Tuesdays (near the beginning of the week), this is not very likely to happen. Still, it may occur for reward availability: high sales (and ensuing redemptions) on Mondays causing low reward availability survey-scores in that same week. To avoid biases from such reversed causation, we use the lag of the reward availability scores in our model. Second, unobserved factors may drive both the level of in-store execution and store sales. To the extent that these common drivers are purely cross-sectional (for instance, store outlets with more experienced or motivated local managers exhibit both higher execution quality and sales) or temporal (e.g. both the level of in-store execution and the sales impact of the reward campaign vary in the course of the program weeks), the store- and time-fixed effects resolve this problem (Germann, Ebbes, & Grewal, 2015). Still, from our descriptives and conceptualization, we know that both the level of execution *during program weeks* and the sales lift during those weeks may differ between stores. To the extent that these effects stem from the size of the store or from the urbanization, density and socio-economic status of its region, this is accounted for by the interactions between these variables and the program pulse and trend variables. Even so, there may be unobserved store features affecting both the level of execution quality and the

TLP sales lift. To alleviate any remaining bias from unobserved common store factors, in the spirit of Skrondal & Rabe-Hesketh (2004), we include the estimated store-specific random components from step 1 as an additional control in the sales model during program weeks. We also consider a refined model version in which we allow the impact of in-store execution for each instrument to vary with the store (location) characteristics (Mundlak, 1978).

3.4.2.2 Model specification

Building on the above, our sales model takes the following form:

$$\begin{aligned}
\text{Sales}_{i,t} = \beta_0 + \; & \beta_1 \text{Signage}_{i,t} + \beta_2 \text{DisplayQuality}_{i,t} + \beta_3 \text{RewardAvailability}_{i,t-1} \\
& + \beta_4 \text{RewardAvailability_Sq}_{i,t-1} + \beta_5 \text{StaffKnowledge}_{i,t} \\
& + \sum_{k=1}^{4} \beta_{6,k} \text{Signage}_{i,t} * \text{StoreChar}_{i,k} \\
& + \sum_{k=1}^{4} \beta_{7,k} \text{DisplayQuality}_{i,t} * \text{StoreChar}_{i,k} \\
& + \sum_{k=1}^{4} \beta_{8,k} \text{RewardAvailabiliy}_{i,t-1} * \text{StoreChar}_{i,k} \\
& + \sum_{k=1}^{4} \beta_{9,k} \text{RewardAvailabiliy_Sq}_{i,t-1} * \text{StoreChar}_{i,k} \\
& + \sum_{k=1}^{4} \beta_{10,k} \text{StaffKnowledge}_{i,t} * \text{StoreChar}_{i,k} \\
& + \sum_{k=1}^{4} \beta_{11,k} \text{Program_Pulse}_{t} * \text{StoreChar}_{i,k} \\
& + \sum_{k=1}^{4} \beta_{12,k} \text{Trend}_{t} * \text{StoreChar}_{i,k} + \gamma_{1,i} + \gamma_{2,t} \\
& + \gamma_3 \text{ReligiousHoliday}_{i,t} + \gamma_4 \hat{\epsilon}_i + \epsilon_{i,t}
\end{aligned} \tag{3.2}$$

where $\text{Sales}_{i,t}$ is the total log sales (in Indonesian Rupiah) of store i in week t. $\text{Signage}_{i,t}$ and $\text{DisplayQuality}_{i,t}$ are stock variables that capture the (proper) use of signage and in-store displays[37]. $\text{RewardAvailability}_{i,t}$($\text{RewardAvailability\_Sq}_{i,t}$) is the (squared) reward availability score for store i in week t, which reflects how many reward items were in stock. $\text{StaffKnowledge}_{i,t}$ measures how knowledgeable store staff is regarding all things program related.[38] To facilitate interpretation, we rescale these execution-quality variables prior to estimation.[39] We include the interactions between these variables and the store (location) characteristics $\text{StoreChar}_{i,k}$ (grand mean-centered) to allow for heterogeneous execution effects depending on the store's size and the socio-economic status, urbanization and population density in its market. $\text{ReligiousHoliday}_{i,t}$ is a weighted pulse dummy to control for the location-specific impact of a religious holiday on store i. $\gamma_{1,i}$ and $\gamma_{2,t}$ represent the store- and time fixed effects. $\text{Program\_Pulse}_{t} * \text{StoreChar}_{i,k}$ and $\text{Trend}_{t} * \text{StoreChar}_{i,k}$ are interactions between a program pulse dummy and trend line on the one hand, and the store (location) characteristics on the other, to control for a direct impact of the latter on the program lift over time. We estimate the model with robust standard errors to account for the panel structure and accommodate heteroscedasticity.

[37] To assess the level of λ, the parameter in the display and signage stock variables ,we conduct a grid search in the range [0, 1] with steps of .1, and retain the value that generates the highest model fit. In both instances, lambda equals 0.9.
[38] Our indication of subscript t for staff knowledge is based on this score equaling zero for pre- and post-program weeks and having the same positive value for all program weeks.
[39] Specifically, we divide the signage stock, display quality stock, reward availability and staff variables by 100, and the squared reward availability variable by 10000.

3.5. Results

3.5.1 Drivers of execution quality

Table 3.5: Estimation results: Impact of store (location) characteristics on execution quality

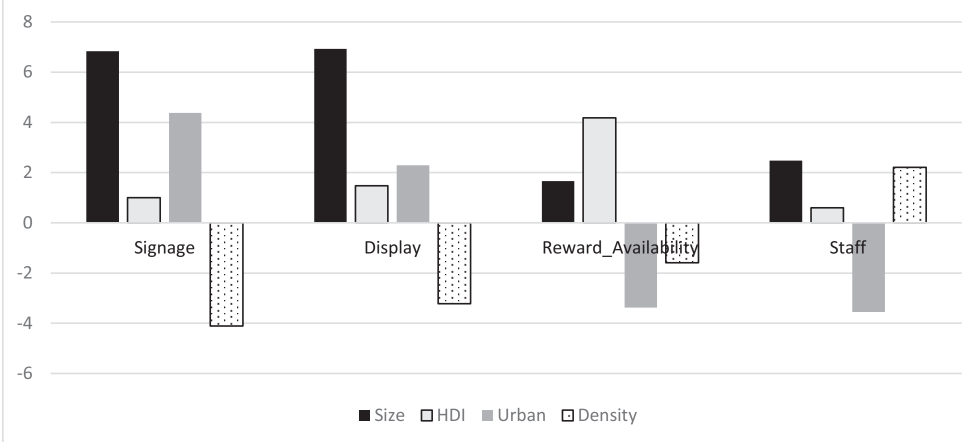
Variable	Expectation	Coefficient	
Constant		52.264	***
Dum_Display		14.795	*
Dum_Reward		-25.608	**
Dum_Staff		11.633	
StoreSize*Dum_Signage	+	4.380	***
HDI*Dum_Signage	+	0.224	**
Urban*Dum_Signage	+	0.096	***
Density*Dum_Signage	+	-0.001	***
Trend*Dum_Signage		-0.332	***
StoreSize*Dum_Display	+	4.441	***
HDI*Dum_Display	+	0.329	***
Urban*Dum_Display	+	0.050	***
Density*Dum_Display	+	-0.000	***
Trend*Dum_Display		-0.410	***
StoreSize*Dum_Reward	+	1.068	**
HDI*Dum_Reward	+	0.934	***
Urban*Dum_Reward	+/-	-0.074	***
Density*Dum_Reward	+/-	-0.000	**
Trend*Dum_Reward		-1.554	***
StoreSize*Dum_Staff	+	1.588	***
HDI*Dum_Staff	+/-	0.133	*
Urban*Dum_Staff	+	-0.078	***
Density*Dum_Staff	+	0.000	***
Distribution center fixed effects		Yes	

Note: N = 627,920, R-square = .1168, * $p < .10$, ** $p < .05$, *** $p < .01$.
Dependent variable is the (raw) instrument score. Estimated with store random effects (reflecting store's overall execution quality, across instruments) and robust standard errors.

Model estimates. Table 3.5 displays the estimation results of Eq. 3.1. As the interactions with the trend variable show, execution quality significantly deteriorates over time, for signage and display, but especially for reward availability. These patterns are consistent with the raw data descriptives above. More interesting for our purposes, we find that store (location) characteristics exert a significant impact on the execution quality of most instruments. The direction of these effects differs between instruments and is largely in line with expectations – as further discussed below.

Effect sizes. Figure 3.4 shows the effect sizes. For each instrument and store (location) characteristic, it displays the difference in execution scores between stores with low and high levels (mean of low and high quartile stores) for that characteristic – based on the model estimates, so after controlling for other factors. As the figure shows, larger stores (i.e., those with higher pre-program sales) have higher execution quality, especially for signage (a 6.83 higher score, $p<.01$) and display (+6.93, $p<.01$). Upscale markets, with higher HDI, exhibit higher reward availability (+4.18, $p<.01$) – in line with the premise that the workforce in higher HDI areas is more educated and aware of the importance of in-store execution. Urbanization has a positive impact on signage and display (+4.37, $p<.01$ and +2.29, $p<.01$, resp.) – consistent with more intense competition and higher commercial stakes. However, it has a negative impact on reward availability (-3.38, $p<.01$) – as it lowers accessibility – and, somewhat surprisingly, staff knowledge (-3.56, $p<.01$) – possibly because urbanization entails less emphasis on personal exchanges and more individualism. Population density, which goes along with less space and more traffic jams, comes with lower execution quality for signage (-4.11, $p<.01$), display (-3.22, $p<.01$) and reward availability (-1.59, $p<.01$). Yet, it is associated with higher staff knowledge (+2.21, $p<.01$), which may get more emphasis because of lower inter-personal distance. Together, these effects entail important store differences. For instance, larger stores in urbanized areas typically score 11.2 points higher on signage; whereas, e.g., small stores in urban, densely populated markets that are low on socio-economic status, score about 10.8 points lower on reward availability. The question is how this affects store sales – a point we turn to next.

Figure 3.4: Impact of store (location) characteristics on level of execution quality



Note: the figure should be read as follows: e.g., for store size, the height of the 'signage' bar is 6.83. This means that large stores have 6.83 points higher execution-quality scores for signage than small stores.

3.5.2 Impact of execution quality on store sales

Model estimates. Table 3.6 shows the estimation results for three nested specifications. Model (1) is a base model that only includes the main effects of execution quality for the four in-store instruments. Model (2) adds the interactions of the execution variables with store size. Model (3) extends Model (2) by adding the interactions between the execution variables and the characteristics of the province in which the store operates. Across the three models, the pattern of effects is quite stable. Henceforth, we focus on Model (3) - the 'full' model.

Table 3.6: Estimation results: (ln) sales as a function of execution quality

	Model (1)		Model (2)		Model (3)	
Constant	16.1102	***	16.1102	***	16.1102	***
ReligiousHoliday	0.0033	***	0.0034	***	0.0034	***
StoreSize*Trend	-0.0029	***	-0.0022	***	-0.0023	***
HDI*Trend	-0.0006	***	-0.0006	***	-0.0006	***
Urban*Trend	-1.73E-05	*	-1.80E-05	*	-5.72E-05	***
Density*Trend	-1.01E-07	*	-9.73E-08	*	7.72E-08	
StoreSize*Program_Pulse	-0.0323	***	0.0084		0.0031	
HDI*Program_Pulse	-0.0008		-0.0008		0.0004	
Urban*Program_Pulse	0.0001		0.0001		-0.0026	***
Density*Program_Pulse	7.40E-06	***	7.49E-06	***	1.81E-05	***
StoreRandomEffect*Program_Pulse	0.0003		-0.0003		-0.0002	
Signage	0.0654	***	0.0692	***	0.0693	***
DisplayQuality	0.0447	***	0.0297	**	0.0302	**
StaffKnowledge	0.0828	***	0.1183	***	0.1109	***
Lag_RewardAvailability	0.1259	***	0.1117	***	0.1156	***
Lag_RewardAvailability_SQ	-0.1130	***	-0.1000	***	-0.1040	***
Signage*StoreSize			-0.0007		0.0099	
DisplayQuality*StoreSize			-0.0469	**	-0.0494	***
StaffKnowledge*StoreSize			0.1249	**	0.1201	*
Lag_RewardAvailability*StoreSize			-0.0439	*	-0.0442	*
Lag_RewardAvailability_SQ*StoreSize			0.0242		0.0278	
Signage*HDI					-0.0057	
DisplayQuality*HDI					0.0067	
StaffKnowledge*HDI					-0.0533	***
Lag_RewardAvailability*HDI					0.0094	
Lag_RewardAvailabilit_SQ*HDI					-0.0136	
Signage*Urban					0.0040	***
DisplayQuality*Urban					-0.0007	
StaffKnowledge*Urban					-0.0016	
Lag_RewardAvailability*Urban					0.0045	***
Lag_RewardAvailability_SQ*Urban					-0.0029	**
Signage*Density					-1.29E-05	**
DisplayQuality*Density					9.09E-08	
StaffKnowledge*Density					1.64E-05	
Lag_RewardAvailability*Density					-3.05E-05	***
Lag_RewardAvailability_SQ*Density					2.62E-05	***
Store fixed effects	Yes		Yes		Yes	
Period fixed effects	Yes		Yes		Yes	

Note: N = 321,809, R-square (model 3) = .8812, * $p < .10$, ** $p < .05$, *** $p < .01$.
Dependent variable is the log of store sales. Standard errors adjusted for 7,849 clusters.

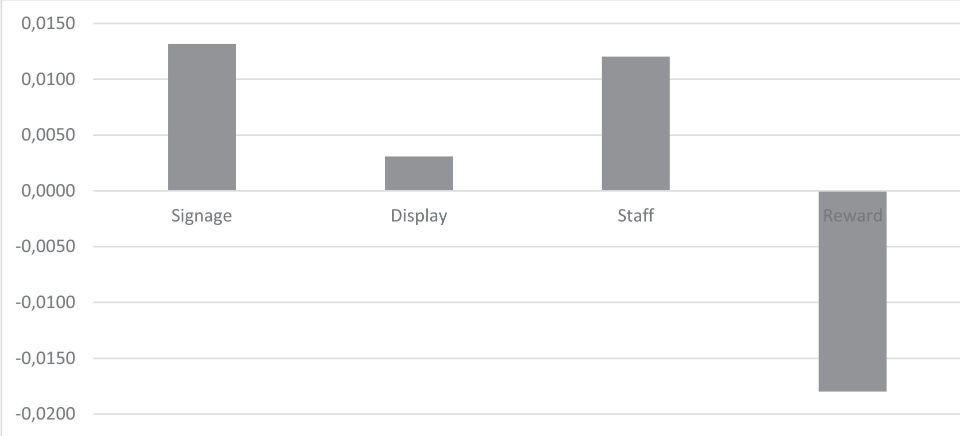
First, as expected, we find positive and significant main effects of the execution of signage (β = 0.069, $p < .01$), display quality (β = 0.030, $p < .05$), and staff knowledge (β =

0.111, $p < .01$) on store sales.[40] As for reward availability, we find a positive impact (β = .116, $p < .01$), but only up to a certain point, as shown by the negative coefficient of the squared availability score (β = -.104, $p < .01$) – consistent with a 'scarcity' effect. Second, turning to the interactions with store size, we find a negative effect of display execution (β = -0.049, $p < .01$) – suggesting that high display quality is more beneficial in smaller stores. Conversely, the impact of staff knowledge is higher in larger stores (β = 0.120, $p < .10$). As for the market characteristics, higher socio-economic status hardly affects the impact of execution quality, it only dampens the role of store personnel (β = -0.053, $p < .01$). Urbanization and population density do exert an important effect, though, especially when it comes to signage and reward availability. Proper signage enhances sales especially in markets that are urbanized (β = 0.004, $p < .01$) yet not too densely populated (β = -1.29E-05, $p < .05$). Likewise, having more rewards in stock lifts sales especially in urban markets (as indicated by the significant interactions for the reward availability score: β = .005, $p < .01$; and its square: β =-.003, $p < .05$) in which population density remains limited (interaction with reward availability score: β = -3.05E-05, $p < .01$; and with its square: β = 2.62E-05, $p < .01$). In all, execution quality significantly affects sales, in a way that depends on store size, and with a somewhat different impact across regions.

Effect sizes. To get a feel for the size of these effects, we use the model estimates, along with commonly observed variations in execution-quality, as inputs for further calculations. For each in-store instrument, we consider low vs. high levels of execution quality, measured as the 25th and 75th percentile of the distribution, respectively. Figure 3.5a shows the percentage increases in sales when changing the level of execution from low to high, for an average store.

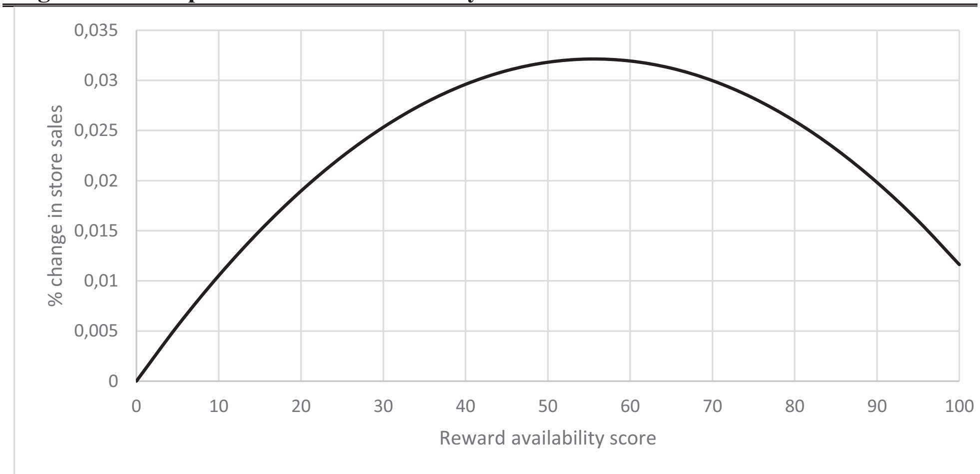
[40] Note that in the sales model, the store (location) variables are mean-centered prior to calculating the interactions, such that the main effects reflect the impact for an 'average' store.

Figure 3.5a: Average difference in store sales between low and high levels of execution scores



Note: the vertical axis gives the relative change in store sales when the execution quality of the instrument changes from low (lower quartile) to high. For instance, the height of the 'signage' bar is .0132. So, increasing the quality of signage from low to high leads to a 1.32 % increase in store sales.

Figure 3.5b: Impact of reward availability on store sales



As can be seen from the figure, we find the strongest effects for signage and staff knowledge. Improving signage increases store sales by about 1.32% on average, while staff knowledge entails a sales lift of about 1.20%. Though these figures may seem small at first, we emphasize that they reflect the change in store sales overall (not the change in sales lift due to the program). The impact of display remains limited (+0.31%), but this partly stems from the fact that display quality is typically high altogether (so its interquartile range is small). The negative impact of reward availability in Figure 3.5a comes from moving toward full availability. The influence of this in-store instrument is actually inverted-U shaped consistent with a scarcity effect (see Figure 3.5b), with an 'optimum' at medium-availability levels (score around 55).[41]

The effect sizes in Figure 3.5 pertain to a store with average characteristics. Zooming in on the impact of execution quality across stores with different profiles, we observe some notable deviations (see Appendix B2 for the corresponding figures). The effect of proper signage is much larger in urban areas but where population density is not too high (+3.4%). Stores in those areas, therefore, have a strong interest in properly putting up signage in support of the TLP. Those are also the areas where reward availability, and hence timely replenishment, is most relevant – at least up to a certain point.[42] The impact of display quality is somewhat lower in large stores (-.8%). Instead, improving program knowledge among store personnel leads to much larger sales increases in large outlets (+2%) and less-developed areas (+2.6%).

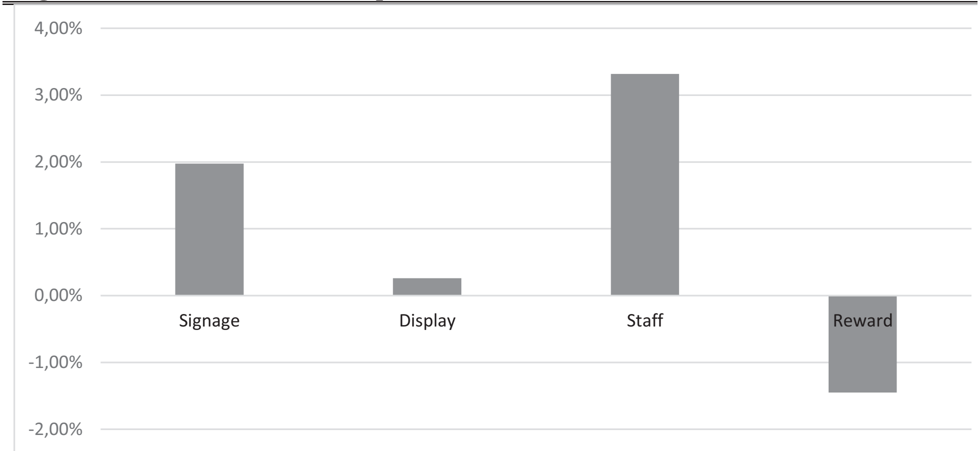
3.5.3 Sales losses from imperfect execution

Our findings so far indicate that in-store execution is often far from perfect, and that lower levels of execution quality significantly hurt sales. Taken together, this begs the question: what are the sales losses from actual (observed) execution deficiencies, and what are the potential gains to be reaped? We again use our model estimates to determine the sales increase due to a change from actual (observed) to perfect execution[43]. We find that perfect

[41] Given that we now find that execution deviations (lower scores for execution quality) decrease sales, we henceforth refer to these deviations as deficiencies.
[42] The 'optimum' level in those areas would be about 65% of items available, higher availability reducing the incentive to spend and save. In urbanized and densely populated areas, the optimum availability level is much higher (about 85%), but reward availability plays less of a role to begin with.
[43] Whereas the previous section (and the corresponding Figure 3.5) quantified the sales changes from typical variations in execution quality among the majority of stores, this section considers the sales effect of moving toward full compliance. Note, also, that multiple stores actually have scores of '100' in several weeks, such that these simulations fall within the data range.

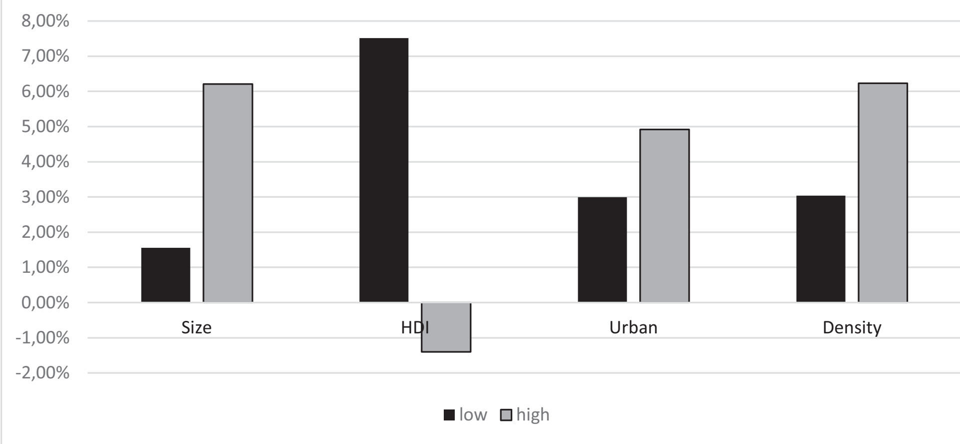
execution across the board would lift sales by no less than 4.1%. Figure 3.6a shows the breakdown by in-store instrument, for an average store. It reveals that deficiencies in signage and especially staff knowledge cause the biggest losses, of up to 1.97% and 3.32%, respectively. Consistent with our estimates above, the figure also shows that ensuring full availability of all reward items at all times actually has an adverse effect – possibly because this signals that the program is not interesting, or that there is no need to speed up item collection to ensure the possibility to redeem.

Figure 3.6a: Sales losses from imperfect execution



Note: for a store with average characteristics, this figure gives the % sales loss with average instead of perfect execution for a given instrument.

Figure 3.6b: Average sales loss during program weeks from imperfect execution, by store (location) characteristic



Note: % sales loss with average instead of perfect execution for all instruments, in stores that are low (average for stores in lower quartile) or high (average for stores in higher quartile) on a store (location) characteristic.

Next, we consider the overall sales losses due to execution deficiencies across all four instruments, for different store (location) characteristics (see Figure 3.6b). The smallest losses are incurred in small stores (where losses occur mainly due to imperfect signage and displays), while the largest losses are incurred in socio-economically less developed, densely populated areas (mainly due to staff deficiencies) – with sales losses of up to 7% and more. Efforts at reducing execution deficiencies should thus be tailored to store and region characteristics – as further discussed below.

3.6 Discussion

With most grocery purchases still occurring in physical stores, in-store execution remains key to the success of promotional campaigns. Yet, anecdotal evidence suggests that proper retailer execution is notoriously difficult to achieve, especially for campaigns laid out by external partners. Our study addresses the call for empirical research on the magnitude, antecedents, and sales impact of execution deviations – using a unique dataset that measures the level of retailer execution over time for several in-store instruments, in the context of a temporary loyalty program. We offer several novel findings, with important managerial implications.

3.6.1 Findings

Execution quality. Though execution scores are not dramatically low, we find that the execution quality of planned in-store support clearly leaves room for improvement. Overall, the quality scores are about 25 to 30% below the ‘optimal’ level, but with large differences across instruments, stores and time. For one, we find that especially the use of in-store signage and staff knowledge is deficient. Signage like banners, ceiling hangers, or shelf tags are often placed in the wrong spot, not well kept, or not used at all. Store personnel are not well informed about the specifics of the program and/or do not actively convey these to customers. In comparison, displays appear to receive much more attention from local store managers. Second, the quality of in-store support clearly changes over time. For reward availability, this is not too surprising: naturally, the program operator may diminish his inventories, and hence store deliveries, of (popular) rewards as the program progresses. Yet, our longitudinal analysis reveals that also the quality of in-store displays deteriorates in the course of the program. This suggests that checking the implementation of promotional support only at the start of the program and/or in larger stores – as is often the case in practice – fails to reveal the true size of execution failures. Third, the quality of in-store execution differs markedly between stores. We find that store size, but also regional differences, systematically influence execution quality, and differently so for different instruments. Smaller stores exhibit lower execution quality across all instruments – but especially for signage and display. As for the market characteristics, implementation of promotional signage and displays is typically worse in densely populated areas where space tends to be scarce. Reward replenishment is more deficient in urban areas – especially those with lower socio-economic development and high population density. Conversely, personnel in densely populated markets appear more knowledgeable of, and communicate more actively about, the promotional program. The underlying logic is that these store and market features tap into the motivation, opportunity and ability to properly implement the promotional support – thereby significantly affecting execution quality.

Impact of execution deviations on store sales. Imperfect compliance with planned support decreases sales in general. Overall, we find that improving in-store execution for a given instrument from low (lowest quartile) to high (highest quartile) levels enhances store sales in the course of the program by up to 1.32%. Given that these are percentages of base sales, these are sizable figures. In general, the effects are relatively stronger for signage and staff knowledge, and weaker for displays. For reward availability, we obtain an inverted-U shaped impact. This is consistent with a scarcity effect (Byun & Sternquist, 2012; Castro, et al., 2013; van Herpen, et al., 2005), where limited stock signals the success of the program and the desirability and popularity of reward items – thereby triggering consumers to step up their reward saving (and thus spending) at the store. The impact of execution quality on store sales also varies with the store and market characteristics. For instance, larger stores in regions with lower socio-economic status suffer more strongly from limited staff knowledge. Interactions with store staff can influence customer engagement (Roggeveen & Grewal, 2018), and our results show that especially in larger stores in less-developed areas, where TLPs are important but displays stand out less, store personnel has a critical role in informing consumers about the program. Combined, knowledge of what drives execution deficiencies and their impact across stores leads to important managerial insights.

3.6.2 Management implications

Awareness. Extant sources indicate that retail employees are often unaware of the magnitude, and even the existence, of execution problems in their chain (Raman et al. 2001). By documenting the execution deficiencies for four different types of in-store instruments, and showing that they are not negligible, our study offers a first step towards resolving the issue.

Importance of monitoring. Industry reports reveal that in current practice, checks on the quality of in-store execution, if any, typically occur at the start of the promotional program only, and are often limited to larger stores (POPAI, 2015). Our findings caution against such practice. We urge managers not to lose sight of smaller outlets – where deficiencies are often larger and can still be substantial. Our findings also warn managers not to fall into the "launch and leave trap" (POPAI, 2015). Instead, managers have a clear interest to keep following up on the execution quality throughout the program, as deterioration often sets in in later weeks.

Sales losses from imperfect execution. Monitoring and improving in-store execution is a costly and effortful endeavor. An important question for managers is thus: What are the stakes?

Previous studies indicate that poor execution of in-store promotions "can be traced back to the fact that the supplier is the main monetary beneficiary of promotion campaigns, and the retailer is responsible for the execution" (Hacker et al. 2010; Srinivasan, Pauwels, Hanssens, & Dekimpe, 2004). As noted by Raman et al. (2001): "*creating awareness of the problem is not enough; awareness of the impact of the problem is also important*" (p. 13). Our paper sheds light on this important issue. We show that a lot of money is left on the table because retailers do not implement the in-store support as the program operator intended. Especially for signage and staff knowledge, significant gains can be reaped by moving toward full compliance – with figures of up to 4.1% from the stores' base sales. Further calculations suggest that about one third of the potential program effect (if perfectly executed) is lost due to execution deficiencies (see Figure B3.1 in Appendix B3). Hence, retailers who accept to participate in a TLP should not merely view in-store execution as a burden to the benefit of the program operator (Hacker et al., 2010), but as an essential endeavor to improve their own performance metrics.

Prioritizing improvements. The findings from this study provide several actionable insights for retailers, store managers and program operators. First and foremost, store managers need to pay attention to signage and especially staff knowledge. Execution for these instruments is often deficient yet they have a strong impact. Reward availability is also far from perfect, but that is less consequential due to a scarcity effect. While a minimum of reward stock is needed, increasing reward availability beyond a certain point (based on our results: 60% of rewards available) typically does not pay off, but rather removes consumer incentives to spend and save.

Second, to get the most bang for their buck, managers should allocate their investments in execution quality differently depending on the size of the store. We find that improvement in the execution of in-store signage is fruitful in either store. Smaller stores are more susceptible to displays and warrant further compliance-investments in display quality. For larger outlets, the biggest sales increase can be obtained by perfecting staff knowledge. These outlets should invest in training to make store staff more knowledgeable about the program and foster their interactions with customers to increase program engagement. Improving reward replenishment, finally, is especially relevant in later periods – reward availability being more likely to drop below the ‘optimum’ point as the program progresses.

Third, regional differences also dictate different priorities. Regional characteristics can be readily observed by chain headquarters and TLP program managers without detailed knowledge on the individual store outlets. Hence, they can be easily acted upon. Our study generates novel insights on which instruments should primarily be improved in which areas (see Appendix B4 for Figures). While staff plays an important role on average, investments in staff knowledge do not pay off in regions with high socio-economic status. In such areas, improvements in signage are rather called for. Stores in urbanized (but not too densely populated) regions benefit most from timely reward replenishment. Reward availability in those stores is lower than average, whereas its impact on sales is higher. Managers can use these insights to optimize their execution efforts.

3.6.3 Limitations and future research

Our study offers several avenues for future research. While our data present a unique opportunity to gauge the size and impact of retail execution deviations, they also have limitations. First, based on the classification of the (experienced) program operator we largely consider that the deviations from the highest obtainable execution quality scores are harmful (and after finding empirical support for this, refer to such deviations as deficiencies). However, some nuance regarding these deviations may still be required, as i) past research shows that strategy standardization might not always be preferred to localized customization (Grewal, Chandrashekaran, & Dwyer, 2008; Steenkamp & Geyskens, 2014), and ii) retailer and store managers also have considerable experience and are notably knowledgeable regarding their own (local) setting. Rather than solely focusing on the standardized recommendations of program operators, it might be useful to exploit combined industry wisdom (Anderson, 1988; Gielens & Dekimpe, 2007). Future research could delve deeper into these nuances, especially by identifying whether store deviations are undesired (simply due to not living up to their own goals), or deliberate (when managers purposely decide to deviate from the suggested strategy).

Second, our data only cover one loyalty program run in one country. Though our pattern of effects may generalize to in-store campaigns in a larger context, the magnitudes may depend on the specific setting. For example, cultural differences may make consumers more (or less) prone to scarcity signals or interaction with store personnel. For third parties who run in-store campaigns at a global scale, or retailers located in other areas of the world, it may be relevant to assess such potential differences. Or, campaign characteristics, such as the type of rewards or the length of the campaign, may affect the level and evolution of execution deficiencies throughout the TLP, and their impact on the sales lift. We leave these as an interesting area for further study.

Third, even though the program operator implemented several checks on the truthfulness of survey responses, we cannot completely rule out that the quality of execution is overstated. As such, our results offer a conservative measure of the room for improvement. Moreover, there may be measurement error. For signage, display quality, and staff knowledge, such error is smoothed out by using stock variables or averages across stores. Any remaining measurement error would likely dampen the estimated effects – still, as they stand, our effects are significant and economically meaningful.

Fourth, our identifying assumption is that we can properly control for differences in the program impact across stores and separate it from differences in in-store execution quality. To the extent that motivation, opportunity, and ability to implement these instruments is inherent to the store (i.e., related to 'fixed' store characteristics, such as the quality of management, the availability of space, or the store size), this is captured by the store fixed effects. To the extent that it is specific to the type of campaigns (TLPs) – in particular, the expected sales returns from such campaigns over time – this is controlled for through the interactions between the program variables and the store (location) features, as well as by the normally-distributed random store component that intervenes in both the execution-quality and the sales equations during program weeks. Still, the latter is identified based on functional form. Hence, we cannot rule out that idiosyncratic store features other than the ones included here affect both execution quality and store sales specifically during program weeks. Some caution thus remains called for.

Fifth, while we are the first to systematically study the antecedents of execution quality for different in-store promotion instruments, our set of drivers was limited – including, next to store size, only characteristics at the province level. Previous studies have shown that regional indicators can be valuable in explaining differences in economic activity and performance, and our study corroborates this in a different setting. Still, a significant portion of the differences in execution between stores was left unexplained by our antecedents. Future studies, with access to more detailed data, could use our MOA framework to uncover additional store-level drivers of execution quality.

Finally, while we could quantify the sales increase from 'perfecting' in-store execution and indicate where time or effort should primarily be allocated, we have no indications of the costs of improving execution quality. These costs could well vary by instrument, and depend on the actions needed (e.g., increase monitoring vs. train staff). Retailer or program managers could use proprietary cost information to gauge the profitability of such actions.

3.7 Appendix B

Appendix B1: Survey Information

The survey comprises a set of binary and multiple-choice question for each in-store instrument. The program operator assigns points to the different answers, where the highest number of points is awarded to the option that contains the correct (or most desirable) answer. Next, based on the answers given by a specific store, a score between 0 and 100 is calculated per in-store instrument, by summing the total earned points. Examples of the survey questions can be found in Table B1.1. Alternatively, some questions may also be “select all that apply”-questions. For example, the questions to determine the reward availability in a store, ask which of the following reward range items are in stock in their store, where each answer corresponds to one of the 5 reward items, and is worth 20 points. If a store has selected all five answers, this means no items are out of stock that week, and the stock availability score of that store is equal to 100.

Table B1.1: Example questions and answers for weekly surveys.

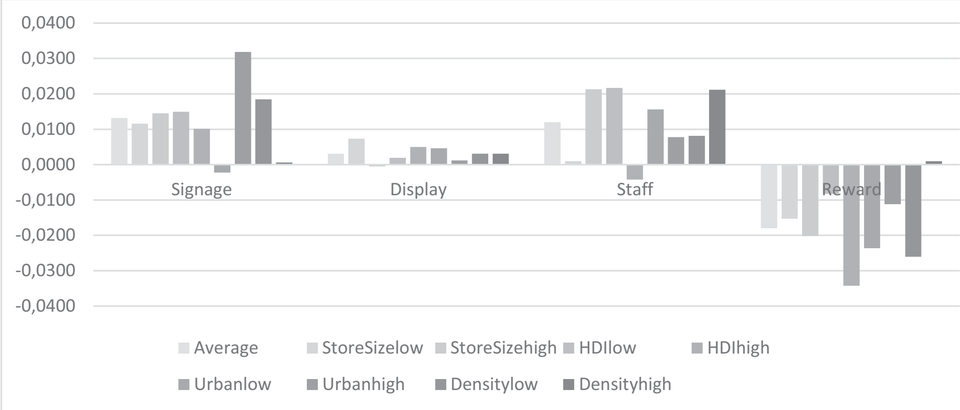
In-store instrument	Example questions	Example answers (and example points per answer)
Signage	Is the banner available outside the store?	Yes – 10p No – 0p
	Are there ceiling hangers present in your store?	Yes – 10p No – 0p
	Where have you placed the stamp cards? (you may select multiple answers)	On the display stand, behind the please take one sign – 5p At checkout, reachable for customers – 10p Behind checkout, unreachable for customers – 5p
	Do you have wobbly shelf tags placed at the product: Silverqueen Chunkybar 100g?	Yes – 15p No, we only placed one – 10p No, because they aren’t attractive – 0p No, we didn’t receive them – 0p

	Do you stil have complete Point of Sales Materials (POSM)? Please select all the POSM that are still in your store.	Banner outside – 10p Display case – 10p Brandliner – 10p Poster – 10p Hanging mobile Alfastamp app – 10p Hanging mobile KAT 2018 – 5p Stamp cards – 10p Stamps – 10p Supplier funded wobblers – 5p Cashier backwall – 10p Point-of-purchase fixture – 10p
Display Quality	In which shelving fixture is the display stand placed in your store?	Promotional shelving fixture – 50p Non-food shelving fixture – 25p Food shelving fixture – 25p I don't know – 0p
	When should the display stand be placed in the assigned shelving fixture?	At all times during program – 50p Only during stamp collection period – 0p
	Please take a photo of the whole shelving fixture with display case and rate the quality.	As new – 50p It's OK – 25p Needs improvement – 0p
	Please take a picture of the cashier desk that shows the reward item and price label.	My store cashier desk has the item and price label – 50p My store cashier desk has the item without price label – 25p My store cashier desk does not have the item and price label – 0p
Reward Availability	Select all that apply. Which of the reward items are available in your store?	Dipping Bowls, 10 cm – 20p Bowls, 15 cm – 20p Serving bowls, 23 cm – 20 p Plates, 24cm – 20p Serving bowl, 30cm – 20p
Staff Knowledge	What material are the reward items made of?	Plastic – 0p Melamine – 6p Ceramic – 0p Glass – 0p
	Where is the reward item brand RoyalVKB from?	England – 0p Germany – 0p United States – 0p The Netherlands – 6p
	Can the reward items be used in the microwave/oven/stove top?	Yes – 0p No – 8p I don't know – 0p

	A customer spends 45,000IDR and purchases one Silverqueen Chunkybar 100g and one Paseo Baby Tissue Pure Soft 130s. How many stamps does the customer get?	1 stamp – 0p 2 stamps – 8p 3 stamps – 0p 4 stamps – 0p
	Will customers get stamps during redemption transactions?	Yes – 0p No – 8p I don’t know – 0p
	What should cashiers do during transactions to help increase reward item sales?	Issue stamps and stamp card – 10p Explain that customers can redeem the reward items for 10 stamps or 30 stamps and a bigger discount – 10p Offer redemption to customers in every transaction – 10p All of the above – 20p

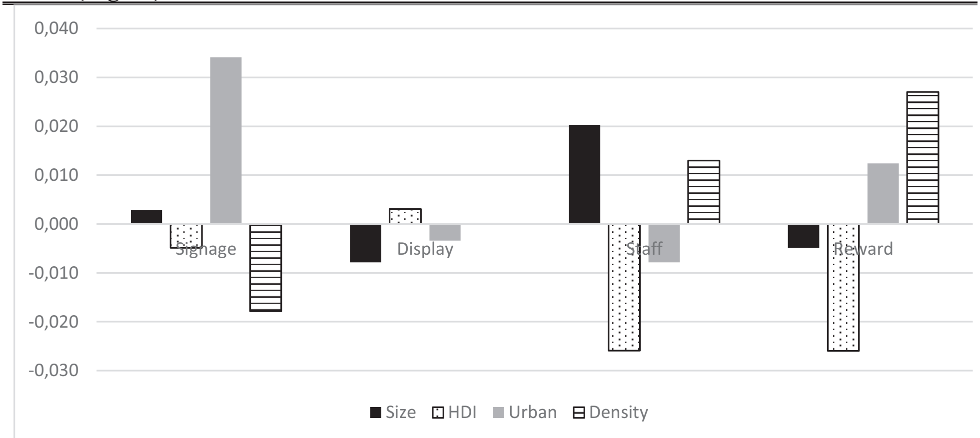
Appendix B2: Impact of execution quality on store sales as a function of store (location) characteristics: Figures

Figure B2.1: Average change in ln store sales between low and high levels of execution scores, for stores with different characteristics



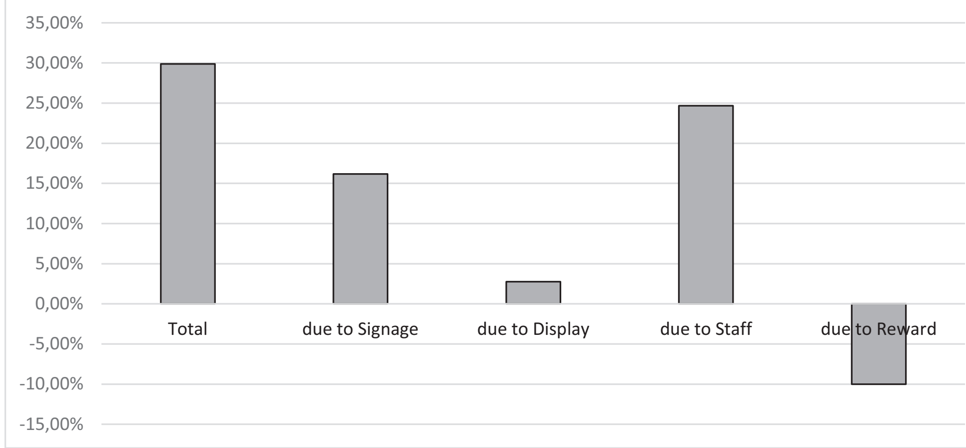
Or, looking at the difference in execution effectiveness across stores:

Figure B2.2: Difference in execution impact on ln sales between high and low levels of store (region) characteristic



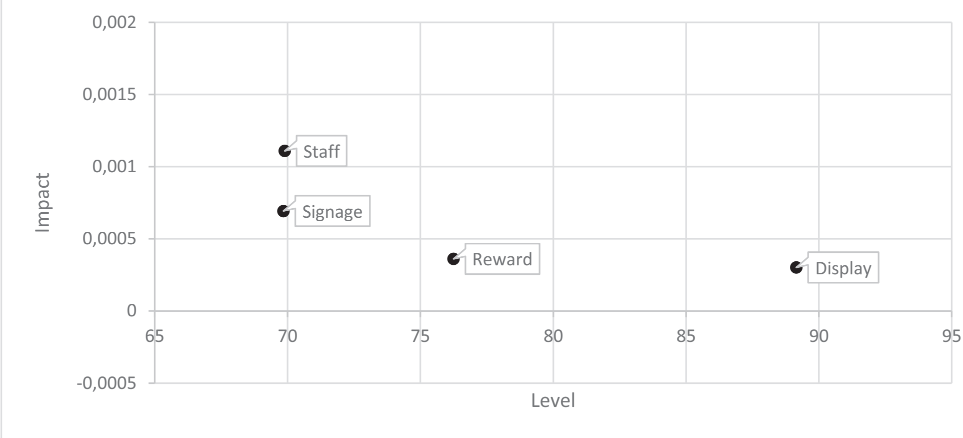
Appendix B3: Losses from imperfect execution

Figure B3.1: Losses from imperfect execution relative to program effect



Appendix B4: Execution Quality: Level vs Impact

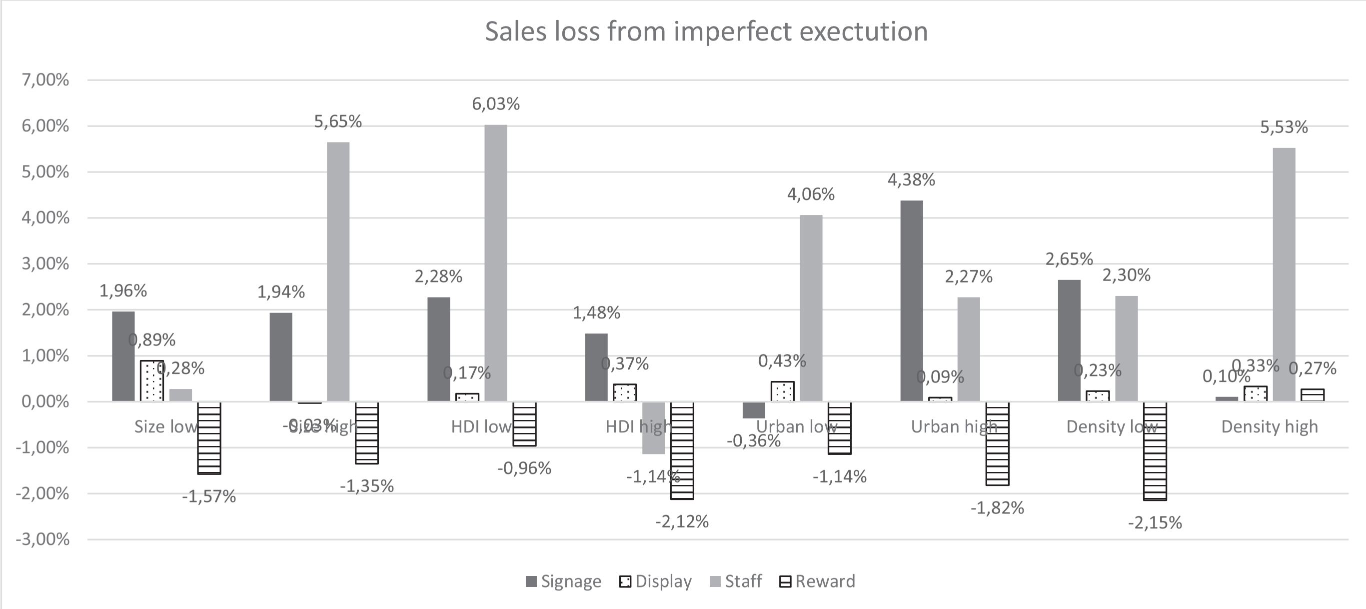
Figure B4.1: Average Store



Note: Upper-left position implies highest potential gain from increased execution quality, because actual level is low and impact is high.

Figure B4.2. Execution level vs. impact as a function of store (region) characteristics

Figure B4.3: Sales loss from imperfect execution, by instrument and store (region) characteristic



4. Determining The Difference in Effectiveness of Different Message Types in Store-Loyalty Programs

4.1 Introduction

While the vast majority of purchases still occurs offline and in-store, the promotion landscape is shifting more and more from offline to online. Retailers are spending more on digital marketing than ever before, as evidenced by industry reports. eMarketer (2021), for instance, found that digital advertising expenses consisted of up to 63% of firms' total advertising budget in 2021, which is even expected to increase to as much as 72% of the total budget by 2025. Furthermore, consumers are also using more and more apps on their smartphones. The number of mobile application downloads worldwide have increased from 141 billion in 2016 to 218 billion in 2020 (Statista, 2021). One illustration of this apparent trend can be found in loyalty programs (LPs), where mobile saving through apps is becoming increasingly common. These mobile applications provide novel opportunities of communications directly to consumers, such as mobile push messaging.

While in Chapter 2 we found these push messages to be effective and elicit positive consumer responses, extant research also suggests that push messaging should be done carefully. It states that sending too many messages could annoy consumers or invade their privacy, which may result in avoidance. Therefore, even though the overall effect of push messaging has been found to be positive, it is unclear whether this is (equally) true for all types of messages, and for different outcome variables. Such knowledge is important for the different LP stakeholders (retailers and program operators). First, if different types of messages have a different impact on spending, this is valuable information for the retailer. Certain messages could be more profitable for them if the increase in spending is larger than the increase in redemption, as more redemptions can be costly. This is the case for investment programs, where retailers pay a small fee to the program operator for each reward consumers redeem. Second, the main interest of the program operator, who makes money for every reward that is redeemed by consumers, is in increasing stamps redeemed. Furthermore, the goals of the program operator and the retailer are not necessarily aligned. If, for instance, one type of message works better on spending while another type has a relatively more positive impact on redemption, this can require considerable planning and different strategies in program or push plan negotiations. However, to the best of our knowledge, no study to date has looked into the impact of message type in the context of mobile push messaging (in general or in the context of LPs).

We address this gap by looking at two distinct message types that are typically used in mobile push messaging during TLPs. *Engagement-oriented* messages aim to increase program salience and awareness to prevent consumers from discontinuing participation in the program. *Promotion-oriented* messages rather aim to draw consumer's attention to a specific promotion on offer within the loyalty program. When purchased in that week, not only do consumers get a price promotion on a given product, but they also receive a free bonus stamp with purchase. This free bonus stamp offers an opportunity for faster progression towards consumers' saving goals. While prior literature has looked at the impact of these types of loyalty program promotions themselves (Dorotic, Fok, Verhoef, & Bijmolt, 2011; Zhang & Breugelmans, 2012; Minnema, Bijmolt, & Non, 2017; Bombaij, 2021), it has not yet determined the impact of messages (or other forms of communication) about them. As such, no comparison between these two distinct message types has been made, which is the goal of this study. In addition, we consider several outcome variables and are interested in how these effects differ for different consumer groups.

Engagement messages can be beneficial by increasing program awareness and reminding consumers of the rewards they stand to earn in case of continued participation and redemption. On top of that, however, promotion messages offer something extra compared to engagement messages: in our program, a discount on a certain product and a free bonus stamp with purchase. A priori, one therefore might always expect a more positive effect for these types of messages. For the retailer, the free bonus stamps offer an opportunity for progression toward the redemption goal that requires less increase in spending. This makes it easier for less-heavy buyers at the chain to reach the redemption threshold, who might otherwise struggle to redeem any rewards. The latter can also be beneficial to the program operator as it helps increase redemption levels. In addition, the promotion messages come with added time pressure, given that the offer on promotion is limited to the week they are sent in, meaning consumers must react rather quickly to actually obtain the free stamp. As a result, these messages may increase the number of shopping trips consumers make, rather than only increasing spending per trip. We consider the latter by including the number of shopping trips as an additional dependent variable in this chapter.

In contrast though, promotion messages may also have downsides that make them less effective. Given that the emphasis of the message is on the specific promotion, the attention that would otherwise go to the program, may be shifted to the promotion on offer and the accompanying products. For the retailer this could mean that consumers purchase the products on offer at a discount, rather than a substitute at full price. This could result in lower consumer spending. In addition, the more attention is brought to the promotions and the extra progression they offer in the program, the less consumers are likely to appreciate the advantages the program itself offers. Hence, they may be less inclined to increase their spending because of these extra opportunities. The latter is also important for the program operator, as consumers are therefore less likely to reach redemption thresholds and redeem. Alternatively, the promotion messages can also trigger, in particular, those households who might have difficulty to reach the redemption threshold regardless, which might mean an increase in spending, but not one in redemption. Hence, it is not clear a priori which forces dominate. Our study sheds light on this issue.

For this purpose, we use a large-scale field experiment during a loyalty program, with data that tracks consumer transactions and redemption behavior at the retailer. In addition, we have information on which types of push messages are sent and when. The randomized assignment in treatment and the panel structure of the data, allow us to avoid common endogeneity issues and cleanly measure the impact of the type of push message. While we use a similar methodology as in Chapter 2, our goal is to provide more insights on (i) how the impact of push messaging differs for distinct message types, (ii) whether the impact of a certain message type is heterogenous, such that separate consumer groups react differently to the message types, and subsequently (iii) how the message types affect consumers, by not only looking at sales and redemption, but also including visit frequency.

Our results show that for the average consumer both engagement and promotion messages increase spending, redemption and store traffic, but that promotion messages are more effective in general. Furthermore, we find that the difference in impact varies with each performance metric. For instance, we find that the strongest advantage over engagement messages is found for spending, where, for the average consumers, we find the impact of promotion messages to be 2.5 times larger compared to engagement messages. This advantage is somewhat smaller for redemption and for traffic, meaning that the increased spending that results from messaging is not only caused by consumers visiting the store more frequently, but that they also spend more per trip. In addition, we look into how this effect varies depending on consumer types and find that promotion messages are particularly effective to enhance spending of heavy buyers at the chain.

In the remainder of this study, section 2 discusses relevant previous literature and provides the conceptual framework. Section 3 introduces the setting and the panel data set.

Section 4 addresses the empirical challenges and explains the model, where section 5 provides the results. Section 6 concludes and discusses the implications of the findings.

4.2 Background literature & conceptual framework

4.2.1 Loyalty program promotions

As mentioned in the introduction, for this study we focus on two distinct types of messages that are sent within our setting of a temporary loyalty program, namely engagement and promotion messages. The promotion messages, which we will explain in more detail in the next section, are messages that aim to draw consumer's attention to a specific promotion on offer within the loyalty program. According to Minnema et al. (2017), these loyalty program promotions (LPPs) are popular ways for brand manufacturers to capitalize on the success of a retailer's loyalty program. With an LPP consumers receive a free bonus stamp or free collectible (bonus premium) with the purchase of a specific promotional item. The benefit for consumers can be considerable, given they not only receive a bonus stamp or collectible with purchase - which makes for an efficient way to obtain more stamps and reach reward saving goals even faster - but that the offer is also often accompanied with an additional price promotion for said item. Though our interest is not in the effect of the LPP itself, a few studies to date have looked into their effectiveness and the processes involved. First, Dorotic et al. (2011) and Zhang and Breugelmans (2012) find a positive effect of LPPs on consumer's spending and store visits during permanent loyalty programs. These studies, however, do not consider the additional pressure that results from the need to reach the redemption threshold prior to the end of the program (i.e., the point-pressure mechanism) that is associated with temporary programs (Taylor & Neslin, 2005; Bijmolt, Dorotic, & Verhoef, 2011). In our study, we consider a temporary program where consumers save stamps (delayed reward), which they can later exchange for a reward if they save enough. Bombaij (2021) considers various types of temporary loyalty programs to identify some of the mechanisms behind when and why an LPP is effective. The author finds that LPPs alleviate the pressure that comes from making the reward threshold before the program ends, because they offer an opportunity to help consumers progress to their goal faster. Though these studies all focus on LPPs and their effectiveness, we do not focus on the effectiveness of the LPPs themselves, but rather on the impact of the push messages that are being sent announcing them, and to compare those to other message types. As such, we are interested in how effective these special types of messages (promotion messages) are compared to the more general engagement messages that are sent throughout the program. In determining this difference in effectiveness, we consider several outcome variables and allow these effects to differ for different consumer groups.

4.2.2 Message type

Previous literature has looked into the impact of message content in a variety of different settings or contexts (e.g., advertising, retailing, online reviews, etc.). One example being Kaul and Wittink (1995), who determine the impact of price vs. non-price-oriented advertising on consumers' price sensitivity and prices. To the best of our knowledge, however, to date, no study has looked at the impact of message type or message content for mobile push messaging (in general or in the context of LPs). Figure 4.1 presents our conceptual framework. In our setting, we distinguish between engagement-oriented messages and promotion-oriented messages, where the former aim to remind consumers of their participation in the loyalty program and prevent consumers from discontinuing participation in the program. A promotion-framed message, on the other hand, aims to make a consumer aware of a specific promotion on offer within the loyalty program, that when purchased offers a free bonus stamp – i.e., offers a desirable opportunity for progression in their goal pursuit.

However, arguing which type of message is more effective in boosting consumer spending, redemption and store visits, is not a straightforward matter. Engagement messages can be very beneficial considering that they can increase program awareness and remind consumers of the rewards they stand to earn. Engagement messages may also remind consumers of the increased effort they have put in up to that point, which will become sunk should they not continuously step up their spending and store visits or fail to see their effort through and actually redeem (Taylor & Neslin, 2005; Kivetz, 2003; Kim, Shi, & Srinivasan, 2001). In line with prospect theory, which has found that losses loom larger than potential gains (Kahneman & Tversky, 2013), engagement messages may therefore be particularly effective in keeping up active consumer participation in the program. On top of that, however, promotion messages offer something extra compared to engagement messages. In our program, promotion messages help make consumers aware of a discount on a certain product and a free bonus stamp with purchase. Taking into account the point-pressure mechanism (Taylor & Neslin, 2005) and the goal-gradient effect (Kivetz, Urminsky, & Zheng, 2006), promotion messages may therefore be more effective, given that the bonus stamps offer an efficient way to obtain more stamps and reach consumers' reward saving goals even faster. In addition, the promotion messages come with added time pressure, given that the offer on promotion is limited to the week they are sent in, meaning consumers have to react to them rather quickly to actually obtain the free stamp. Furthermore, drawing attention to a chance at increased progression towards the goal, may increase consumers' awareness of their current saving status and make the program even more salient. This increased program salience may even be stronger than for engagement messages, as the promotion-oriented message forces consumers to think more about how much they need to make use of this offer to begin with. As such, one might expect a more positive effect for promotion messages. On the other hand, however, promotion messages also have downsides.

Nowlis and Simonson (1996) found that when LPPs are accompanied with a price promotion they were found to be less effective than the sum of each separate promotion. Similarly, given that the promotion message also emphasizes the specific promotion on offer, the attention that would otherwise go to the program and the free stamp, may be shifted to this promotion and the accompanying products, thus decreasing their effectiveness (Dorotic et al, 2011).

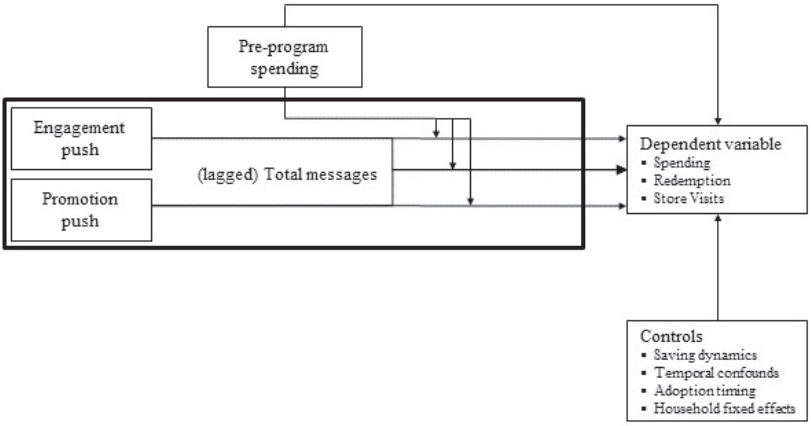
Considering all of the above, which of these forces dominates is not necessarily clear a priori. Nevertheless, we argue that increasing consumers' awareness of the potential benefit of using bonus stamps to reach their goal faster, outweighs the potential lessened effect that comes from the addition of the price promotion. In addition, given that the promotion messages come with added time pressure that requires consumers to act quickly, they may increase the number of shopping trips consumers make, rather than only increasing spending per trip. As such, we expect promotion messages to be more effective compared to engagement messages in increasing consumer spending and store visits at the retail chain.

For redemption on the other hand, we do not expect the same. Given that previous literature has pointed out the problem of unredeemed stamp piling up and never being used (Dorotic, Verhoef, Fok, & Bijmolt, 2014), we expect engagement messages to be a more effective tool to increase redemption, because they remind consumers of their goal and what increased effort is lost should they fail to redeem. Whereas the promotion message, that offers faster progression towards their saving goal, does not necessarily remind, or invite consumers to redeem the stamps they have collected thus far. Hence, we expect promotion messages to be less impactful to trigger redemption.

4.2.3 Consumer type

Next, as mentioned in the introduction and shown in Figure 4.1, we also allow the impact of messages to differ depending on whether consumers are more- vs. less-heavy buyers at the chain, based on their pre-program spending levels. We opt to do so, because we expect that the effectiveness of certain message types may depend on what customer group receives that message. When it comes to promotion messages, the free bonus stamps offer an opportunity for progression toward the redemption goal that requires less increase in spending. This may make it easier for less-heavy buyers at the chain, who might otherwise struggle to reach the redemption threshold, to reach it. Making consumers aware of this offer, promotion messages might therefore be more effective for less-heavy buyers at the chain. On the other hand, though, lighter buyers may not be in the habit of visiting the chain on a frequent basis, and thus have less opportunity or might be less willing to (immediately) step up their expenditures or incur an extra visit. As we mentioned, however, this is necessary given that the promotions are usually only on offer for a limited time. This added pressure may even lead to a reactance effect where lighter buyers get annoyed and spend or frequent less as a result (Stauss, Schmidt, & Schoeler, 2005). In contrast, while heavier buyers at the chain likely have more opportunity to act on the promotion messages, they may also have less leeway to increase their spending or incur an extra visit (i.e., ceiling effect, Bijmolt et al., 2011). For instance, heavy buyers who receive a promotion message might be more likely to purchase the offer on promotion rather than a full price substitute, which could lower overall spending. Since it is unclear a priori which of these forces dominates, we leave this as an empirical issue.

Figure 4.1: Conceptual framework



4.3 Data

4.3.1 Setting

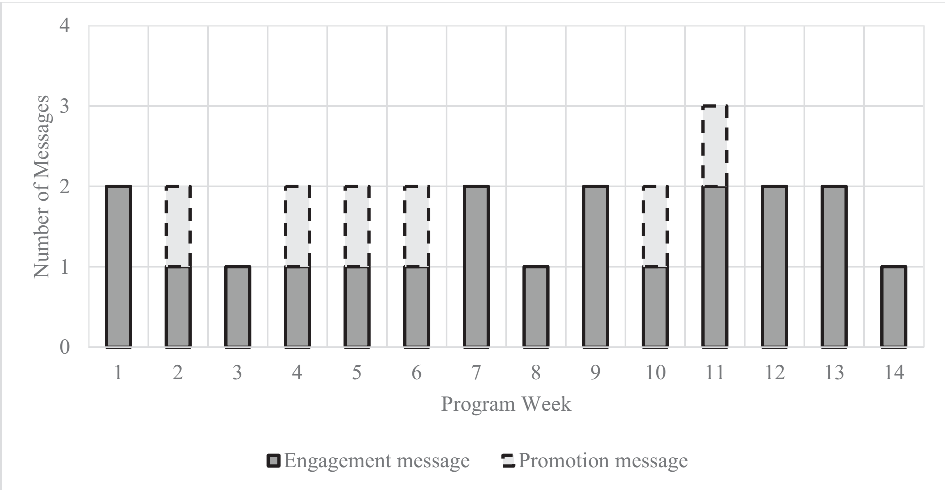
For this study, we consider the same large-scale field experiment run by global program operator BrandLoyalty in the fall of 2016, as in the second chapter of this dissertation. The program ran for 18 weeks at a large retailer in Southeast Asia. During the program, stamps can be collected physically, as well as digitally by means of a mobile application that requires installation. Installation is not mandatory, but it does allow for automatic updating of consumers' stamp balance by linking the app to a consumer's customer card, which is scanned at checkout. In addition, consumers only start to receive push messages within the app upon installation. The reward items that consumers can save for during this program are plastic food storage containers that can be redeemed for a pre-determined number of stamps and are offered at a large discount (60-98%). Consumers can collect and redeem during the first 14 weeks of the program, but only redeem stamps during the final 4 clean-up weeks. During these clean-up weeks, however, the program operator only sends push messages to those consumers that still have enough stamps to make redemption possible. Therefore, to avoid any potential bias that may result from this, we estimate the effects of promotion push messaging based on the 14 regular program weeks only and exclude the clean-up weeks from our analyses. During these regular weeks, no targeting occurs, and all treated consumers are sent the same number and type of messages.

As explained in earlier sections, when it comes to the push messages sent during this program, we make a distinction between engagement messages and promotion messages, two types mobile push messages typically used in TLPs. In this study, we are especially interested in comparing the impact of engagement messages to that of promotion messages. Engagement messages are more general messages that aim to remind consumers of the program and try to enhance consumer participation therein. An example of an engagement message that is sent during this program is: *"Keep on collecting! We are sure you will redeem a reward soon."*. In contrast, promotion messages typically alert consumers to certain promotional items that, when purchased in that week, offer a free bonus stamp. A promotion message takes the following form: "<P*roduct name> special price at <promotional price*> *and get an extra stamp for the purchase of this product."*. Which products and what exact promotion is offered differs from one week to the next. It is clear, however, that these promotion messages contain a very clear promotional offer. In addition, they also explain clearly to consumers that they additionally receive a free bonus stamp with the promoted products.

To assess the impact of engagement vs. promotion messages sent within the program app (which requires consumers to adopt it first), we use two sources of exogenous variation in the push messages, as was the case in the second chapter. First, the program operator created a randomized control group of 2,305 households who install the app during week 3 of the program. This control group thus selects on app installation in the first 2 weeks of the program, and the households in it only receive push messages during the first 3 weeks of the program, and none after that. Following the same procedure as in Chapter 2, we only consider those treatment households who install the app in the exact same period as the control households, i.e., the first 2 weeks of the program. This way we ensure that both groups are comparable on unobservables that correlate with app installation timing. As a result, we are left with 44.2K treated households who received push messages throughout the program. Second, as you may recall from Chapter 2, the program operator withheld individual push messages from randomly selected subsamples of treated consumers. These subsamples were drawn independently each time and occurred for both engagement and promotion messages. Hence, the resulting variation in push messages is exogenous.

The push schedule of engagement and promotion messages as used by the program operator is shown in Figure 4.2. Approximately 23% of the messages that are sent throughout the program are promotion messages, meaning far more engagement messages are sent in general. As the Figure shows, the number of engagement messages sent per week varies from 1 to 2, but never more than 1 promotion message is sent in a given week. Furthermore, promotion messages are sent in different program weeks, however, none are sent in the final weeks. We discuss possible limitations thereof in the final section of this chapter.

Figure 4.2: Number of push messages sent per program week



4.3.2 Data sources

Our study merges two main data sources. The first contains information from the mobile LP application at the household-day level. It indicates when consumers visit the retailer, whether they redeem any stamps and if so, how many, and when they receive engagement and promotion push messages. This data only contains this information from the point of app adoption onwards.

The second dataset consists of transactional (sales) data for the same households for 18 weeks before and 18 weeks during the program. To allow for the merging of these two datasets we aggregate the first dataset to the household-week level. Because our interest is in the effect of engagement vs. promotion push messages, which can only be sent during the program itself, we use only the weeks of the program for our estimation. We do, however, use the pre-program sales levels to allow for heterogeneity and distinguish between more- and less-heavy buyers. The final merged dataset contains information for those households that subscribe to the app during the first 2 weeks of the program and contains a total of 651K observations at the household-week level for estimation.

4.4 Model

As we explained in section 3.1, our estimation sample includes households who, like the control households, install the app in the first two weeks of the program. This way, we ensure we use the exogenous variation from random assignment to treatment and control conditions only.

The model we use to determine the effect of engagement vs. promotion push messaging on consumers' spending (stamp collection), takes the following form:

$$\begin{aligned}\text{SPEND}_{i,t} = {} & \alpha_i + \gamma_t + \beta_0 + \beta_1\text{DIST}_{i,t} + \beta_2\text{BALANCE}_{i,t} + \beta_3\text{PREVREDEMP}_{i,t} + \beta_4\text{ADOPT}_{i,t} + \\ & \beta_5\text{DIST}_{i,t} \times \text{PREVSPEND}_i + \beta_6\text{BALANCE}_{i,t} \times \text{PREVSPEND}_i + \\ & \beta_7\text{PREVREDEMP}_{i,t} \times \text{PREVSPEND}_i + \beta_8\text{ADOPT}_{i,t} \times \text{PREVSPEND}_i + \\ & \beta_9\text{ENG\_MESSAGE}_{i,t} + \beta_{10}\text{PROMO\_MESSAGE}_{i,t} + \beta_{11}\text{ENG\_MESSAGE}_{i,t} \times \text{PREVSPEND}_i + \\ & \beta_{12}\text{PROMO\_MESSAGE}_{i,t} \times \text{PREVSPEND}_i + \beta_{13}\text{MESSAGES\_TOT}_{i,t-1} + \\ & \beta_{14}\text{MESSAGES\_TOT}_{i,t-2} + \beta_{15}\text{MESSAGES\_TOT}_{i,t-1} \times \text{PREVSPEND}_i + \\ & \beta_{16}\text{MESSAGES\_TOT}_{i,t-2} \times \text{PREVSPEND}_i + \beta_{17}\text{MESSAGES\_TOT}_{i,t} \times \text{DIST}_{i,t} + \\ & \beta_{18}\text{MESSAGES\_TOT}_{i,t} \times \text{BALANCE}_{i,t} + \epsilon_{i,t},\end{aligned} \tag{4.1}$$

where the dependent variable $\text{SPEND}_{i,t}$ is the total amount of money spent at the retailer by household i in week t. First, we include several variables as controls. We include variables to control for the saving dynamics that result from consumer participation in the program ($\text{BALANCE}_{i,t}$, $\text{DIST}_{i,t}$, $\text{PREVREDEMP}_{i,t}$), and allow these saving dynamics to differ between more-or-less-heavy buyers at the chain ($\text{BALANCE}_{i,t} \times \text{PREVSPEND}_i$, $\text{DIST}_{i,t} \times \text{PREVSPEND}_i$, and $\text{PREVREDEMP}_{i,t} \times \text{PREVSPEND}_i$). We also control for a possible program effect or other temporal factors using time fixed effects (γ_t), an effect of app adoption timing using an app adoption step dummy ($\text{ADOPT}_{i,t}$), and allow the impact thereof to vary with household's pre-program spending levels at the chain ($\text{ADOPT}_{i,t} \times \text{PREVSPEND}_i$). Finally, we control for any household differences that may remain – even though our control group conditions are randomized – by including household fixed effects (α_i).

Our main interest though, is in the push message effects. The variable $\text{ENG\_MESSAGE}_{i,t}$ captures the number of engagement-oriented messages, and the variable $\text{PROMO\_MESSAGE}_{i,t}$ captures the number of promotion-oriented messages received by the household in a given week. Furthermore, we allow the effects of these messages to differ depending on consumers' (mean-centered) pre-program spending levels at the chain ($\text{ENG\_MESSAGE}_{i,t} \times \text{PREVSPEND}_i$, $\text{PROMO\_MESSAGE}_{i,t} \times \text{PREVSPEND}_i$).

To avoid overfitting, we opt to not make this split in message types for the lagged message variables or the interactions with the saving dynamics, where we use the total number of messages (i.e., the sum of the engagement and promotion messages, $\text{MESSAGES\_TOT}_{i,t}$), instead.[44] As such, we include interactions of total messages and distance and balance ($\text{MESSAGES\_TOT}_{i,t} \times \text{BALANCE}_{i,t}$, $\text{MESSAGES\_TOT}_{i,t} \times \text{DIST}_{i,t}$), the lags of the total message variable ($\text{MESSAGES\_TOT}_{i,t-1}$, $\text{MESSAGES\_TOT}_{i,t-2}$), and their interactions with household's pre-program spending ($\text{MESSAGES\_TOT}_{i,t-1} \times \text{PREVSPEND}_i$, $\text{MESSAGES\_TOT}_{i,t-2} \times \text{PREVSPEND}_i$). This way, we additionally capture the possible post-message effects that occur due to carryover or purchase acceleration. For a complete overview of these variables, see Table 4.1.

[44] Note that our operationalization here is different, than our model in Chapter 2, where we included the logs of the messaging variables to account for diminishing returns of message effects. For this study though, considering we split the main message variables into engagement and promotion messages which is at most 2 in a given week, we use no log transformations of our messaging variables.

Table 4.1: Variable descriptions

Variable	**Description**
Spending/Redemption variables	
$\mathrm{SPEND}_{i,t}$	*Total spend (in 1000 Rupiah)*. Measured as the total amount of money spent at the retailer by household i in week t. Equal to zero if no spending takes place that week.
$\mathrm{REDEEM}_{i,t}$	*Stamps redeemed*. Measured as the number of stamps that are redeemed by household i in week t. Equal to zero if no redemption takes place that week.
$\mathrm{PREVREDEMP}_{i,t}$	*Previous redemption.* Dummy variable (0/1) that indicates whether household i has made a redemption in the previous week.
Message variables	
$\mathrm{ENG\_MESSAGE}_{i,t}$	*Engagement message.* This variable equals the number of engagement messages that are received by household i in week t.
$\mathrm{PROMO\_MESSAGE}_{i,t}$	*Promotion message.* This variable equals the number of promotion messages that are received by household i in week t.
$\mathrm{MESSAGES\_TOT}_{i,t}$	*Total messages.* This variable equals the total number of messages that are received by household i in week t, and is equal to the sum of engagement and promotion messages ($\mathrm{MESSAGES\_TOT}_{i,t} = \mathrm{ENG\_MESSAGE}_{i,t} + \mathrm{PROMO\_MESSAGE}_{i,t}$).
$\mathrm{MESSAGES\_TOT}_{i,t-x}$	*Lagged total messages.* $\mathrm{MESSAGES\_TOT}_{i,t-x}$ equals the number of messages received by household i in the previous week $(t-1)$ or the week before that $(t-2)$.
Moderators	
$\mathrm{BALANCE}_{i,t}$	*Previous stamp balance.* Measured as the stamp balance of household i, prior to (before the start of) week t. It is calculated as $\mathrm{BALANCE}_{i,t-1} + \mathrm{COLLECT}_{i,t-1} - \mathrm{REDEEM}_{i,t-1}$. Stamp collection during a given week is reflected in the balance of week $t+1$.
$\mathrm{DIST}_{i,t}$	*Distance to the reward threshold.* The number of stamps household i requires to reach the redemption threshold, prior to (before the start of) week t. In this case we consider the target/threshold to be ten stamps, as this is the modal number of stamps required, for which they can redeem a reward. If at least ten stamps have been collected, or in other words, if the previous stamp balance is equal to at least ten, distance to the target is equal to zero.
$\mathrm{PREVSPEND}_{i}$	*Average spending prior to the program (in 1000 Rupiah).* Equal to the average (grand mean-centered) spending of household i in the weeks prior to the start of the program.
Controls	
$\mathrm{ADOPT}_{i,t}$	*Adoption timing.* Step dummy that is equal to 1 from the app adoption time of household i onwards, 0 otherwise.
γ_t	*Week fixed effects.*
α_i	*Household fixed effects.*
$\mathrm{CF}_{i,t}$	*Correction factor.* Included in the 'number of stamps redeemed' layer of the hurdle model, to capture the correlation with the redemption incidence layer (see McFadden & Dubin 1984); it is measured as $\left(1-\hat{P}_{i,t}\right) * \frac{\ln\left(1-\hat{P}_{i,t}\right)}{\hat{P}_{i,t}} + \ln\left(\hat{P}_{i,t}\right)$, where $\hat{P}_{i,t}$ is the predicted redemption incidence probability for household i in week t.

To measure the effect of engagement vs. promotion push messaging on stamps-redemption behavior, we use a hurdle specification that uses the same regressors as in Equation 4.1. The first layer of the hurdle model captures the probability of redemption incidence (REDEMP_INC$_{i,t}$) using a binary-logit specification, while the second layer captures the number of stamps redeemed given incidence (REDEEM$_{i,t}$).[45] In line with Chapter 2, we follow McFadden and Dubin (1984) and estimate the two layers sequentially but capture their interdependence by using a correction factor in the second layer (see Table 4.1).

Next, we measure the effect of engagement vs. promotion push messaging on traffic, or the number of transactions of household *i* in week *t* (TRANSACTIONS$_{i,t}$). This is a count variable, for which a Poisson model is considered appropriate (van Nierop, Leeflang, Teerling, & Huizingh, 2011). As such, we opt for a fixed effects Poisson specification that uses the same regressors as our spending model. We note that in our spending model, in the 'number of stamps redeemed' layer of the hurdle model, and in the transactions model, the main effect of PREVSPEND$_i$ is subsumed in the fixed effects. This is not the case for our 'redemption incidence' layer (logit) of the hurdle model, where we include PREVSPEND$_i$ itself.

To accommodate heteroscedasticity, we use the sandwich estimator of variance (White-Huber standard errors) in our models.

[45] We opt for a continuous (linear) specification for the number of stamps redeemed given redemption incidence, because the number of stamps required for different rewards vary from one reward to the next, and given that consumers redeem any number of products at once, they can (thus) redeem almost any number of stamps on a given occasion.

4.5 Results

4.5.1 Spending

In Panel A of Table 4.2 we show the results for spending, which as we explained before, directly translates to stamp collection. When it comes to the saving dynamics, we find that they play a role and that their signs are in line with previous literature (and with the results in Chapter 2).

Our main interest though, is in the effect of push messaging, and engagement vs. promotion messages in particular. The results shows that the main effect of engagement messages on consumer spending and stamp collection is not significant ($p > .10$), but that promotion messages have a strong positive impact on spending ($\beta = 12.36$, $p < .001$). Moreover, when looking at the interactions with pre-program spending, we find that engagement messages do appear to have a small positive impact on spending for heavier buyers ($\beta = .0141$, $p < .01$). Furthermore, we find negative post-message effects (1-week lag: $\beta = -4.016$, $p < .001$) that are also stronger for these heavier buyers (1-week lag: $\beta = -.0408$, $p < .001$). In addition, similar to engagement messages, we find that the impact of promotion messages is strengthened for more-heavy buyers at the chain ($\beta = .0683$, $p < .001$). It seems heavier buyers are more likely to increase their spending as a result of promotion messages than lighter buyers, as they have more opportunity to do so.

4.5.2 Redemption

Table 4.2, Panel B, shows the estimation results for redemption incidence (left column) and number of stamps redeemed (right column). First, in the right column we can see that only stamp balance has a significant impact on the number of stamps that are redeemed. It seems that once redemption is set to occur, it is only the available stamp balance that determines how many stamps are redeemed. For redemption incidence, however, we find many significant effects. First, the main effects of the saving dynamics are what we would expect based on previous research. When it comes to the impact of messaging, we see that both engagement and promotion messages increase redemption incidence (β = .222, β = .399, $p < .001$), showing that messages in the same week help create program awareness and remind consumers that they have the option to redeem. However, this difference in effect size is significant[46], meaning that promotion messages are more effective in increasing the likelihood of redemption than engagement messages. This is an interesting and unexpected finding, given that promotion messages are mainly used to remind consumers of the special promotion that offers faster progression towards their saving goal. It seems that promotion messages may increase program salience in general, perhaps making consumers more aware of their progress in the program to begin with. This increased salience in turn can increase redemptions if consumers are reminded they had enough stamps to redeem. The stronger positive impact of promotion messages for heavier (than lighter) buyers at the chain (β = .000104, $p < .001$), also points to this. Alternatively, it could also be that for heavier buyers, the strong same-week increase in spending added enough stamps to consumers' balance to allow redemption. In contrast, however, we find no difference in the impact of engagement messages between heavier and lighter buyers ($p > .10$).

4.5.3 Traffic

Panel C of Table 4.2 reports the estimation results for the number of transactions (store visits). In terms of the savings dynamics, our results are in line with expectations based on previous findings. When it comes to the impact of messaging, we find that both engagement and promotion messages in the same week positively influence the number of transactions (β = .0143, β = .0527, $p < .001$). In line with our expectations, however, we do find that the effect of promotion messages is larger. Interestingly, the coefficient of

[46] We use the Delta Method to infer whether these coefficients are statistically different, and find that they are.

promotion messages for traffic is over 3 times larger than that of engagement messages, whereas for redemption incidence it was only twice as large. When comparing the main effects across the three dependent variables, the difference in impact is larger for traffic than it is for redemption, but the difference for spending is largest by far. Furthermore, when looking at the interaction with pre-program spending levels, we find that the positive effect of engagement messages and promotion messages are both strengthened for heavier buyers at the chain (β = .0000236, p < .01, β = .0000841, p < .001). These results are similar to what we found for our spending model, the main difference being that the main effect of engagement messages was insignificant there. It seems that promotion messages are more effective in increasing the number of store visits, and that for either message heavier buyers at the chain see more opportunity to incur an extra visit than lighter buyers. However, unlike for spending, we do not subsequently see fewer trips in the weeks thereafter (p > .10). In fact, we even find an increase in visits for the 2-week lag (β = .0139, p < .001), though these effects are slightly weaker for more heavy buyers of the chain (1-week lag: β = -.0000287, 2-week lag: β = -.0000784, p < .001).

Table 4.2: Estimation results

	Panel A	Panel B		Panel C
	Spending	Redemption Incidence	# of Stamps Redeemed	# of Transactions
$\text{BALANCE}_{i,t}$	-2.520*** (0.182)	0.0134*** (0.000393)	0.304*** (0.0277)	-0.000893*** (0.000225)
$\text{DIST}_{i,t}$	-3.502*** (0.378)	-0.187*** (0.00297)	-0.174 (0.362)	-0.0122*** (0.00106)
$\text{PREVREDEMP}_{i,t}$	30.11*** (1.790)	1.241*** (0.0108)	1.828 (2.219)	0.164*** (0.00481)
PREVSPEND_{i}		0.00150*** (0.0000389)		
$\text{ADOPT}_{i,t}$	113.4*** (3.110)			0.543*** (0.0128)
$\text{BALANCE}_{i,t} \times \text{PREVSPEND}_{i}$	-0.00219*** (0.000326)	-0.0000151*** (0.000000503)	-0.0000347 (0.0000359)	-0.00000221*** (0.000000394)

$\text{DIST}_{i,t} \times \text{PREVSPEND}_i$	-0.00432*** (0.00127)	0.000137*** (0.00000472)	-0.000120 (0.000297)	0.00000189 (0.00000214)
$\text{PREVREDEMP}_{i,t} \times \text{PREVSPEND}_i$	0.0520*** (0.00840)	-0.000314*** (0.0000323)	0.00116 (0.00104)	0.0000116 (0.0000139)
$\text{ADOPT}_{i,t} \times \text{PREVSPEND}_i$	0.320*** (0.0253)			-0.0000299 (0.0000523)
$\text{ENG\_MESSAGE}_{i,t}$	-0.141 (0.784)	0.222*** (0.0103)	0.664 (0.455)	0.0143*** (0.00311)
$\text{PROMO\_MESSAGE}_{i,t}$	12.36*** (1.426)	0.399*** (0.0193)	0.751 (0.810)	0.0527*** (0.00606)
$\text{ENG\_MESSAGE}_{i,t} \times \text{PREVSPEND}_i$	0.0141** (0.00486)	-0.0000174 (0.0000233)	-0.000206 (0.000397)	0.0000236** (0.00000867)
$\text{PROMO\_MESSAGE}_{i,t} \times \text{PREVSPEND}_i$	0.0683*** (0.00670)	0.000104*** (0.0000303)	-0.000338 (0.000494)	0.0000841*** (0.0000108)
$\text{MESSAGES\_TOT}_{i,t-1}$	-4.016*** (0.642)	0.122*** (0.00824)	-0.102 (0.273)	-0.00176 (0.00255)
$\text{MESSAGES\_TOT}_{i,t-2}$	1.285 (0.691)	0.0721*** (0.00767)	0.150 (0.175)	0.0139*** (0.00272)
$\text{MESSAGES\_TOT}_{i,t-1} \times \text{PREVSPEND}_i$	-0.0408*** (0.00365)	-0.000105*** (0.0000184)	0.0000328 (0.000381)	-0.0000287*** (0.00000638)
$\text{MESSAGES\_TOT}_{i,t-2} \times \text{PREVSPEND}_i$	-0.0252*** (0.00417)	-0.0000525** (0.0000186)	-0.000253 (0.000302)	-0.0000784*** (0.00000678)
$\text{MESSAGES\_TOT}_{i,t} \times \text{BALANCE}_{i,t}$	-0.132* (0.0612)	-0.00109*** (0.000201)	0.00319 (0.00572)	-0.000374*** (0.0000910)
$\text{MESSAGES\_TOT}_{i,t} \times \text{DIST}_{i,t}$	-0.911*** (0.170)	0.0141*** (0.00162)	0.0291 (0.0425)	-0.00194*** (0.000534)
$\text{CorrectionFactor}_{i,t}$			0.370 (2.100)	
Constant	154.0*** (4.503)	-2.689*** (0.0502)	15.72* (7.722)	
Observations	651056	636981	93530	646618
R^2 (pseudo-R^2)	0.042	(0.167)	0.166	

All models include time fixed effects, the (linear) models for spending and # stamps redeemed and the (Poisson) model for # transactions also include household fixed effects, the (logit) model for redemption incidence does not. Standard errors in parentheses. Reported R-squares net of fixed effects.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

4.5.4 Effect sizes

As the previous sections show, for the average consumer, both types of push messages have a significant and positive impact on redemption incidence and the number of transactions. For spending, we find a large positive impact of promotion messages, which is strengthened for heavy buyers, but we only find a positive impact of engagement messages for heavy buyers. Furthermore, we find that the impact of promotion messages is larger than that of engagement messages for all 3 dependent variables, though the size of this difference in impact varies. However, given that we include many interactions and lags in our models, it is difficult to gauge the (over-time) impact of sending an additional push message, be it engagement or promotion, based on the coefficients alone. In addition, it is also not straightforward to determine how effective one type of message is compared to the other in terms of exact effect size. Hence, to properly assess this impact, and to see how it varies between more- vs. less-heavy buyers of the chain, we use the results of our estimation as inputs for dynamic simulations. Following the same procedure as in Chapter 2, we predict household spending, reward redemption and number of trips in the course of the program for various different push plans and compare these predictions. As in Chapter 2, we do so for different levels of households' pre-program spending, where we simulate 1000 trajectories for each corresponding level and average the outcomes across those 1000 draws. Recall from Chapter 2 that we account for uncertainty in the estimated parameters, by repeating these simulations for 1000 draws from the parameter sampling distribution. This allows us to assess the significance of the differences between the different push plans.

As mentioned, our main focus is on comparing the impact of engagement messages to that of promotion messages. To this end, our simulation results in Table 4.3, show the impact of sending one additional message (in different weeks of the program) on top of the actual push plan. We furthermore make a distinction between the immediate (same week) effect only, and the total effect over all weeks of the program. The results show that both engagement and promotion messages have a positive impact on our three dependent variables, but also provide interesting additional insights.

First, we find that promotion messages are more effective than engagement messages across the board. To illustrate, we find that the total impact of promotion messages on spending is, on average, about 2.5 times larger.

Second, we find that while promotion messages are more effective for all three performance metrics, the difference is larger for spending than for redemption. This means that the gap between available stamps and stamps that were actually redeemed (the redemption gap) – which previous literature has indicated to be a problematic phenomenon – remains larger when sending promotion messages rather than engagement ones. In addition, when looking more closely at the changes in store spending and traffic that result from sending additional messages, we find that the relative increase in spending is larger than the relative increase in traffic, for both engagement and promotion messages[47]. This means that the increased spending that results from messaging is not only caused by consumers visiting the store more frequently, but rather that they also spend more per trip.

Third, we find that the difference in impact also varies depending on which consumer type we look at. For heavy buyers at the chain, we see that engagement messages have a positive immediate impact on spending and traffic, but that the total impact is less positive (or even negative). We see a similar pattern when we look at the impact of promotion messages. For traffic, though, we see largely the same increase regardless of which customer group or week we look at.

[47] The only exception here being engagement messages for heavy buyers, which do not increase spending.

Table 4.3: Impact of adding extra messages

	Panel A: One additional push in week 4				Panel B: One additional push in week 7				Panel C: One additional push in week 10			
	Consumer Type				**Consumer Type**				**Consumer Type**			
	Average	Low 25%	Median	High 25%	Average	Low 25%	Median	High 25%	Average	Low 25%	Median	High 25%
Spending												
Engagement message												
Immediate effect	2.426	0.755	1.543	6.131	1.589	-0.293	0.643	5.253	1.086	-0.700	0.189	4.805
Total effect	9.840	13.479	11.683	3.771	10.269	13.455	11.385	1.705	8.812	14.514	11.293	-2.499
Promotion message												
Immediate effect	14.924	5.296	10.014	34.640	14.087	4.247	9.114	33.761	13.584	3.840	8.660	33.314
Total effect	24.429	20.014	21.638	38.671	26.286	21.650	23.101	36.889	25.579	22.364	23.891	32.147
Stamps Redeemed												
Engagement message												
Immediate effect	0.393	0.415	0.426	0.634	0.638	0.438	0.517	0.637	0.631	0.632	0.617	0.689
Total effect	0.377	0.421	0.398	0.250	0.511	0.532	0.481	0.401	1.128	1.157	1.116	0.944
Promotion message												
Immediate effect	0.919	0.774	0.817	1.684	1.247	1.025	1.097	1.677	1.441	1.195	1.337	1.873
Total effect	0.749	0.603	0.655	1.164	0.971	0.808	0.830	1.384	1.830	1.594	1.730	1.991
Traffic												
Engagement message												
Immediate effect	0.038	0.034	0.035	0.047	0.032	0.027	0.030	0.040	0.030	0.025	0.028	0.040
Total effect	0.067	0.087	0.077	0.031	0.064	0.085	0.074	0.028	0.067	0.090	0.078	0.025
Promotion message												
Immediate effect	0.113	0.092	0.103	0.155	0.104	0.084	0.094	0.140	0.104	0.082	0.093	0.146
Total effect	0.144	0.145	0.145	0.154	0.142	0.146	0.143	0.143	0.148	0.152	0.150	0.147

4.6 Discussion

With the recent shift to a digital age, retailers have been searching for ways to adapt their marketing strategies to fit an online promotion landscape. For loyalty programs this trend is also apparent, where digital saving through a mobile application is becoming increasingly common. Yet, while digital saving via a mobile application paves the way for new forms of direct communication with consumers it is still unclear what the impact of message type is in the context of mobile push messaging. Our study addresses this gap. Using a unique dataset we measure the impact of engagement- vs. promotion-oriented messages during a TLP. We offer several interesting findings, with important managerial implications.

4.6.1 Findings

We find that both engagement and promotion messages, have a positive overall impact on spending, redemption incidence and store traffic[48]. One of the key findings of this study, however, is that promotion messages are more effective than engagement messages but that the difference in impact varies with our three variables of interest. We find that the increase in impact is largest for spending, where promotion messages on average are about 2.5 times more effective than their engagement counterparts. It seems that drawing attention to the added benefit the promotion message offers, is more likely to make consumers act than sending engagement messages. In addition, though we expected the promotion message might draw attention away from the program to the discounted product on offer as well, we find this does not distract consumers from their saving or redemption goals. Furthermore, for redemption and traffic, we find a smaller difference between engagement and promotion messages, meaning that sending promotion messages not only causes consumers to visit the chain more often, but that it also increases their basket size. In addition, it also indicates the

[48] Note that engagement messages directed at heavy buyers do not increase spending throughout the weeks.

redemption gap between available stamps and stamps that were actually redeemed, remains larger when sending promotion messages compared to engagement ones.

When taking into account the different consumer groups, we find that, in terms of total impact, engagement messages are most beneficial for less-heavy buyers. For promotion messages we find the opposite: they are more effective for heavier buyers at the chain. For traffic, however, we do not find large differences between lighter and heavier buyers at the chain. As such, we can conclude that heavy buyers in particular, have more opportunity to increase their spending and incur extra visits to the retailer when needing to react to promotion messages in a timely manner.

4.6.2 Managerial implications

One of the main findings of our study, is that different types of messages primarily affect different consumer types. While we find that engagement and promotion messages can both have a positive total impact on consumer spending, redemptions and store visits, we find that for heavy buyers in particular, we are better off not sending engagement messages. In addition, we find that pushing promotion messages is more effective than pushing engagement messages for all consumer types, but particularly so for these same heavy buyers. A clear managerial recommendation therefore seems to be to send more promotion messages. However, before we can conclude such a thing, we should recall from Chapter 2, that these extra redemptions also come at a cost to the retailer. In practice, the LPP itself is often used as a supplier collaboration, where brand manufacturers pay to be the brand on offer in a given week. This means that it is common practice that the cost of the price cut and of the free bonus stamp, are carried by the brand manufacturer, rather than the retailer. As such, the costs we are referring to here, are the costs of more redemptions to the retailer, who pay a price to the program operator per redeemed item. The question is then whether these costs are covered by the increase in spending (and the extra payments consumers pay to the retailer to obtain the reward). To determine this, we do a profitability calculation using the following formula:

$$\Delta PROFIT = MARGIN * \Delta SPEND - LOSS\_REWARD * (\Delta REDEEM/STAMPS\_REWARD)$$

Table 4.4: Profitability of push messaging (1 additional message in week 4 of the program)

		Change in		
Consumer Type	**Baseline spending (in IDR, across program weeks)**	**Margin on amount spent (in IDR, across program weeks)**	**Cost on redeemed stamps (in IDR, across program weeks)**	**Profit (in IDR, across program weeks)**
Engagement messages				
Average	3,303,700	2,460.0	301.6	2,158.4
Low Quartile	2,868,100	3,369.8	336.8	3,033.0
Median	3,084,000	2,920.8	318.4	2,602.4
Upper Quartile	4,186,700	942.8	200.0	742.8
Promotion messages				
Average	3,303,700	6,107.3	599.2	5,508.1
Low Quartile	2,868,100	5,003.5	482.4	4,521.1
Median	3,084,000	5,409.5	524.0	4,885.5
Upper Quartile	4,186,700	9,667.8	931.2	8,736.6

Note: 1000 IDR is approximately equal to 0.07 USD. Margin obtained as change in spending (see Table 4.3) times .25, a typical retailer margin for groceries. Cost obtained as change in redeemed stamps (see Table 4.3) divided by average number of stamps per reward (11.938) times average retailer loss per reward (9,550.32 IDR) (Figures obtained from the company). The numbers needed to calculate profit are the same as in Chapter 2.

To get an idea for the profitability of one type of message compared to the other, we calculate the profit of sending one additional push message in week 4 of the program. We find that (see Table 4.4), though redemption comes at a cost to the retailer, this is more than offset by the increase in spending, both for engagement messages (though to a lesser extent for heavy buyers) as well as for promotion messages – the latter being particularly profitable. We find that sending one additional message can increase retailer profit by .26% for engagement and even .67% for promotion messages on average, a very substantial effect. During program negotiations with the retailer, program operators can take into account this strong increase in retailer profit, to help them sell more digital programs that facilitate push messaging. This is also beneficial to the program operator themselves, considering that push messaging also has a positive impact on redemption.

Overall, we can conclude that push messages – and especially promotion messages – are a very effective tool to enhance consumer participation in temporary loyalty programs. For retailers and program operators involved in these programs, we thus recommend favoring promotion messages over engagement messages. Our data show that this stands in contrast with what is currently industry practice, as we found that throughout the TLP only 6 promotion messages were sent, compared to 20 engagement messages. How many promotion messages are sent during a program can be limited by the number of LPPs on offer throughout the program. As such, retailers (and program operators) should facilitate brand manufacturers to have their product on offer as an LPP, to ensure ample opportunity to send promotion messages. For the programs considered in past studies however (Dorotic et al., 2011; Zhang & Breugelmans, 2012; Minnema et al., 2017), enough LPPs are on offer. If ample LPPs are available, retailers (or program operators) could also attempt to earn even more from sending promotion messages. By asking the brand manufacturer to pay an additional fee for featuring their product in the message, promotion messages could be even more beneficial to the retailer or program operator sending them.

4.6.3 Limitations and future research

While this study offers valuable new insights in the effectiveness of engagement vs. promotion types of mobile push messages, it also comes with limitations that pave the way for future research. First, though the randomized control group(s) and the panel structure of our data enable us to avoid endogeneity concerns, our data is limited in other ways. One such way, is that we are unable to allow for a full distinction in engagement and promotion messages in our moderator and lags. For these variables we opt for using a total message variable to avoid overfitting that would result from including separate variables for each message type. Though we still find ample evidence of different effectiveness for the two message types, we may have found larger or smaller differences had we been able to include this distinction even further. Therefore, we may be under- or overestimating the difference in impact to some extent.

Second, while our study provides an interesting finding about how promotion messages in particular are effective in increasing spending, store visits, and to a slightly lesser extent also redemption, we are only able to analyze one specific program. Though LPPs help alleviate consumers perceived pressure of having to make the reward threshold in a short time frame (by offering a free bonus stamp with purchase), the LPP too can come with its own design elements. This means that the promotional conditions of the LPPs that are the focus of our promotion messages may differ. For instance, in our program, LPPs are certain products that are only on offer for a week, meaning that considerable pressure is on consumers to act on the promotion message in the same week. This perceived pressure could potentially influence how effective the messaging is. For instance, the impact might be mitigated (enhanced) if the offer is valid for two weeks (only a weekend), rather than one week. Therefore, future research may want to extend this research to determine whether the effects (or the effect sizes) of promotion-type messages are different depending on the design of the LPP itself. In addition, in the program we study, no promotion messages were sent in later program weeks. Though we take this into account in our simulations, by not simulating push messages for those weeks where we did not observe them, we still advise caution in interpreting the results and effect sizes as even for all stages of the program. Finally, in the program we analyze, the number of promotion messages being sent per week, never exceeds one, whereas the number of engagement messages does. In addition, the number of engagement messages sent throughout this program far exceeds the number of promotion messages. While promotion messages tend to be less common in TLPs in general (based on anecdotal evidence from the program operator), one might consider that consumers become accustomed to the more commonly sent engagement messages, which could make promotion messages stand out more, thus increasing their effectiveness. As such, this may impact the generalizability of our findings, which future research may wish to shed more light on.

Third, though we are able to provide a profitability calculation for the two message types we consider, we cannot, for lack of data, take into account that LPPs are in essence often used as supplier collaborations. In these collaborations, brand manufacturers often pay a fee to the retailer to be the brand on offer in a given week. Given that we have no knowledge of a potential fee that brand manufacturers are paying to the retailer, or know exactly how the retailer or brand manufacturer is paying for offered discount of the product or the extra stamp that is given away, we cannot take this into account in our profitability calculations. Future research might want to take this into account, given that this likely affects how profitable promotion messages are for the retailer.

Finally, while we find valuable insights on the impact of different message types, we do not take into account that the precise content of promotion messages may differ slightly from one week to the next. While the general point of the message is the same, the content will vary depending on how many and which products are on offer that week, which may impact their effectiveness.

5. Conclusion

5.1 Findings & Implications

In this dissertation, we set out to identify key mechanisms to activate consumers' engagement in temporary loyalty programs (TLPs). The activation of consumers is especially important in these programs, as their finite duration only allows a small window of opportunity to spark a consumer's interest. While some activation techniques (in general) are covered in past literature, there are still many forms that received little to no attention (in a general context, or in a TLP setting). Therefore, in this dissertation, we assess the impact of several activation techniques, but also identify when, or for whom, such techniques work better, and do so for several important industry performance metrics that are relevant for various key players (retailers, program operators, and reward manufacturers). Specifically, in Chapter 2, we show that push messaging has a positive impact on both redemption and sales, although the effect is stronger for heavier buyers at the chain. In Chapter 3, we illustrate that retailers have four in-store instruments that can influence how successful programs are in terms of increasing sales, while they are not all equally important for all types of stores. Finally, in Chapter 4, we show that the content of activation techniques also matters, in the sense that promotion messages, which alert consumers to the opportunity to earn bonus stamps through the purchase of specific loyalty program promotions, have a stronger impact on redemption and sales behavior than engagement messages.

Combined, these essays furthermore offer us several other important insights. First, recent literature tends to have a strong focus on the online and/or digital environment. Yet, we show that activation mechanisms in both an online setting (Chapters 2 + 4) and in an offline setting (Chapter 3) are important. While online and digital methods are getting more and more integrated in everyday life, it does not mean that the offline methods no longer have an impact or should be neglected. Second, we show that the impact of activation techniques is not universal, but rather dependent on a variety of contingency factors. Overall, we illustrate that there is considerable heterogeneity in the impact of activation techniques that needs to be considered when devising an activation strategy. Explicitly, we find important differences on a consumer level (Chapter 2 + 4), a store level (Chapter 3), and on message type (Chapter 4). Finally, we shed more light on the impact on several performance metrics that are relevant for a diverse set of TLP players. While previous literature predominantly focuses on standard performance metrics (i.e., typically sales), Breugelmans et al. (2015) indicates that there is a wide variety of loyalty program performance measures. Along the same line, van Heerde, Moorman, Moreau and Palmatier (2021) argue that it is important not to lose track of performance metrics that are relevant to all key stakeholders, in order to maintain high ecological validity. Therefore, we do not only focus on sales (Chapters 2 + 3 + 4), but also study the impact on redemption (Chapters 2 + 4) and shopping frequency (Chapter 4). Our insights are thus not only relevant for retailers (who typically mainly focus on sales), but also for other parties in a TLP setting, namely program operators and reward manufacturers (who typically also focus on redemption).

5.2 Areas for future research

Even though we provide insights into novel and underresearched activation techniques under different circumstances, there are still some boundary conditions to our essays, which provide fruitful areas for future research. These themes are more general and go beyond the topics for future research in the individual chapters. Specifically, we discuss (i) the combination of activation techniques, (ii) the power and compliance between industry players, and (iii) the changing retail landscape.

5.2.1 Combination of activation techniques

Each of our essays investigated a TLP activation technique in isolation (i.e., push messaging in Chapters 2 + 4, and four in-store metrics in Chapter 3). However, retailers or program operators can utilise several of these mechanisms simultaneously, and the question arises whether such combinations will lead to synergy effects, or whether they result in diminishing returns. For example, proper execution of in-store activation might raise awareness of the temporary program, which might be a necessary condition for consumers to download the app, which is a requirement for sending push messages. On the other hand, both proper execution of in-store instruments as well as push notifications can be considered as tools reminding consumers to save or redeem in the TLP, making a combination of the two unnecessary. The similarities in types of channels (i.e., online and offline) might play a role here as well. For example, the digital push messaging might complement the proper execution of in-store activation, as they might speak to, and be used by, different types of consumers, whereas push messaging in combination with emails could speak to the same type of (digitally oriented) consumer, and be seen as redundant. One notable study in regard to the combination of activation techniques is that of Dorotic et al. (2011), who find that the joint usage of email and post leads to a 13% increase in issued loyalty points. Of course, there are a wide variety of combinations possible that are left for future research. In addition, if our essays are any indication, each of these combinations might not display a uniform effect, but rather an effect that is furthermore dependent on the type of consumer, store, or message content.

5.2.2 The power and compliance between industry players

While we cover several performance metrics for various TLP players, the goals of each player might not always be aligned. Retailers are mostly interested in a revenue lift, whereas the redemptions are the main source of income for program operators and reward manufacturers. Despite initial evidence that activation techniques (such as push messaging in Chapters 2 + 4) improve both sales and redemptions, their effects are not always equal or equally strong. For example, we show that back-end loading has a considerable effect on redemption, but less so on sales, where messages sent in mid-program weeks have the largest impact. In addition, since the retailer might incur a cost for each redemption, retailers ideally want to raise sales without increasing redemption. However, this might be difficult, as we find that the mere act of redemption subsequently increases sales, in line with the rewarded behavior effect found in for example Dorotic et al. (2014) and Taylor & Neslin (2005). Overall, the impact of activation techniques might not always be similar between the different performance metrics, and which mechanisms eventually get utilised might partly depend on which party has the most power. For example, if a large program operator runs a program at a small retailer, they might push strategies that have a stronger impact on redemption, whereas small operators working for large retailers might not be able to do so. We know very little about the impact of channel power regarding loyalty programs, even though past literature (e.g., Geyskens, Gielens, & Dekimpe, 2002) shows that power in general can play a role. Hence, there is still a rich area for future research in investigating whether power differences impact the use of activation techniques, and their subsequent impact on performance metrics.

Another source of power dynamics might exist within retailers themselves. Stores (or franchisees) might not find that the central retailer strategy is optimal for them. Indeed, the execution strategies covered by our essays are largely centralised (i.e., similar to global integration, where the same strategy is utilised everywhere), while local adaptations or more hybrid forms might be more useful in some situations (Steenkamp & Geyskens, 2014). Hence, it might be interesting to test whether stores of centralised or decentralised retailers comply better with the retailer's or operator's guidelines, but also whether they are indeed equally effective.

5.2.3 Changing retail landscape

The retailer landscape is ever evolving, and the new trends might have an impact on the success of loyalty programs in general, as well as on the subsequent use of activation techniques. One obvious trend is the increasing use of online purchasing of groceries (which became particularly popular during the COVID-19 pandemic). They can be delivered at home, or using a click & collect method, where consumers either pick up their purchases in-store, near-store or at a separate location altogether (Gielens, Gijsbrechts, & Geyskens, 2021). The shift to more online purchases can provide both opportunities and challenges to the activation techniques. For example, it might require less effort for consumers to react to mobile push messaging, as they can immediately act and order products online at any time or place, which is far more convenient (and less time consuming) than having to visit a physical store. However, investing in proper execution of in-store activation is likely to become less useful when more and more consumers decide to purchase online. While in-store instruments can be replaced by online counterparts (e.g., online banners) in an attempt to still increase program salience, it is unclear whether they grab equal attention and are thus equally effective. For example, the physical displays in the store give consumers an opportunity to see and feel the quality of the reward items, which can not easily be replicated online. In addition, for online banners, it is even more difficult to ascertain what can be considered successful execution, as this may be consumer- rather than store-specific.

Another recent trend is the use of self-checkout or even no-checkout zones in stores (Dekimpe, Geyskens, & Gielens, 2020). The former includes self-scanning desks typically located at the regular checkout, or scanning-on-the-go with a special store-scanner or app via your mobile phone. The latter can use pressure sensors with a combination of cameras and biometric authentication to provide a fully autonomous and low-effort transaction convenience. All such methods decrease the amount of contact between consumers and store employees, making personal communication and information about loyalty programs difficult. Also, the distribution of stamps is currently not included in either the self-checkout of no-checkout machines, meaning consumers have to spend more effort to visit an information desk to obtain them. Hence, program operators might have to find new ways to adapt the loyalty programs and activation mechanisms to these new circumstances, and future research can subsequently identify the effect of these adaptations.

References

Abel, Jaison R. and Gabe, Todd M. and Stolarick, Kevin, Workforce Skills across the Urban-Rural Hierarchy (February 1, 2012). FRB of New York Staff Report No. 552, Available at SSRN: https://ssrn.com/abstract=2010646 or http://dx.doi.org/10.2139/ssrn.2010646.

Ailawadi, K. L., Beauchamp, J. P., Donthu, N., Gauri, D. K., & Shankar, V. (2009). Communication and Promotion Decisions in Retailing: A Review and Directions for Future Research. *Journal of Retailing, 85*, 42-55.

Anderson, E. (1988). Strategic implications of Darwinian economics for selling efficiency and choice of integrated or independent sales forces. *Management Science*, *34*(5), 599-618.

Andrews, M., Goehring, J., Hui, S., Pancras, J., & Thornswood, L. (2016). Mobile promotions: A framework and research priorities. *Journal of Interactive Marketing*, *34*, 15-24.

Appel, G., Libai, B., Muller, E. & Shachar, R. (2020). On the monetization of mobile apps. *International Journal of Research in Marketing*, *37*(1), 93-107.

Baker, J., Levy, M., & Grewal, D. (1992). An experimental approach to making retail store environmental decisions. *Journal of Retailing, 68*, 445.

Baker, J., Parasuraman, A., Grewal, D., & Voss, G. B. (2002). The influence of multiple store environment cues on perceived merchandise value and patronage intentions. *Journal of Marketing, 66*, 120-141.

Bart, Y., Stephen A.T. & Sarvary, M. (2014). Which products are best suited to mobile advertising? A field study of mobile display advertising effects on consumer attitudes and intentions. *Journal of Marketing Research, 51*(June), 270–285.

Bellman, S., Potter, R. F., Treleaven-Hassard, S., Robinson, J. A., & Varan, D. (2011). The effectiveness of branded mobile phone apps. *Journal of Interactive Marketing*, *25*(4), 191-200.

Bemmaor, A. C., & Mouchoux, D. (1991). Measuring the Short-Term Effect of In-Store Promotion and Retail Advertising on Brand Sales: A Factorial Experiment. *Journal of Marketing Research, 28*, 202-214.

Bijmolt, T. H., Dorotic, M., & Verhoef, P. C. (2011). Loyalty programs: Generalizations on their adoption, effectiveness and design. *Foundations and Trends® in Marketing*, *5*(4), 197-258.

Bijmolt, T. H., & Verhoef, P. C. (2017). Loyalty programs: Current insights, research challenges, and emerging trends. In *Handbook of marketing decision models* (pp. 143-165). Springer, Cham.

Blattberg, R. C., Kim, B. D., & Neslin, S. A. (2008). Frequency reward programs. In: *Database Marketing* (pp. 549-578). Springer, New York, NY.

Bolton, R.N., Kannan, P. K. & Bramlett, M. D. (2000). Implications of loyalty program membership and service experiences for customer retention. *Journal of the Academy of Marketing Science*, *28*, 95-108.

Bombaij, N. J. F. & Dekimpe, M. G. (2020). When do loyalty programs work? The moderating role of design, retailer strategy, and country characteristics. *International Journal of Research in Marketing, 37*(1), 175-195.

Bombaij, N. J. F. (2021). Effectiveness of Loyalty Programs. Tilburg: Tilburg University, CentER Graduate School.

Bombaij, N. J. F. Gelper, S. & Dekimpe, M. G. (2022). Designing Successful Temporary Loyalty Programs: An Exploratory Study on Retailer and Country Differences. *International Journal of Research in Marketing,* forthcoming.

Braun, M., & Moe, W. W. (2013). Online display advertising: Modeling the effects of multiple creatives and individual impression histories. *Marketing Science*, *32*(5), 753-767.
Breugelmans, E., Bijmolt, T. H., Zhang, J., Basso, L. J., Dorotic, M., Kopalle, P., & Wünderlich, N. V. (2015). Advancing research on loyalty programs: a future research agenda. *Marketing Letters*, *26*(2), 127-139.
Breugelmans, E., Gijsbrechts, E., & Campo, K. (2018). Product unavailability. In *Handbook of Research on Retailing*: Edward Elgar Publishing.
Brogie, M. (2019). What Is Retail Execution? (+Effective Strategies to Manage It). Retrieved from: https://learn.g2.com/retail-execution.
Byun, S.-E., & Sternquist, B. (2012). Here today, gone tomorrow: Consumer reactions to perceived limited availability. *Journal of Marketing Theory and Practice, 20*, 223-234.
Cândido, C. J. F., & Santos, S. P. (2015). Strategy implementation: What is the failure rate? *Journal of Management & Organization, 21*, 237-262.
Castro, I. A., Morales, A. C., & Nowlis, S. M. (2013). The influence of disorganized shelf displays and limited product quantity on consumer purchase. *Journal of Marketing, 77*, 118-133.
Chain Store Age (2007). RFID reconciles CPG and retailer relationships- HP reduced phantom inventory and backroom discrepancies. 83 (9), September, p. 6A.
Chauduri, M., Voorhees, C. M. & Beck, J. M. (2019). The effects of loyalty program introduction and design on short- and long-term sales and gross profits. *Journal of the Academy of Marketing Science*, *47*, 640–658.
Corsten, D., & Gruen, T. (2003). Desperately seeking shelf availability: an examination of the extent, the causes, and the efforts to address retail out-of-stocks. *International Journal of Retail & Distribution Management*.
Danaher, P. J., Sajtos, L., & Danaher, T. S. (2020). Tactical use of rewards to enhance loyalty program effectiveness. *International Journal of Research in Marketing, 37*, 505-520.
Danaher, P. J., Smith, M. S., Ranasinghe, K. & Danaher, T. S. (2015). Where, when, and how long: Factors that influence the redemption of mobile phone coupons. *Journal of Marketing Research, 52*(5), 710–72.
Datta, H., van Heerde, H. J., Dekimpe, M. G., & Steenkamp, J. B. E. (2022). Cross-National Differences in Market Response: Line-Length, Price, and Distribution Elasticities in Fourteen Indo-Pacific Rim Economies. *Journal of Marketing Research*, *59*(2), 251-270.
Dekimpe, M. G., Geyskens, I., & Gielens, K. (2020). Using technology to bring online convenience to offline shopping. *Marketing Letters*, *31*(1), 25-29.
Dhar, S. K., Hoch, S. J., & Kumar, N. (2001). Effective category management depends on the role of the category☆. *Journal of Retailing, 77*, 165-184.
Donovan, R., & Rossiter, J. (1982). Store atmosphere: an environmental psychology approach. *Journal of Retailing, 58*, 34-57.
Dorotic, M. (2010). The influence of marketing instruments and rewarding on cardholders' behavior in coalition loyalty programs. Groningen: University of Groningen, SOM research school.
Dorotic, M., Fok, D., Verhoef, P. C., & Bijmolt, T. H. (2011). Do vendors benefit from promotions in a multi-vendor loyalty program?. *Marketing Letters*, *22*(4), 341-356.
Dorotic, M., Verhoef, P. C., Fok, D., & Bijmolt, T. H. (2014). Reward redemption effects in a loyalty program when customers choose how much and when to redeem. *International Journal of Research in Marketing*, *31*(4), 339-355.
Drèze, X., & Hoch, S. J. (1998). Exploiting the installed base using cross-merchandising and category destination programs. *International Journal of Research in Marketing*, *15*(5), 459-471.

Drèze, X., & Nunes, J. C. (2011). Recurring goals and learning: The impact of successful reward attainment on purchase behavior. *Journal of Marketing Research*, *48*(2), 268-281.
Dubé, J. P., Fang, Z., Fong, N. & Luo, X. (2017). Competitive price targeting with smartphone coupons. *Marketing Science*, *36*(6), 944-975.
East, R., Eftichiadou, V., & Williamson, M. (2003). Research note: point-of-purchase display and brand sales. *The International Review of Retail, Distribution and Consumer Research, 13*, 77-98.
eMarketer (2021). Worldwide Digital Ad Spending Year-End Update. Retrieved from: https://www.emarketer.com/content/worldwide-digital-ad-spending-year-end-update.
Fang, Z., B. Gu, X. Luo & Y. Xu (2015). Contemporaneous and delayed sales impact of location-based mobile promotions. *Information Systems Research*, 26(3), 552–564.
Fong, N., Z. Fang, & X. Luo (2015). Geo-conquesting: competitive locational targeting of mobile promotions. *Journal of Marketing Research*, *52*(October), 726-735.
Germann, F., Ebbes, P., & Grewal, R. (2015). The chief marketing officer matters! *Journal of Marketing, 79*, 1-22.
Geyskens, I., Gielens, K., & Dekimpe, M. G. (2002). The market valuation of internet channel additions. *Journal of Marketing*, *66*(2), 102-119.
GfK (2015). Meer dan de helft van Nederland spaart mee voor huishoudelijke artikelen. Retrieved from https://www.gfk.com/nl/insights/news/meer-dan-de-helft-van-nederland-spaart-mee-voor-huishoudelijke-artikelen/.
Ghose, Li and Liu (2019). Mobile targeting using customer trajectory patterns. *Management Science*, *65*(1), 5027-5049.
Gielens, K., & Dekimpe, M. G. (2007). The entry strategy of retail firms into transition economies. *Journal of Marketing*, *71*(2), 196-212.
Gielens, K., Gijsbrechts, E., & Geyskens, I. (2021). Navigating the Last Mile: The Demand Effects of Click-and-Collect Order Fulfillment. *Journal of Marketing*, *85*(4), 158-178.
Gomez, L., & Sides, R. (2015). *Retail Execution: the New Differentiator.* Deloitte.
Grewal, D., Bart, Y., Spann, M., & Zubcsek, P. P. (2016). Mobile advertising: A framework and research agenda. *Journal of Interactive Marketing*, *34*, 3-14.
Grewal, R., Chandrashekaran, M., & Dwyer, F. R. (2008). Navigating local environments with global strategies: A contingency model of multinational subsidiary performance. *Marketing Science*, *27*(5), 886-902.
Grewal, D., Roggeveen, A. L., Sisodia, R., & Nordfält, J. (2017). Enhancing Customer Engagement Through Consciousness. *Journal of Retailing, 93*, 55-64.
Gu, X. & Kannan, P. K. (2021). The dark side of mobile app adoption: Examining the impact on customers' multichannel purchase. *Journal of Marketing Research*, forthcoming.
Hacker, P. A., Floerkemeier, C., Sarma, S., & Schuh, G. (2010, April). RFID diagnostics of promotion execution. In 2010 IEEE International Conference on RFID (IEEE RFID 2010) (pp. 134-140). IEEE.
Helmefalk, M., & Hultén, B. (2017). Multi-sensory congruent cues in designing retail store atmosphere: Effects on shoppers' emotions and purchase behavior. *Journal of Retailing and Consumer Services, 38*, 1-11.
Inman, J. J., McAlister, L., & Hoyer, W. D. (1990). Promotion signal: proxy for a price cut? *Journal of consumer research, 17*, 74-81.
Kahneman, D., & Tversky, A. (2013). Prospect theory: An analysis of decision under risk. In *Handbook of the fundamentals of financial decision making: Part I* (pp. 99-127).
Kaul, A., & Wittink, D. R. (1995). Empirical generalizations about the impact of advertising on price sensitivity and price. *Marketing Science*, *14*(3_supplement), G151-G160.

Kim, B. D., Shi, M., & Srinivasan, K. (2001). Reward programs and tacit collusion. *Marketing Science*, *20*(2), 99-120.
Kim, J. J., Steinhoff, L., & Palmatier, R. W. (2021). An emerging theory of loyalty program dynamics. *Journal of the Academy of Marketing Science*, *49*(1), 71-95.
Kim, S.J., Wang, R.J.H, & Malthouse, E.C. (2015). The effects of adopting and using a brand's mobile application on customers' subsequent purchase behavior, *Journal of Interactive Marketing*, *31*, 28-41.
Kivetz, R. (2003). The effects of effort and intrinsic motivation on risky choice. *Marketing Science*, *22*(4), 477-502.
Kivetz, R., Urminsky, O., & Zheng, Y. (2006). The goal-gradient hypothesis resurrected: Purchase acceleration, illusionary goal progress, and customer retention. *Journal of Marketing Research*, *43*(1), 39-58.
Koo, M., & Fishbach, A. (2008). Dynamics of self-regulation: How (un) accomplished goal actions affect motivation. *Journal of Personality and Social Psychology*, *94*(2), 183.
Kopalle, P. K., Sun, Y., Neslin, S. A., Sun, B., & Swaminathan, V. (2012). The joint sales impact of frequency reward and customer tier components of loyalty programs. *Marketing Science*, *31*(2), 216-235.
Kotler, P. (1973). Atmospherics as a marketing tool. *Journal of Retailing*, *49*(4), 48-64.
Krishna, M. (2019). Retail Execution Basics: What Every CPG Manager Should Know. Retrieved from https://traxretail.com/blog/retail-execution-basics-every-cpg-manager-know/.
Kwong, J. Y., Soman, D., & Ho, C. K. (2011). The role of computational ease on the decision to spend loyalty program points. *Journal of Consumer Psychology*, *21*(2), 146-156.
Lal, R., & Bell, D. E. (2003). The impact of frequent shopper programs in grocery retailing. *Quantitative Marketing and Economics*, *1*(2), 179-202.
Lamberton, C. & Stephen, A. T. (2016). A thematic exploration of digital, social media, and mobile marketing: Research evolution from 2000 to 2015 and an agenda for future inquiry. *Journal of Marketing*, *80*(November), 146–172.
Leenheer, J., Van Heerde, H. J., Bijmolt, T. H., & Smidts, A. (2007). Do loyalty programs really enhance behavioral loyalty? An empirical analysis accounting for self-selecting members. *International Journal of Research in Marketing*, *24*(1), 31-47.
Lewis, M. (2004). The influence of loyalty programs and short-term promotions on customer retention. *Journal of Marketing Research*, *41*(3), 281-292.
Li, C., Luo, X., Zhang, C. & Wang, X. (2017). Sunny, rainy, and cloudy with a chance of mobile promotion effectiveness. *Marketing Science, 36*(5), 762-779.
Liu, Y. (2007). The long-term impact of loyalty programs on consumer purchase behavior and loyalty. *Journal of Marketing*, *71*(4), 19-35.
Liu, H., Lobschat, L., Verhoef, P. C., & Zhao, H. (2019). App adoption: The effect on purchasing of customers who have used a mobile website previously. *Journal of Interactive Marketing*, 47, 16–3.
Liu, Y. & Yang, R. (2009). Competing loyalty programs: Impact of market saturation, market share, and category expandability. *Journal of Marketing*, *73*(January), 93–108.
Luo, X., Andrews, M., Fang, Z., & Phang, C. W. (2014). Mobile targeting. *Management Science*, *60*(7), 1738-1756.
Mägi, A. W. (2003). Share of wallet in retailing: The effects of customer satisfaction, loyalty cards and shopper characteristics. *Journal of Retailing*, *79*, 97–106.
Mankins, M., & Steele, R. (2017). Turning great strategy into great performance. Retrieved from: https://www.bain.com/insights/turning-great-strategy-into-great-performance/.
McFadden, D. L. & Dubin, J. A. (1984). An econometric analysis of residential electric appliance holdings and consumption. *Econometrica, 52*(2), 45-362.

Mehrabian, A., & Russell, J. A. (1974). The Basic Emotional Impact of Environments. *Perceptual and Motor Skills, 38*, 283-301.

Meyer-Waarden, L. (2007). The effects of loyalty programs on customer lifetime duration and share of wallet. *Journal of Retailing*, *83*(2), 223-236.

Meyer-Waarden, L. & C. Benavent (2009). Grocery retail loyalty program effects: Self-selection or purchase behavior change? *Journal of the Academy of Marketing Science*, *37*, 345–358.

Mills, P. & Zamudio, C. (2018). Scanning for discounts: Examining the redemption of competing mobile coupons. *Journal of the Academy of Marketing Science*, *46*, 964–982.

Minnema, A., Bijmolt, T. H., & Non, M. C. (2017). The impact of instant reward programs and bonus premiums on consumer purchase behavior. *International Journal of Research in Marketing*, *34*(1), 194-211.

Mundlak, Y. (1978). On the Pooling of Time Series and Cross Section Data. *Econometrica, 46*, 69-85.

Narang, U. & Shankar, V. (2019). Mobile app adoption and online and offline purchases and product returns. *Marketing Science*, *38*(5), 756-772.

Neslin, S.A. & Van Heerde, H. (2008). Promotion dynamics. *Foundations and Trends in Marketing*, *3*(4), 177-268.

Nordfält, J. (2011). Improving the attention-capturing ability of special displays with the combination effect and the design effect. *Journal of Retailing and Consumer Services, 18*, 169-173.

Nordfält, J., Grewal, D., Roggeveen, A. L., & Hill, K. M. (2014). Insights from In-Store Marketing Experiments. In *Shopper Marketing and the Role of In-Store Marketing* (Vol. 11, pp. 127-146): Emerald Group Publishing Limited.

Nowlis, S. M., & Simonson, I. (1996). The effect of new product features on brand choice. *Journal of Marketing Research*, *33*(1), 36-46.

Osinga, E.C.; Zevenbergen, M. & van Zuijlen, M.W.G. (2019). Do mobile banner ads increase sales? Yes, in the offline channel. *International Journal of Research in Marketing*, *36*(3), 439-453.

Park, C., Park, Y. & Schweidel, D.A. (2018). The effects of mobile promotions on customer purchase dynamics. *International Journal of Research in Marketing*, *35*(3), 453-470.

Phelps, J. E., D'Souza, G., & Nowak, G. J. (2001). Antecedents and consequences of consumer privacy concerns: An empirical investigation. *Journal of Interactive Marketing*, *15*(4), 2-17.

POPAI. (2015). In-Store Insights Dedicated to Enhancing the Total Shopper Experience.

Raman, A., DeHoratius, N., & Ton, Z. (2001). Execution: The missing link in retail operations. *California Management Review, 43*, 136-152.

Reinartz, W., & Linzbach, P. (2018). Customer loyalty and reward programs in retail in the digital age. In: *Handbook of Research on Retailing*, Gielens, K. & Gijsbrechts, E. (Eds), Edward Elgar Publishing, Cheltenham, UK, p. 296.

Roggeveen, A. L., & Grewal, D. (2018). In-store marketing: existing and emerging elements. In K. G. Gielens, Els (Ed.), *Handbook of Research on Retailing*. Cheltenham, UK: Elgar.

Rossi, F. (2018). Lower price or higher reward? Measuring the effect of consumers' preferences on reward programs. *Management Science*, *64*(9), 4451-4470.

Sato, Y. & Y. Zenou (2014). How Urbanization Affects Employment and Social Interactions. Institute for the Study of Labor, Discussion Paper IZA DP N0. 7914, January.

Shankar, V., Kleijnen, M., Ramanathan, S., Rizley, R., Holland, S., & Morrissey, S. (2016). Mobile shopper marketing: Key issues, current insights, and future research avenues. *Journal of Interactive Marketing*, *34*, 37-48.
Sharp, B. & Sharp, A. (1997). Loyalty programs and their impact on repeat-purchase loyalty patterns. *International Journal of Research in Marketing*, *14*, 473-486.
Siebert, W. S., & Zubanov, N. (2009). Searching for the optimal level of employee turnover: A study of a large UK retail organization. *Academy of management journal*, *52*(2), 294-313.
Statista (2021). Annual number of global mobile app downloads 2016-2021. Retrieved from: https://www.statista.com/statistics/271644/worldwide-free-and-paid-mobile-app-store-downloads/.
Skorupa, J. (2018). Store Execution Versus Chaos Battle Plan. In: RISNEWS.COM.
Skrondal, A., & Rabe-Hesketh, S. (2004). Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models. Chapman and Hall/CRC.
Spence, C., Puccinelli, N. M., Grewal, D., & Roggeveen, A. L. (2014). Store Atmospherics: A Multisensory Perspective. *Psychology & Marketing, 31*, 472-488.
Srinivasan, S., Pauwels, K., Hanssens, D. M., & Dekimpe, M. G. (2004). Do Promotions Benefit Manufacturers, Retailers, or Both? *Management Science*, *50*(5), 617–629.
Stauss, B., Schmidt, M., & Schoeler, A. (2005). Customer frustration in loyalty programs. *International Journal of Service Industry Management*, *16*(3), 229-252.
Steenkamp, J. B. E., & Geyskens, I. (2014). Manufacturer and retailer strategies to impact store brand share: Global integration, local adaptation, and worldwide learning. *Marketing Science*, *33*(1), 6-26.
Stilley, K. M., Inman, J. J., & Wakefield, K. L. (2010). Spending on the fly: Mental budgets, promotions, and spending behavior. *Journal of Marketing, 74*, 34-47.
Stourm, V., Bradlow, E. T., & Fader, P. S. (2015). Stockpiling points in linear loyalty programs. *Journal of Marketing Research*, *52*(2), 253-267.
Sull, D., Homkes, R., & Sull, C. (2015). Why Strategy Execution Unravels-and What to Do About It. *Harvard Business Review, 93*, 58-66.
Taylor, G. A., & Neslin, S. A. (2005). The current and future sales impact of a retail frequency reward program. *Journal of Retailing*, *81*(4), 293-305.
Tsang, M. M., Ho, S. C., & Liang, T. P. (2004). Consumer attitudes toward mobile advertising: An empirical study. *International journal of electronic commerce*, *8*(3), 65-78.
Unni, R., & Harmon, R. (2007). Perceived effectiveness of push vs. pull mobile location based advertising. *Journal of Interactive Advertising*, *7*(2), 28-40.
Van Heerde, H.J., Dinner, I.M. and Neslin, S.C. (2019), Engaging the unengaged customer: The value of a retailer mobile app. *International Journal of Research in Marketing*, *36*(3), 420-438.
Van Heerde, H. J., Moorman, C., Moreau, C. P., & Palmatier, R. W. (2021). Reality check: Infusing ecological value into academic marketing research. *Journal of Marketing*, *85*(2), 1-13.
van Herpen, E., Pieters, R., & Zeelenberg, M. (2005). How Product Scarcity Impacts on Choice: Snob and Bandwagon Effects. In G. Menon & A. R. Rao (Eds.), *Advances in Consumer Research* (Vol. 32, pp. 623-624). Duluth, MN: Association for Consumer Research.
Van Nierop, J. E., Leeflang, P. S., Teerling, M. L., & Huizingh, K. E. (2011). The impact of the introduction and use of an informational website on offline customer buying behavior. *International Journal of Research in Marketing*, *28*(2), 155-165.

Venkatesan, R. & Kumar, V. (2004). A customer lifetime value framework for customer selection and resource allocation strategy. *Journal of Marketing*, *68*(4), 106-125.

Wang, Y., Lewis, M., Cryder, C. & Sprigg, J. (2016). Enduring effects of goal achievement and failure within customer loyalty programs: A large-scale field experiment. *Marketing Science*, *35*(4), 565-75.

Wang, R.J.H., Krishnamurthi, L. & Malthouse, E.C. (2018). When reward convenience meets a mobile app: Increasing customer participation in a coalition loyalty program. *Journal of the Association for Consumer Research*, *3*(3), 314-329.

Yoobic. (2020). How Poor Task Execution Impacts Restaurant Guest Experience. https://blog.yoobic.com/how-poor-task-execution-impacts-restaurant-guest-experience.

Zhang, J. & Breugelmans, E. (2012). The impact of an item-based loyalty program on consumer purchase behaviour. *Journal of Marketing Research*, *49*(1), 50-65.

Zubcsek, P. P., Katona, Z. & Sarvary, M. (2017). Predicting mobile advertising response using consumer colocation networks. *Journal of Marketing*, *81*(July), 109–126.

CENTER DISSERTATION SERIES

CentER for Economic Research, Tilburg University, the Netherlands

No.	Author	Title	ISBN	Published
638	Pranav Desai	Essays in Corporate Finance and Innovation	978 90 5668 639 0	January 2021
639	Kristy Jansen	Essays on Institutional Investors, Asset Allocation Decisions, and Asset Prices	978 90 5668 640 6	January 2021
640	Riley Badenbroek	Interior Point Methods and Simulated Annealing for Nonsymmetric Conic Optimization	978 90 5668 641 3	February 2021
641	Stephanie Koornneef	It's about time: Essays on temporal anchoring devices	978 90 5668 642 0	February 2021
642	Vilma Chila	Knowledge Dynamics in Employee Entrepreneurship: Implications for parents and offspring	978 90 5668 643 7	March 2021
643	Minke Remmerswaal	Essays on Financial Incentives in the Dutch Healthcare System	978 90 5668 644 4	July 2021
644	Tse-Min Wang	Voluntary Contributions to Public Goods: A multi-disciplinary examination of prosocial behavior and its antecedents	978 90 5668 645 1	March 2021
645	Manwei Liu	Interdependent individuals: how aggregation, observation, and persuasion affect economic behavior and judgment	978 90 5668 646 8	March 2021
646	Nick Bombaij	Effectiveness of Loyalty Programs	978 90 5668 647 5	April 2021
647	Xiaoyu Wang	Essays in Microeconomics Theory	978 90 5668 648 2	April 2021
648	Thijs Brouwer	Essays on Behavioral Responses to Dishonest and Anti-Social Decision-Making	978 90 5668 649 9	May 2021
649	Yadi Yang	Experiments on hold-up problem and delegation	978 90 5668 650 5	May 2021
650	Tao Han	Imperfect information in firm growth strategy: Three essays on M&A and FDI activities	978 90 5668 651 2	June 2021

No.	Author	Title	ISBN	Published
651	Johan Bonekamp	Studies on labour supply, spending and saving before and after retirement	978 90 5668 652 9	June 2021
652	Hugo van Buggenum	Banks and Financial Markets in Microfounded Models of Money	978 90 5668 653 6	August 2021
653	Arthur Beddock	Asset Pricing with Heterogeneous Agents and Non-normal Return Distributions	978 90 5668 654 3	September 2021
654	Mirron Adriana Boomsma	On the transition to a sustainable economy: Field experimental evidence on behavioral interventions	978 90 5668 655 0	September 2021
655	Roweno Heijmans	On Environmental Externalities and Global Games	978 90 5668 656 7	August 2021
656	Lenka Fiala	Essays in the economics of education	978 90 5668 657 4	September 2021
657	Yuexin Li	Pricing Art: Returns, Trust, and Crises	978 90 5668 658 1	September 2021
658	Ernst Roos	Robust Approaches for Optimization Problems with Convex Uncertainty	978 90 5668 659 8	September 2021
659	Joren Koëter	Essays on asset pricing, investor preferences and derivative markets	978 90 5668 660 4	September 2021
660	Ricardo Barahona	Investor Behavior and Financial Markets	978 90 5668 661 1	October 2021
660	Stefan ten Eikelder	Biologically-based radiation therapy planning and adjustable robust optimization	978 90 5668 662 8	October 2021
661	Maciej Husiatyński	Three essays on Individual Behavior and New Technologies	978 90 5668 663 5	October 2021
662	Hasan Apakan	Essays on Two-Dimensional Signaling Games	978 90 5668 664 2	October 2021
663	Ana Moura	Essays in Health Economics	978 90 5668 665 9	November 2021
664	Frederik Verplancke	Essays on Corporate Finance: Insights on Aspects of the General Business Environment	978 90 5668 666 6	October 2021

No.	Author	Title	ISBN	Published
665	Zhaneta Tancheva	Essays on Macro-Finance and Market Anomalies	978 90 5668 667 3	November 2021
666	Claudio Baccianti	Essays in Economic Growth and Climate Policy	978 90 5668 668 0	November 2021
667	Hongwei Zhang	Empirical Asset Pricing and Ensemble Machine Learning	978 90 5668 669 7	November 2021
668	Bart van der Burgt	Splitsing in de Wet op de vennootschapsbelasting 1969 Een evaluatie van de Nederlandse winstbelastingregels voor splitsingen ten aanzien van lichamen	978 90 5668 670 3	December 2021
669	Martin Kapons	Essays on Capital Markets Research in Accounting	978 90 5668 671 0	December 2021
670	Xolani Nghona	From one dominant growth mode to another: Switching between strategic expansion modes	978 90 5668 672 7	December 2021
671	Yang Ding	Antecedents and Implications of Legacy Divestitures	978 90 5668 673 4	December 2021
672	Joobin Ordoobody	The Interplay of Structural and Individual Characteristics	978 90 5668 674 1	February 2022
673	Lucas Avezum	Essays on Bank Regulation and Supervision	978 90 5668 675 8	March 2022
674	Oliver Wichert	Unit-Root Tests in High-Dimensional Panels	978 90 5668 676 5	April 2022
675	Martijn de Vries	Theoretical Asset Pricing under Behavioral Decision Making	978 90 5668 677 2	June 2022
676	Hanan Ahmed	Extreme Value Statistics using Related Variables	978 90 5668 678 9	June 2022
677	Jan Paulick	Financial Market Information Infrastructures: Essays on Liquidity, Participant Behavior, and Information Extraction	978 90 5668 679 6	June 2022
678	Freek van Gils	Essays on Social Media and Democracy	978 90 5668 680 2	June 2022

No.	Author	Title	ISBN	Published
679	Suzanne Bies	Examining the Effectiveness of Activation Techniques on Consumer Behavior in Temporary Loyalty Programs	978 90 5668 681 9	July 2022

The use of temporary loyalty programs has become increasingly common in recent years. However, much like permanent programs, temporary programs also deal with limited or declining consumer saving behavior throughout the program. Considering the limited time span and small window of opportunity during which consumers can act, it is crucial to determine how to maintain program salience and increase customer engagement in the temporary program. This dissertation aims to provide new insights on how this can be done, by looking into several activation techniques commonly used in such programs. The first essay studies the impact of mobile push messaging and determines heterogeneous treatment effects thereof. In addition, it investigates how saving dynamics and message timing influence the effectiveness of push messaging. The second essay focuses on the impact of in-store execution quality, and to what extent deviations from planned support plans are common, and how they influence sales. The third essay looks at the differences in effectiveness of different message types, and determines how they vary for different consumer types and outcome variables.

Suzanne M.T.A. Bies (Eindhoven, The Netherlands, 1991) received her bachelor's degree in International Business Administration from Tilburg University in 2014. She obtained her master's degree in Marketing Research (2015) and research master's degree in Business (2016) at the same university. Upon this completion, she joined the Ph.D. program of the Marketing Department of Tilburg University, working in collaboration with BrandLoyalty. As of 2021, she is employed as an assistant professor at the University of Groningen

ISBN: 978 90 5668 681 9
DOI: 10.26116/kzfe-y651