Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Proactive Customer Education, Customer Retention, and Demand for Technology Support: Evidence from a Field Experiment

German F. Retana, Chris Forman, D. J. Wu

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 18, No. 1, Winter 2016, pp. 34–50
ISSN 1523-4614 (print) | ISSN 1526-5498 (online)

http://dx.doi.org/10.1287/msom.2015.0547

Proactive Customer Education, Customer Retention, and Demand for Technology Support: Evidence from a Field Experiment

German F. Retana
INCAE Business School, Alajuela, Costa Rica, gf@incae.edu

Chris Forman, D. J. Wu
Georgia Institute of Technology, Scheller College of Business, Atlanta, Georgia 30308
{chris.forman@scheller.gatech.edu, dj.wu@scheller.gatech.edu}

Do service provider efforts to educate customers influence customer outcomes? We analyze the outcome of a field experiment executed by a major public cloud infrastructure services provider in 2011. Out of 2,673 customers who adopted the service during the experiment, 366 received a service intervention: an engagement through which the provider offered initial guidance on how to use basic features of the service. Before execution, it was unclear if this proactive customer education would have positive or negative effects on customer retention and demand for technology support. We show the treatment reduces by half the number of customers who churn from the service during the first week. Further, treated customers ask 19.55% fewer questions during the first week of their tenure than the controls. Although the treatment's effects decay within one week, we show that such proactive customer education can have significant economic benefits for the provider. In particular, we find that treated customers increase their accumulated usage of the service by 46.57% in the eight months after sign-up. Finally, we provide evidence that the effects of the treatment are strongest among customers who have less experience with the provider.

Keywords: field experiment; proactive service; service coproduction; customer retention; technology support; cloud computing
History: Received: February 25, 2014; accepted: April 2, 2015. Published online in *Articles in Advance* September 28, 2015.

1. Introduction

Academics and practitioners have long recognized service customers' role as both recipients and producers, or coproducers, of the service delivered, particularly in the context of high contact services where customers are deeply involved in the creation of the service (Chase 1978). Examples of such services are online self-service technologies (SSTs), for which research has consistently shown that customers' knowledge, skills, and abilities in coproducing the service are a key determinant of their adoption and continued usage (e.g., Xue et al. 2011). Surveys have also shown that customers' expertise is positively associated with their loyalty in contexts that require intensive involvement from customers, such as financial services (Bell and Eisingerich 2007).

Given the positive relationship between a customer's capabilities and its adoption and use of SSTs, providers make themselves available to assist customers in their service coproduction efforts. A common channel for this is *reactive* technical support—customer-initiated interactions with the provider in which the latter assists customers in deriving greater utility from the service. Prior research has shown that reactive support through face-to-face interactions accelerates customers' coproduction learning rates when using SSTs such as online banking portals (Field et al. 2012). More commonly, reactive support is offered through contact centers where agents use phone calls, chats, or other media to interact with customers (Gans et al. 2003). Ongoing research in the context of cloud infrastructure services—a SST with a relatively high level of technical sophistication—has shown that this type of support has a positive impact on the consumption of the service (Retana et al. 2015). However, much less is known about the potential benefits for providers who offer *proactive* (i.e., provider-initiated) assistance.

In this research, we explore one type of proactive engagement, *proactive customer education* (PCE). Our goal is to understand how PCE influences customer behavior in the context of public cloud infrastructure services. Public cloud infrastructure services, or public infrastructure as a service, are a very high-contact SST in which on-demand computing and storage resources (i.e., servers) are offered on a pay-as-you-go basis (Mell and Grance 2011). We define PCE as any provider-initiated effort to increase customers' knowledge and skills immediately after signing up for the service. We distinguish PCE from reactive education, which could be offered through reactive technical support channels (e.g., a contact center), and which has been the focus of prior work (e.g., Field et al. 2012, Retana et al. 2015). We also distinguish PCE from proactive sales or cross-selling engagements (e.g., Akşin and Harker 1999, Armony and Gurvich 2010), because in our context the education is offered by technical staff and not by sales representatives. We specifically examine proactive education (Challagalla et al. 2009) that is offered after customers initially sign up for the service. Typically, PCE occurs as customers are taking their initial steps in coproducing the service (i.e., adapting the service to their idiosyncratic needs). We attempt to empirically study the following research question: *What are the effects of PCE on customers' retention and demand for technology support during the early stage of their coproduction processes*? We also empirically explore *the rate at which PCE's effects decay*. Our focus on the period immediately following sign-up to the service is motivated by practice: this is when customers' risk of churning and demand for technology support are strongest. It is also when PCE is most likely to have an effect on customer behavior.

The potential influence of PCE on customer retention is not a trivial matter, especially in contexts such as ours where switching costs are initially very low. It is unclear ex ante whether PCE will improve or decrease retention. PCE can foster retention by increasing customers' perceived service quality (Eisingerich and Bell 2008, Sharma and Patterson 1999), setting realistic expectations about service features and performance (Bhattacherjee 2001, McKinney et al. 2002), and making customers more efficient in their use of the service (Xue et al. 2007). However, educating customers may also make them more capable and willing to consider alternate options in the market, implying a negative impact on retention (Fodness et al. 1993, Nayyar 1990).

A similar set of opposing forces exists in regards to customer education and its effect on customer demand for (reactive) technology support. Education can make customers more efficient (i.e., they need less input to produce the service output), and in turn reduce the costs to the provider of serving them (Xue and Harker 2002). However, proactive education can also lead to escalated expectations, whereby customers continue expecting and seeking constant assistance from the provider (Challagalla et al. 2009). Education could also lead to increased demand for support if the information presented to customers is unstructured (Kumar and Telang 2012).

We collect unique data from a field experiment run by a major public cloud computing infrastructure services provider. The provider ran the experiment during October and November 2011 and we observe customer use of the service until August 2012. Upon sign-up, 366 customers selected out of 2,673 customers that opened an account during this period received the field experiment's treatment: PCE. The treatment consisted of a short phone call followed up by a support ticket through which the provider offered initial guidance on how to use the basic features of the service. After the proactive engagement, the treated customers could continue interacting with the provider's contact center through reactive support, which was the only channel for technology support available for the nontreated control customers since sign-up. Our empirical strategy controls for potential treatment assignment issues. We employ survival analysis and count data models to examine the differences in retention and demand for reactive technology support between the two customer groups early in their tenure as customers of the provider. Our robustness checks provide evidence that the treatment assignment is independent of any customer attributes and the observed outcomes, an aspect critical to our identification strategy.

We find that treated customers' are 3.1 percentage points more likely to survive through their first week. This represents a significant effect on customer retention, consistent with a 49.60% reduction in the hazard rate of leaving during the first week after sign-up. We argue that customers' exposure to PCE enables them to derive more value from the service and gives customers provider-specific knowledge that increases the value of the provider to the customer and may create switching costs that lower the benefits from churning. These findings have important implications for the provider. On average, 34.3% of new adopters abandon the service within the first eight months of use. However, 18.8% of those who abandon (or 6.4% of all adopters) do so during the first week, which is much more than in any other week. By improving customer retention during this stage in customer tenure, PCE has a significant positive impact on the overall size of the customer base.

We also test the effect of PCE on customers' early demand for technology support, measured by the number of questions they ask to the provider through online live chat sessions and support tickets. PCE reduces the average number of questions asked during customers' first week after sign-up by 19.55%. We argue that this occurs because in the early stages of the coproduction process the provider can preempt customers' most frequently asked questions. This is, again, an important economic benefit for the provider. Customers' demand for support is strongest when they are starting to use the service and the drop in the number of questions implies a reduction in one of the provider's major operational costs: labor-intensive reactive technology support.

Our results have significant economic implications for the provider. The treatment's effects decay quickly and influence customer behavior primarily in the first week, however this is also the time in customers' tenure when their propensity to churn and needs for technology support are highest. In the long run (e.g., eight months after sign-up), in part thanks to its positive impact on early retention, by one estimate PCE increases usage of the provider's service by 46.57%.

Last, we examine circumstances under which PCE will have the greatest implications for provider outcomes. Our results suggest that PCE is most effective for customers that have not previously used the provider's services. That is, our results are consistent with the view that PCE plays an important role in educating customers who may not have alternative capabilities and resources that can aid them in using the service effectively. These results will inform service providers on the types of customers whom will have the greatest response to such a service intervention.

In addition to contributing to the service operations literature by informing whether and when PCE will influence retention and demand for technology support, our work contributes to existing literature that explores providers' support costs in contexts with a high level of customer involvement in the service delivery process (e.g., Kumar and Telang 2011). A common mechanism discussed in the call center operations literature to reduce support costs is to combine or blend customer-initiated and provider-initiated calls. It has been suggested that outbound calls can be used to call back customers (e.g., Armony and Maglaras 2004a, b), to attend to low priority customers whose service may be delayed (e.g., Jouini et al. 2011, 2009), or to cross sell to customers (e.g., Akşin and Harker 1999, Armony and Gurvich 2010). However, to our knowledge, there has been no prior work studying outbound (proactive) education as an alternative for traditional outbound calls.

Our study has significant managerial implications for providers. Cloud services are generally viewed as being fully self-serviced, on-demand offerings with minimal interaction between customers and service providers (Mell and Grance 2011). Our research suggests that cloud offerings and other SSTs that require a certain level of technical skill from customers may actually benefit from not being exclusively "self-service." Proactive education engagements could yield benefits to providers of SSTs. As we discuss in further detail below, we believe these findings may also generalize to other contexts where customers play an important coproduction role, such as online banking and e-learning.

2. Theoretical Background

In what follows, we examine PCE's potential influence on customer behavior. Ex ante, it is unclear if PCE will have positive or negative effects on customer retention and demand for technology support, so here we limit ourselves to discussing the mechanisms driving the potential outcomes. The intervention we study was applied immediately after customer sign-up. We focus on the effects of the treatment immediately after it was applied because churn rates and support workloads are highest in the periods after customers initially adopt the service (we provide empirical evidence of this in §3.1).

2.1. PCE and Customer Retention

A recurring result in the service operations literature is the positive effect of customer education on perceived service quality and, in turn, on customer loyalty (e.g., Bell and Eisingerich 2007, Sharma and Patterson 1999, Zeithaml et al. 1996). In the particular context of information technology, PCE can increase satisfaction and loyalty by helping customers match their expectations regarding the features of the SST and their early experiences with the service (Bhattacherjee 2001, McKinney et al. 2002), which in turn motivates them to continue using a service (Bhattacherjee 2001, Staples et al. 2002). Moreover, in the context of Internet-based services, there is an added complication in managing customers' expectations given how rapidly technologies evolve and thus how fast experiences may differ from expectations (Khalifa and Liu 2003). In the context of such rapidly changing environments, it is especially valuable to engage customers early in their tenure through programs such as PCE.

PCE can also incentivize customers to use a service by making them more efficient (Xue et al. 2007) through a reduction in their early service coproduction costs. Rather than requiring customers to invest in experimenting and learning how to use the basic functionalities of the service on their own, via PCE a provider can take that burden off customers, or at least make their initial ramp-up process less cumbersome. The skills acquired, even if basic, will increase the value of the service provider to the customer relative to alternatives. In other words, valuable nontransferable investments made in learning how to use the provider's services can increase the costs to switching providers (e.g., Johnson et al. 2003, Klemperer 1995), increasing expected customer retention.

However, education can also lead to attrition rather than retention (Bell and Eisingerich 2007), particularly if it is offered soon after sign-up and when there are near zero switching costs. When customers learn from the provider, the information asymmetry between them gets reduced and the former may be motivated to evaluate other alternatives in the market (Fodness et al. 1993, Nayyar 1990). For example, PCE can make customers aware of limitations of the service they did not know before the treatment. The risks of attrition will be greater if PCE is not effective in driving satisfaction. Customers who are not necessarily satisfied with an SST continue using it because of switching costs (Buell et al. 2010, Jones and Sasser 1995). However, in our setting, there are no contracts that lock customers in for a certain period of time (e.g., a subscription). Further, new customers have not yet incurred any large provider-specific coproduction efforts (e.g., invested in deploying an application in the cloud service) that increase the value of the provider relative to alternatives. In this environment, the risks of PCE driving attrition may be relatively high.

2.2. PCE and Demand for Technology Support

Educating and improving the efficiency with which customers use the service can lead to a reduction in costs for the provider since it will employ less labor and other resources when delivering the service (Xue and Harker 2002). In the particular context of technology support contact centers, the provider's initial investment in PCE could potentially lead to a reduction in later reactive support costs by reducing the number of questions asked by customers through the reactive support channel (e.g., customers submit fewer tickets). For example, by guiding customers on how to navigate through the service control panel, the provider can preempt future questions regarding its functionality. In the particular case of cloud infrastructure services, even seasoned system administrators will be unfamiliar with the provider's Web-based control panel until they see it for the first time.

Nevertheless, education, and in particular PCE, could also have the opposite effect. PCE can lead customers to realize early on that the provider is a reliable, fast, and easy-to-access knowledge source, especially if, as in our context, there are no additional fees associated with contacting the provider. Thus, customers who have received PCE may become more aware of the provider's support capabilities and realize that it is much more convenient for them to contact the provider for assistance, instead of searching knowledge bases or experimenting to solve their issues on their own. As a result, customers may become overly dependent on the provider (Challagalla et al. 2009) and increase their demand for technology support. Kumar and Telang (2012) found a similar phenomenon in the context of insurance services. They showed that presenting customers with more information, especially if it was unstructured, increased the number of calls they made to the insurance provider. Relatedly, Campbell and Frei (2010) found in the context of consumer banking that customers who better understand their service also use offline assisted-service channels (e.g., call centers) more. In summary, customers who receive PCE might demand more reactive support.

3. Research Setting, Field Experiment, and Data

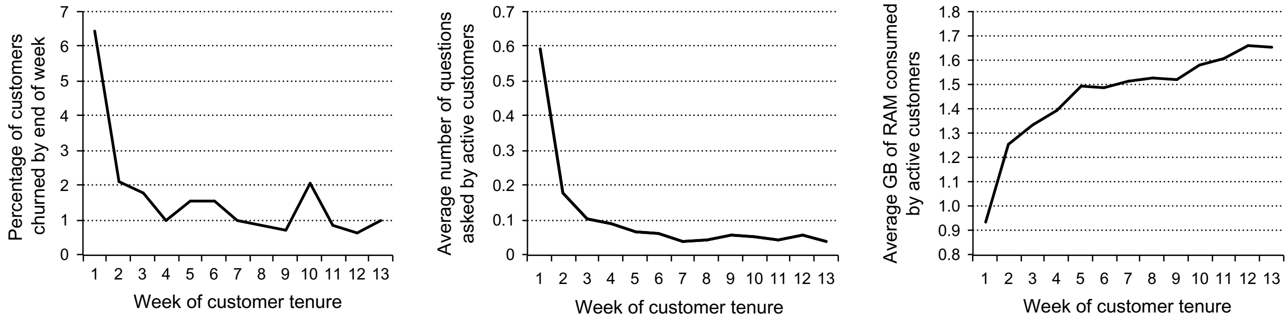
Our research examines the effects of PCE on customer churn and demand for technology support by analyzing the outcome of a field experiment executed by a major cloud infrastructure services provider during October and November, 2011. In this section we present our sample and data as we describe our research context and the field experiment.

3.1. Cloud Infrastructure Services and Customer Behavior

An essential characteristic of cloud infrastructure services is that they are offered on demand and are self-serviced (Mell and Grance 2011). Customers can unilaterally provision as much computing resources (i.e., processing cores, memory, and storage space) as they want, when they want. Moreover, in infrastructure-as-a-service offerings, "the consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications; and possibly limited control of select networking components" (Mell and Grance 2011, p. 3). Thus, customers confront a significant technical burden and not all of them will find it easy to start using these services.

Figure 1 presents data from customers' first trimester of tenure in our empirical setting. The left panel shows most customers who churn from the service do so during the first couple of weeks after sign-up. Conditional on survival, each data point represents the percentage who churn (exit) from the service in that week, or the hazard rate. Moreover, this is also the period when customers ask the most questions to the provider through reactive support channels (see center panel). In other words, the risk of churn and the demand for technology support are both particularly strong during this period. This evidence is suggestive that customer skills and experience influence their early behavior. Finally, once customers overcome their ramp up phase, they consume much more of the service (see right panel); as explained in §6.2 we measure usage in terms of gigabytes (GB) of random access

Figure 1 Hazard Rate of Churn (Left), Demand for Technology Support (Center), and Service Usage (Right) over Active Customers' First 13 Weeks (91 Days)



memory (RAM) per hour consumed. These statistics motivate our empirical approach to study the effects of PCE on customer churn and demand for technology support during the early weeks of customer tenure.

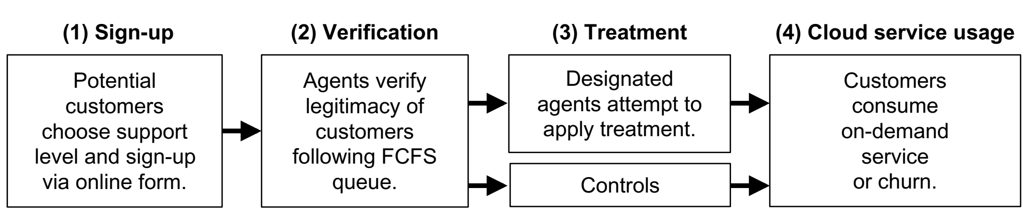
3.2. Field Experiment and Sample Construction

In our setting, new customers go through a process that includes sign-up and customer verification. During this process some customers receive the PCE treatment. The steps are illustrated in Figure 2, and we discuss each in turn as we also describe how we constructed our sample.

3.2.1. Sign-up. New customers sign up for the service and open an account through an online form without payment; they only pay when they start using computing resources. From October 11 to November 28, 2011—the time span of the field experiment—4,739 new accounts were opened. We exclude 71 accounts that were opened by the provider's staff (to test the functionality of their system or for support agent training purposes) and 134 accounts that were opened by customers who had already opened an earlier account during the field experiment (we consider a customer's first account as its sign-up). Additionally, 702 were opened by customers with prior accounts with the provider (customer had signed up prior to the start of the field experiment on November 28, 2011). Since experienced customers opening a new account presumably already achieved a basic level of proficiency with the service, PCE should have a systematically different (i.e., smaller) effect on them relative to new customers who are the target of the treatment. We test this assumption later in §6.3, yet we exclude them from our baseline sample to avoid an attenuation bias and are left with 3,832 new customers.

Customers can sign up under two different levels of service support: *full* and *basic*. Of the 3,832 new customers, 3,416 chose basic support, 344 chose full support, and 72 upgraded from basic to full after sign-up. Even though the cloud infrastructure service offering is identical under each support regime (customers use the same servers and server management tools), full support involves frequent interactions of customers with their account managers about their particular server configurations and needs, whereas basic support is limited to addressing the general quality of service issues. Thus, it is unlikely that PCE will have a substantial effect on full support customers, as the treatment would be considered just the first of many rich and frequent interactions. Empirical tests, discussed in Online Appendix A (available as supplemental material at http://dx.doi.org/10.1287/msom.2015.0547), confirmed PCE does not alter full support customers' behavior. Since our goal is to estimate PCE's impact on customers that are not receiving other service interventions, we focus on customers that exclusively used basic support. Although the exclusion of the full support customers reduces certain types of noise in our sample that may aid statistical inference, our results for basic support customers remain consistent if full support customers are retained in the sample. Also, the treatment has no measurable influence on the likelihood of a customer upgrading from basic to full support.

Figure 2 Customer Timeline of Events



3.2.2. Verification. Just a few minutes after sign-up, customers receive a call by an agent of a verification team who attempts to ensure that the new account was opened by a legitimate customer (e.g., a customer that will not use the service to spam). Agents of the provider's verification team call prospective customers following a simple first-come first-serve (FCFS) queue. If they pass the verification process, customers can start using the service. The verification process does not entail any starting guidance from the provider aside from online documentation and manuals. This is the case for the control customers in our field experiment.

3.2.3. Treatment. For the field experiment, a few designated agents of the verification team performed additional tasks beyond verifying the legitimacy of the new customers: they applied the PCE treatment. The FCFS queue determined if a new customer was called by any of the designated or nondesignated agents. Therefore, even though the provider chose which agents would be applying the treatment, it had no control regarding which agents would call which customers. The designated agents who applied the treatment prolonged the verification call and followed it up with a support ticket. Out of the 3,175 basic support customers who passed the verification process, 476 (15.0%) of them were verified by the designated agents and consequently were offered the PCE treatment.

The treatment had three components: confirming product fit, setting expectations, and educating customers. The first two prevent potential dissonances between prior expectations and experiences that might drive customer churn. To educate customers, during the call and through the support ticket, the agent sought to teach the customer how to access and use the online control panel, how to set up and access its first server, and how to make a backup of that server, among other topics. Although these constitute only basic functionalities of the service, PCE prevented customers from having to investigate and learn them on their own, thus lowering their coproduction costs and increasing their efficiency, as well as providing them with skills that they would relinquish, at least partially, if they opted to switch to another provider.

3.2.4. Service Usage. Customers that pass the verification stage can start consuming cloud resources and requesting reactive technology support from the provider. We observe each customer's use of the on-demand infrastructure services, and the timing and content of all support interactions through online live chat sessions and support tickets between each customer and the provider, up to August 15, 2012. Therefore, our data have between eight and nine months of history per customer depending on day of sign-up; this is relevant to our identification of churn as will be discussed shortly. Customers can also request support via phone calls. Unfortunately, we only observe phone call data at the aggregate level, so we cannot study how PCE influences customer-level phone support. However, as we discuss in §3.3.3, PCE does not alter the aggregate volume of phone support relative to the other support channels.

Worth noting, not all accounts are opened with the intention of using the service. Some of the accounts have very short tenure (e.g., less than one day) or never launch a server. Through interviews with the provider, we learned it is often the case that a customer opens an account simply to check if the provider's platform supports some particular feature. The customer opens the account, checks for the availability of the feature, and very often never launches a server (i.e., uses computing resources). Since they are systematically different in ways that would bias our estimates, we exclude customers with less than one day of tenure or who never launch a server from our sample. However, our results are robust to the inclusion of the customers with less than a day of tenure. Our final sample has 2,673 customers, with 366 (13.7%) who were offered PCE.

3.3. Measurement of Variables and Descriptive Statistics

Table 1 shows the descriptive statistics of the variables used in our analysis. We next discuss how we construct these variables and related issues for identifying the treatment effect of interest.

3.3.1. Identification of the Treatment Effect. Our main covariate of interest is the treatment indicator, captured in PCE_i. Given its critical role for the validity

Table 1 Descriptive Statistics

Customer group:	All customers				Controls				Treated			
Number of customers:	2,673				2,307				366			
Variable	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max	Mean	S.D.	Min	Max
PCE_i	0.137	0.344	0	1	0	0	0	0	1	0	1	1
$SurvivedWeek1_i$	0.936	0.245	0	1	0.931	0.253	0	1	0.964	0.185	0	1
$SurvivedMonth1_i$	0.888	0.316	0	1	0.881	0.324	0	1	0.932	0.253	0	1
$QuestionsWeek1_i$	0.591	1.445	0	18	0.609	1.497	0	18	0.481	1.054	0	8
$QuestionsWeek2_i$	0.759	1.851	0	25	0.777	1.896	0	25	0.647	1.536	0	15

of our analyses, we discuss its adequacy in identifying the treatment effect of interest.

The assignment of treatment is determined by the FCFS verification queue, is not based on information provided by the customers, and so is independent of customer attributes. In Online Appendix B we show treated and controls are very similar in their sizes, industry, and type of application they intend to deploy, based on the information available to the provider upon sign-up. There are, however, small variations in the proportion of agents applying the treatment at different times during the field experiment. The number of designated agents applying the treatment relative to the total size of the verification team varies across work shifts, days of the week, and weeks of the year. This, in turn, slightly alters the likelihood of receiving PCE depending on the time of sign-up. We control for these variations in our analyses.

In our data we observe which customers are designated to receive the PCE treatment, however we do not observe whether the treatment has actually been received by the customer. In particular, the agent may be unable to reach the customer to apply the phone call aspect of the treatment (all treated customers receive the support ticket with information that was not being transmitted to the controls). Thus, in the sense of the Rubin Causal Model (e.g., Angrist et al. 1996, Rubin 1974) we observe the *intention to treat* customers with a phone call rather than the phone call treatment itself. The intention to treat may differ from the treatment itself based upon customer characteristics (e.g., some customers may be more amenable to receiving the phone call than others) or potentially agent characteristics (e.g., some agents may be more persuasive in their attempts to reach customers). Under some common assumptions, the average causal effect of the intention to treat will be proportional to the average causal effect of the treatment (Angrist et al. 1996). Intention to treat is sometimes worthy of independent study (e.g., Angrist 1990, Hearst et al. 1986), and this is the case in our setting: While the cloud provider controls the intention to treat, it is unable to directly control whether or not the treatment is received. Thus, the efficacy of the intention to treat is in fact the quantity of interest for the cloud provider.

One potential concern in our experimental design may arise from heterogeneity in application of the treatment. If there exist unobserved differences in how the treatment is applied and these are correlated with outcomes, then our approach may not yield an unbiased estimate of the average treatment effect. This problem could potentially arise because of differences in the treating agent's expertise, which varied from new hires to experts, and how it may influence the treatment application. However, when offering PCE, all agents were following a preestablished script with well-codified information, and the follow-up support ticket they sent was based on a template with just minor variations from customer to customer. This procedure significantly increased the homogeneity of the treatment application.

3.3.2. Customer Retention Variables. We next describe the variables associated with customer retention, which depend on the accurate identification of a churn event. This can be potentially difficult in our on-demand cloud services context as customers may cease use of the service without necessarily closing their account. One potential approach to estimate the timing of the churn event is to use the frequency of purchase transactions (Fader et al. 2005). However, cloud customers usually make continuous consumption of the services, which makes this approach unfeasible. Instead, we identify the moment at which a customer stops using the service (the churn) using the later of two events: the customer's last observed usage of a cloud server or last observed support interaction with the provider. Our results are also robust to using only the last server usage to identify churn. We acknowledge that some of the customers marked as churned may return after our observed period. However, we have no reason to believe that the incidence of this behavior will be systematically associated with the treatment, particularly since our observation period ends between eight and nine months after sign-up. Let the variables $SurvivedWeek1_i$ and $SurvivedMonth1_i$ indicate if customer i uses the service for at least one week (i.e., 7 days) or one month (i.e., 30 days), respectively. During our sample period, 37.4% (999) of the customers in the sample churned. The mean of our survival indicators suggest 93.6% of customers survive past the first week and 88.8% do so past the first month. The Kaplan and Meier (1958) survivor function estimate indicates 5% of customers churn by day 20 of their tenure, yet after this the churn rate is only an average of 3.2%'per month.

3.3.3. Demand for Technology Support. We operationalize customers' demand for technology support through the variables $QuestionsWeek1_i$ and $QuestionsWeek2_i$, which represent the count of the number of questions asked by customers during their first week and first two weeks since signing up for the service. To consider a support interaction as a question, it must satisfy two requirements.

First, the customer must have initiated it. Although all chats can only be initiated by a customer, some support tickets are announcements or alerts sent out by the provider through the ticketing system. To identify such announcements, we scanned the tickets' content and excluded from our count those that were either identical (i.e., exact same subject and content) or that followed a certain template (e.g., automated messages where only the customers' names are changed).

Second, the support interaction must not represent a response to an exogenous and unexpected failure in the service offering (e.g., a physical component of the provider's hardware fails). Specifically, we seek to identify exogenous failures that were not due to some action, mistake, or lack of knowledge by the customer. Such events are generally due to some failures at the provider's end that were completely unexpected by the customer. We used text-matching techniques to identify these interactions and then exclude them from our count. Online Appendix C includes further details concerning the procedures employed to identify questions from the support interactions. Our results are qualitatively similar if we use the count of all customer-initiated support interactions as the dependent variable.

Before discussing the descriptive statistics of the number of questions, we comment on two issues concerning their adequacy in capturing customers' demand for technology support. First, the provider interacts with its customers through three support channels of which we only observe two: we observe all online live chat sessions and support tickets in customers' tenure, but we do not observe their phone calls. More precisely, we observe the aggregate volume of phone calls but are unable to link each individual call to a specific customer. However, analysis of the aggregate data indicates that roughly 60% of support interactions occur through chats, 20% through tickets, and 20% through phone calls. Thus, although we do not observe phone calls, we are only missing a minority of all support interactions. One potential concern is that customers used the phone call support channel more intensively after the introduction of PCE; if customers substitute (unobserved) phone support for other (observed) support channels, then our analysis of the effects of PCE on support will be misleading. As noted above, we performed an analysis on the aggregate number of support requests per channel before and after the PCE field experiment. This analysis showed that the support channel mix (i.e., relative proportions of support requests through the three channels) did not vary significantly in the months soon before and after the execution of the field experiment. Also, there are no statistically significant differences between treated and controls in their channel use preferences following the field experiment regarding the observed chats and tickets channels. These facts mitigate the substitution concern. Finally, we also know that some phone calls are followed up by a support ticket, such as when the support agent wants to transmit some information to the customer (e.g., some step-by-step guide on how to configure some component of the infrastructure), which in turn means we do capture the phone-initiated interaction through the resulting support ticket.

Second, the count of support interactions does not offer insight into the complexity or topic of the questions asked, attributes that may affect the provider's cost of offering the reactive support. Although we have made an effort to cleanly identify support interactions that constitute questions, our counts consider all questions to be equally costly to answer.

As noted in Figure 1, the distribution of the questions asked is frontloaded during a customers' tenure, and most customers do not ask any questions at all. The mean number of questions during the first and second weeks of customers' tenure are 0.591 and 0.179, respectively, whereas the metric drops below 0.103 for all other weeks. Additionally, during the first week 72.2% of customers do not ask any questions at all, 91.4% do not ask any during the second week, and thereafter at least 94.1% per week refrain from asking questions.

4. Empirical Models

Our empirical strategy employs survival analysis to examine the effects of PCE on customer retention and employs count data models to study its effect on customers' demand for technology support. The models are also used to measure the rate of decay of the treatment's effects. We first discuss the effects of PCE on customer retention and then study the implications for technology support.

4.1. Customer Retention

To test the effects of PCE on customer retention we employ nonparametric and semiparametric survival analysis methods. However, we start with simpler linear probability and probit models that will facilitate the economic interpretation of our findings.

We first examine the effect of the treatment on the likelihood of a customer surviving up to a certain age. We use $SurvivedWeek1_i$ and $SurvivedMonth1_i$ as our dependent variables in linear probability and probit models as follows (we use $SurvivedWeek1_i$ below, yet the model is the same with $SurvivedMonth1_i$):

$$SurvivedWeek1_i = \alpha + \beta PCE_i + \delta SignupControls_i + \varepsilon_i \quad \text{and} \tag{1a}$$

$$\Pr(SurvivedWeek1_i = 1) = \Phi(\alpha + \beta PCE_i + \delta SignupControls_i). \tag{1b}$$

Since PCE_i indicates if customer i received the treatment, the coefficient β identifies PCE's effect and will be positive if PCE improves retention. If the estimates of β in the models with $SurvivedMonth1_i$ are not significantly larger than those with $SurvivedWeek1_i$, we can infer that the treatment's effect on customer survival decayed since most of the effect occurred during the first week of tenure. We further explore the decay of the treatment's effect on retention later in §6.1.

Our identification strategy relies on the assumption that the treatment assignment is independent of any unobserved customer or agent attributes that may influence outcomes. As noted above, the incidence of customer treatment is independent of customer attributes, however the likelihood of treatment varies over the course of a day, over the days of the week, and over the weeks of the field experiment. Customers may differ in unobservable ways depending on their time of sign-up; similarly, the number of agents applying the treatment and the fraction of treated customers vary over time (Online Appendix D offers further details on this latter phenomena). Controlling for these differences is critical to ensure the validity of our identification strategy. We correspondingly implement the vector of controls, $SignupControls_i$, to account for potential issues associated with the time of sign-up.

Our first controls are in the vector $SignupHour_i$, which consists of 23 dummies, one for each hour of the day at the provider's time zone (we leave the 24th hour as the base level). The variable $SignupHour_i$ controls for both differences in the profile of customers that may sign up at different times of the day (e.g., those signing up during office hours may be working at a firm, whereas those signing up after hours may be individuals working on personal projects) as well as differences in the likelihood of receiving the treatment across hours. Although our data are from a global provider, the concept of "office hours" remains valid as both the provider and the vast majority of its customers are located in the United States.

The binary indicator $SignupWeekday_i$ is equal to one if sign-up occurred from Monday through Friday and is zero otherwise. It controls for potential differences in customers who sign up during a weekday or on the weekend.

Finally, we have a vector of weekly dummies, $SignupWeek_i$. We add this to control for time shocks such as how close the time of sign-up (which occurs between October and November 2011) was to the 2011 holiday season, when firms in several sectors (e.g., retail) may be drawn to cloud services for their ability to handle uncertain peaks in demand. The vector also controls for a change on the provider's end whereby starting on November 13 (week 47 of the year) a greater proportion of agents in the verification team were applying the treatment than before.

Next, we employ nonparametric survival analysis to determine the overall effect of the treatment on customer retention. For this, we study the rate at which customers churn at time t through the hazard function $h(t)$. We use the log-rank (Mantel and Haenszel 1959) and Wilcoxon (Breslow 1970) tests for the equality of hazard functions between the treated and control customer groups. The latter test places more weight on earlier failure times (Cleves et al. 2010), which is important for us since in our context the hazard rate of failure is highest during the early stages of customers' tenure.

Nevertheless, neither of these nonparametric approaches tests for the equality of the survivor functions at some point in time; they test for the equality across the entire timespan of the data. To cleanly distinguish the time-varying effects of the treatment (e.g., its decay) we must make some parametric assumptions. In the Cox (1972) proportional hazard model, the hazard for the ith customer at time t is $h(t \mid X_i) = h_0(t)\exp(X_i\beta_X)$. In this model, we assume that all individuals are subject to the same underlying baseline hazard, $h_0(t)$, yet we make no assumptions regarding its functional form. Instead, we simply assume the treatment and other covariates in the vector X_i influence the baseline hazard in a multiplicative (proportional) way:

$$h(t \mid X_i) = h_0(t)\exp(\beta PCE_i + \delta\, SignupControls_i). \quad (2)$$

With this model, $e^{\hat{\beta}} - 1$ will be the estimated percentage change in the hazard (churn) rate caused by the treatment. A finding that $\hat{\beta} < 0$ would imply a decrease in the hazard rate and hence an increase in customer retention.

4.2. Demand for Technology Support

To estimate the effect of the treatment on the demand for technology support, we employ count data models that have the number of questions asked by customers as the dependent variable. We use the variables $QuestionsWeek1_i$ and $QuestionsWeek2_i$ described in §3.3.3. The number of questions is a good proxy for customers' demand for reactive technology support and represents a very important cost driver for the provider.

Count data models, such as the Poisson and negative binomial models, are appropriate for our setting since the number of questions asked is a nonnegative integer value. However, since most customers do not ask any questions at all, our distribution has a large number of zeroes and hence suffers from overdispersion. To account for this, we relax the equivariance assumption of the Poisson model and employ the quasi-maximum likelihood approach that uses a robust variance-covariance matrix for the Poisson maximum likelihood estimator (Wooldridge 2010). We also use the negative binomial model (with quadratic variance), which despite making more assumptions on the functional form of the distribution than the Poisson model, may fit our data better as it explicitly models overdispersion as well as a longer right tail in the probability distribution (Cameron and Trivedi 2010).

Since both the Poisson and the negative binomial models have the same conditional means, we present the same model for both estimation methods (we show the model with $QuestionWeek1_i$, yet the model is the same with $QuestionsWeek2_i$):

$$E(QuestionsWeek1_i \mid X_i) = \exp(\alpha + \beta PCE_i + \delta\, SignupControls_i + \varepsilon_i). \quad (3)$$

Parameter β identifies PCE's effect on the demand for technology support, and if the treatment reduces the demand we will find that $\hat{\beta} < 0$. Our specification models the cumulative number of questions over a specific time period, a quantity that will be lower in expectation when a customer churns from the service before the end of the period. Moreover, since PCE (negatively) influences attrition, parameter β in model (3) captures the treatment's effect on both the number of questions and on attrition. This will make it more difficult for us to obtain a finding of $\hat{\beta} < 0$; i.e., to find evidence that PCE reduces the demand for technology support. Finally, similar to our prior approach, we could use the relative magnitudes of the estimates of β in the models with $QuestionsWeek1_i$ and $QuestionsWeek2_i$ (i.e., the latter being closer to zero than the former) to make inferences about the decay of the treatment effect. However, we again acknowledge that, since PCE will also influence attrition, our cross-sectional model cannot cleanly distinguish the magnitude of this time-varying effect.

5. Results

5.1. Customer Retention

The results attained using our various models all show that the PCE treatment has a positive effect on customer retention. We suggest that PCE has this effect because (i) customers will derive more value using a service they understand better because of the treatment and, additionally, (ii) the treatment generates a small yet important switching cost that motivates customers to continue using the service.

We present our results with the linear probability model (1a) and the probit model (1b) in columns (1) through (4) of Table 2. Because it is difficult to interpret the magnitude of the coefficient estimates for the probit model, we report the marginal effect of the treatment for each model type in the lower section of the table. Columns (1) and (2) suggest the treatment increases the likelihood of a customer surviving at least its first week of using the service between 3.1 and 3.2 percentage points. These results show that the treatment is effective in increasing customer retention during the early days after sign-up. To put this estimate in perspective, recall that mean retention for the sample after the first week (i.e., mean $SurvivedWeek1_i$) is 93.6% (see Table 1), so PCE brings survival rate much closer to 100% during the period in which it is most likely for customers to churn.

If we extend our analysis to survival through at least the first month, we get very similar results. Columns (3) and (4) indicate treated customers are between 5.2 and 5.3 percentage points more likely to survive past their first month. Although the effect is greater in magnitude than that for the first week, the relative magnitudes of the estimates may suggest the treatment effect has decayed since most of the effect (i.e., churn prevention) already occurred during the first week. We find similar results if we use dummy indicators for survival over longer periods of time (e.g., six or eight months) as our dependent variables. We investigate these claims in greater detail below using the estimates from the hazard models.

Our result for the overall effect of PCE on the hazard rate employing the Cox proportional hazard model (2) is presented in column (5) of Table 2. The model implicitly assumes PCE has a constant marginal effect throughout customers' tenure, an assumption that we will relax in §6.1. Our result suggests PCE reduces the hazard rate by 22.46%, again demonstrating that customers treated with PCE had lower churn.

We followed the recommendations by Cleves et al. (2010) and performed various tests for our proportional-hazards assumption. We performed a link test, interacted our treatment variable with time and confirmed insignificance of the interaction, and confirmed

Table 2 Survival Results

Column:	(1)	(2)	(3)	(4)	(5)
Dependent variable:	$SurvivedWeek1_i$		$SurvivedMonth1_i$		Firm survival
Model:	LPM	Probit	LPM	Probit	Cox proportional hazard
PCE_i	0.032***	0.308**	0.053***	0.323***	−0.254**
	(0.011)	(0.133)	(0.015)	(0.109)	(0.102)
Customers that churned					999
Marginal effect of PCE_i	0.032	0.031	0.053	0.052	
Percentage change in hazard ($e^{\hat{\beta}} - 1$)					−22.46%

Notes. All regressions use the 2,673 customers in the sample and include hourly, weekday, and weekly dummies. Robust standard errors in parentheses. $^{**}p < 0.05$; $^{***}p < 0.01$.

that our scaled Schoenfeld (1982) residuals have a zero slope over time. We also ran our models using alternative and more aggregate (e.g., eight-hour shift instead of hourly dummies) sets of time-of-sign-up controls. We do this to explore whether the initial vector with 31 controls (i.e., 23 hours, one weekday, and seven weeks) is absorbing too much of the variance and making it difficult to identify the effect of the covariate of interest (Hall et al. 2007). We also experimented with interacting the timing controls, allowing the effects of time of day controls to vary by day of week (e.g., $SignupHour_i \times SignupWeekday_i$). The results of all of these models are consistent with our main findings.

5.2. Demand for Technology Support

We now examine whether treated customers ask fewer questions during the initial stages of their service coproduction processes. The results of model (3) are presented in Table 3. In addition to reporting the coefficient for PCE_i, we also report the marginal effects of PCE on the percentage change in the number of questions asked and on the number of questions asked. The Poisson specification in column (1) indicates that the treatment reduces the number of questions asked by customers during their first week by 19.55%, an average of 0.119 questions less. The negative binomial specification in column (2) suggests a reduction of 23.89% in the number of questions, or 0.146 questions less, a slightly stronger yet qualitatively consistent estimate of the treatment's effect relative to that in column (1).

Columns (5) and (6) repeat the same analyses for $QuestionsWeek2_i$. We do not find any measurable effect of PCE using the Poisson model in column (5). Similarly, the economic and statistical significance of the negative binomial results in column (6) are weaker than those in column (2). In sum, we do not find conclusive evidence that PCE is effective in reducing the number of questions asked by customers during their first two weeks. This result is consistent with our previous findings that the effect of the treatment is primarily taking place during the first week.

We explore whether unobserved factors that drive customers' demand for technology support might bias our estimates of the effects of PCE. For example, as we explore in additional detail below, PCE may increase service usage either through its effects on churn or by increasing service usage per unit of time. Customers who use the service more might also have more questions; thus, service use is an unobserved variable whose presence could bias our estimates of β. Although this would likely create a positive bias in our estimates—making it more difficult to show a reduction in demand for support—we nonetheless explore how our results change when we control for service use.

In our data we observe, at any point in time, the number of different servers being used by customers. The provider believes (and we have empirically confirmed) this variable is positively correlated with customer demand for technology support. We calculated the number of different servers used by customers over their first week and first two weeks (i.e., $ServersWeek1_i$ and $ServersWeek2_i$), and because of the strong positive skew in the distribution, we use the log of (one plus) these variables as controls in our models. The results using these additional controls are shown in columns (3), (4), (7), and (8) of Table 3 and are very consistent with our main results. They are also similar if we use a $\ln(X_i)$ transformation instead of the shown $\ln(X_i + 1)$ transformation for the number of servers. The $\ln(X_i + 1)$ specification prevents us

Table 3 **Results for Number of Questions Asked**

Column:	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dependent variable:	$QuestionsWeek1_i$				$QuestionsWeek2_i$			
Model:	Poisson	Negative binomial	Poisson	Negative binomial	Poisson	Negative binomial	Poisson	Negative binomial
PCE_i	-0.218^{*}	-0.273^{**}	-0.232^{*}	-0.301^{**}	-0.164	-0.211^{*}	-0.180	-0.244^{**}
	(0.123)	(0.122)	(0.120)	(0.121)	(0.132)	(0.127)	(0.130)	(0.124)
$\ln(ServersWeek1_i + 1)$			0.542^{***}	0.548^{***}				
			(0.056)	(0.055)				
$\ln(ServersWeek2_i + 1)$							0.381^{***}	0.366^{***}
							(0.047)	(0.041)
Percentage change $e^{\hat{\beta}} - 1$ (%)	-19.55	-23.89	-20.73	-25.96	-15.12	-19.04	-16.51	-21.66
Discrete change[a]	-0.119	-0.146	-0.126	-0.160	-0.117	-0.149	-0.128	-0.169

Notes. All regressions use Model (3), consider the 2,673 customers in the sample, and include hourly, weekday, and weekly dummies. Robust standard errors in parentheses.

[a] $E(QuestionsWeekn \mid PCE_i = 1) - E(QuestionsWeekn \mid PCE_i = 0)$, $n = 1, 2$, holding all other covariates' values at their means.

$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

from loosing zero-valued observations of customers who had not yet launched a server during the first one or two weeks, and the $\ln(X_i)$ transformation prevents potential concerns with the aforementioned transformation of the variable.

Finally, to better understand PCE's effects on number of questions separately from its implications for survival, we ran the models in Table 3 on the subsample of customers who have lived at least as long as the period during which we count the number of questions (i.e., one and two weeks). The results were qualitatively similar to those described above. All models are robust to the use of the alternate sets of time-of-sign-up controls described near the end of §5.1.

6. Extensions

In this section, we develop a series of complementary analyses to our main findings.

6.1. Decay of Treatment Effect on Retention

We explore an interesting question for service operations: how long after sign-up does the treatment still influence customer retention. Standard tests for the equality of hazard functions on subsamples that gradually leave out early churners from the sample (e.g., sequentially drop from the sample customers who churn on day one, two, three, and so forth) do not find statistical differences between the treated and controls' hazard functions for subsamples beginning three or more days after sign-up. In other words, if we only consider customers who continue using the service at least this long, their hazard functions are very similar, suggesting the treatment's effect has decayed. These results are available in Online Appendix E.

Turning to a semiparametric approach, since it appears that the treatment's effect decays over time, the marginal effect of β in Model (2) should be smaller as customer tenure increases. We will identify this effect by interacting our treatment indicator with weekly tenure dummies (Cleves et al. 2010). We will use PCE_Weekn_{it} to represent the interaction with the nth week indicator, and $PCE\_OtherWeeks_{it}$ is turned on for all other weeks not considered in the model with the weekly dummies. We expect that only the estimates of the coefficients for the interactions with the early weeks (i.e., low n) will be negative and significant.

The results with this model are in Table 4. Column (1) uses a single indicator for the first week (i.e., $n = 1$). The coefficient for PCE_Week1_{it} represents a 49.60% (i.e., $e^{-0.685} - 1$) drop in the hazard rate, implying PCE causes treated customers to fail about half as fast as the controls during the first week of their tenure.

Variable $PCE\_OtherWeeks_{it}$ in column (1) is also negative and statistically significant, albeit it is only significant at the 10% level. The coefficient suggests that from the second week onward the treatment still reduces the hazard rate yet only by 16.71%, an effect much smaller than that found during the first week. Moreover, once we include an indicator for the second week, the effect vanishes. In other words, the treatment has no measurable effect during the second week nor afterward.

Table 4 **Decay of Treatment Effect on Survival Results**

Column:	(1)	(2)	(3)	(4)	(5)	(6)
PCE_Week1_{it}	−0.685**	−0.685**	−0.685**	−0.685**	−0.685**	−0.685**
	(0.287)	(0.287)	(0.287)	(0.287)	(0.287)	(0.287)
PCE_Week2_{it}		−0.269	−0.269	−0.269	−0.269	−0.269
		(0.434)	(0.434)	(0.434)	(0.434)	(0.434)
PCE_Week3_{it}			−0.273	−0.273	−0.273	−0.273
			(0.474)	(0.474)	(0.474)	(0.474)
PCE_Week4_{it}				−1.358	−1.358	−1.358
				(1.021)	(1.021)	(1.021)
PCE_Week5_{it}					−0.087	−0.087
					(0.479)	(0.479)
PCE_Week6_{it}					−0.634	−0.634
					(0.602)	(0.602)
PCE_Week7_{it}						−1.330
						(1.023)
PCE_Week8_{it}						0.022
						(0.627)
$PCE\_OtherWeeks_{it}$	−0.183*	−0.177	−0.172	−0.146	−0.126	−0.102
	(0.109)	(0.112)	(0.116)	(0.117)	(0.123)	(0.126)

Notes. All regressions employ the Cox proportional hazard model described in §6.1, use the 2,673 customers in the sample, and include hourly, weekday, and weekly dummies. There are 999 customers that churn. Robust standard errors, clustered on customers in parentheses.

*$p < 0.10$; **$p < 0.05$.

In sum, both our nonparametric and our semiparametric analyses indicate that the decay in PCE's effect is very fast and does not seem to last more than a week.

6.2. Influence of PCE on Service Usage

Thus far we have investigated the implications of PCE for customer retention and number of questions over the short run. These are important metrics that will influence both revenues and costs for the provider. However, another key performance metric that any service provider would like to improve is service usage. In this section we discuss PCE's short-run and long-run effects on service usage.

Our strategy for measuring the implications of PCE for service usage is shaped by several considerations. First, it is difficult to discern the effects of PCE on service usage over the short-run horizon that we used for retention and questions. The deployment of an application and its correct configuration in a cloud infrastructure service will usually take more than a couple of weeks, especially if the customer has never used a cloud infrastructure service before. Usage grows rapidly over the first several weeks of customer tenure (Figure 1). Further, usage patterns vary widely initially, depending upon such things as the type of application being deployed, the time available to deploy it, and customers' capabilities, among other things. This combination of factors makes it very difficult to measure the effects of PCE on usage over the very short run.

In Online Appendix F we describe the results of tests to measure the effects of PCE on short-run usage. We find that treated customers increase their usage over the first two weeks by 21.98% and over the first two months by 34.20%, however in some specifications the results are not statistically significant. In short, the results provide mixed evidence that PCE influences short-run service usage. These findings are consistent with the view that PCE provides benefits for customers, but also with the particular challenges of measuring the effects of PCE on usage over the short run.

However, over time the treatment can potentially influence long-run service usage. Over any fixed window of time (say, eight months), cumulative service usage per customer will increase if churn declines, as retained customers use the service over a longer window. This will be true even if PCE causes some marginal users of the service to now be included in the sample (i.e., those on the lower tail of the per-period usage distribution). Further, holding churn constant, PCE will also increase cumulative usage if it increases per period usage rates.

In our empirical models, we are unable to separately identify whether PCE influences cumulative long-run usage through churn or through per-period usage. To separately identify such effects would require an exclusion restriction that influences churn but does not influence use (Wooldridge 2010). We were unable to identify such an exclusion restriction. As a result, we simply explore the managerial implications of PCE for long-run use, whatever the causal mechanism might be.

To test if PCE influences long-run usage, we employ a linear model (e.g., $y=\alpha+\beta PCE_i+\delta SignupControls_i+\varepsilon_i$) with a metric of service usage as the dependent variable. Customers' server sizing and the provider's pricing decisions are both based on the amount of memory (i.e., GB of RAM) consumed per hour. Therefore, we can directly capture the service usage by aggregating the amount of GB RAM-hours consumed by users over time. In particular, we compute it over customers' first three, six, and eight months of tenure; eight months is the longest history we can observe of a customer who signed up for the service on the last day of the field experiment. Given the positive skew in its distribution we use the log as our dependent variable. As an additional robustness check, we also aggregate (and log) the number of different servers used per day by customers over the same time periods. We use $MemoryMonthT_i$ and $ServersMonthT_i$, $T=3,6,8$, to denote these metrics, and show their descriptive statistics in Table 5.

Table 5 Descriptive Statistics of Accumulated Service Usage Variables

Customer group:	All customers					Controls					Treated				
Number of customers:	2,673					2,307					366				
Variable	Obs.[a]	Mean	S.D.	Min	Max	Obs.	Mean	S.D.	Min	Max	Obs.	Mean	S.D.	Min	Max
$\ln(MemoryMonth3_i)$	2,611	6.38	2.48	−5.55	12.65	2,257	6.34	2.54	−5.55	12.65	354	6.67	2.07	−3.14	10.83
$\ln(MemoryMonth6_i)$	2,654	7.00	2.64	−5.55	12.85	2,291	6.95	2.70	−5.55	12.85	363	7.30	2.22	−3.14	12.14
$\ln(MemoryMonth8_i)$	2,665	7.24	2.73	−5.55	13.14	2,300	7.19	2.78	−5.55	13.14	365	7.55	2.33	−3.14	12.48
$\ln(ServersMonth3_i)$	2,611	3.93	1.45	0.00	7.62	2,257	3.90	1.48	0.00	7.62	354	4.15	1.22	0.00	6.45
$\ln(ServersMonth6_i)$	2,654	4.50	1.67	0.00	8.32	2,291	4.46	1.70	0.00	8.32	363	4.74	1.44	0.00	7.68
$\ln(ServersMonth8_i)$	2,665	4.72	1.77	0.00	8.61	2,300	4.68	1.80	0.00	8.61	365	4.96	1.55	0.00	7.90

[a]When performing the log transformation our sample loses a few observations that have zero values in their GB of RAM consumption and server count. All these customers eventually launched at least one server and had some consumption, yet this had not occurred by the time each variable aggregates total usage (e.g., more than eight months after sign-up).

Table 6 Results for Accumulated Service Usage

Column:	(1)	(2)	(3)	(4)	(5)	(6)
Dependent variable:		$\ln(MemoryMonthT_i)$			$\ln(ServersMonthT_i)$	
Accumulation period (T)	Three months	Six months	Eight months	Three months	Six months	Eight months
PCE_i	0.362***	0.373***	0.382***	0.278***	0.303***	0.312***
	(0.127)	(0.136)	(0.141)	(0.075)	(0.087)	(0.093)
Customers	2,611	2,654	2,665	2,611	2,654	2,665
Percentage change $e^{\hat{\beta}}-1$ (%)	43.65***	45.21***	46.57***	32.06***	35.34***	36.61***

Notes. All linear regressions include hourly, weekday, and weekly dummies. Robust standard errors in parentheses.
$^{***}p<0.01$.

The linear regression results, shown in Table 6, suggest the treatment increases service consumption. In particular, the point estimate of PCE's effect on service usage over eight months (column (3)) implies a percentage increase of 46.57%; its 95% confidence interval goes from 0.106 to 0.659, which implies a percentage increase in consumption between 11.15% and 93.27%. We show results with a $\ln(X_i)$ transformation, yet results are robust to using $\ln(X_i+1)$.

6.3. Role of Prior Experience

We provide further evidence that PCE plays a role in educating customers. Because of space limitations, the descriptive statistics and results discussed here are included in Online Appendix G. When constructing our sample (see §3.2.1) we had excluded customers with prior accounts with the provider because these might be different in systematic ways than the remainder of our sample. Here we investigate whether the effects of PCE on these customers might be smaller, as would be the case if prior experience in using cloud infrastructure services renders the early education less useful.

Including customers with prior relationships adds 419 customers to our baseline sample, 349 controls and 70 treated. We create a new dummy variable, $PriorAccounts_i$, that is turned on if customer i had any prior accounts with the provider in addition to the account opened during the time frame of the field experiment. Then, we add the new variable and its interaction with the treatment indicator to all our models. If having prior accounts mitigates the effect of the treatment, then PCE's marginal effect should be greater for new customers than for existing customers.

Using the modified version of our probit model (1b), we find that both the treatment and the condition of having a prior account increase the likelihood of survival through the first week. However, the interaction of the two is negative and significant, indicating that the effects of PCE are smaller for customers with prior experience. These results imply that PCE increases the likelihood of survival through the first week by 3.7 percentage points for customers without prior accounts, yet has no measurable impact on those with prior accounts. This is consistent with the view that having prior experience reduces the value of the treatment. We find qualitatively similar results in the Cox proportional hazard model: both PCE and prior accounts decrease the hazard rate and hence increase survival, yet their interaction term is positive and significant at the 5% level. Although PCE reduces the hazard of churn by 22.21% for customers without prior experience, it has no statistically significant impact on customers with prior experience.

We also find, using the modified version of the Poisson model (3), that PCE has no measurable effect on the number of questions asked during the first week by customers that have prior accounts. Meanwhile, it reduces the number of questions asked by customers without prior accounts by an average of 0.129 (19.63%), consistent with our result in Table 3.

In conclusion, PCE has no measurable impact on the behavior of customers with prior experience in using the cloud provider's services. This supports the view that PCE educates new adopters since they are the ones facing the steepest learning curves.

7. Managerial Implications

When asked about the rationale for offering PCE, an executive from the provider noted that their reactive technology support agents were often asked very basic questions regarding the service's features by customers who had signed up some months ago. This implied that customers were very likely self-servicing themselves a degraded service experience that could be having negative effects on their satisfaction and increasing their risk of churning. Moreover, any question that can be preempted and addressed in a proactive manner represents a reduction in the demand for reactive technical support, which given its uncertainty requires additional staffing and is more costly to offer (Akşin et al. 2007). After the provider concluded the field experiment, it decided to apply the PCE treatment to all new customers.

To gain a better perspective on the provider's incentives to offer PCE, we calculate an estimate of the economic payoff per customer over eight months—the maximum period of time that we observe customers in our data—from offering PCE. We first estimate the cost of treating a customer and then the marginal profit gains from the increased service usage and the support cost savings.

The provider indicated to us that a PCE treatment could last at most 25 minutes and that a verification agent costs about \$18.30/hour. Other than the labor costs of the verification agent, there are no other direct costs to treatment. Thus, the direct costs of offering PCE are approximately $\$18.30 \times (25/60) = \$7.63$ per treatment.

Before estimating the profit increases from PCE, recall from §3.2.4 that not all customers who sign up for the service actually use it (e.g., run a server for more than 24 hours). As a result, although the costs of PCE are borne by the provider for each customer who receives the treatment, the benefits to the provider will only be realized for customers who actually use the service. Based on our data, we estimate that 82.3% of legitimate customers use the service and therefore we will normalize the per customer profit gains by that number.

The median control customer consumes 2,498 GB RAM-hours over the first eight months of its tenure (we do not use the mean because of the skew in the distribution), which is roughly equivalent to running a 512 MB (or 0.5 GB) RAM server all the time. Then, the lower bound of the confidence interval for PCE's impact on service usage after eight months was a 11.15% increase (see §6). Finally, at the time of the field experiment, the provider charged \$0.06 per GB RAM-hour, so the additional consumption yields $2{,}498 \times 11.15\% \times 0.06 = \$16.71$ more in revenues for the median customer over eight months. To have a more realistic estimate, we also consider the variable cost of the usage of the cloud servers.

We learned from the provider's strategic finance group that their variable costs are around 20%. These include server and data center depreciation expenses, data center rent, power and cooling, and noninfrastructure related items like credit card fees and bad debt expenses. After considering the 80% profit margin on the cloud servers offering, the prior revenue yields \$13.37 in additional profits. We note there are additional costs that vary with demand, but respond to it in a delayed manner or in the longer term. These "semivariable" costs (the term used by the provider) include, for example, the need to invest in new data centers and hire new staff to manage them as demand grows. However, we will only focus on variable costs since our estimates of the effects of PCE span eight months, a period during which longer-term investments such as those mentioned would not come into play.

Regarding support costs savings, the provider estimates that, during the duration of the field experiment it cost them on average \$36.83 to address each support ticket and \$7.46 to hold each chat session. To attain a conservative cost savings estimate and since most (60%) questions came through chats, we assume that all questions come through chats. PCE reduced the number of questions asked during the first two weeks by approximately 0.149 (see column (6) of Table 3). Since we believe—and tested—that PCE has no or little effect on the demand for support past the second week, we may estimate the support cost savings per customer as $\$7.46 \times 0.149 = \$1.11$. There is an indirect effect of PCE that increases overall support costs by means of having more customers who may demand support, yet since customers demand so little support later in their tenure this effect will be negligible.

Considering the likelihood of a customer using the service, the estimated profit gains are then $(\$13.37 + \$1.11) \times 82.3\% = \$11.92$. If we additionally consider the cost of applying PCE, we have a net gain of $\$11.92 - \$7.63 = \$4.29$, which represents a return on investment of around $\$4.29/\$7.63 = 56\%$. Note that this estimate is an approximation and based on a series of assumptions, and so it should be interpreted with care. However, many of our assumptions are conservative. Specifically, we assume the lower bound of the confidence interval for PCE's effect on increasing accumulated usage, we assumed all support interactions are through the less costly chat channel, and our estimate would grow if we considered a longer lifetime value (e.g., two years).

8. Conclusion

Leveraging a field experiment executed by a public cloud infrastructure services provider, our study is the first to quantify the effects of customer education, and in particular PCE, on customer retention and demand for technology support. In doing so we contribute to the still scarce literature that explores providers' customer support costs in service coproduction environments (Kumar and Telang 2011). In a broader context, again to our knowledge, we are also the first to measure education's effects on retention based on actual usage of a service and not just on customers' forward-looking intentions to continue using a service captured through surveys (e.g., Bell and Eisingerich 2007).

Our estimates of PCE's effect on customer behavior are economically significant. During the first week, which is when customers are most likely to abandon the service, customers who receive PCE are retained about twice as much as customers who do not. Additionally, on average, the treated customers ask 19.55% fewer questions during their first week since signup relative to the controls. Since the offering of technology support is a very costly and labor-intensive endeavor, reducing the number of customer-initiated support requests represents an important cost reduction. We further show that PCE also increases overall service usage over a longer time horizon, perhaps in part because of its effects on reducing customer churn. In sum, by offering PCE, the provider affects its profits through both increases in revenues and reductions in support costs. A rough estimate of these economic benefits suggests over an eight-month period the provider could earn a 56% return on the costs of offering PCE.

We believe our findings regarding PCE's positive impact on customer behavior can be generalized to other service settings where the following conditions are met: (i) customers can enroll in and freely abandon the service, (ii) customers play an important role in coproducing the service, (iii) there are common starting skill and knowledge requirements (e.g., frequently asked questions), and (iv) it is possible to proactively engage customers when they sign up. An example that may satisfy all of these conditions and where customer support has improved customer efficiency is online banking (Field et al. 2012). Online learning programs (e.g., distance education courses) are known to suffer from early attrition problems (Muilenburg and Berge 2005, Tyler-Smith 2006) and also meet these criteria.

Despite this paper's contributions, it is still subject to some limitations. For example, our model for the number of questions asked by customers captures PCE's effect on both attrition and the number of questions. Similarly, we are unable to identify PCE's effects on per-period service consumption. Furthermore, we have assumed that all questions are equally costly to address. Future work could measure the effects on per-period consumption and different types of support interactions through additional data collection.

Future research may address questions associated with the value of a PCE-based business strategy across various settings. For example, although PCE's payoff at the individual level is positive, there are challenges at large scales that may limit its feasibility (e.g., ability to staff enough agents). A potential way of addressing this issue is by examining varying levels of PCE and determining less costly treatments that still produce the desired outcomes. Future field experiments, similar to the one used in this study, can serve this purpose.

Supplemental Material

Supplemental material to this paper is available at http://dx.doi.org/10.1287/msom.2015.0547.

Acknowledgments

The authors thank Ryan W. Buell, Manpreet Hora, Jérémie Gallien, and Alan Scheller-Wolf (the coeditors), the associate editor, and three anonymous reviewers. This research would not have been possible without the data and constructive feedback from the staff of the cloud provider in this study, to whom the authors are sincerely grateful. The authors are particularly grateful for all of the support received from the cloud provider's business analytics, customer tech support, finance, and strategy teams. The authors also thank seminar participants at the Georgia Institute of Technology, INCAE Business School, and Fudan University, as well as conference attendees and reviewers from the Production and Operations Management (POMS) 25th Annual Conference, the Conference on Digital Experimentation @ MIT (CODE@MIT) 2014, and the Conference of Information Systems and Technology (CIST) 2014 for their helpful comments and suggestions. The first author acknowledges funding support received through the Ashford Watson Stalnaker Memorial Prize for Student Excellence in the Ph.D. Program, from the Scheller School of Business, Georgia Institute of Technology. All errors or omissions are the authors' own.

References

Akşin OZ, Harker PT (1999) To sell or not to sell: Determining the trade-offs between service and sales in retail banking phone centers. *J. Service Res.* 2(1):19–33.

Akşin Z, Armony M, Mehrotra V (2007) The modern call center: A multi-disciplinary perspective on operations management research. *Production Oper. Management* 16(6):665–688.

Angrist JD (1990) Lifetime earnings and the vietnam era draft lottery: Evidence from social security administrative records. *Amer. Econom. Rev.* 80(3):313–336.

Angrist JD, Imbens GW, Rubin DB (1996) Identification of causal effects using instrumental variables. *J. Amer. Statist. Assoc.* 91(434):444–455.

Armony M, Gurvich I (2010) When promotions meet operations: Cross-selling and its effect on call center performance. *Manufacturing Service Oper. Management* 12(3):470–488.

Armony M, Maglaras C (2004a) Contact centers with a call-back option and real-time delay information. *Oper. Res.* 52(4): 527–545.

Armony M, Maglaras C (2004b) On customer contact centers with a call-back option: Customer decisions, routing rules, and system design. *Oper. Res.* 52(2):271–292.

Bell SJ, Eisingerich AB (2007) The paradox of customer education: Customer expertise and loyalty in the financial services industry. *Eur. J. Marketing* 41(5/6):466–486.

Bhattacherjee A (2001) Understanding information systems continuance: An expectation-confirmation model. *MIS Quart.* 25(3):351–370.

Breslow N (1970) A generalized Kruskal-Wallis test for comparing K samples subject to unequal patterns of censorship. *Biometrika* 57(3):579–594.

Buell RW, Campbell D, Frei FX (2010) Are self-service customers satisfied or stuck? *Production Oper. Management* 19(6):679–697.

Cameron AC, Trivedi PK (2010) *Microeconometrics Using Stata, Revised Edition* (Stata Press, College Station, TX).

Campbell D, Frei F (2010) Cost structure, customer profitability, and retention implications of self-service distribution channels: Evidence from customer behavior in an online banking channel. *Management Sci.* 56(1):4–24.

Challagalla G, Venkatesh R, Kohli AK (2009) Proactive postsales service: When and why does it pay off? *J. Marketing* 73(2): 70–87.

Chase RB (1978) Where does the customer fit in a service operation? *Harvard Bus. Rev.* 56(6):137–142.

Cleves MA, Gould WW, Gutierrez RG (2010) *An Introduction to Survival Analysis Using Stata*, 3rd ed. (Stata Press, College Station, TX).

Cox DR (1972) Regression models and life-tables (with discussion). *J. Roy. Statist. Soc., Ser. B* 34(2):187–220.

Eisingerich AB, Bell SJ (2008) Perceived service quality and customer trust: Does enhancing customers' service knowledge matter? *J. Service Res.* 10(3):256–268.

Fader PS, Hardie BGS, Lee KL (2005) "Counting your customers" the easy way: An alternative to the Pareto/NBD model. *Marketing Sci.* 24(2):275–284.

Field JM, Xue M, Hitt LM (2012) Learning by customers as coproducers in financial services: An empirical study of the effects of learning channels and customer characteristics. *Oper. Management Res.* 5(1–2):43–56.

Fodness D, Pitegoff BE, Sautter ET (1993) From customer to competitor: Consumer cooption in the service sector. *J. Services Marketing* 7(3):18–25.

Gans N, Koole G, Mandelbaum A (2003) Telephone call centers: Tutorial, review, and research prospects. *Manufacturing Service Oper. Management* 5(2):79–141.

Hall BH, Mairesse J, Turner L (2007) Identifying age, cohort, and period effects in scientific research productivity: Discussion and illustration using simulated and actual data on French physicists. *Econom. Innovation New Tech.* 16(2):159–177.

Hearst N, Newman TB, Hulley SB (1986) Delayed effects of the military draft on mortality: A randomized natural experiment. *New England J. Medicine* 314(10):620–624.

Johnson EJ, Bellman S, Lohse GL (2003) Cognitive lock-in and the power law of practice. *J. Marketing* 67(2):62–75.

Jones TO, Sasser WE (1995) Why satisfied customers defect. *Harvard Bus. Rev.* 73(6):88–91.

Jouini O, Akşin Z, Dallery Y (2011) Call centers with delay information: Models and insights. *Manufacturing Service Oper. Management* 13(4):534–548.

Jouini O, Dallery Y, Akşin Z (2009) Queueing models for full-flexible multi-class call centers with real-time anticipated delays. *Internat. J. Production Econom.* 120(2):389–399.

Kaplan EL, Meier P (1958) Nonparametric estimation from incomplete observations. *J. Amer. Statist. Assoc.* 53(282):457–481.

Khalifa M, Liu V (2003) Determinants of satisfaction at different adoption stages of Internet-based services. *J. Assoc. Inform. Systems* 4(5):206–232.

Klemperer P (1995) Competition when consumers have switching costs: An overview with applications to industrial organization, macroeconomics, and international trade. *Rev. Econom. Stud.* 62(4):515–539.

Kumar A, Telang R (2011) Product customization and customer service costs: An empirical analysis. *Manufacturing Service Oper. Management* 13(3):347–360.

Kumar A, Telang R (2012) Does the Web reduce customer service cost? Empirical evidence from a call center. *Inform. Systems Res.* 23(3):721–737.

Mantel N, Haenszel W (1959) Statistical aspects of the analysis of data from retrospective studies of disease. *J. National Cancer Inst.* 22(4):719–748.

McKinney V, Yoon K, Zahedi FM (2002) The measurement of Web-customer satisfaction: An expectation and disconfirmation approach. *Inform. Systems Res.* 13(3):296–315.

Mell P, Grance T (2011) The NIST definition of cloud computing. NIST Special Publication 800-145, National Institute of Standards and Technology, Gaithersburg, MD. http://csrc.nist.gov/publications/nistpubs/800-145/SP800-145.pdf.

Muilenburg LY, Berge ZL (2005) Student barriers to online learning: A factor analytic study. *Distance Ed.* 26(1):29–48.

Nayyar PR (1990) Information asymmetries: A source of competitive advantage for diversified service firms. *Strategic Management J.* 11(7):513–519.

Retana GF, Forman C, Narasimhan S, Niculescu MF, Wu DJ (2015) Technology support and post-adoption IT service use: Evidence from the cloud. Working paper, INCAE Business School, Alajuela, Costa Rica.

Rubin DB (1974) Estimating causal effects of treatments in randomized and nonrandomized studies. *J. Educational Psychol.* 66(5):688–701.

Schoenfeld D (1982) Partial residuals for the proportional hazards regression model. *Biometrika* 69(1):239–241.

Sharma N, Patterson PG (1999) The impact of communication effectiveness and service quality on relationship commitment in consumer, professional services. *J. Services Marketing* 13(2):151–170.

Staples DS, Wong I, Seddon PB (2002) Having expectations of information systems benefits that match received benefits: Does it really matter? *Inform. Management* 40(2):115–131.

Tyler-Smith K (2006) Early attrition among first time eLearners: A review of factors that contribute to drop-out, withdrawal and non-completion rates of adult learners undertaking eLearning programmes. *J. Online Learn. Teaching* 2(2):73–85.

Wooldridge JM (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. (MIT Press, Cambridge, MA).

Xue M, Harker PT (2002) Customer efficiency. *J. Service Res.* 4(4):253–267.

Xue M, Hitt LM, Chen P-Y (2011) Determinants and outcomes of Internet banking adoption. *Management Sci.* 57(2):291–307.

Xue M, Hitt LM, Harker PT (2007) Customer efficiency, channel usage, and firm performance in retail banking. *Manufacturing Service Oper. Management* 9(4):535–558.

Zeithaml VA, Berry LL, Parasuraman A (1996) The behavioral consequences of service quality. *J. Marketing* 60(2):31–46.