Mitigating the Cold-start Problem in Reputation Systems: Evidence from a Field Experiment*

Xiang Hui[†] Washington University

Zekun Liu[‡] New York University

Weiqing Zhang[§] New York University

November 15, 2020

Abstract

Reputation systems are typically used in markets with asymmetric information, but they can cause the cold-start problem for young sellers who lack historical sales. Exploiting a field experiment on eBay, we show that in the presence of a long-run quality signal, introducing a less history-dependent quality signal mitigates the cold-start problem: it increases demand for high-quality young sellers, incentivizes their quality provision, and increases their chance of obtaining the long-run quality signal. Moreover, it prompts established sellers to re-optimize their effort decision. Therefore, the net impact of introducing a less history-dependent signal on quality provision depends on underlying market fundamentals.

*We thank Dan Elfenbein, Andrey Fradkin, Ginger Jin, Tobias Klein, Meng Liu, Maryam Saeedi, Konrad Stahl, Tianxin Zou, and many seminar and conference participants for helpful discussions and comments. We are grateful to eBay for providing access to the data.

[†] hui@wustl.edu

[‡] zl3126@stern.nyu.edu

[§] wz1571@stern.nyu.edu

1 Introduction

Markets in which seller quality is imperfectly observed by buyers are subject to both adverse selection and moral hazard. To reduce asymmetric information in these markets, buyers typically rely on reputation systems, including seller ratings and third-party certifications.[1] A large literature has demonstrated the effectiveness of online reputation systems (summarized in Tadelis (2016)). However, reputation is based on sellers' past performance, and can therefore lead to the cold-start problem: young sellers need enough sales to build their reputation, but in markets where consumers automatically flock to reputable sellers, even high-quality, young sellers can be prevented from building their reputation. Anticipating the problem, sellers may not even enter the market in the presence of an entry cost. This concern is shared in many contexts, including the argument over whether information on past performance should be used when selecting vendors in public procurement auctions (Butler et al. (2020)), to the justification of infant-industry protection as a way to overcome the informational barrier to entry (Grossman and Horn (1988)), to the various programs by digital platforms to grow their network of service providers.

The cold-start problem is informational: without established reputation, high-quality young sellers can be indistinguishable from low-quality ones in the eyes of consumers (Klein and Leffler (1981)). One natural question arises: Could market regulators mitigate the problem by increasing market transparency *for young sellers*? Our answer is yes, based on a field experiment on one of the largest e-commerce marketplaces. In our setting, there is a long-run quality certification for sellers and the increase in market transparency for young sellers is achieved by the introduction of a second quality signal that is less history-dependent than the existing certification. Specifically, a seller's certification status is evaluated monthly based on whether her historical sales and measured quality in various aspects are above some marketplace-determined thresholds. Compared to the certification, the second quality signal has no requirement on sellers' historical sales and focuses on a particular aspect of quality, i.e., fast shipping. The two features make the new signal less history-dependent (and less demanding to achieve) than the existing certification, as a young seller is eligible for the new signal if she offers fast shipping even without having accumulated many sales. Therefore, the second signal is much more attainable by young sellers than the existing certification.

To guide our empirical analyses, we build a stylized asymmetric information framework. Sellers differ in their historical sales and quality, and their effort choice affects their measured quality

[1] Examples of reputation: Yelp ratings on businesses and seller ratings on various other digital platforms. Examples of certification: investment grades in financial markets, Better Business Bureau ratings, and occupational licensing.

for the coming evaluation. Buyers do not observe sellers' quality, but observe quality signals that are credibly issued by the platform. Because the second signal is less history-dependent than the first, high-quality, young sellers can quickly obtain it upon entry. This allows them to separate themselves from low-quality sellers in the eyes of consumers, thereby increasing their demand and sales. The lift in demand in turn increases their chance of meeting the certification's requirement on historical sales for the coming evaluation, and hence their chance of becoming certified. The better career perspective incentivizes them to exert effort to also meet the certification's requirement on measured quality for the coming evaluation. Because the second signal is less demanding to obtain, established, certified sellers re-evaluate the cost and benefit of being certified with the introduction of the new signal, and those with higher effort costs of staying certified decide to reduce their effort and opt out of the selective certification, and adopt the less demanding signal instead. We take these testable predictions to the data and the experiment.

Our empirical context is eBay. Since 2009 the e-commerce platform has had a renowned certification system, the eBay Top Rated Seller (eTRS) program. Specifically, the platform evaluates sellers' performance every month and privately labels those who meet the program's requirement based on historical sales and various aspects of quality as eTRS sellers. When eTRS sellers offer one-day handling and 30-day returns on a listing, a public eTRS Plus badge is displayed on this listing.[2] We consider the eTRS Plus badge the existing certification, i.e., the first signal. In August 2017, eBay introduced the eBay Guaranteed Delivery (eGD) program, which we regard as the second signal. Specifically, eBay evaluates sellers' eligibility for the program every month based on their shipping quality. After enrollment, sellers need to offer one-day handling on a listing for it to receive a public eGD signal. The eGD signal reads "Guaranteed Delivery By X," where X is a date calculated by eBay based on the shipping service and buyer and seller locations. If the item arrives later than eBay's guaranteed date, the buyer can choose either getting a $5 coupon towards future purchases on eBay or returning the item for free.[3] Although buyers may regard the eGD signal as insurance against late delivery, we provide evidence that the insurance value of the eGD signal is limited.

We use a field experiment to identify the effect of the second signal. Before rolling out the eGD

[2]eBay also displays feedback scores and percentage positive (pp) ratings on the item description page. Nosko and Tadelis (2015) show that pp ratings have a median of 100% and a mean of 99.3% and have limited informativeness of quality. The skewed distribution of ratings is also documented by other authors, including Resnick and Zeckhauser (2002), Dellarocas and Wood (2008), Bolton et al. (2013), Mayzlin et al. (2014), Zervas et al. (2015), Luca and Zervas (2016), Filippas et al. (2018), Fradkin et al. (2019), and Schoenmüller et al. (2019).

[3]The cost is borne by eBay if the seller promptly shipped the item as promised, but borne by the seller otherwise.

program on the entire marketplace in May 2018, eBay invited a selected group of sellers, who are on average larger sellers with higher quality than an average seller on eBay, to join the waitlist for the pilot eGD program. In August 2017, eBay randomly enrolled half of the sellers on the waitlist in the pilot program, while keeping the other half status quo. Our identification of the program effects comes from comparing the two groups of sellers, who both wanted in, but only one group got enrolled for a random reason. This comparison gives us the complier average causal effect—the policy effect on sellers who will enroll in the eGD program when they are offered the opportunity to do so.

Our data includes listings from the six months before and the six months after the experiment (Feb 2017–Jan 2018), which we refer to as the pre- and post-period, respectively. We start our analyses by showing that enrolling in the eGD program increases sellers' sales quantity by 3.6% and price by 1.7% in the post-period. The increase in price is larger for sellers who are never eTRS (or non-eTRS sellers) in the pre-period, and is statistically insignificant for eTRS sellers, suggesting that the signalling value of the eGD signal is smaller for certified sellers. We also find that non-eTRS sellers (resp., eTRS sellers) in the pre-period are more likely to become (resp., stay) eTRS in the post-period. Lastly, enrolled non-eTRS sellers are more likely to ship items on time and receive fewer negative ratings and buyer complaints in the post-period.

Next, we study the effect of the second signal on seller dynamics and quality provision. First, among non-eTRS sellers, young sellers who have registered on eBay for less than one year are more likely to step up and become certified in the post-period after enrolling in the eGD program. Additionally, stepping up happens more for sales-constrained sellers, namely those who did not meet the certification's requirement on historical sales in the pre-period. Moreover, young and sales-constrained non-eTRS sellers provide higher quality after enrolling in the eGD program. Second, since we do not observe eTRS sellers' cost of being certified as eTRS, we construct (imperfect) proxies for it. For example, one proxy is a seller's share of voluntary refunds: one key eTRS requirement is to have few cases of claim escalations to eBay, and eTRS sellers sometimes prevent such escalations by voluntarily offering partial refunds to buyers even though these refunds are not mandated by eBay; a seller with no interest in being eTRS has less incentive to do so. Using the cost proxies, we find that while eTRS sellers with low effort costs are more likely to maintain their certification in the post-period, those with high effort costs choose to step down and opt for the eGD signal instead. Furthermore, the sellers who step down decrease their quality provision. Lastly, we perform a back-of-the-envelope calculation on consumer welfare and find that it increases by 5.2% for transactions from eTRS sellers and by 9.7% for transactions from non-eTRS sellers.

Could the results be driven by mechanisms other than the signalling value of the eGD signal? One alternative mechanism is search: sellers are more likely to offer one-day handling, and eBay's search algorithm prioritizes these listings. Alternatively, it is possible that the eGD has no signalling value, but sellers wrongly believe otherwise and increase effort, and the results are entirely driven by the increased effort. To mitigate these concerns, we estimate the demand for the eGD signal using the matched listing approach developed by Elfenbein et al. (2012) and Einav et al. (2015). The listings are matched based on seller identity, listing title, listing subtitle, product catalog, listing start week, and price plus shipping fees. The granular matching aims at reducing the effect of differences in unobserved seller effort on sales outcomes. We then estimate that offering one-day handling (i.e., having the eGD signal) increases the sales probability by 1.8% within matched listings. Additionally, this estimate becomes 1.6% when controlling for the number of times that a listing is seen by buyers on search result pages. The small decrease in the demand estimate for the eGD signal suggests that the search effect is small. Lastly, another alternative mechanism is that buyers may view the eGD singal as insurance against late delivery. However, conditional on late delivery, consumers return only 0.014% of the items, and claim the voucher 18% of the time; the unconditional average for the two measures is 0.0008% and 1.04%, respectively. These numbers suggest that the insurance value of the eGD signal is limited.

Lastly, to understand the effect of having a second quality signal in different markets, we use the full rollout of the eGD program in May 2018 as a natural experiment, and use eBay's more than 400 subcategories of products to define the markets. Our model predicts that stepping up should happen more frequently in markets where the certification's requirement on historical sales is more binding. To test this prediction, we construct our measure of bindingness using the share of sellers who fail to meet the sales requirement of the certification. We find that in the markets where the sales requirement is more binding, sellers are more likely to become eTRS in the six months after the policy change, and the average quality provision is higher.

Our results help inform the design of reputation systems in markets with asymmetric information. We demonstrate that, on the one hand, having a second quality signal that is less history-dependent than an existing long-run quality signal could serve as a stepping stone for young sellers and mitigate their cold-start problem, thereby increasing their incentive of exerting effort and the likelihood of obtaining the long-run signal. On the other hand, established sellers with high effort costs may step down and provide lower quality. The overall change in the share of certified sellers and quality provision depends on the distribution of the bindingness of the certification and the distribution of the effort costs in the market. Therefore, market designers should consider these forces when designing reputation systems.

1.1 Related Literature

Our paper is closely related to an emerging literature that empirically studies how the design of reputation systems affects the entry and growth of young sellers.[4] Using a major update of the reputation-based certification on eBay, Hui et al. (2017) show that the stringency of the certification changes consumers' belief about the average quality of sellers with and without the certification, and hence the price they pay. This in turn has a large effect on the entry rate and quality distribution of entrants in the market. Farronato et al. (2020) show in the context of Thumbtack that occupational licensing, which is a certification by the government, adds little informational value to peer-generated reviews on the platform but is associated with less competition. Both papers find evidence consistent with the fact that certification could become an entry barrier for new sellers. In the context of Taobao, Li et al. (Forthcoming) study a program that allows sellers to pay for feedback. They find that sellers who enroll in this program receive more feedback and experience an increase in their sales especially on new products, because the enrollment is a signal of their quality. Our paper contributes to this novel literature in three ways. First, we show that the introduction of a less history-dependent and more focused quality signal can mitigate the cold-start problem induced by a reputation system that signals long-run quality. Second, our rare field-experiment setting mitigates a key identification issue that sellers with high latent quality self-select into adopting a quality signal, which would bias the estimation. Lastly, we document both step-up and step-down dynamics from sellers due to the introduction of the new quality signal, which makes its net impact dependent on underlying market fundamentals.

Our paper is also related to the broad literature that studies the effect of reputation systems both as a motivational device for sellers and as an informational device for buyers. Starting from the motivational role of reputation systems, Klein et al. (2016) show that increased market transparency due to a reputation policy change on eBay leads to higher quality provision from sellers. Besides leading to reduced moral hazard, the policy change also leads to less adverse selection due to a redistribution of sales from low-quality to high-quality sellers, according to Hui et al. (2018). More generally, the finding that quality disclosure reduces moral hazard has also been documented

[4]There is also a growing theoretical literature on this topic, e.g., Hopenhayn and Saeedi (2018) and Vellodi (2018).

in other contexts, and Dranove and Jin (2010) provide a comprehensive summary of this literature. Moving on to the informational role of reputation systems, many papers show that consumers respond to seller information in predictable ways. Chevalier and Mayzlin (2006) show that an increase in book reviews increases book sales. Luca (2016) estimates large economic impacts of reviews in the case of Yelp. Barach et al. (2020) show that platforms can steer demand by recommending or giving signals to sellers. Bollinger et al. (2011) show that consumers alter their food choice after mandatory calorie posting. Resnick et al. (2006), Cabral and Hortacsu (2010), Saeedi (2019), Elfenbein et al. (2015), and Hui et al. (2016) study the effect of ratings and certification status on eBay and show that they lead to variations in price variations, higher sales probabilities, and larger consumer welfare across a spectrum of products. In the presence of reputation premiums, sellers may initially charge a lower price to gain reputation, as shown by Fan et al. (2016). Dellarocas (2003), Einav et al. (2016), and Tadelis (2016) provide helpful summaries of this literature.

2 Background and Policy Change

eBay is among the first e-commerce platforms in the world. To facilitate trade among strangers, eBay designed a feedback rating system where a buyer could rate the experience as positive, neutral, or negative after each transaction. This rating is then aggregated and presented to future buyers in two measures, the percentage positive score and the feedback score (the number of positive minus negative feedback ratings). Additionally, buyers can give anonymous detailed seller ratings (DSR) to sellers in four dimensions: item as described, communication, shipping speed, and shipping charge.

In September 2009, eBay introduced the eBay Top Rated Seller (eTRS) certification program. Sellers are evaluated on the 20th of each month based on a set of requirements on sales, transaction quality, and shipping quality set by eBay. Sellers who qualify receive an eTRS badge that is displayed on all their listings. In November 2012, eBay changed its seller-level certification system to a listing-level certification. In particular, a listing from an eTRS seller can get a Top Rated Seller Plus (eTRS Plus) badge if and only if the listing offers short handling time and allows for returns. The old eTRS badge became obsolete and does not appear on the search result or on the item description page.[5]

[5]Theoretically, a potential buyer can click on a seller's user name on the item description page and go to the seller's profile page to see if a seller is "Top Rated", "Above standard", or "Below standard". However, using clickstream logs from eBay, Nosko and Tadelis (2015) show that less than 1% of buyers ever click on a seller's user name and on the seller profile page. Therefore, one can essentially treat the old eTRS signal as retired and obsolete.

The requirements for becoming an eTRS seller are listed below:

- Sales: At least 100 transactions and $1,000 in sales in the past 12 months;

- Transaction quality: (1) claims without seller resolution[6] are less than 0.3%; and (2) the combined rate of seller-initialized cancellation and unresolved claims is less than 0.5%;

- Shipping quality: (1) the late shipment rate does not exceed 3% and (2) tracking information[7] is uploaded on time for at least 95% of transactions.

The rates are calculated based on a seller's transactions in the past three months if the seller has fewer than 400 transactions in that period, or based on the past 12 months otherwise. If a seller meets the above requirements in a month, she is categorized by eBay as an eTRS for that month. In addition, she receives the Top Rated Plus (eTRS Plus) badge on listings that offer one day or less handling time and 30-day or longer returns with money back option. Besides the signalling value of the eTRS Plus badge, sellers also receive a 10% discount on final value fees on listings with an eTRS Plus badge. In our sample period, around 70% of listings from eTRS sellers have the eTRS Plus badge.

In May 2018, eBay rolled out the eBay Guaranteed Delivery (eGD) on its U.S. marketplace after a few pilot studies. With this program, eBay wants to inform buyers that many sellers already offer fast shipping and encourage fast shipping from other sellers. The program is free to join for sellers who satisfy the following requirements: (1) the late shipment rate does not exceed 5% and (2) tracking information is uploaded on time for at least 95% of transactions.[8] The evaluation date and look-back period for the eGD program are the same as those of the eTRS program. Conditional on enrollment, a seller can obtain an eGD signal on a listing that reads "Guaranteed Delivery By X"—where X is a date calculated by eBay based on the shipping service and buyer and seller locations—if the seller chooses one of the two options for that listing: (1) handling time—offer 1 day or less handling time and (2) door-to-door—guarantee delivery by the seller herself. In the analysis, we do not distinguish between the two options, and the handling time option is much more popular (almost 90% of enrolled sellers choose this option). In the case of late delivery, buyers can choose to receive a $5 coupon towards future purchases on eBay or return the item for free. The

[6] When a buyer files a claim on an order and the seller cannot resolve it in three days, the case is escalated to eBay. It is counted towards a claim without seller resolution if eBay decides the seller is at fault.

[7] eBay requires sellers to upload orders' tracking information before the estimated delivery date, which is calculated based on zip codes and shipping methods.

[8] After the full rollout of eGD in May 2018, eligible sellers must also have at least 100 transactions over the past 12 months. However, this requirement is absent during the pilot runs that we study. Also, even after the full rollout, the eGD still has lower sales requirements than the eTRS because the eGD program does not require past sales of $1,000.

cost is borne by eBay if the tracking shows on-time shipping, but borne by the seller otherwise. Besides the insurance value of the eGD signal in the case of late delivery, the eGD signal can also have search benefits: Buyers can choose to see eGD listings in the search results page. A screenshot of the eGD signal and the eTRS Plus badge in the search results page is shown in Figure 5.

Comparing the new signal and the certification, the eGD signal has no requirements on historical sales and focuses on a specific quality aspect (i.e., shipping quality), making it strictly easier to get than the eTRS certification.[9] Also, the eGD signal may have insurance and search benefits.

3 A Theoretical Framework

We present a stylized framework to illustrate how sellers' historical sales level and total effort cost to reach a certain reputation level affect their certification dynamics and effort decision (quality provision) following the introduction of the eGD program.

Supply: Consider a model in which a seller chooses an effort level to maximize her total profit. First, the seller is endowed with a historical sales volume level V_0 and a reputation level R_0, which are the performance metrics that eBay evaluates. The seller then determines the effort level e. We normalize the per unit effort cost to 1, and therefore, the initial reputation level R_0 determines the total effort cost of reaching a certain reputation level. Hence, a greater initial reputation level implies a lower total cost.

Information: Consumers do not observe the reputation level R.[10] Instead, a marketplace regulator observes this and issues a credible badge B that signals if a seller's reputation and/or sales level is above a certain threshold. We consider two certification mechanisms to represent what we observe in the empirical setting. Prior to the introduction of the eGD program, sellers are either eTRS or have no badge. After the introduction of the eGD program, sellers that are rewarded with the eGD signal serve as the intermediate tier. Formally, we use $\mathcal{M}$ to denote the mechanism: (1) a two-tier mechanism ($\mathcal{M} = II$), where $B \in \{T, NB\}$, to reflect the situation before the introduction of the eGD program where a seller is either T (eTRS) or NB (non-badged seller), and (2) a three-tier mechanism ($\mathcal{M} = III$), where $B \in \{T, G, NB\}$ and G corresponds to an intermediate reputation level, which is the eGD signal in our context. A badge is awarded to a seller if she satisfies certain requirements. The requirements for the T badge are (1) historical

[9] Even though it is possible for an eTRS seller to have both the eTRS Plus badge and the eGD signal on the same listing, we later show that the signalling value of the eGD signal is small in these cases, as expected.

[10] For example, consumers do not observe the number of claims that buyers have filed against a seller.

sales level $V \geq \bar{V}$ and (2) reputation level $R \geq \bar{R}$. Meanwhile, the G badge has requirements only on a seller's reputation level, i.e., $R \geq \underline{R}$, where $\underline{R} < \bar{R}$. Formally, we define $B^{\mathcal{M}}$ as a function of historical sales and reputation levels under mechanism $\mathcal{M}$. For the two-tier mechanism, we have

$$B^{II}(R,V) = \begin{cases} T, & \text{if } R \geq \bar{R} \text{ and } V \geq \bar{V}, \\ NB, & \text{otherwise.} \end{cases}$$

For the three-tier mechanism, we have

$$B^{III}(R,V) = \begin{cases} T, & \text{if } R \geq \bar{R} \text{ and } V \geq \bar{V}, \\ G, & \text{if } \bar{R} > R \geq \underline{R}, \text{ or } R \geq \bar{R} \text{ and } V < \bar{V}, \\ NB, & \text{otherwise.} \end{cases}$$

Note that sellers who are qualified for the T badge are also qualified for the G badge. However, for illustration purpose, we use the G badge to denote the situation where a seller meets the requirements of the eGD program but not those of the eTRS program.[11]

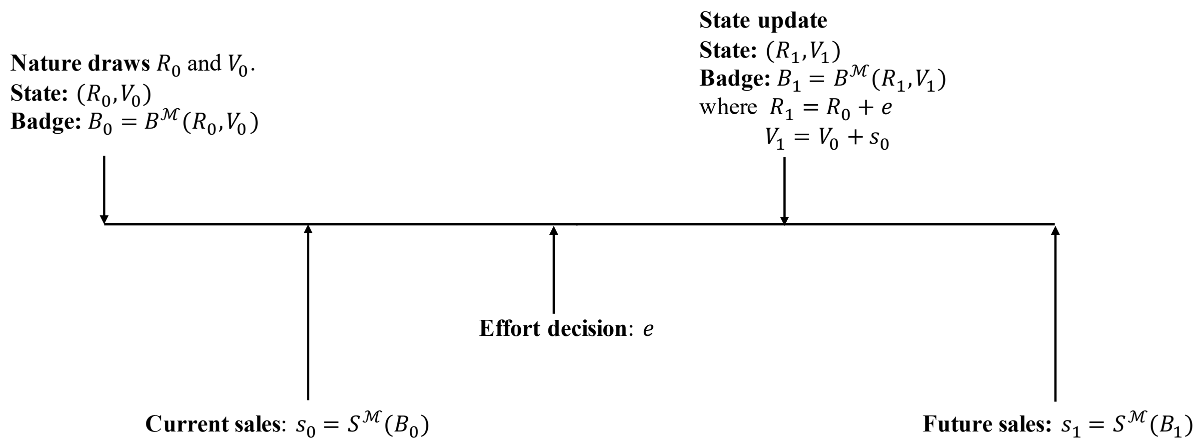
Demand: Buyers are homogeneous and observe a seller's badge. The sales that a seller with badge B can obtain under mechanism $\mathcal{M}$ are $s = S^{\mathcal{M}}(B)$, where $\mathcal{M} \in \{II, III\}$. Since the T badge has stricter requirements than the G badge, we assume that demand increases in the badge tier: i.e., $S^{II}(NB) < S^{II}(T)$ and $S^{III}(NB) < S^{III}(G) < S^{III}(T)$. Furthermore, we assume $S^{III}(G) > S^{II}(NB)$.[12]

Sequence of Events: First, nature draws a seller's initial historical sales level V_0 and reputation level R_0, which immediately determines the seller's badge status $B_0 = B^{\mathcal{M}}(R_0, V_0)$. Once the badge status is determined, the seller realizes the current sales, $s_0 = S^{\mathcal{M}}(B_0)$. The seller then decides the effort level e and incurs the effort cost. Afterwards, the seller's states in historical sales

[11] For simplicity, we assume that the badge is awarded on a seller-level instead of on a listing-level in the theoretical model.

[12] These are reasonable assumptions. For example, suppose that consumers' willingness-to-pay for the products of sellers in each tier equals the average quality of the sellers in that tier; then, the first two inequalities immediately follow. The last inequality means that NB sellers that obtain the G badge experience a price increase, which would follow if the average quality of the G sellers was higher than the average quality of the NB sellers under a two-tier system. To solve the equilibrium, note that for the marginal seller that exerts effort, her total effort cost to achieve G equals the difference in the average quality of sellers with the G badge and those without any badge. This calculation requires that the seller consider others' effort and reputation information in equilibrium.

Figure 1: Sequence of Events



Notes: The figure illustrates the sequence of events of our model.

and reputation are updated following the platform's evaluation, and specifically, we have

$$V_1 = V_0 + s_0 \text{ and } R_1 = R_0 + e.$$

Further, the seller's badge information is updated as $B_1 = B^{\mathcal{M}}(R_1, V_1)$, and her future sales are realized as $s_1 = S^{\mathcal{M}}(B_1)$.

The seller's effort e affects only her future sales s_1 through the updated reputation level R_1 and does not affect the current sales s_0. Therefore, the seller's objective to maximize the total payoff can be written as follows:

$$\max_e S^{\mathcal{M}}(B_1) - e.$$

3.1 Optimal Strategy

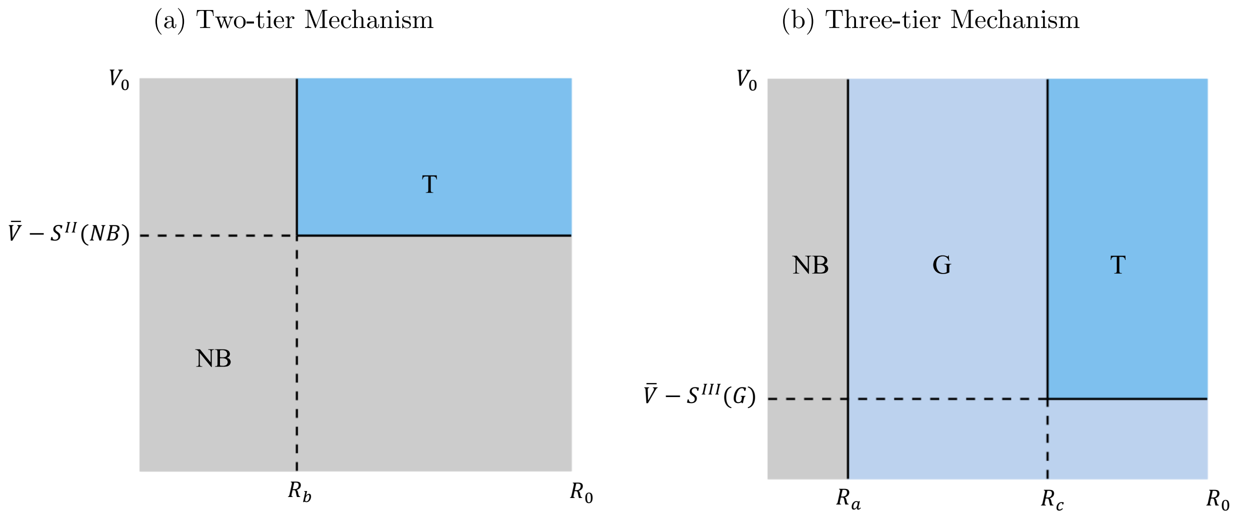
Lemma 1 characterizes the seller's optimal effort decision and her updated badge status after evaluation under both reputation mechanisms.

Lemma 1. *Under the two-tier mechanism,*

1. *If $R_0 < \bar{R} - (S^{II}(T) - S^{II}(NB))$ or $V_0 < \bar{V} - S^{II}(NB)$, the optimal effort decision is $e^* = 0$ and the updated badge status is NB.*

2. *If $R_0 \geq \bar{R} - (S^{II}(T) - S^{II}(NB))$ and $V_0 \geq \bar{V} - S^{II}(NB)$, the optimal effort decision is $e^* = \max\{0, \bar{R} - R_0\}$ and the updated badge status is T.*

Figure 2: Updated Badge under Different Reputation Mechanisms



Notes: The figure illustrates sellers' updated badge under two-tier and three-tier reputation mechanisms in a two-dimensional space. The horizontal axis represents the initial reputation level R_0, and the vertical axis represents the initial sales level V_0. We define $R_a = \underline{R} - (S^{III}(G) - S^{III}(NB))$, $R_b = \bar{R} - (S^{II}(T) - S^{II}(NB))$, and $R_c = \bar{R} - (S^{III}(T) - S^{III}(G))$.

Under the three-tier mechanism,

1. *If* $R_0 < \underline{R} - (S^{III}(G) - S^{III}(NB))$*, the optimal effort decision is* $e^* = 0$ *and the updated badge status is* NB*.*

2. *If* $R_0 \geq \underline{R} - (S^{III}(G) - S^{III}(NB))$*, and* $R_0 < \bar{R} - (S^{III}(T) - S^{III}(G))$ *or* $V_0 < \bar{V} - S^{III}(G)$*, the optimal effort decision is* $e^* = \max\{0, \underline{R} - R_0\}$ *and the updated badge status is* G*.*

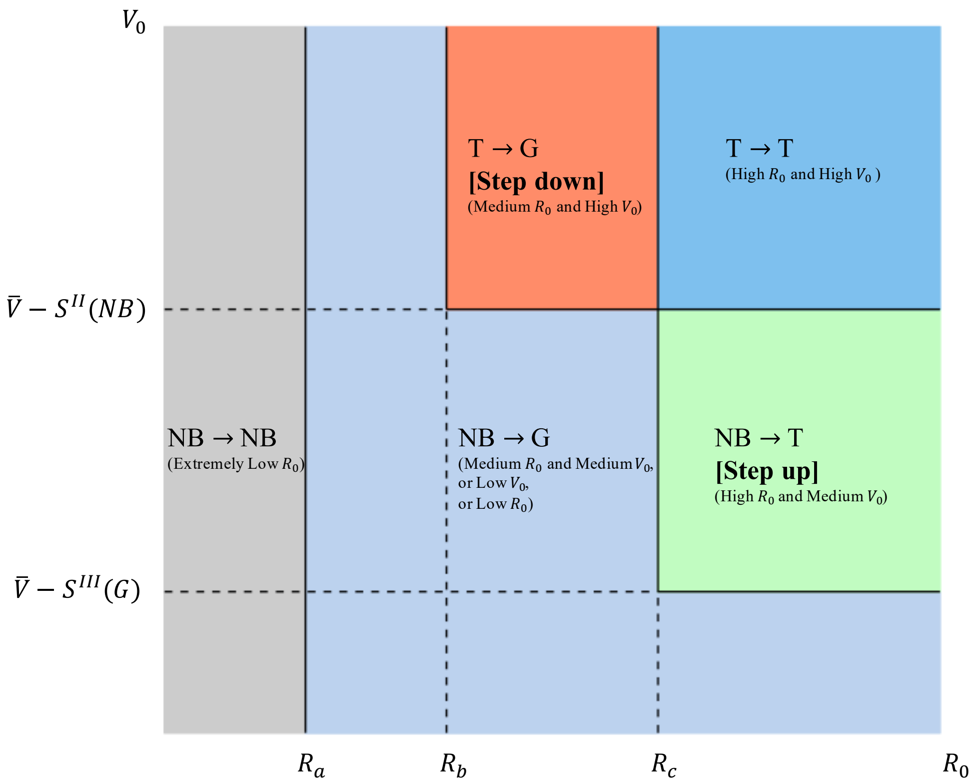
3. *If* $R_0 \geq \bar{R} - (S^{III}(T) - S^{III}(G))$ *and* $V_0 \geq \bar{V} - S^{III}(G)$*, the optimal effort decision is* $e^* = \max\{0, \bar{R} - R_0\}$ *and the updated badge status is* T*.*

All proofs in this section are provided in the appendix. As illustrated in Lemma 1, under the two-tier mechanism, sellers who lack sufficient historical sales cannot receive the T badge, but under the three-tier mechanism, the high-quality ones will receive the G badge instead. Figure 2a and 2b show how sellers' updated badge status B_1 varies with endowed R_0 and V_0.

3.2 Comparative Statics

We are interested in the comparative statics of moving from a two-tier to a three-tier mechanism. In particular, we explore how sellers' certification dynamics and quality provision change. Figure 3 illustrates the certification dynamics. With extremely low R_0, i.e., high total effort cost, sellers

Figure 3: Updated Badge Status Change with the Introduction of the G Badge



Notes: The figure illustrates sellers' updated badge change when moving from a two-tier reputation mechanism to a three-tier one. The horizontal axis represents the initial reputation level R_0, and the vertical axis represents the initial historical sales level V_0.

prefer to stay non-badged under both mechanisms. Sellers with either medium R_0 and medium V_0, or low V_0, or low R_0 would pursue the G badge under the three-tier mechanism but settle with no badge under the two-tier mechanism. Meanwhile, sellers with high V_0 and high R_0 would pursue the T badge under both mechanisms.

The total cost of reaching $\bar{R}$ (or $\underline{R}$) to get the T (or G) badge is $\max\{\bar{R} - R_0, 0\}$ (or $\max\{\underline{R} - R_0, 0\}$). We want to highlight two regions in Figure 3. The green region illustrates the certification dynamics for sellers with relatively low V_0 and high R_0. Under the two-tier mechanism, these sellers' updated badge status is NB, but under the three-tier mechanism, with relative low effort cost in addition to the current demand boost of the G badge, these sellers are able to step up and obtain the T badge after the evaluation. The red region indicates the certification dynamics for sellers with high V_0 and medium R_0. These sellers would obtain the T badge after their badge is updated under the two-tier mechanism, but under the three-tier mechanism, due to the non-negligible effort cost, they are better off when they step down and adopt the G badge.

Together with the updated badge status change, sellers' effort decisions (quality provision) also differ under different mechanisms. The sellers who step up with $R_0 < \bar{R}$ exert more effort so that their $R_1 = \bar{R}$, in order to qualify for the T badge. The sellers who step down exert less effort so that $R_1 = \underline{R}$ instead of $\bar{R}$. Proposition 1 and 2 summarize sellers' certification dynamics and corresponding effort decisions.

Proposition 1 (Step up)**.** *Sellers with a low historical sales level V_0 and a relatively high reputation level R_0 will (1) exert more effort (if $R_0 < \bar{R}$) and (2) step up from no badge to the T badge with the introduction of the G badge.*

Proposition 2 (Step down)**.** *Sellers with a high historical sales level V_0 but a relatively low reputation level R_0 will (1) exert less effort and (2) step down from the T badge to the G badge with the introduction of the G badge.*

Lastly, to understand the heterogeneous impact of a three-tier mechanism on certification dynamics and quality provision in markets with different characteristics, we study an illustrative example in Appendix B.2. We find that the impacts on both outcomes are larger in markets where a larger share of sellers do not meet the requirements on historical sales for the T badge.

In summary, our model has three predictions: (1) the eGD program can help sellers with low total effort costs and insufficient sales to become eTRS, and their overall quality provision increases; (2) eTRS sellers with relatively high effort costs may exert less effort and choose to step down from eTRS to eGD with the introduction of the eGD program; and (3) the net increase in the share of eTRS sellers and the increase in average quality provision are more pronounced in markets where the sales requirement of the certification is more binding, i.e., sellers are not eTRS due to their lack of sufficient historic sales.

4 Experiment, Empirical Strategy, and Data

Before fully rolling out the eGD program for all sellers, eBay invited a selected group of sellers to sign up for the waitlist of a pilot eGD program in the summer of 2017. These invited sellers are larger and have higher measured quality than an average eBay seller. In August 2017, $22,560$ of the invited sellers signed up on the waitlist, and eBay randomly enrolled half of them into the eGD program, while keeping the other half status quo.

This experiment allows us to obtain an unbiased estimate of the program effect by simply comparing the outcome variables across the treatment and control groups in the post-period, because the random assignment of program status removes systematic differences across the two groups. To minimize the potential impact of any difference caused by chance across the two groups and increase the precision of our treatment effect estimate, we also use the pre-period data and adopt the following difference-in-differences (DiD) specification:

$$y_{it} = \alpha Enrolled_i \times Post_t + \eta_i + \tau_t + \epsilon_{it}, \tag{1}$$

where y_{it} are outcome variables of seller i in month t (e.g., average sales price, share of claims, etc.); $Enrolled_i$ is an indicator variable which equals 1 if seller i is in the treatment group; $Post_t$ indicates whether month t is after the experiment starts; η_i and τ_t are seller and month fixed effects, respectively; and ϵ_{it} is an idiosyncratic error term. The coefficient of interest is α, which represents the complier average causal effect (CACE): the policy effect on sellers who will enroll in eGD when they are offered the opportunity to do so.

To study the policy effect on certification dynamics and quality provision for different types of sellers, we perform the following regressions:

$$y_i = \beta_0 + \beta_1 Enrolled_i + \beta_2 subgroup_i + \beta_3 Enrolled_i \times subgroup_i + \nu_i, \tag{2}$$

$$y_{it} = \beta_0 + \beta_1 Enrolled_i + \beta_2 subgroup_i + \beta_3 Enrolled_i \times subgroup_i + \mu_t + \xi_{it}, \tag{3}$$

where y_i indicates seller i's change in certification status; y_{it} indicates seller i's behavior at time t; $subgroup_i$ is a dummy variable that equals 1 if seller i belongs to the subgroup (e.g., young seller); μ_t are month fixed effects; and ξ_{it} is an error term. Also, β_1 indicates the CACE estimate for sellers not in the subgroup, and β_3 measures the additional program effect for sellers in the subgroup. We use equations 2 and 3 instead of a DiD specification such as equation 1 to avoid somewhat convoluted interpretations of triple interaction terms.

Our data includes listings from the six months before and the six months after the experiment (Feb 2017–Jan 2018), which we refer to as the pre- and post-period, respectively. We choose this time window because in February 2018, eBay selected more sellers from the waitlist to join the program. We report summary statistics and check for baseline balance between the treatment and control groups in Table 1. The statistics are obtained by first creating one observation per seller and then averaging across sellers. Column (1) reports the control group mean; column (2) shows the difference across the two groups; and column (3) reports the t-statistics of the difference. All

Table 1: Sample Statistics and Baseline Balance

	(1) Control Group Mean	(2) Difference in Treatment	(3) t-stat of the Difference
Panel A. Seller Performance			
Monthly Sales Volume	162.17	-13.25	-0.87
Sales Price	50.76	-1.36	-0.57
Monthly Number of Listings	925.11	-41.19	-0.18
Panel B. eBay Top Rated Seller (eTRS)-Related Measures			
Share of Top Rated Months	0.67	-0.005	-0.85
Share of Sellers' Sales Requirement Not Met	0.30	0.003	0.45
Share of Sellers' Quality Requirement Not Met	0.23	0.001	0.19
Share of Sellers' Shipment Requirement Not Met	0.56	-0.008	-0.93
Share of Young Sellers (Selling<1yr on eBay)	0.14	0.006	1.10
Panel C. Ratings and Feedback			
Percentage Negative Feedback	0.003	-0.0002	-1.11
Share of Low DSR: Item Description	0.003	-0.0003	-0.94
Share of Low DSR: Communication	0.002	-0.0003	-1.597
Share of Low DSR: Shipping Time	0.001	0.0002	0.998
Share of Low DSR: Shipping and Handling Charges	0.002	-0.0002	-0.921
Share of Claims	0.006	-0.0001	-0.284
Panel D. Quality Provision			
Share of Voluntary Refunds	0.006	-0.0002	-1.188
Share of Late Shipments	0.05	0.001	1.248
Share of Transactions with 30-day Returns Option	0.52	-0.003	0.341
Share of Transactions with Specified 1-day Handling	0.49	0.0006	0.109

Notes: The table reports sample statistics and baseline balance. One observation is one seller. Each row represents a regression of a variable on a constant and an enrolled dummy, which is 1 if a seller is in the treatment group. Column (1) reports the constant; columns (2) and (3) report the estimated coefficient of the dummy and its t-statistics, respectively. In panels A, C, and D, the variables are defined based on the six months before the experiment. In panel B, the variable is defined based on the month before the experiment. See the main text for variable definitions.

variables in this table are based on listings and transactions in the pre-period.

There are four panels in Table 1. Panel A reports seller performance variables. The monthly sales volume of sellers in our sample in the pre-period is 162 units, with an average sales price of around \$50. Their monthly average number of listings is 925, which indicates that these sellers have a 17.5% conversion rate. As discussed previously, the sellers in this sample sell more units than an average seller on eBay and have a higher conversion rate.[13] In columns (2) and (3), we see that even though the averages are different across the treatment and control groups, the differences are not statistically significant at the 10% level.

Panel B reports variables that are related to sellers' eligibility for the eTRS program. The average share of months in which a seller is eTRS is 67% in our six-month pre-period. Recall that the eTRS status has requirements on historical sales, transaction quality, and shipping. The proportion of sellers that do not meet these three requirements is 30%, 23%, and 56%, respectively. Given that the sellers in our sample are larger and have higher quality than an average seller, these three variables have higher means for all sellers. Additionally, the share of young sellers, namely those who have been selling on eBay for less than one year, is 14%. Lastly, t-statistics reported in column (3) are consistent with the baseline balance.

In panel C, we focus on variables about ratings and feedback. The first measure, percentage negative feedback, is calculated as the ratio of the number of negative feedback ratings received divided by the total number of transactions. The next four measures reference eBay's 5-point scale detailed seller ratings (DSR) in the following order: item description, communication, shipping time, and shipping and handling charges. Specifically, the share of low DSR is defined as the ratio of DSRs of 1 or 2 points received divided by the total number of transactions. Claims are buyers' complaints to eBay when they do not receive the item or the item is not as described. According to the feedback measures in this panel, the average share of bad buyer experiences is 0.1%–0.6%.

Panel D reports the variables related to sellers' quality provision. The first measure concerns voluntary refunds, which are partial or full refunds that sellers offer to buyers without being required to do so by eBay. Voluntary refunds could happen when sellers accept buyer returns or when they want to avoid consumers' escalating claims to eBay. The second measure is the share of late shipments, which is used by both the eTRS and eGD programs. A transaction is counted as a late shipment case if the tracking shows that the item is received by the post office later than the specified handling date after payment has been cleared. Lastly, we report the share of transactions

[13] We cannot report summary statistics of an average seller on eBay due to our non-disclosure agreement with eBay.

that provide 30-day returns and 1-day handling, which determines the eTRS Plus badge and the eGD signal on a given listing. These shares are around 50%, consistent with a high share of eTRS sellers in our sample. The results in column (3) are again consistent with the baseline balance at the 10% level.

5 Results

We start by evaluating the overall policy impact on seller performance and other variables in Section 5.1. We then focus on studying the policy effect certification dynamics and quality provision separately for non-eTRS sellers in Section 5.2 and for eTRS sellers in Section 5.3. Next, we study the mechanism behind the policy effect and discuss potential alternative explanations in Section 5.4. We then provide an analysis of consumer welfare in Section 5.5. Lastly, we conclude this section by analyzing aggregate-level policy effects in Section 5.6.

5.1 Overall Policy Impact on Seller Performance

According to our theory, the eGD signal allows high-quality young sellers to separate themselves from low-quality ones. Therefore, enrolling in the eGD program should affect sellers' demand, their likelihood of being certified, and their quality provision. To study this, we run regression (1) and report estimated α in Table 2. Each observation used in the regression is a seller–month level statistic, and we have an imbalanced panel in that not every seller has sales in each of the 12 months.[14] The outcome variables include six measures: (1) sales volume, (2) average sales price, (3) number of listings, (4) share of Top Rated months, (5) share of claims, and (6) share of late shipments. We cluster standard errors at the seller level and account for heteroskedasticity in this table and throughout the paper when there are multiple observations per seller.

Panel A presents the results for all sellers. Columns (1) and (2) show that enrolling in the eGD program increases the monthly sales volume by 3.6% and the average sales price by 1.7%. These increases are consistent with a demand increase for eGD sellers, as consumers appreciate fast shipping. Column (3) shows that the number of listings also increases by 3.4%, suggesting little change in sellers' conversion rate. Column (4) shows that enrolled sellers are badged for 3.1% more months compared to non-enrolled sellers in the post-period, indicating that the eGD program helps sellers climb the reputation ladder. In column (5), we see no significant difference in the share of

[14]The results remain qualitatively unchanged when we fill in zeros when appropriate.

Table 2: Overall Impact

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. All Sellers						
	log(Sales Volume)	log(Price)	log(Number of Listings)	Share of Top Rated Mn.	Share of Claims	Share of Late Shipment
Enrolled×Post	0.036***	0.017**	0.034**	0.031***	-0.00001	-0.007***
	(0.012)	(0.007)	(0.015)	(0.005)	(0.0003)	(0.001)
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	168,958	168,958	168,958	168,958	168,958	168,958
Adjusted R^2	0.840	0.793	0.797	0.618	0.147	0.191
Panel B. Sample: Sellers Who Were Never Top Rated (Non-eTRS)						
	log(Sales Volume)	log(Price)	log(Number of Listings)	Share of Top Rated Mn.	Share of Claims	Share of Late Shipment
Enrolled×Post	0.083***	0.099***	0.060*	0.071***	-0.002*	-0.037***
	(0.030)	(0.017)	(0.037)	(0.007)	(0.001)	(0.003)
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	33,505	33,505	33,505	33,505	33,505	33,505
Adjusted R^2	0.809	0.767	0.782	0.206	0.172	0.275
Panel C. Sample: Sellers Who Were Top Rated (eTRS)						
	log(Sales Volume)	log(Price)	log(Number of Listings)	Share of Top Rated Mn.	Share of Claims	Share of Late Shipment
Enrolled×Post	0.025*	-0.002	0.027*	0.025***	0.0004	0.00004
	(0.013)	(0.007)	(0.016)	(0.006)	(0.0003)	(0.001)
Seller FE	Yes	Yes	Yes	Yes	Yes	Yes
Month FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	134,824	134,824	134,824	134,824	134,824	134,824
Adjusted R^2	0.847	0.800	0.800	0.374	0.122	0.149

Notes: The table reports treatment effect estimates on measures of seller performance, certification status, and quality provision using equation 1. The regressions are at the seller-month level and are based on the six months before and after the experiment. Panel A reports the results for all sellers. Panel B reports the results for non-eTRS, namely sellers who were never badged in the six months before the experiment. Panel C reports the results for eTRS, namely sellers who were badged for at least one month in the six months before the experiment. Standard errors are clustered at the seller level and account for heteroskedasticity.

*** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

claims between the treatment and control groups. Column (6) shows that enrolling in the eGD reduces the share of late shipments by 0.7%, indicating that sellers exert more effort in terms of fast shipping after enrolling in the program.

In panels B and C, we investigate the impact of introducing the eGD program on eTRS and non-eTRS sellers. We define non-eTRS sellers as those who were never qualified for eTRS in the pre-period, and eTRS sellers as those who were eTRS for at least one month in that period. From column (1) in the two panels we see that the sales volume increases for both types of sellers, and the percentage increase is three times larger for non-certified sellers (8.3%) than for certified ones (2.5%). Column (2) across panels shows a similar result: the price increase for non-eTRS sellers is large, whereas price essentially stays the same for eTRS sellers. Column (3) shows that the number of listings increases for both non-eTRS and eTRS sellers, and that the conversion rate increases for non-eTRS sellers, but not for eTRS sellers. These estimates suggest that the demand increase from enrolling in the eGD program is smaller for eTRS sellers, because of the small signalling value of the eGD conditional on having the eTRS Plus badge. Column (4) shows that both non-eTRS and eTRS sellers are badged for more months in the post-period, although the increase for the former group is more than twice as large. Lastly, we study the changes in quality provision in columns (5) and (6). Specifically, the outcome variable in column (6) is the share of late shipments. This information is based on the tracking information rather than user-generated feedback, and therefore is a direct measure of a seller's effort. Besides predicting quality improvement in shipping, our theoretical framework also predicts that enrolled sellers are incentivised to exert effort in general. We do not directly observe sellers' effort in other dimensions, and therefore we use buyer claims as a proxy for it.[15] Both columns show that non-eTRS sellers increase their quality provision after enrolling in the eGD program.

5.2 Effects on Non-eTRS Sellers

Proposition 1 of the model predicts that sellers who do not meet the certification's requirements on historical sales are more likely to step up and become certified and provide higher quality. Because young sellers are more likely to be sales-constrained due to their lack of sales history, we study the program effects on both young sellers and established sellers, and report the results in panel A of

[15] We use buyer claims as a proxy because eBay critically relies on this quality measure for seller performance, as can be seen from the requirements for eTRS sellers. The result on the changes in buyer feedback is qualitatively the same if we use the share of low negative feedback ratings or share of low DSRs as the outcome variable. For example, non-eTRS sellers see a decrease of 0.001 in the share of negative feedback ratings and a decrease of 0.0007 in the average share of low DSRs, both of which are statistically significant at the 5% level.

Table 3. Next, we study the heterogeneous program effects depending on whether sellers meet the certification's requirements on past sales, and report the results in panel B. Note that in panels A2 and B2, we test if the program effects on the subgroup are statistically different from zero.

We start by studying the changes in certification dynamics using equation 2 and report the results in panel A, column (1). The sample in this analysis contains all non-eTRS sellers in the pre-period, and one observation is one seller in the regression. The outcome variable is a dummy for whether a seller becomes eTRS in the post-period, i.e., is certified for at least one month. The subgroup variable 'Young' is a dummy for whether a seller has been selling on eBay for less than a year. We find that enrolling in the eGD program increases established sellers' chance of becoming certified by 24.9%, and for young sellers, this step-up effect is 9.4% higher. The larger effect for young sellers is consistent with our prediction in Proposition 1.

Besides certification dynamics, our model also predicts that young sellers increase their quality provision because the introduction of the second signal increases their chance of getting certified in the future. To study this, we use equation 3, and the sample used in columns (2)–(6) contains the behavior of non-eTRS sellers (seller type defined based on the pre-period) in each month in the post-period, and each observation in the regression is a seller–month average. In column (2), we study sellers' likelihood of offering 30-day returns, which is a requirement for the eTRS Plus badge. Enrolled sellers increase this offering by 3.6% in the post-period, and young sellers see an additional increase of 1.8%. In column (3), we study sellers' likelihood of offering voluntary refunds. As mentioned before, a seller may want to prevent buyer escalations of claims to eBay–a key requirement of the eTRS certification–by offering voluntary refunds even if not required by eBay. We see that young sellers increase voluntary refunds by 0.1%, while the increase for established sellers is statistically insignificant. Next, column (4) shows that the claim rate of young sellers decreases by 0.5% compared to that of the control group. High consumer satisfaction suggests that young sellers improve their quality provision after enrolling in the program. Lastly, columns (5) and (6) show that both young and established sellers improve their handling time and lower their late shipment rate by similar magnitudes. This is because high shipping performance is needed from all sellers to benefit from the second signal, regardless of whether they become eTRS sellers or not.

In panel B of Table 3, we examine the heterogeneous program effects by whether a seller meets the sales requirement of the certification—a key parameter in our model that determines the effect on certification dynamics. Column (1) shows that enrolling in the eGD program increases the chance

Table 3: Non-eTRS Sellers: Seller Dynamics and Quality Provision

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Subgroup: Young Sellers						
	Gain Badge	Share 30-day Returns	Share Vol. Refunds	Share Claims	Share 1-day Handling	Share Late Shipments
Enrolled	0.249***	0.036***	0.0002	0.001	0.092***	-0.029***
	(0.020)	(0.010)	(0.0002)	(0.001)	(0.013)	(0.003)
Subgroup	-0.019**	-0.022***	-0.001***	0.002	-0.027***	0.006*
	(0.009)	(0.006)	(0.0002)	(0.001)	(0.010)	(0.003)
Interaction	0.094**	0.018*	0.001**	-0.006***	-0.005	0.004
	(0.038)	(0.010)	(0.0004)	(0.002)	(0.015)	(0.005)
Month FE		Yes	Yes	Yes	Yes	Yes
Observations	2,667	15,358	15,362	15,366	15,359	15,360
Adjusted R^2	0.094	0.006	0.004	0.001	0.016	0.015
Panel A2.Sample: Sellers in the Subgroup						
Enrolled	0.343***	0.053***	0.001***	-0.005***	0.087***	-0.025***
	(0.032)	(0.012)	(0.0003)	(0.001)	(0.017)	(0.005)
Month FE		Yes	Yes	Yes	Yes	Yes
Panel B. Subgroup: Whether Sellers Fail to Meet the Sales Requirements of Certification						
	Gain Badge	Share 30-day Returns	Share Vol. Refunds	Share Claims	Share 1-day Handling	Share Late Shipments
Enrolled	0.238***	0.027**	0.0002	0.0003	0.102***	-0.032***
	(0.024)	(0.011)	(0.0002)	(0.001)	(0.016)	(0.003)
Subgroup	0.031	-0.028***	-0.001***	-0.001	0.0003	0.002
	(0.023)	(0.008)	(0.0002)	(0.001)	(0.013)	(0.003)
Interaction	0.070**	0.029**	0.0005	-0.002	-0.023	0.008*
	(0.033)	(0.012)	(0.0003)	(0.002)	(0.017)	(0.004)
Month FE		Yes	Yes	Yes	Yes	Yes
Observations	2,666	15,358	15,364	15,365	15,360	15,359
Adjusted R^2	0.099	0.007	0.002	0.0003	0.015	0.015
Panel B2. Sample: Sellers in the Subgroup						
Enrolled	0.308***	0.056***	0.001**	-0.002*	0.080***	-0.023***
	(0.024)	(0.011)	(0.0003)	(0.001)	(0.015)	(0.004)
Month FE		Yes	Yes	Yes	Yes	Yes

Notes: Panel A reports the heterogeneous program effects on young vs. established sellers. Panel B reports the effects depending on whether sellers meet the certification's sales requirement. Column (1) reports the regression results of whether a non-eTRS seller becomes eTRS (i.e., badged for at least one month) in the post-period on a constant, the enrolled dummy, subgroup variables, and their interactions using equation 2. One observation is a seller. Columns (2)–(6) report the results of changes in non-eTRS sellers' behavior in the post-period on the month fixed effects, the enrolled dummy, subgroup variables, and their interactions using equation 3. One observation is a seller–month average. Panels A2 and B2 report the regression results only for sellers in the subgroups.

*** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

of becoming eTRS by 23.8% for non-eTRS sellers who meet the sales requirements. Additionally, the effect for sales-constrained sellers is 7% higher, consistent with Proposition 1. Next, columns (2) and (3) show that sales-constrained sellers increasingly offer 30-day returns (by 5.6%) and voluntary refunds (by 0.1%), and these increases are larger than those of sellers who are not sales constrained. Correspondingly, sales-constrained sellers receive 0.2% fewer claims compared to the control group. Lastly, columns (5) and (6) show that all sellers increase their offering of one-day handling and have a lower late-shipment rate after enrolling in the eGD program.

To summarize, we find that compared to established sellers, young sellers and sales-constrained sellers are more likely to step up after enrolling in the eGD program, and they increase their quality provision more. These results are consistent with Proposition 1 of our model.

5.3 Effects on eTRS Sellers

Proposition 2 of our model predicts that eTRS sellers with higher effort costs of maintaining their certification are more likely to step down and exert less effort after enrolling in the eGD program. Since we do not observe the costs of maintaining the certification, we use two proxies for them. Our first proxy is the share of voluntary refunds out of all transactions. When a buyer files a claim against a seller, the seller could immediately resolve the case with the buyer (normally by offering a partial or full refund). Alternatively, the seller could wait for the buyer to escalate the case to eBay, and the seller will have to offer refunds only if she is considered at fault by eBay. The cost of letting buyers escalate a claim is that this will count towards the number of "claims without seller resolution" if the seller is found at fault. eBay takes this number into account when evaluating the seller for the eTRS certification. Therefore, a larger share of voluntary refunds can be thought of as sellers exerting higher effort to maintain the eTRS certification, which induces higher effort costs. Our second proxy for effort costs is the share of late shipments. While all eTRS sellers meet the certification's requirement on shipping by definition, some sellers' measured shipping qualities are closer to the eTRS threshold than others. We assume that these "marginal" sellers have higher costs of providing fast shipping, which is another source of effort costs.[16]

We now study the heterogeneous program effects for eTRS sellers by proxies of effort costs, and report the results in panels A and B of Table 4. To study the changes in eTRS sellers' certification dynamics, we estimate equation 2 and report the results in column 1. The sample used in this

[16] Neither cost measure is perfect as both depend on seller behavior. An ideal cost measure would reflect seller type, which we unfortunately do not observe.

Table 4: eTRS Sellers: Seller Dynamics and Quality Provision

	(1)	(2)	(3)	(4)	(5)	(6)
Panel A. Subgroup: Whether Sellers' Share of Voluntary Refunds Is in Top Decile						
	Lose Badge	Share 30-day Returns	Share Vol. Refunds	Share Claims	Share 1-day Handling	Share Late Shipment
Enrolled	-0.022***	0.004	0.00004	0.0004	0.009	-0.00003
	(0.007)	(0.004)	(0.0001)	(0.0003)	(0.008)	(0.001)
Subgroup	-0.029*	-0.020**	0.012***	0.007***	0.028	-0.002
	(0.016)	(0.010)	(0.0004)	(0.001)	(0.018)	(0.002)
Interaction	0.059***	-0.015	-0.001	0.004***	0.032	-0.002
	(0.022)	(0.012)	(0.001)	(0.002)	(0.025)	(0.003)
Observations	13,120	63,116	63,116	63,116	63,116	63,115
Adjusted R^2	0.001	0.002	0.122	0.007	0.001	0.008
Panel A2. Sample: Sellers in the Subgroup						
Enrolled	0.036*	-0.011	-0.001	0.004***	0.041*	-0.002
	(0.021)	(0.012)	(0.001)	(0.001)	(0.023)	(0.003)
	(1)	(2)	(3)	(4)	(5)	(6)
Panel B. Subgroup: Whether Sellers' Share of Late Shipments Is in Top Decile						
	Lose Badge	Share 30-day Returns	Share Vol. Refunds	Share Claims	Share 1-day Handling	Share Late Shipment
Enrolled	-0.025***	0.007*	0.00004	0.001**	0.009	-0.00001
	(0.007)	(0.004)	(0.0001)	(0.0002)	(0.008)	(0.001)
Subgroup	0.197***	-0.012	-0.001**	0.004***	0.004	0.124***
	(0.016)	(0.010)	(0.0003)	(0.001)	(0.019)	(0.005)
Interaction	0.097***	-0.055***	-0.001**	0.004**	0.039	0.001
	(0.022)	(0.013)	(0.0004)	(0.002)	(0.026)	(0.007)
Observations	13,120	63,116	63,115	63,116	63,115	63,116
Adjusted R^2	0.038	0.004	0.001	0.003	0.001	0.124
Panel B2. Sample: Sellers in the Subgroup						
Enrolled	0.072**	-0.048***	-0.001**	0.004**	0.049**	0.001
	(0.028)	(0.013)	(0.0004)	(0.002)	(0.025)	(0.007)

Notes: Panels A and B report the heterogeneous program effects on eTRS sellers with different effort costs based on two measures. Column (1) reports the regression results of whether an eTRS seller becomes non-eTRS (i.e., badged for half or fewer months) in the post-period on a constant, the enrolled dummy, subgroup variables, and their interactions using equation 2. One observation is a seller. Columns (2)–(6) report the results of changes in eTRS sellers' behavior in the post-period on the month fixed effects, the enrolled dummy, subgroup variables, and their interactions using equation 3. One observation is a seller–month average. Panels A2 and B2 report the regression results only for sellers in the subgroups.

*** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

analysis contains all eTRS sellers, namely sellers who were certified for at least one month in the pre-period, and one observation in the regression is one seller. The outcome variable is a dummy variable for whether the eTRS seller steps down in the post-period, where the dummy equals 1 if a seller is certified for half or fewer months in the post-period compared to the pre-period. The subgroup variable in panel A is whether a seller's share of voluntary refunds is in the top decile among all eTRS sellers in our sample. In panel B, we use the other effort cost proxy to define subgroups, which is whether a seller's share of late shipments is in the top decile.[17]

The results in column (1) of panel A show that while the sellers with lower effort costs measured in terms of voluntary refunds are 2.2% less likely to lose their certification, the sellers with effort cost in the top decile are 3.6% (estimate marginally significant) more likely to step down. Panel B shows a similar pattern, and, in particular, the sellers with higher effort costs measured by share of late shipments are 7.2% more likely to step down and lose their badge. These results are consistent with Proposition 2 that sellers with high effort costs are better off adopting the new eGD signal and not incurring the effort cost of staying eTRS. On the other hand, the sellers with lower marginal costs are more likely to maintain their eTRS status in the post-period, consistent with the results in column (4) of Table 2.

When eTRS sellers step down, they should decrease their quality provision because they no longer receive the eTRS Plus badge regardless of their effort. We use equation 3 to study the changes in eTRS sellers' quality provision. Similar to our analyses for non-eTRS sellers, the sample used in columns (2)–(6) contains the behavior of eTRS sellers (seller type defined based on the pre-period) in each month in the post-period, and each observation in the regression is a seller–month average. Column (2) reveals that sellers with high effort costs measured in share of late shipments are 4.8% less likely to offer the 30-day return option in their transactions. This is because after they lose their eTRS status, offering 30-day returns no longer earns them an eTRS Plus badge on their listings, and therefore the incentive to offer 30-day returns is smaller. In the same column, we see that the sellers with low effort costs either do not change or marginally increase their offering of 30-day returns. Next, column (3) shows that these sellers are also 0.1% less likely to offer voluntary refunds, and we do not observe such changes for sellers with low effort costs. Consistent with lower quality provision, column (4) shows that the claim rate for sellers with high effort costs increases by 0.4% using either measure, and it either does not change or increases by 0.1% for sellers with

[17] We have also split the sellers into top quartile and bottom three quartiles, and replicated the qualitative findings in this table.

low effort costs. Lastly, according to columns (5) and (6), even though sellers with high effort costs decrease their quality provision in general, they increase their offering of one-day handling in their transactions, whereas their share of late shipments barely changes. This is because providing high shipping performance allows them to benefit from the new eGD signal.

To conclude, in this section we find that the eTRS sellers whose effort costs, as proxied in two ways, are higher are more likely to step down after enrolling in the eGD program, and they shirk in aspects that do not affect their eligibility for the eGD signal. We do not observe these changes in behavior from the eTRS sellers with low effort costs. These results are consistent with Proposition 2 of our model.

5.4 Mechanism and Alternative Explanations

The results presented so far are consistent with the signalling mechanism: the introduction of a second signal increases the demand of sales-constrained sellers with high shipping quality, therefore increasing their chance of getting the certification and their incentive for quality provision. On the other hand, certified sellers with high effort costs find the certification less attractive and therefore step down to adopt the new signal instead.

Could our results be driven by alternative mechanisms? We consider three potential candidates. First, perhaps the eGD has no signalling value, but sellers wrongly believe otherwise and increase effort, and the results are entirely driven by the increased effort. Second, enrolling in the eGD could have search benefits, because buyers can filter out listings with the eGD option or eBay can prioritize eGD listings on the search results page. Third, buyers may regard eGD as insurance against late delivery because they can get a $5 voucher towards future purchases or return the item for free if the item arrives later than the guaranteed delivered date.

To mitigate the first concern, we estimate consumer demand for the eGD signal. For this analysis, we cannot simply regress outcome variables on whether a listing has the eGD signal because the two groups of listings may differ in unobservable ways other than the signal. For example, maybe items with lower shipping weights are more popular on eBay and sellers are more likely to handle these items fast. In this case, we cannot interpret the correlation between sales outcome and the eGD signal as causal. Additionally, sellers can change their unobserved effort level over time, which correlates both the adoption of the eGD and sales. Therefore, it is important to control for seller and product attributes, and product listing time when estimating the demand for the eGD signal.

To do so, we exploit the within-seller-product-listing week-triplet variation of the eGD signal across listings. Recall that a listing has the eGD signal only if an enrolled seller offers one-day returns and provides the zip code of the item's location. We therefore take the matching approach à la Elfenbein et al. (2012) and Einav et al. (2015), and match listings based on seller ID, listing title, listing subtitle, leaf category ID on eBay, posted price, and week that the listing starts. Within each matched set, we compare the sales probability for listings which have the eGD signal and that for listings which do not. The assumption for this approach to identify demand is that conditional on the set of matching variables, whether or not a listing has the eGD signal is random, which according to these authors, corresponds to instances where sellers experiment with sales parameters (the eGD feature in our case). The econometric specification is as follows:

$$Success_{ij} = \rho \; eGD_{ij} + \phi_j + \iota_{ij},$$

where $Success_{ij}$ is a binary variable for whether the listing i in matched set j gets sold, eGD_{ij} is a dummy variable for whether listing i in matched set j has the eGD signal, ϕ_j is a fixed effect for the set of matched listings, and ι_{ij} is an error term. The coefficient ρ captures consumers' preferences for eGD, ceteris paribus. The results from this analysis are not directly comparable with our seller-level analyses in the previous sections due to a selected sample based on matched listings. The focus here is to exploit the rich listing-level data and obtain a finely matched sample to better identify the mechanism.

The results are reported in panel A in Table 5. Column (1) shows that the eGD signal increases the sales probability by 1.8%. In column (2), we include the interaction term of the eGD signal and the eTRS Plus badge in the regression. We find that while the eGD signal increases the sales probability of the listings without the eTRS Plus badge by 3.1%, the effect is an order of magnitude smaller, i.e., 0.3%, for the listings with the eTRS Plus badge. These results indicate that the eGD signal has a sizable signalling value for non-eTRS sellers, but the signalling value is small if a listing already has the eTRS certification.

Next, to study the search benefit of the eGD signal, we control for the logarithm of the number of impressions on the search result pages that a listing receives before the item gets sold or until the listing ends, in columns (3) and (4). The estimated benefit of the eGD signal becomes slightly smaller but is still positive and highly significant. This indicates that better search ranking due to the eGD signal partially explains the increase in demand, but is not the main mechanism at play.

Table 5: Mechanisms

	(1)	(2)	(3)	(4)
Panel A. Matched Listings – Demand for eGD				
	Success (0/1)	Success (0/1)	Success (0/1)	Success (0/1)
eGD	0.018***	0.031***	0.017***	0.028***
	(0.002)	(0.004)	(0.003)	(0.004)
eGD × eTRS		-0.028***		-0.026***
		(0.005)		(0.005)
log(Impressions)			0.009***	0.009***
			(0.002)	(0.002)
Matched Listing FE	Yes	Yes	Yes	Yes
Observations	80,209	80,209	80,163	80,163
Adjusted R^2	0.515	0.516	0.518	0.518
Panel B. Buyers' Use of Insurance				
	Pct. Returns	Pct. Returns Cndtn on Late Delivery	Pct. Vouchers	Pct. Vouchers Cndtn on Late Delivery
Mean	0.0008%	0.0139%	1.04%	17.89%

Notes: Panel A reports the regression results of whether a listing sells on the dummy for the eGD signal, its interaction with the dummy for eTRS, the logarithm of number of times that a listing appears in search results, and matched listing FE in the following six months. Standard errors are clustered at the matched listing level. The matching is based on seller ID, listing title, listing subtitle, leaf category, listing start week, and price. One observation is a listing. Panel B reports the average returns and the rates of claimed vouchers based on either all eGD transactions (columns 1 and 3) or eGD transactions conditional on late delivery (columns 2 and 4).

*** indicates significance at p $\leq$ 0.01; ** p $\leq$ 0.05; * p $\leq$ 0.10.

Lastly, we look into the potential insurance value of the eGD signal. We find that among the transactions with the eGD signal, only 0.0008% of them are returned. Conditional on late delivery, this measure becomes 1.39%. This finding indicates that buyers often do not return an item that offers eGD when it arrives late, perhaps because dropping off a return to the post office is costly. Additionally, buyers request the $5 voucher in 1.04% of transactions with the eGD signal. This number increases to 18% conditional on delayed deliveries. Given the magnitudes of the cases, insurance value does not seem to be the primary mechanism behind our results.

5.5 Consumer Welfare

Our analyses so far have focused on sellers, and have shown that both their sales and price increase as a result of the new eGD signal. Therefore, seller surplus likely increases unless their cost of offering one-day handling is very large. The changes in consumer welfare, however, are more ambiguous. A large literature on product differentiation argues that while consumers can benefit from a larger selection of products, product differentiation can soften price competition and reduce consumer welfare.

Formally estimating consumer welfare requires a structural model of consumer demand with additional assumptions. Instead, we perform a simple estimation on the changes in consumer welfare by exploiting auctions on eBay. Since eBay's auction format is similar to a second-price auction, a bidder's dominant strategy is to bid her valuation. Therefore, we proxy for the winning bidder's consumer surplus by the difference between the winning bid, which we observe from eBay's internal data, and the sales price she pays. We use the following equation to estimate the changes in consumer welfare:

$$log(Y_{ijt}) = \gamma Enrolled_i \times Post_t + \eta_j + \tau_t + \epsilon_{ijt}, \tag{4}$$

where $log(Y_{ijt})$ is the logged consumer surplus in transaction i of product j in month t, where Y_{ijt} equals the winner bidder's bid minus the sales price; $Enrolled_i$ indicates whether transaction i is from an enrolled seller from the experiment; η_j and τ_t are product ID and month fixed effects, respectively; and ϵ_{ijt} is idiosyncratic errors. In this analysis, we restrict our sample to products with a product ID defined by eBay, which are eBay's finest category of items. For example, an iPhone 11, black, 256GB, unlocked, has a unique product ID that is different from that of another version of iPhone. Even though this sample restriction leaves us with only standardized products in the analyses, it allows us to compare the changes in consumer surplus of identical items.

We estimate equation 4 using auctions from the sample in the six months before and six month after the experiment. For non-eTRS sellers, the estimated γ is 0.097 and is statistically significant at the 1% level, indicating that consumer welfare increases by 9.7% after sellers enroll in the eGD program. For eTRS sellers, this estimate is 0.052 and is also statistically significant at the 1% level. These results suggest that consumer welfare increases after eBay's introduction of the second signalling opportunity.

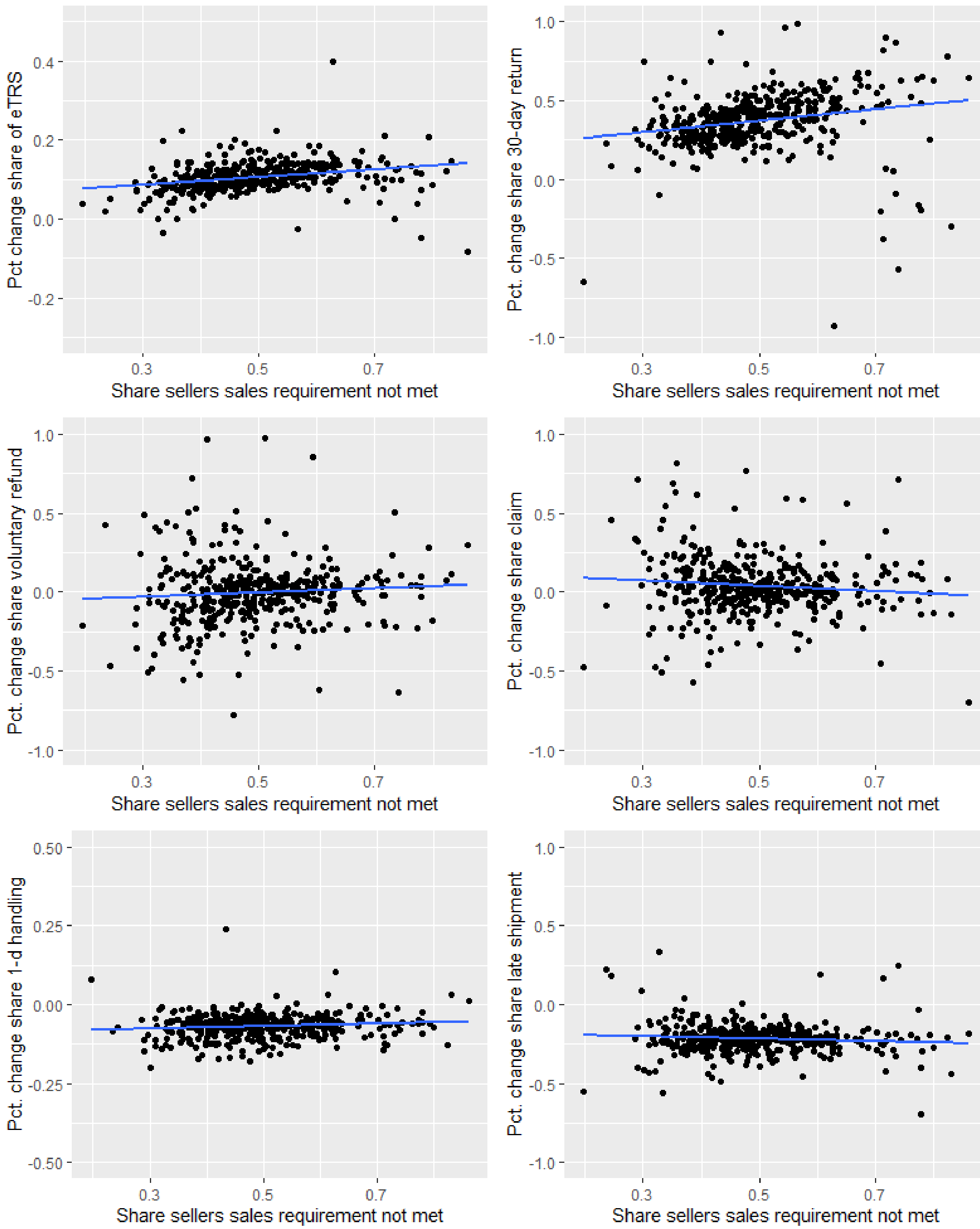
5.6 Aggregate-level Analyses

When designing reputation systems, market regulators may wonder if having an additional less history-dependent and less demanding signal is always better than having one certification with strict requirements. Also, does the answer depend on the underlying market characteristics? To answer these questions, we perform market-level analyses. While aggregate correlations can be affected by unobservable factors, this information can nevertheless be useful when guided by our theoretical framework and combined with the experimental results.

Our model predicts that stepping up should happen more frequently in markets where the sales requirement of certification is more binding, i.e., markets where more sellers struggle to satisfy the sales requirements. To test this, we use eBay's more than 400 subcategories of products to define the markets, and use the full rollout of the eGD program in May 2018 as a natural experiment. We plot the changes in outcomes and the bindingness measures in different markets in Figure 4. We measure the bindingness in each market using the share of sellers who do not meet the sales requirement of the certification. The outcome variables are the percentage changes in outcome variables in the six months before and the six months after the full rollout of the program, and the bindingness measures are based on the six months before the full rollout. All the outcome variables are obtained by first taking averages within each seller and then taking another average across sellers in the market.

In the first sub-figure, we study how certification dynamics differ across markets. We find that the changes in the share of eTRS sellers are mostly positive after the introduction of the eGD program, suggesting that the new signal is associated with the step-up effect in most markets. Additionally, we find a positive correlation between the bindingness of the sales requirements in a market and the percentage change in the share of eTRS sellers. This is consistent with our model's prediction that the step-up effect due to a second signal is larger in markets where sellers are more

Figure 4: Market-level Effects



Notes: The figure plots the average share of certified sellers and seller behavior in markets on eBay as a function of the bindingness of the certification's sales requirements, fitted with a regression line. The regression coefficients and sample sizes are reported in Table 6. The variable definitions are described in the main text.

Table 6: Market-level Effects

	(1)	(2)	(3)	(4)	(5)	(6)
	ΔShare eTRS	ΔShare 30-day Returns	ΔShare Vol. Refunds	ΔShare Claims	ΔShare 1-day Handling	ΔShare Late Shipment
Share Sellers' Sales	0.098***	0.364***	0.134	-0.167**	0.040**	-0.079*
Req. Not Met	(0.017)	(0.081)	(0.089)	(0.081)	(0.017)	(0.043)
Observations	415	415	415	415	415	415
Adjusted R^2	0.076	0.045	0.003	0.008	0.010	0.006

Notes: The table reports the regression results of certification dynamics and seller behavior on the bindingness of the certification's sales requirements in each market. One observation is a market (a subcategory on eBay). All market-level variables are calculated by first computing the average at seller level and then at market level across sellers. Outcome variables are the percentage differences of variables in the six months before and after the full rollout of the eGD program. The bindingness measure is based on the six months before the full rollout of the program.

*** indicates significance at $p \leq 0.01$; ** $p \leq 0.05$; * $p \leq 0.10$.

sales constrained or, equivalently, the sales requirements are more binding.

In the other sub-figures, we investigate how quality provision changes across markets. The measures of quality provision are the same as those used in the previous analyses, including share of 30-day returns, share of voluntary refunds, share of claims, share of 1-day handling, and share of late shipments. For all five measures, we observe positive correlations between the bindingness of the sales requirements in a market and the improvement in quality provision in that market. This positive correlation is consistent with our model's prediction.

To further test whether the correlations in the graphs are statistically significant, we run the following regression:

$$y_m = \kappa_0 + \kappa_1 x_m + \omega_m,$$

where y_m is the percentage change in the outcome variable in market m, x_m measures market m's bindingness of the sales requirements, and ω_m is an error term.

Table 6 reports the results, and each column corresponds to a sub-figure in Figure 4. Column (1) shows a statistically significant correlation between the bindingness measure and the change in the share of eTRS sellers after the introduction of the eGD program. Specifically, in markets with 10% more sellers that do not meet the certification's sales requirement, the increase in the share of eTRS sellers is about 0.98% larger after the introduction of the second signal.

In columns (2)–(6), we explore the market-level changes in quality provision. We find that markets with higher shares of sales-constrained sellers are associated with a higher increase in quality provision as measured by the offering of 30-day returns, lower claim rates, the offering of

1-day handling, and reduced late shipments. For instance, column (2) shows that adding the second signal is associated with a 3.6% increase in the offering of 30-day returns in markets with 10% more sellers that do not meet the certification's sales requirement. Through the lens of the model, in markets where the certification's requirement on past sales is more binding for sellers, adding a second signal will encourage more sellers to exert effort and step up.

To recap, our market-level analyses show that in markets where a larger share of sellers cannot meet the certification's sales requirements, the introduction of the eGD signal is correlated with a larger increase in the share of eTRS sellers and more quality provision. These correlations are consistent with our model prediction. A logical extension of this result suggests that in markets where sellers are not sales constrained, adding a second signal may lead to a net decrease in the share of certified sellers and lower average quality if the step-up effect from young sellers is dominated by the step-down effect from established sellers.

6 Conclusion

The use of reputation and certification mechanisms is ubiquitous in markets with asymmetric information. However, this practice may lead to the cold-start problem and discourage the entry and growth of young sellers. In this paper, we study a market design change on a large e-commerce platform to illustrate a possible mitigation of this problem. Specifically, in the presence of an existing quality certification, the platform introduces a second signal that has no requirements on historical sales and focuses on a specific quality dimension. We find that the new signal increases the demand for high-quality young sellers, incentivises their quality provision, and increases their likelihood of becoming certified. On the other hand, established, certified sellers with high effort costs withdraw from the selective certification and adopt the new signal instead. Our market-level analyses suggest that markets with more binding sales requirements for certification experience larger increases in the share of certified sellers and in average quality provision.

Although in our setting, the step-up effect is largely driven by the second signal not requiring historical sales, unlike the existing certification, our logic applies to more general reputation systems that may not have requirements on past sales to begin with. Consider a common reputation system that displays the number of transactions (or the number of feedback ratings received) and an average star rating, such as that of Yelp or Amazon. Given any star rating, a Bayesian buyer is still unsure of the quality of a seller with few transactions relative to a seller with many transactions, which creates the cold-start problem for young sellers. The problem is again that this rating system gives a long-run quality signal that is intrinsically difficult for new sellers to obtain. Our results suggest that if market regulators want to mitigate the cold-start problem induced by reputation systems, they should introduce a short-run signal that focuses on a particular aspect of seller quality. For example, Upwork introduced the Rising Talent (RT) badge in addition to the Top Rated (TR) badge in recent years.[18] Unlike the TR badge, the new RT badge does not require a long history of job success with multiple clients, and focuses on more short-term aspects, such as passing a screening test by Upwork, completing projects on time, maintaining a detailed profile, and keeping the availability status up to date. Based on our findings, the introduction of the RT badge should encourage high-quality providers to exert effort and increase their chance of obtaining the TR badge.[19]

Moreover, our results suggest that market regulators should also realize the pros and cons of adding a second signal to a reputation system. On the one hand, being able to obtain a quality signal that is less history-dependent and less demanding encourages the entry and participation of high-quality sellers, leading to an increase in average quality provision from them. On the other hand, established sellers may choose to opt for the less demanding signal instead, leading to a decrease in quality. The overall impact of a second signal on the number of certified sellers and average quality provision depends on the relative magnitudes of the step-up and step-down effects, which in turn depend on the bindingness of the sales requirement for the certification and the distribution of effort costs in a market.

[18] https://support.upwork.com/hc/en-us/articles/211063228-Rising-Talent; accessed on 10/22/2020.

[19] Since established sellers are not eligible for the RT badge, we do not expect to see a step-down effect in the case of Upwork.

References

Barach, Moshe A, Joseph M Golden, and John J Horton, "Steering in online markets: the role of platform incentives and credibility," *Management Science*, 2020.

Bollinger, Bryan, Phillip Leslie, and Alan Sorensen, "Calorie posting in chain restaurants," *American Economic Journal: Economic Policy*, 2011, *3* (1), 91–128.

Bolton, Gary, Ben Greiner, and Axel Ockenfels, "Engineering trust: reciprocity in the production of reputation information," *Management science*, 2013, *59* (2), 265–285.

Butler, Jeffrey V, Enrica Carbone, Pierluigi Conzo, and Giancarlo Spagnolo, "Past performance and entry in procurement: An experimental investigation," *Journal of Economic Behavior & Organization*, 2020, *173*, 179–195.

Cabral, Luis and Ali Hortacsu, "The dynamics of seller reputation: Evidence from eBay," *The Journal of Industrial Economics*, 2010, *58* (1), 54–78.

Chevalier, Judith A and Dina Mayzlin, "The effect of word of mouth on sales: Online book reviews," *Journal of marketing research*, 2006, *43* (3), 345–354.

Dellarocas, Chrysanthos, "The digitization of word of mouth: Promise and challenges of online feedback mechanisms," *Management science*, 2003, *49* (10), 1407–1424.

_ **and Charles A Wood**, "The sound of silence in online feedback: Estimating trading risks in the presence of reporting bias," *Management science*, 2008, *54* (3), 460–476.

Dranove, David and Ginger Zhe Jin, "Quality disclosure and certification: Theory and practice," *Journal of Economic Literature*, 2010, *48* (4), 935–63.

Einav, Liran, Chiara Farronato, and Jonathan Levin, "Peer-to-peer markets," *Annual Review of Economics*, 2016, *8*, 615–635.

_ **, Theresa Kuchler, Jonathan Levin, and Neel Sundaresan**, "Assessing sale strategies in online markets using matched listings," *American Economic Journal: Microeconomics*, 2015, *7* (2), 215–247.

Elfenbein, Daniel W, Ray Fisman, and Brian McManus, "Charity as a substitute for reputation: Evidence from an online marketplace," *Review of Economic Studies*, 2012, *79* (4), 1441–1468.

_ , **Raymond Fisman, and Brian McManus**, "Market structure, reputation, and the value of quality certification," *American Economic Journal: Microeconomics*, 2015, *7* (4), 83–108.

Fan, Ying, Jiandong Ju, and Mo Xiao, "Reputation premium and reputation management: Evidence from the largest e-commerce platform in China," *International Journal of Industrial Organization*, 2016, *46*, 63–76.

Farronato, Chiara, Andrey Fradkin, Bradley Larsen, and Erik Brynjolfsson, "Consumer Protection in an Online World: An Analysis of Occupational Licensing," Technical Report, National Bureau of Economic Research 2020.

Filippas, Apostolos, John Joseph Horton, and Joseph Golden, "Reputation inflation," in "Proceedings of the 2018 ACM Conference on Economics and Computation" 2018, pp. 483–484.

Fradkin, Andrey, Elena Grewal, and David Holtz, "Reciprocity in Two-sided Reputation Systems: Evidence from an Experiment on Airbnb," 2019.

Grossman, Gene M and Henrik Horn, "Infant-industry protection reconsidered: the case of informational barriers to entry," *The Quarterly Journal of Economics*, 1988, *103* (4), 767–787.

Hopenhayn, Hugo and Maryam Saeedi, "Reputation Signals and Market Outcomes," Working Paper 2018.

Hui, Xiang, Maryam Saeedi, and Neel Sundaresan, "Adverse Selection or Moral Hazard, An Empirical Study," *The Journal of Industrial Economics*, 2018, *66* (3), 610–649.

_ , _ , **Giancarlo Spagnolo, and Steve Tadelis**, "Certification, Reputation and Entry: An Empirical Analysis," *Unpublished Manuscript*, 2017.

_ , _ , **Zeqian Shen, and Neel Sundaresan**, "Reputation and regulations: evidence from eBay," *Management Science*, 2016, *62* (12), 3604–3616.

Klein, Benjamin and Keith B Leffler, "The role of market forces in assuring contractual performance," *Journal of Political Economy*, 1981, *89* (4), 615–641.

Klein, Tobias J, Christian Lambertz, and Konrad O Stahl, "Market transparency, adverse selection, and moral hazard," *Journal of Political Economy*, 2016, *124* (6), 1677–1713.

Li, Lingfang Ivy, Steven Tadelis, and Xiaolan Zhou, "Buying reputation as a signal of quality: Evidence from an online marketplace," *RAND Journal of Economics*, Forthcoming.

Luca, Michael, "Reviews, reputation, and revenue: The case of Yelp. com," *Com (March 15, 2016). Harvard Business School NOM Unit Working Paper*, 2016, (12-016).

_ **and Georgios Zervas**, "Fake it till you make it: Reputation, competition, and Yelp review fraud," *Management Science*, 2016, *62* (12), 3412–3427.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier, "Promotional reviews: An empirical investigation of online review manipulation," *American Economic Review*, 2014, *104* (8), 2421–55.

Nosko, Chris and Steven Tadelis, "The limits of reputation in platform markets: An empirical analysis and field experiment," Technical Report, National Bureau of Economic Research 2015.

Resnick, Paul and Richard Zeckhauser, "Trust among strangers in Internet transactions: Empirical analysis of eBay's reputation system.," *The Economics of the Internet and E-commerce*, 2002, *11* (2), 23–25.

_ **,** _ **, John Swanson, and Kate Lockwood**, "The value of reputation on eBay: A controlled experiment," *Experimental economics*, 2006, *9* (2), 79–101.

Saeedi, Maryam, "Reputation and adverse selection, theory and evidence from eBay," *RAND Journal of Economics*, 2019.

Schoenmüller, Verena, Oded Netzer, and Florian Stahl, "The extreme distribution of online reviews: Prevalence, drivers and implications," *Columbia Business School Research Paper*, 2019, (18-10).

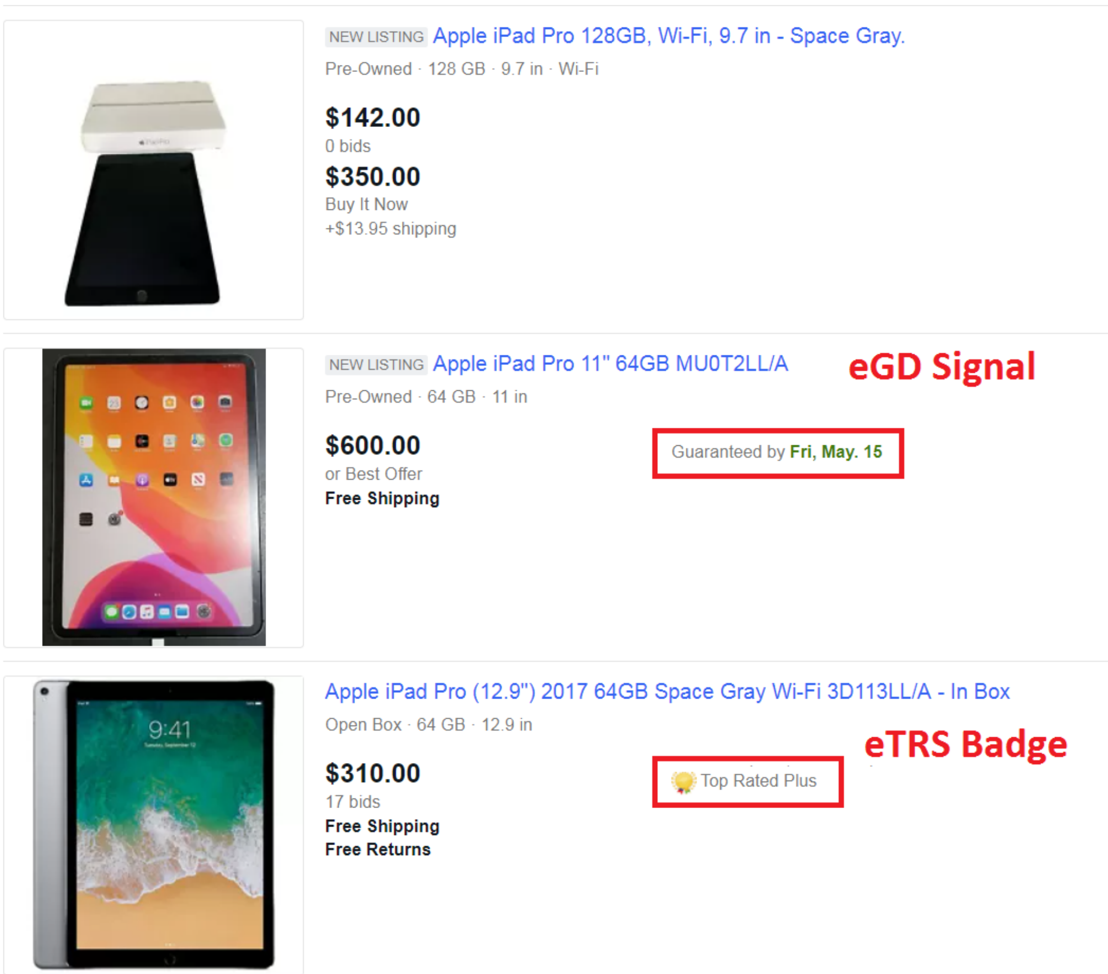
Tadelis, Steven, "Reputation and feedback systems in online platform markets," *Annual Review of Economics*, 2016, *8*, 321–340.

Vellodi, Nikhil, "Ratings design and barriers to entry," *Available at SSRN 3267061*, 2018.

Zervas, Georgios, Davide Proserpio, and John Byers, "A first look at online reputation on Airbnb, where every stay is above average," *Where Every Stay is Above Average (January 28, 2015)*, 2015.

Appendix A Figures

Figure 5: eBay Guaranteed Delivery (eGD) Signal and eBay Top Rated Seller (eTRS) Badge



Notes: The figure provides examples of listings from a non-badged seller, an eGD seller, and an eTRS seller with a Top Rated Plus badge.

Appendix B Supplements for the Theoretical Framework

In this appendix, we give the proofs and assumptions for the lemma and propositions in the theoretical framework. In addition, we provide one illustrative example in an attempt to investigate the market-level effects of the eGD program.

B.1 Proofs

Assumptions in the Theoretical Framework. We define $R_a = \underline{R} - (S^{III}(G) - S^{III}(NB))$, $R_b = \bar{R} - (S^{II}(T) - S^{II}(NB))$, and $R_c = \bar{R} - (S^{III}(T) - S^{III}(G))$ and make the following assumptions:

1. $\underline{R} < \bar{R} - \max\{S^{III}(T) - S^{III}(G), (S^{II}(T) - S^{II}(NB)) - (S^{III}(G) - S^{III}(NB))\}$,
2. $S^{II}(T) - S^{II}(NB) < S^{III}(T) - S^{II}(G)$.

The first assumption implies that the reputation requirement for the G badge is not too high. Furthermore, we allow the sales of sellers with the same badge under different mechanisms to be different, and the second assumption indicates that the G badge does boost sales compared with the no-badge situation. With these two mild assumptions, we have $R_a < R_b < R_c$ and $R_a < \underline{R} < R_c$.

Proof for Lemma 1. Each seller solves the following maximization problem:

$$\max_e \Pi^{\mathcal{M}}(e|R_0, V_0) = \max_e S^{\mathcal{M}}(B_1) - e.$$

Under the two-tier mechanism, i.e., $\mathcal{M} = II$,

$$\begin{aligned}\Pi^{II}(e|R_0, V_0) &= S^{II}(B_1) - e \\ &= \begin{cases} S^{II}(T) - e, & \text{if } e \geq \bar{R} - R_0 \text{ and } V_0 + S^{II}(B_0) \geq \bar{V}, \\ S^{II}(NB) - e, & \text{if } e < \bar{R} - R_0 \text{ or } V_0 + S^{II}(B_0) < \bar{V}. \end{cases}\end{aligned}$$

This implies

$$e^* = \begin{cases} \max\{0, \bar{R} - R_0\} & \text{if } R_0 \geq R_b \text{ and } V_0 \geq \bar{V} - S^{II}(NB), \\ 0, & \text{if } R_0 < R_b \text{ or } V_0 < \bar{V} - S^{II}(NB). \end{cases}$$

Under the three-tier mechanism, i.e., $\mathcal{M} = III$,

$$\begin{aligned}\Pi^{III}(e|R_0, V_0) &= S^{III}(B_1) - e \\ &= \begin{cases} S^{III}(T) - e, & \text{if } e \geq \bar{R} - R_0 \text{ and } V_0 + S^{III}(B_0) \geq \bar{V}, \\ S^{III}(G) - e, & \text{if } \underline{R} - R_0 \leq e < \bar{R} - R_0, \\ S^{III}(NB) - e, & \text{if } e < \underline{R} - R_0. \end{cases}\end{aligned}$$

This implies

$$e^* = \begin{cases} \max\{0, \bar{R} - R_0\}, & \text{if } R_0 \geq R_c \text{ and } V_0 \geq \bar{V} - S^{III}(G), \\ \max\{0, \underline{R} - R_0\} & \text{if } R_a \leq R_0 < R_c \text{ or } V_0 < \bar{V} - S^{III}(G) \text{ and } R_0 \geq R_c, \\ 0, & \text{if } R_0 < R_a. \end{cases}$$

□

Proofs for Proposition 1 and 2. Proposition 1 and 2 follow by comparing the optimal strategies under two different mechanisms.

Specifically, for the sellers who step up, their updated badge status under the two-tier and three-tier mechanisms is NB and T, respectively. According to Lemma 1, this implies

$$R_0 < \bar{R} - (S^{II}(T) - S^{II}(NB)) \text{ or } V_0 < \bar{V} - S^{II}(NB)$$
$$\text{and } R_0 \geq \bar{R} - (S^{III}(T) - S^{III}(G)) \text{ and } V_0 \geq \bar{V} - S^{III}(G).$$

According to our assumptions, we have $S^{II}(NB) < S^{III}(G)$ and $\bar{R} - (S^{II}(T) - S^{II}(NB)) < R_0 \leq \bar{R} - (S^{III}(T) - S^{III}(G))$. Therefore, the sellers who step up can be characterized by the following set of inequalities based on their initial reputation level R_0 and initial historical sales level V_0:

$$R_0 \geq \bar{R} - (S^{III}(T) - S^{III}(G)),$$
$$\bar{V} - S^{III}(G) \leq V_0 < \bar{V} - S^{II}(NB).$$

Similarly, for the sellers who step down, their updated badge status under the two-tier and three-tier mechanisms are T and G, respectively. According to Lemma 1, this implies

$$R_0 \geq \bar{R} - (S^{II}(T) - S^{II}(NB)) \text{ and } V_0 \geq \bar{V} - S^{II}(NB)$$
$$\text{and } R_0 \geq \underline{R} - (S^{III}(G) - S^{III}(NB)), \text{ and } R_0 < \bar{R} - (S^{III}(T) - S^{III}(G)) \text{ or } V_0 < \bar{V} - S^{III}(G).$$

According to our assumptions, we have $S^{II}(NB) < S^{III}(G)$ and $\bar{R} - (S^{II}(T) - S^{II}(NB)) < R_0 \leq \bar{R} - (S^{III}(T) - S^{III}(G))$. Therefore, the sellers who step down can be characterized by the following set of inequalities based on their initial reputation level R_0 and initial historical sales level V_0:

$$\bar{R} - (S^{II}(T) - S^{II}(NB)) \leq R_0 < \bar{R} - (S^{III}(T) - S^{III}(G)),$$
$$V_0 \geq \bar{V} - S^{II}(NB).$$

$\square$

B.2 An Illustrative Example for Market-level Effects

We examine the market-level effects of introducing an intermediate tier to the two-tier mechanism. Specifically, we are interested in understanding how different markets react to the eGD program in terms of certification dynamics and quality provision. We discuss a special case here.

Let us consider a market where all sellers have the same initial reputation level R_0, where $R_0 > \bar{R} - (S^{III}(G) - S^{III}(NB))$. The initial sales level V_0 follows a uniform distribution with support $[V_0^a, V_0^a + \Delta V]$. Without loss of generality, we assume $\Delta V > S^{III}(G) - S^{II}(NB)$. We fix R_0 and ΔV, and investigate how seller dynamics and quality provision change as we vary V_0^a from $\bar{V} - S^{III}(G)$ to $\bar{V} - S^{II}(NB)$.

With the introduction of the G badge, sellers with $V_0^a \leq V_0 \leq \bar{V} - S^{II}(NB)$ will step up. This indicates that the ratio of sellers who step up is $\frac{\bar{V} - S^{II}(NB) - V_0^a}{\Delta V}$, which decreases in V_0^a. Accordingly, the change in quality provision (represented by R_1) with the introduction of the intermediate tier badge G is $\frac{\bar{V} - S^{II}(NB) - V_0^a}{\Delta V}\bar{R}$, which also decreases in V_0^a. With lower V_0^a, the sellers in this market have lower initial sales levels. In a market with lower average sales levels, the step-up ratio with the introduction of the G badge increases, and so does the change of quality provision. Since the G badge has less strict sales requirements, a market which consists of many sellers who cannot satisfy the sales requirement of the new signal witnesses a larger step-up effect. In addition, the increase of the change in quality provision results from a positive shift of sellers' effort level. This suggests that the introduction of this intermediate badge induces more sellers to exert more effort and provide higher quality in a market with more sellers with an insufficient sales history.