Quality Unobserved: Can Information Provision Unlock Demand-Side Incentives for Upgrading in Low-income Countries?*

Julia Cajal-Grossi   Lore Vandewalle   Christopher Woodruff†

March 4, 2026

Abstract

Can improving consumers' ability to discern quality increase demand-side incentives for quality provision? We conduct a framed field experiment in Uganda's furniture market, where quality dispersion is high but the price-quality gradient is flat. Nearly 900 prospective buyers ranked and priced tables spanning the quality distribution; a random subset received information on quality markers. At baseline, individual consumers perform worse than industry insiders at discerning quality. Information provision closes this gap: treated consumers' likelihood of correct rankings increases by 23 percentage points, and their price-quality gradient steepens significantly. Extrapolated market-wide, this would raise markups for top-quality producers by approximately 7.5%.

Keywords: information frictions, quality discernment, upgrading, willingness to pay, developing countries
JEL Codes: D83, L15, O12, C93

*This paper has benefited from valuable feedback from Jie Bai, Vittorio Bassi, Lauren Falcao Bergquist, Lorenzo Casaburi, Kevin Donovan, Amit Khandelwal, Stefan Klonner, Rocco Macchiavello, Ritwika Sen, Eric Verhoogen, Ana Vitali. We thank Lucía Maldonado, England Can and Marco Matani for outstanding research assistance. We are grateful to participants at seminars, workshops, and conferences at Aix-Marseille School of Economics, Ghent University, IHEID, International Finance Corporation, Ludwig Maximilian University of Munich, Newcastle University, Stanford University, ULB, University of Bergen, University of Birmingham, University of Exeter, University of Oslo, University of Zurich, Yale University. We are grateful for funding for this project from Argidius Foundation, BMGF, Mastercard Impact Fund for the Small Firm Diaries project, IGC, the Weiss Foundation, and PEDL. The project has human subjects approval from the Research and Ethics Committee of Makerere University (MAK SPHREC), under study number SPH-2021-23. The experiment discussed in this paper was submitted to the AEA RCT Registry under number AEARCTR-0015833.

†Julia Cajal-Grossi: Geneva Graduate Institute (IHEID), Switzerland; Lore Vandewalle: KU Leuven, Belgium, and Geneva Graduate Institute (IHEID), Switzerland; Christopher Woodruff: Oxford University, United Kingdom.

1 Introduction

Access to large, rich consumer markets is a key pathway to quality upgrading and growth for firms in low-income countries (Verhoogen, 2023). In practice, policies aimed at providing firms with access to these markets have involved moving firms from domestic consumers to foreign buyers in higher-income countries (see, for example, Verhoogen, 2008; Atkin et al., 2017; Alfaro-Ureña et al., 2022). However, this route is only available to a small fraction of firms –those large enough to overcome the barriers to entry into distant markets. The private sector in low-income countries is instead dominated by micro and small enterprises that are limited in their reach beyond immediate local markets (Hsieh and Olken, 2014). This restricts the impact that cost-effective market access policies can have at scale.

Whether other interventions can replicate the positive effects of foreign market access is unclear. Selling to foreign markets exposes firms to a bundle of 'treatments': access to consumers with higher incomes and willingness to pay for quality, *and* to consumers better able to discern quality. While income levels and preferences in local markets are difficult to treat at scale, consumers' ability to discern quality appears as an accessible policy instrument for changing demand-side incentives for the mass of firms in low-income countries. Beyond its practical feasibility, this policy lever addresses a market failure central to economic theory: information asymmetries drive adverse selection (Akerlof, 1970). In particular, markets with some proportion of undiscerning consumers –those observing price but not quality– generate equilibria where firms provide low quality since competing on unobserved dimensions proves unprofitable (Armstrong and Chen, 2009; Gabaix and Laibson, 2006).

We carry out a framed field experiment measuring domestic buyers' sophistication and assessing whether simple information provision improves their ability to discern quality. The empirical setting for our study is the market for artisanal wooden furniture in Uganda, a sector that draws significant numbers of small-scale enterprises and the self-employed throughout the country. Besides its intrinsic interest, the carpentry market features characteristics ubiquitous to manufactured consumables in low-income countries. First, furniture production is largely artisanal, yielding substantial variation in output quality, beyond horizontal differentiation. Workmanship differences translate into measurable variation in quality in a vertical sense –joint construction, measurement precision, and wood seasoning all affect durability and load capacity. This quality dispersion persists even among spatially clustered competitors: proximity generates neither quality nor price convergence, pointing to weak competitive pressure through customer arbitrage. Second, customers engage in modest search and base purchase decisions on price and horizontal differentiation, rather than vertical quality. Price negotiations feature centrally. Disagreements over price account for more than four-fifths of unsuccessful sales, and most transactions settle below initial asking prices. Third, prices increase only modestly –and markups not at all– with quality, offering producers scant returns to upgrading. Whether this reflects consumer indifference or difficulty evaluating quality has important implications for addressing the potential underperformance of the market in the provision of quality.

Our experiment tests whether information on observable markers of quality improves customers' ability to identify quality differences and increases their willingness to pay for higher-quality items. We commissioned identical small tables (same sizing, design specifications, and wood species) from carpentry workshops spanning the quality distribution in Uganda's furniture sector, purposefully eliminating horizontal differentiation to isolate vertical quality variation. Three master carpenters –formally qualified experts who train apprentices and teach at vocational institutes– independently evaluated each table, providing quality ratings we take, in an ordinal sense, as ground truth. Operating from kiosks in two furniture markets in Greater Kampala, we recruited actual furniture shoppers. Each participant ranked five tables by quality –incentivized to match expert assessments– and stated the purchase price they would offer for each item. A randomly selected group received a brochure discussing the physical markers of high-quality wooden furniture, covering joinery, stability, and warping, among others. The control group was presented with neutral content promoting Ugandan-made products. All participants then revised their quality rankings and stated willingness to pay.

Over three months, 897 customers participated in our experiment. Approximately 80% of participants were *individual consumers* intending to buy furniture for their private use, while the remaining 20% constituted *industry insiders* –that is, wholesalers, middlemen, and/or carpenters. We obtain three key results from the intervention. First, at baseline individual consumers are worse at assessing quality, relative to industry insiders. Consequently, the price-quality gradient of industry insiders at baseline is steeper than that of individual consumers, which in turn matches the gradient observed in the Ugandan furniture market at large. Second, information provision significantly improves the ability of customers to discern quality on all outcomes. Relative to the control, customers receiving the information treatment see a 23 percentage point increase in their likelihood of ordering all items correctly. This improvement is driven by individual consumers, whose ability to assess quality 'catches up' with that of industry insiders, for whom the treatment does not change quality discernment. Third, we find a significant steepening of the price-quality gradient for individual consumers after treatment. Treatment rotates stated willingness to pay around intermediate quality levels: consumers lower their valuations for low-quality items and raise them for high-quality items, steepening the price-quality gradient by approximately 1.4 percentage points per quality decile. Consistent with the results on quality discernment, industry insiders show no significant change in their price-quality gradient.

Taken together, these results indicate that the provision of information is effective at improving quality discernment by uninformed consumers, who in turn state a higher willingness to pay for quality. Because markups in the furniture sector are essentially flat in quality, producers currently face weak incentives to upgrade. If the treatment-induced steepening obtained market-wide, a producer moving from median to top-decile quality would earn a markup roughly 7.5% higher. While our design does not allow us to speak to general equilibrium effects, it points to the potential that this type of intervention carries, as a low-cost policy to improve demand-side incentives for quality upgrading in local markets in low-income countries.

This paper relates to an empirical literature on quality disclosure, surveyed by Dranove and Jin (2010), documenting how information provision affects market outcomes. One strand of this literature establishes that third-party certification and quality ratings improve quality provision and intensify competition once performance becomes observable (Jin and Leslie, 2003; Anderson and Magruder, 2012; Andrabi et al., 2017). Relatedly, a growing body of work examines information problems in agricultural input markets, where farmers struggle to assess quality at the point of purchase, stifling technology adoption (Michelson et al., 2021; Hsu and Wambugu, 2022). In these settings, training farmers to detect quality-verified products shifts purchasing toward higher-quality inputs. Unlike these studies, we focus on a manufactured final product that typifies consumer markets in low-income countries: buyers are by definition non-experts, purchase frequency may be too low to develop discernment through experience, certification labels are infeasible at market scale and horizontal differentiation further complicates quality assessment. While these differences pose methodological challenges, the distinction is consequential. Non-food manufactured goods constitute a substantial share of household consumption among the poor and, thus, distortions in these markets may have large welfare implications (Banerjee and Duflo, 2007).

Our paper also relates to a smaller literature on demand-side constraints to quality upgrading. Bold et al. (2017) and Bold et al. (2022) show that in Uganda's maize seed market, farmers cannot distinguish seed quality, eliminating returns to quality provision; critically, supply-side training proves ineffective without concurrent demand-side change. Jensen and Miller (2018) find that Kerala's boat-building sector benefits from market integration as fishermen learn over time the identity of high-quality producers. Studying melon sellers in China, Bai (2025) shows that costly signals (labels) can induce producer effort to provide higher-quality goods, though effects dissipate as low-quality sellers eventually pool. In contrast to these mechanisms, our intervention directly trains non-expert individual consumers to assess quality in one-off purchases without relying on certifications or repeated interactions. Our focus on domestic consumers also contrasts with research emphasizing access to foreign, sophisticated buyers. We focus on a scalable intervention that enhances the discernment of local 'lay' consumers, offering a complementary pathway to quality upgrading—one with potentially large income effects given the dominance of small businesses in low-income countries.

The rest of this paper is organized as follows. Section 2 characterizes the Ugandan furniture market. Section 3 describes the experimental design and Section 4 introduces outcomes and baseline patterns. Section 5 presents the results of our experiment and discusses implications, and Section 6 concludes.

2 Empirical Setting

The wood furniture sector in Uganda shares features common to light-manufacturing consumer goods across low-income countries. This section documents four key features, drawing on two data sources that characterize carpentry production and customer behavior. First, we leverage surveys administered to the 897 participants in our experiment, who are representative of typical buyers in large furniture markets in Uganda. The presentation of our sampling is left for Section 3. Second, we use data from a related study covering a random sample of 748 carpentry workshops across five districts—Kampala, Wakiso, Mukono, Kabarole, and Mbarara (Cajal-Grossi et al., 2025).[1]

First, **furniture production is largely artisanal, leading to substantial dispersion in product quality across workshops**. Production involves labor-intensive cutting, joining, and finishing of lumber using hand tools and rented machines (Bassi et al., 2022). Beyond the high scope for horizontal differentiation, variable workmanship creates measurable dif-

[1] Cajal-Grossi et al. (2025) administer different types of data collection instruments to the 748 randomly selected small workshops, including (i) high-frequency fortnightly surveys with quality assessments of furniture items at each workshop, and (ii) ledgers tracking sales, orders, and customer visits over two to three months. The characterization of workshops and quality dispersion draws from the former, while the materials in Table 1 leverage the latter. In this paper, we focus on the set of characteristics most related to the study at hand. A summary of broader characteristics of the carpentry workshops is presented in Supplemental Appendix Table A1 for completeness, and a more systematic presentation of the sector's profile can be found in Bassi et al. (2022).

ferences in the structural integrity of furniture pieces. Joinery characteristics, for example, determine an item's load-bearing capacity and susceptibility to insect infestations. Precise measurements and symmetry ensure stability and, in turn, durability. Superior workmanship, thus, increases utility for all consumers in a vertical sense, even in the presence of consumers' heterogeneous tastes for horizontal attributes, discount rates, and budget constraints. In practice, the market exhibits significant dispersion in vertical quality. For concreteness, physical examination of 3,207 items from the 748 workshops in Cajal-Grossi et al. (2025) shows that only 46% of chairs meet angle symmetry requirements, 67% of the wood was adequately seasoned, and only 26% of table tops were completely free of visible damage, including cracks, knots, and traces of insects.

Second, **quality and price dispersion persist across nearby workshops despite dense spatial agglomeration**. Workshops cluster primarily within cities and along main roads and arteries connecting towns – a map is provided in Supplemental Appendix Figure A1. Most carpenters locate their workshop within walking distance of their most important competitor, and the average (geocoded) distance to the closest competitor is 210 meters. Despite this proximity, both quality and prices show only modest spatial correlation. Residual prices exhibit weak correlation (0.05–0.25) among carpenters within 1 km, disappearing beyond that radius; residual vertical quality shows even weaker spatial correlation (see Supplemental Appendix Figure A2). This suggests limited arbitrage of local differences in prices and quality through customers' mobility and competitive pressure.

Third, **customers engage in search before purchasing but assess options primarily on horizontal characteristics and price**. Most customers (65%) interviewed in our study purchased furniture within the last year, visiting an average of 2.8 outlets before buying (Supplemental Appendix Table A2). Drawing on over 6,000 interactions between carpenters and customers, Table 1 shows that referrals and return visits account for 66% of customer traffic (Panel A). In virtually all visits the customer seeks a specific type of item and always discusses prices (Panel D). When no purchase occurs, 90% of the cases cite disagreement over price as the reason for the failed transaction (Panel C). This is consistent with strong bargaining in furniture purchases, where 88-92% of items sell below their original asking price.

Fourth, **markups remain flat across the quality spectrum, providing carpenters with little incentive to upgrade**. Table 2 shows that wooden tables command no quality-increasing premia in the market. Moving up one decile in the quality distribution correlates with a 3.9% increase in asking prices, commensurate with the 3.8% rise in unit costs associated with such an improvement. As a result, the markup factor is flat in quality. This suggests weak demand-side incentives for quality provision.

Importantly, if the relationship between prices and quality documented here reflects consumer preferences, no intervention is warranted. However, if the flat markup-quality gradient stems from buyers' inability to discern quality, this could drive the underprovision of quality in the market. The following section presents a framed field experiment testing this channel directly: can furniture buyers learn to better discern quality, and if so, does this increase their willingness to pay for it?

3 Study Design

Our experiment presents furniture buyers with information on how to detect quality, and examines whether this changes quality discernment and willingness to pay. To capture decision-making as close to real market conditions as possible, participants had to be actual furniture customers, presented with real furniture pieces, in their typical purchasing environment. We ran our experiment in two prominent furniture markets in Greater Kampala: Nateete and Kigo. Both markets feature extensive product variety across the quality spectrum and maintain high customer traffic. Operating from rented kiosks, we collected data over seven weeks between March and May 2023.

We selected ten carpenters, from different quality strata and not operating in Nateete or Kigo, and commissioned four identical side tables from each, using standardized sizing specifications and a fixed design template. This ensured that the items displayed in the kiosks would feature no horizontal differentiation and vary only on vertical quality. Three master carpenters evaluated all tables using a comprehensive quality assessment tool and rated them on a 1-10 scale, where 10 represents the highest quality produced in Uganda.[2] The resulting quality ratings showed substantial variation across items: the lowest rating awarded by the master carpenters was 2 and the highest was 9, with an average (median) score across master carpenters and items of 6.3 (6.0). A series of validation exercises shows a high within-item correlation of ratings, both across master carpenters (0.88-0.93) and within master carpenters (0.85-0.97) when reassessing a random subsample of items several weeks later.

Enumerators recruited participants by approaching individuals browsing or shopping for furniture. Participants joined enumerators in a kiosk, where a pseudo-random set of five tables

[2] Supplemental Appendix B discusses the quality assessment tool, as well as the validation exercises with master carpenters' ratings.

was displayed. The choice of five tables –large enough for meaningful quality comparisons yet small enough to permit careful assessment of each item– reflected three design objectives: comparability across participants, collective coverage of the market quality distribution, and randomness in allocation. For each kiosk, we pre-specified 72 eligible five-table combinations satisfying these criteria, yielding 900 possible pairings of table set and display order; one pairing was drawn at random for each participant. How a respondent assesses relative quality likely depends on several features of the table mix, such as average quality or the magnitude of pairwise differences. The random assignment of menus to participants and of participants to treatment therefore does not guarantee balance in all plausibly relevant features of the quality distribution that participants faced.[3] This feature of our design means that identification of treatment effects must account for chance differences in the quality composition participants faced –we address this in the next section.

After completing a brief demographic survey, participants provided their willingness to pay and ordered all five tables by their perceived quality. These rankings were incentivized based on the participant's ability to match the *ordering* of the quality assessments of the master carpenters. Specifically, participants received a show-up fee of UGX 8,000 ($2) plus an additional UGX 1,000 for each table they ranked in the same relative position as the master carpenters. Those perfectly matching the entire ranking received, in addition, a bonus "Furniture Study Champion" cap. In turn, the willingness to pay was not incentivized, given that our interest lies in the price-quality gradient rather than price levels.[4] Willingness to pay was elicited as a stated value in a hypothetical purchase, in response to the question *"Suppose that you were to buy a table of this type, and that you have the necessary budget for it. What is the maximum price you would be willing to pay for each of these items?"*

We randomly assigned participants to treatment or control, and both groups received informational brochures available in English and the local language. The random treatment assignment was automated and only revealed to the enumerator at the point where brochures needed to be produced, after the baseline survey. Control participants were presented with a

[3] Indeed, a Monte Carlo exercise that simulates 10,000 different experiments with 900 participants using our tables shows that in 22% of the experiments, the treatment and control groups face significant differences (at 5%) in at least one of average quality, quality range, smallest pairwise gap or largest pairwise gap.

[4] We avoid incentivized elicitation mechanisms like Becker-DeGroot-Marschak (BDM) for two reasons (Becker et al., 1964). First, wooden furniture is purchased infrequently and only when genuine need arises. Our specific table design means most participants may have no immediate demand for it at the time of the experiment –some may even view owning an additional piece of furniture as a *bad*– leading to imprecision in BDM responses (Dizon-Ross and Jayachandran, 2022). Second, to characterize price-quality gradients rather than price levels, we elicit valuations for multiple close substitutes. This comparative framing heightens concerns about violations of incentive-compatibility in BDM designs under reference dependent preferences (Horowitz, 2006).

brochure emphasizing the benefits of purchasing locally-produced items instead of imported furniture, directing attention to quality broadly, without highlighting any physical markers.[5] Treated participants received a brochure detailing how to assess quality, based on observable characteristics of an item. This included the detection of bending or cracks due to poor wood seasoning, the examination of joinery techniques, the measurement of angles involved in structural stability, and the protection conferred by the finishing of the item.

4 Outcomes and Baseline Patterns

Our analysis focuses on two types of outcomes. The first is *quality discernment*, intended as participants' ability to correctly identify the quality ordering of tables, as established by the master carpenters' consensus ranking. We capture discernment using four measures: (i) the respondent matching the ordering of all five tables (indicator $\in \{0,1\}$), (ii) the count of items ordered correctly (numeric $\in \{0,1,2,3,4\}$), (iii) the respondent getting the *easiest* pairwise comparison right (indicator $\in \{0,1\}$), and (iv) the respondent getting the *hardest* pairwise comparison correctly (indicator $\in \{0,1\}$). The easiest ranking among the tables presented to the respondent corresponds to the comparison of the two tables farthest apart in quality. For concreteness, these comparisons are akin to telling apart an item below the 5^{th} percentile of the quality distribution in the market, and an item in the 99^{th} percentile (Supplemental Appendix Figure A3). Analogously, the hardest ranking is the comparison of the two tables with the smallest rating difference. In the market, this would be as hard as producing the right quality ordering for an item in the 20^{th} percentile of the quality distribution and one in the 35^{th} percentile. The second key outcome of our study, stated willingness to pay, captures the participants' hypothetical valuation for each item in Ugandan Shillings (UGX).

Before turning to the results of our experiment, we present the key patterns at baseline, and contrast our sample of respondents and products with the wider furniture market in Uganda.

Following the recruitment protocol discussed above, 897 individuals consented to participate: 727 individual consumers and 170 industry insiders. Table 3 presents summary statistics and balance checks. The randomization allocated 49% of individual consumers and 46% of industry insiders to the treatment condition. The table also provides the standard difference

[5] For concreteness, a key placebo statement in the control brochure reads *"The first aspect we want to discuss is the quality of products. A study in Tanzania has shown that one of the main reasons why customers prefer imported furniture is because the designs are nicer [...] They are made cheaply by cutting on construction costs and on materials. These products may not stand the test of time, and are thus less durable."*

by treatment status in the overall sample (columns (1)-(3)), the subsample of individual consumers (columns (4)-(6)), and that of industry insiders (columns (7)-(9)), largely revealing the expected balance in individuals' characteristics between treatment and control in the whole sample and both subsamples. As anticipated, the combination of tables presented to individuals leads to imbalance in quality distributions –and hence outcomes– at baseline. Reassuringly, balance is largely restored when we condition on the configuration of tables presented to respondents (see Supplemental Appendix Table A3). We address this incidental imbalance econometrically by using an ANCOVA structure to obtain the effects of the treatment in Section 5.

Consistent with the discussions in Section 2, when asked about the dimensions they attend to when buying wooden furniture, respondents tended to emphasize attributes that are connected to horizontal characteristics of furniture, including finishing, ornamentation, and wood species (see Supplemental Appendix Table A2). Accordingly, when asked to explain what markers led them to rank a table as lowest or highest in the ranking of quality at baseline, participants most commonly mentioned aspects of ornamentation and finishing of the item.

As a result, quality discernment at baseline is rather poor. Table 3 shows that only 26% of the respondents in the control group rank all five tables correctly according to their quality. The easiest ordering is successfully identified by the majority of the control respondents (96%), consistent with the fact that this outcome represents comparisons between bottom-quality (5^{th} percentile of the quality distribution) and top-quality (99^{th} percentile) items. Instead, the hardest comparison is solved correctly by 59% of the respondents, close to the 50% that would attain the correct result by a random guess. Notably, we find that individual consumers exhibit significantly worse quality discernment compared to the more knowledgeable industry insiders.

Respondents' stated willingness to pay in the experiment at baseline mirrors the weak price-quality relationship in the broader market. Returning to Table 2, column (4) shows that moving up one decile in the quality distribution leads individual consumers to offer a 4.2% higher price, consistent with the 3.9% market gradient (column (3)). Notably, industry insiders –who appear more quality-discerning– exhibit a 6.4% gradient, substantially steeper than consumers'. Furthermore, master carpenters, who are reputable experts in furniture construction and assessment, report suitable prices that imply a 12.4% increase per decile of quality –three times the gradient observed among typical consumers. This suggests that individuals better able to discern quality also better assess its value. These patterns, both in the market and in our experiment, remain unchanged when a wide array of alternative approaches to measure quality are used (Supplemental Appendix Table A4), as well as when the relationship between prices and quality is captured non-parametrically (Supplemental Appendix Figure A4).

5 Experimental Results

This section presents the results of our information provision experiment. We examine treatment effects on quality discernment and stated willingness to pay, finding that individual consumers substantially improve their ranking ability and steepen their price-quality gradient, while industry insiders remain largely unaffected. We close with a discussion of identification, market implications and external validity.

Quality Discernment. We run two sets of regressions that are embedded in the following ANCOVA specification:

$$Y_i^1 = \alpha_1 Individual\ Consumer_i + \alpha_2 Treated_i + \\ \alpha_3 Individual\ Consumer_i \times Treated_i + \alpha_4 Y_i^0 + \Gamma_{k(i)} + \epsilon_i, \quad (1)$$

where Y_i^1 is the quality discernment outcome of interest for respondent i after treatment and Y_i^0 controls for the same outcome at baseline. $Individual\ Consumer_i$ is a dummy that distinguishes whether i is an individual consumer or an industry insider, and $Treated_i$ is an indicator for treatment assignment. Finally, $\Gamma_{k(i)}$ is an indicator for the market k (Nateete versus Kigo) in which individual i participated in the experiment. Table 4 reports the results for the four quality discernment outcomes introduced in Section 4. Odd-numbered columns show overall treatment effects and the even-numbered columns allow effects to differ by customer type.[6]

Column (1) shows that treatment increases the share of participants placing all five tables in the correct order by 23 percentage points, relative to a control group mean of 30%. Similarly, while the control on average orders 2.8 tables correctly, the treatment delivers an additional increase of 0.5 in the count of correct rankings (column (3)). The richer specifications of columns (2) and (4) indicate industry insiders outperform individual consumers in the control group, and that treatment slightly more than catches the consumers up to

[6] The treatment is randomized at the individual level and standard errors are robust to heteroskedasticity. Supplemental Appendix Table A5 reproduces all experimental results in this paper, using a Randomization Inference approach.

the insiders' discernment. Indeed, the treatment has no significant effect on the number of items correctly ranked by insiders (column (4)) and only a marginally significant effect for a completely correct ranking (column (2)). In the same vein, the information treatment does not significantly improve the discernment of industry insiders in the easiest and hardest comparisons, where control insiders already perform relatively well (columns (6) and (8)). Treatment does, however, improve individual consumers' assessment of both the easy and hard comparisons, leading to a discernment comparable to that of industry insiders.

Willingness to Pay. We study the willingness-to-pay outcome, with a specification analogous to the one discussed above:

$$Y_{ij}^1 = \beta_1 Quality_j + \beta_2 Treated_i + \beta_3 Quality_j \times Treated_i + \beta_4 Y_{ij}^0 + \Gamma_{k(i)} + \varepsilon_{ij}, \tag{2}$$

where Y_{ij}^1 measures the willingness to pay of respondent i for table j, after treatment. On the right-hand side, Y_{ij}^0 controls for willingness to pay at baseline. $Quality_j$ corresponds to the median quality rating that master carpenters assign to item j and, as above, $Treated_i$ is an indicator of treatment status. As such, the interaction term measures the treatment impact on the price-quality gradient in an ANCOVA specification. To aid the quantitative interpretation of the overall quality slope (here β_1) we also present results for alternative specifications that study the outcome before treatment (Y_{ij}^0) and after treatment (Y_{ij}^1) separately. The first of these characterizes the price-quality gradient at baseline and, the second, the 'pure' treatment effect on post-treatment willingness to pay. Across all specifications, individual i states their willingness to pay for the five tables they are presented with, so we cluster the standard errors ε_{ij} at the individual level. Table 5 collates the results of the estimation of equation (2) –and its variations– on different samples: we study all respondents pooled together (columns (1), (4), (7)), individual consumers only (columns (2), (5), (8)), and industry insiders only (columns (3), (6), (9)).

The central hyper-column (*Before Treatment*) in Panel A of Table 5 reports the specification with pre-treatment outcomes, showing that individual consumers increase their willingness to pay by UGX 6,309 for each quality point at baseline (column (5)). Industry insiders are more sensitive to quality, awarding an additional UGX 10,856 for each additional quality point (column (6)). There is no significant difference between the control and treatment groups in the price-quality gradient at baseline. The rightmost hyper-column (*After Treatment*) shows that this gradient becomes steeper for the control group, particularly so for industry insiders (column (9)). We posit that this reflects the opportunity of reconsideration given to all respondents, with no differential shift in the price-quality gradient for the treated insiders. However, treatment leads to a significant increase in the price-quality gradient only for consumers.

We report the results of the ANCOVA (our preferred specification) in the leftmost hyper-column in Panel A of Table 5. The results show that the information provision treatment rotates the price-quality gradient, by reducing the willingness to pay for low-quality items, while increasing the slope of prices on quality. For concreteness, the linear specification in column (1) would imply a decrease of UGX 4,885 in the willingness to pay for the lowest quality item (rating of 2) and an increase of UGX 4,775 in the willingness to pay for the highest quality item (rating of 9) after treatment. The control and treatment price-quality lines implied by the ANCOVA intersect each other at intermediate quality levels: on average, items with a quality lower than 5.5 in rating see revised prices below the control willingness to pay following the treatment, while those of quality above 5.5 see higher stated prices. Columns (2) and (3) show that the rotation in the price-quality gradient is entirely driven by individual consumers, with no significant effect of the treatment on the price-quality gradient of industry insiders.[7]

Discussion. We elaborate on four aspects of our results that require careful examination.

First, the brochure presented to the treated group can change baseline quality discernment either by directing attention to physical markers (like joint angles or coplanarity) previously ignored, or by teaching individuals how to assess these markers. In contrast, the control brochure directs attention to quality broadly, without highlighting specific physical markers. Thus, improved discernment among treated respondents might arise either from increased salience of the physical characteristics that determine quality or from increased knowledge of how to evaluate them. Our experiment does not allow us to separately identify the effect of salience vis-à-vis knowledge, but we conjecture that both play a role in improving discernment.

Second, the reader may be concerned about experimenter demand effects on prices: an individual who restates their quality ranking may feel compelled to change their prices accordingly, even if their underlying willingness to pay has not changed. Our hypothesis is that consumers value quality but cannot identify it at baseline; treatment provides information enabling them to express pre-existing preferences in stated prices. Under this account,

[7] Supplemental Appendix Tables A6 and A7 show that the treatment effects reported in this section, on both quality discernment and willingness to pay, obtain as well in specifications that condition on the configuration of tables presented to the respondent.

rankings and prices move together precisely *because* information allows preference revelation. The concern, rather, is that treated respondents may adjust prices to match rankings simply because they believe this is what the experimenter expects, with demand effects operating differentially in the treatment group. Note that the desire for coherence between rankings and prices is present for all respondents throughout the exercise, at baseline and for both treatment and control. Both groups received brochures mentioning quality and were invited to reconsider their responses. The question is whether the treatment brochure –which taught specific quality markers– differentially amplified this coherence motive. Several patterns suggest it did not.

Conditional on revising a ranking, the probability of correspondingly adjusting prices is similar across treatment and control. Among item-respondent pairs where the ranking was revised upward, 91% in treatment also revised prices upward, compared to 98% in control; for downward revisions, the figures are 95% and 93%, respectively. The difference between groups lies in the unconditional probability of revision –treatment induces more ranking changes (51% versus 18% of respondents)– not in a heightened propensity to adjust prices conditional on revising. In addition, treatment induces price adjustments at both ends of the quality distribution: 14% of treated respondents lowered their minimum stated price while 19% raised their maximum, compared to 2% and 5% in control. This rotation is consistent with revaluation across the quality distribution. Finally, we replicate the baseline analysis on Panel A of Table 5 using respondents' own quality ratings rather than master carpenters' ratings (Panel B of Table 5). If demand effects were primarily responsible for our results –with prices mechanically following own rankings– the treatment effect should be stronger in this specification – if anything, it is weaker (a point estimate of 1.245 compared to one of 1.604 in Panel A of Table 5).

Third, we consider what the experimental results would imply for market outcomes. To benchmark the effects in terms comparable to Table 2, we replicate the willingness-to-pay analysis using the physical quality index (Panel C of Table 5). In percentage terms, treatment steepens consumers' price-quality gradient from approximately 4.5% to 5.9% per decile—an increase of 1.4 percentage points.[8] Two features of our design support cautious extrapolation to the market. First, experiment participants were actual furniture shoppers. Second, as shown in Table 2, the baseline price-quality gradient among consumers in the experiment (4.2% per decile) closely matches that in market transactions (3.9% per decile), suggesting

[8]To see this, note that column (8) of Panel C of Table 5 implies a price-quality gradient, in percentage terms, of approximately 4.5% (2.639/59.15) for the control group and 5.9% ((2.639 + 0.837)/59.15) for the treatment group.

experimental responses reflect broader demand conditions.

Table 2 documents that markups are essentially flat in quality: prices and costs both rise by approximately 3.9% per decile, so a producer at median quality earns virtually the same margin as one at the top of the distribution. If the treatment-induced steepening of 1.4 percentage points obtained in the market—with costs unchanged—a producer moving from median to top-decile quality would earn a markup roughly 7.5% higher. Whether this shift is enough to incentivize upgrading is a complex empirical question beyond the scope of this paper. However, this back-of-the-envelope quantifies a potentially non-negligible contribution of information frictions to the compression of returns to quality.

Fourth, we address two dimensions of external validity. The first is retention: would consumers retain this information if purchasing weeks or months later? Our design does not speak to this question, but the policy-relevant intervention need not require long-term retention. Point-of-sale information provision (through signage, brochures, or brief seller explanations) could equip consumers with quality markers precisely when purchase decisions are made. The second concern is whether quality discernment and stated willingness to pay translate into actual purchasing behavior. On discernment, the ranking task was incentivized, participants evaluated real furniture, and treated consumers converged to insider-level performance, suggesting genuine skill acquisition. On willingness to pay, the outcome is hypothetical and participants were not committing real resources. However, the tight correspondence between baseline experimental responses and market price patterns (4.2% vs. 3.9% per quality decile), as well as our focus on price gradients rather than levels, provides reassurance that participants approached the task as they would a real transaction.

6 Conclusions

This paper characterizes a setting where producers' markups do not increase with quality, limiting the incentives for quality upgrading. We posit that information frictions are in part responsible for this lack of incentives. In particular, if customers do not possess the relevant information to tell high-quality and low-quality products apart, even with strong preferences for quality, the market carries inefficiently low-quality. We show that a simple information provision intervention is effective at improving the ability of non-expert consumers to discern quality in the market for wooden furniture. Importantly, we show that, following the provision of information, individuals' willingness to pay for low-quality items goes down, and that for high-quality items goes up. As a result, the intervention steepens the price-quality gradient, shifting demand-side incentives favorably for quality upgrading.

References

Akerlof, George A., "The Market for "Lemons": Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 1970, *84* (3), 488–500.

Alfaro-Ureña, Alonso, Isabela Manelici, and Jose P. Vasquez, "The Effects of Joining Multinational Supply Chains: New Evidence from Firm-to-Firm Linkages," *The Quarterly Journal of Economics*, 2022, *137* (3), 1495–1552.

Anderson, Michael, "Multiple Inference and Gender Differences in the Effects of Early Intervention: A Reevaluation of the Abecedarian, Perry Preschool, and Early Training Projects," *Journal of the American Statistical Association*, 2008, *103* (484), 1481–1495.

_ **and Jeremy Magruder**, "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database," *The Economic Journal*, 2012, *122* (563), 957–989.

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja, "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets," *American Economic Review*, June 2017, *107* (6), 1535–63.

Armstrong, Mark and Yongmin Chen, "Inattentive Consumers and Product Quality," *Journal of the European Economic Association*, 2009, *7* (2-3), 411–422.

Atkin, David, Amit K. Khandelwal, and Adam Osman, "Exporting and Firm Performance: Evidence from a Randomized Experiment," *The Quarterly Journal of Economics*, 2017, *132* (2), 551–615.

Bai, Jie, "Melons as Lemons: Asymmetric Information, Consumer Learning and Seller Reputation," *The Review of Economic Studies*, 2025, *92* (6), 3574–3610.

Banerjee, Abhijit V. and Esther Duflo, "The Economic Lives of the Poor," *Journal of Economic Perspectives*, 2007, *21* (1), 141–167.

Bassi, Vittorio, Raffaela Muoio, Tommaso Porzio, Ritwika Sen, and Esau Tugume, "Achieving Scale Collectively," *Econometrica*, 2022, *90* (6), 2937–2978.

Becker, Gordon M, Morris H DeGroot, and Jacob Marschak, "Measuring Utility by a Single-Response Sequential Method," *Behavioral Science*, 1964, *9* (3), 226–232.

Bold, Tessa, Kayuki C. Kaizzi, Jakob Svensson, and David Yanagizawa-Drott, "Lemon Technologies and Adoption: Measurement, Theory and Evidence from Agricultural Markets in Uganda," *The Quarterly Journal of Economics*, 2017, *132* (3), 1055–1100.

_, Selene Ghisolfi, Frances Nsonzi, and Jakob Svensson, "Market Access and Quality Upgrading: Evidence from Four Field Experiments," *American Economic Review*, 2022, *112* (8), 2518–52.

Cajal-Grossi, Julia, England Can, Jonathan Morduch, Lore Vandewalle, and Christopher Woodruff, "Quality Upgrading in Ugandan Carpentry," AEA RCT Registry April 2025.

Dizon-Ross, Rebecca and Seema Jayachandran, "Improving Willingness-to-Pay Elicitation by Including a Benchmark Good," *AEA Papers and Proceedings*, 2022, *112*, 551–55.

Dranove, David and Ginger Zhe Jin, "Quality Disclosure and Certification: Theory and Practice," *Journal of Economic Literature*, 2010, *48* (4), 935–963.

Gabaix, Xavier and David Laibson, "Shrouded Attributes, Consumer Myopia, and Information Suppression in Competitive Markets," *The Quarterly Journal of Economics*, 2006, *121* (2), 505–540.

Horowitz, John K., "The Becker-Degroot-Marschak Mechanism Is Not Necessarily Incentive Compatible, Even for Non-Random Goods," *Economics Letters*, 2006, *93* (1), 6–11.

Hsieh, Chang-Tai and Benjamin A. Olken, "The Missing "Missing Middle"," *Journal of Economic Perspectives*, 2014, *28* (3), 89–108.

Hsu, Eric and Anne Wambugu, "Can Informed Buyers Improve Goods Quality? Experimental Evidence from Crop Seeds," *G2LM— LIC Working Paper*, 2022, *74.*

Imbens, Guido W. and Jeffrey M. Wooldridge, "Recent Developments in the Econometrics of Program Evaluation," *Journal of Economic Literature*, 2009, *47* (1), 5–86.

Jensen, Robert and Nolan H. Miller, "Market Integration, Demand, and the Growth of Firms: Evidence from a Natural Experiment in India," *American Economic Review*, 2018, *108* (12), 3583–3625.

Jin, Ginger Zhe and Phillip Leslie, "The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards," *The Quarterly Journal of Economics*, 2003, *118* (2), 409–451.

Michelson, Hope, Anna Fairbairn, Brenna Ellison, Annemie Maertens, and Victor Manyong, "Misperceived Quality: Fertilizer in Tanzania," *Journal of Development Economics*, 2021, *148*, 102579.

Schwab, Benjamin, Sarah Janzen, Nicholas P. Magnan, and William M. Thompson, "Constructing a Summary Index Using the Standardized Inverse-Covariance Weighted Average of Indicators," *The Stata Journal*, 2020, *20* (4), 952–964.

Verhoogen, Eric, "Trade, Quality Upgrading, and Wage Inequality in the Mexican Manufacturing Sector," *The Quarterly Journal of Economics*, 2008, *123* (2), 489–530.

_ , "Firm-Level Upgrading in Developing Countries," *Journal of Economic Literature*, 2023, *61* (4), 1410–64.

Tables and Figures

Table 1: Records of Interactions with Customers: Carpenters' Ledgers

	All Ledgers		Sales from Stock		Order Placement		Order Collection		Visits	
	Count (1)	Mean (2)	Count (3)	Mean (4)	Count (5)	Mean (6)	Count (7)	Mean (8)	Count (9)	Mean (10)
Panel A: How did the Customer Find your Business										
Returning customer	6,223	0.39	1,627	0.32	2,011	0.49	1,318	0.47	1,267	0.22
Referral or Family/Friend	6,223	0.27	1,627	0.22	2,011	0.30	1,318	0.32	1,267	0.22
Passerby or Others	6,223	0.36	1,627	0.47	2,011	0.23	1,318	0.24	1,267	0.56
Panel B: Customer Type										
Individual consumer	5,733	0.88	1,359	0.87	1,950	0.87	1,267	0.88	1,157	0.92
Industry Insider	5,733	0.12	1,359	0.13	1,950	0.13	1,267	0.12	1,157	0.08
Panel C: Bargaining										
Final < Ask price	3,618	0.90	1,613	0.92	2,005	0.88				
Ratio Final/Ask price	3,618	0.84	1,613	0.84	2,005	0.84				
Customer bargained price at collection	1,415	0.31					1,415	0.31		
Ratio Final/Agreed price	393	0.86					393	0.86		
Share paid today	4,790	0.65	1,518	0.88	1,992	0.56	1,280	0.53		
Disagreement on price	548	0.90							548	0.90
Disagreement on any other dimension	548	0.15							548	0.15
Panel D: Features of Visits										
Duration (min) of customer visit	1,199	26.26							1,199	26.26
Customer looking for a specific product	1,288	0.96							1,288	0.96
Customer visited beforehand	1,241	0.80							1,241	0.80
Discussed price	722	0.94							722	0.94
Discussed modifications	732	0.15							732	0.15

The table shows statistics from ledgers collected as part of the study in Cajal-Grossi et al. (2025). A random sample of 748 carpenters across five Ugandan districts filled in ledgers fortnightly for a period of two to three months, recording a (random) selection of interactions with customers. In particular, a group of 250 carpenters was asked to record the first four sales, the first four orders picked up, and the first four orders placed on the day after the interview. Another group of 500 carpenters was asked to record the first two sales, the first three orders picked up, and the first two orders placed. In total, 598 carpenters submitted ledgers, with an average of 11 entries. Coverage varies because information was collected only for the ledger type shown in each hypercolumn and because some questions were not presented to all customers for each ledger type. Panel A reports mutually exclusive responses to the associated question about how customers found the business. In this panel, *Referral or Family/Friend* groups "*referral by other customers*" and "*the customer is a family member, friend or acquaintance*", while *Passerby or Others* groups "*the customer had seen an advertisement I had done for the workshop*", "*saw shop/window display from the street*", and "*other*". Panel B classifies the type of customer the ledger reports on. Panel C summarizes bargaining outcomes, price deviations, payment timing, and transactional disagreements. *Final < Ask price* is a dummy for the final price agreed upon being lower than the initial asking price set by the carpenter. *Ratio Final/Ask price* is the ratio between the former and the latter prices, both made conditional on the final price being lower than or equal to the asking price (bottom 1% winsorized). Within the same panel, *Customer bargained price at collection* is a dummy for the specified occurrence, and *Ratio Final/Agreed price* is the ratio between the final price agreed upon and the price agreed during placement, conditional on the occurrence of *Customer bargained price at collection* and the final price being lower than or equal to the agreed one. *Share paid today* is the ratio between the part of the price that was paid on the same day the ledger was filled out and the final price agreed upon. *Disagree on price* and *Disagree on any other dimension* report non-mutually exclusive responses to the question regarding the reasons customers decided not to buy, conditional on at least one reason being provided. The latter includes disagreement on date, product specification, or other reasons different from price. Panel D summarizes customer engagement and interaction dynamics. *Discussed price* and *Discussed modifications* are only available for 722 and 732 of the 1,288 ledgers that are *Visits* with no sale. The other variables in this panel have self-explanatory labels.

Table 2: Price-Quality Gradient in the Market and the Experiment

	Market for Tables in Workshops			Experiment		
				Pre-Treatment WTP		
	Unit Cost (1)	Markup Factor (2)	Unit Price (3)	Consumers (4)	Insiders (5)	Master Carpenters (6)
Quality (1-10)	0.038***	0.001	0.039***	0.042***	0.064***	0.124***
	(0.014)	(0.006)	(0.013)	(0.002)	(0.005)	(0.011)
Mean Outcome (levels, th.)	124.92	1.88	224.95	58.77	71.99	173.38
Mean Quality	5.56	5.56	5.56	4.98	5.02	4.72
R^2	0.60	0.25	0.58	0.10	0.24	0.56
Obs.	315	315	315	3,635	850	120

Bootstrapped standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Bootstrapped standard errors are clustered at the carpenter, respondent, and master carpenter level for the *Market*, the *Experiment*, and the *Master Carpenters* regressions, respectively. Bootstrapping is used as a conservative approach to account for the additional sampling variability introduced by the prior estimation of the quality index. The table shows the outcome of linear regressions of different price-related outcomes on the quality of carpentry items. The panel *Market for Tables in Workshops* restricts attention to tables truly offered in the market by carpenters in the study by Cajal-Grossi et al. (2025). The panel *Experiment* considers tables presented to customers in the context of our experiment. *Quality (1-10)* is a continuous quality score based on the physical assessment of the items (see Supplemental Appendix B), recast into ten bins. In the *Experiment*, these correspond to deciles of the distribution of the quality score of tables in the market, such that a table in the experiment that is assigned to bin 1 corresponds to a table comparable to tables in the lowest decile of the quality distribution in the market. Analogously, a table in the experiment assigned to bin 10 has a quality comparable to that of the top decile in the market. A unit of observation is an item in the *Market for Tables in Workshops* and a combination of a table and a respondent in the *Experiment*, except for column (6) under the header *Master Carpenters* for which a unit of observation is a table and master carpenter combination. All outcomes are expressed in logarithms. In the *Market for Tables in Workshops*, column (1) reports the logarithmic unit production cost per table, column (2) the logarithmic markup factor, and column (3) the logarithmic price listed by the carpenter. The production costs (including wood, other material inputs, labor, etc.) and prices of each item are collected in fortnightly surveys. The top and bottom 1% of the price distribution are excluded in columns (1), (2), and (3). In the *Experiment*, in columns (4) and (5), the outcome variable is the baseline willingness to pay by the respondent for the table, i.e., the answer expressed before treatment in logarithmic UGX from the 897 respondents to the question on what maximum price they would be willing to pay for each table. Column (4) studies *Consumers*, i.e., respondents that are not *Insiders*, and column (5) considers *Insiders*, i.e., the 170 respondents that are carpenters, furniture middlemen, or re-sellers. Column (6) considers the logarithmic price that master carpenters consider as a suitable market price for the item at hand. The regressions include fixed effects for varieties –item type and size, wood species, ornamentation, finishing type– and interviewer. The bottom panel reports the mean of the outcome in UGX 1,000 levels (except for the markup factor), and the mean of the main regressor, the quality score. We note that while the carpenters' survey used for the analysis in columns (1)–(3) specifically collects information on unit production costs, such data are not available for our experimental items. The items were commissioned by our research team from actual carpenters at generous prices that likely influenced their input choices. While this does not affect our experimental design or results, it prevents us from constructing meaningful markups for the experimental items. The results in this table are reported in the Supplemental Appendix A4 using alternative measures of quality, and in Supplemental Appendix Figure A4 allowing for a non-linear relationship between prices and quality.

Table 3: Baseline Descriptives and Balance

	All Customers			Consumers (81.05%)			Industry Insiders (18.95%)		
	Mean (SD) (1)	Estimate (SE) (2)	Normalized Difference (3)	Mean (SD) (4)	Estimate (SE) (5)	Normalized Difference (6)	Mean (SD) (7)	Estimate (SE) (8)	Normalized Difference (9)
Panel A. Demographics & Household									
Female *(0,1)*	0.238 (0.426)	0.055* (0.029)	0.088	0.283 (0.451)	0.062* (0.034)	0.095	0.054 (0.228)	-0.003 (0.035)	-0.010
Age *(#)*	31.458 (8.942)	-0.539 (0.589)	-0.043	30.722 (8.737)	-0.015 (0.641)	-0.001	34.424 (9.191)	-2.539* (1.421)	-0.194
Household Size *(#)*	3.421 (1.938)	0.118 (0.138)	0.040	3.353 (1.929)	0.268* (0.154)	0.091	3.696 (1.960)	-0.529* (0.308)	-0.187
Completed Primary Ed. *(0,1)*	0.657 (0.475)	0.029 (0.031)	0.044	0.687 (0.464)	0.008 (0.034)	0.013	0.533 (0.502)	0.108 (0.076)	0.156
Household Food Exp. > Med. *(0,1)*	0.477 (0.500)	0.002 (0.033)	0.003	0.466 (0.500)	0.008 (0.037)	0.012	0.522 (0.502)	-0.022 (0.077)	-0.031
High Income Household *(0,1)*	0.135 (0.342)	-0.016 (0.022)	-0.034	0.117 (0.321)	-0.009 (0.024)	-0.019	0.209 (0.409)	-0.042 (0.061)	-0.076
Wage Employment *(0,1)*	0.261 (0.440)	0.004 (0.029)	0.007	0.313 (0.464)	-0.006 (0.034)	-0.009	0.054 (0.228)	0.023 (0.038)	0.064
Self-Employed *(0,1)*	0.518 (0.500)	0.001 (0.033)	0.002	0.445 (0.498)	0.031 (0.037)	0.044	0.815 (0.390)	-0.097 (0.065)	-0.163
Unemployed *(0,1)*	0.028 (0.165)	-0.014 (0.010)	-0.070	0.035 (0.184)	-0.018 (0.012)	-0.081			
Panel B. Experiment Outcomes									
Complete Ranking Correct *(0,1)*	0.261 (0.440)	-0.075*** (0.028)	-0.127	0.210 (0.408)	-0.036 (0.029)	-0.065	0.467 (0.502)	-0.224*** (0.072)	-0.338
Count Correct Rankings *(#, 0-4)*	2.672 (1.145)	-0.273*** (0.078)	-0.165	2.555 (1.150)	-0.238*** (0.088)	-0.142	3.141 (1.001)	-0.372** (0.158)	-0.256
Easiest Ranking Correct *(0,1)*	0.961 (0.194)	0.002 (0.013)	0.007	0.957 (0.203)	0.001 (0.015)	0.003	0.978 (0.147)	0.009 (0.020)	0.048
Hardest Ranking Correct *(0,1)*	0.590 (0.492)	-0.062* (0.033)	-0.088	0.561 (0.497)	-0.049 (0.037)	-0.070	0.707 (0.458)	-0.104 (0.073)	-0.155
Willingness to Pay (UGX 1,000)	58.707 (37.365)	5.319** (2.127)	0.096	55.878 (36.199)	5.915** (2.372)	0.108	70.113 (39.788)	4.100 (4.392)	0.073
Obs.	463	897	897	371	727	727	92	170	170

Standard deviations (columns (1), (4), and (7)) and robust standard errors (columns (2), (5), and (8)) are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table shows statistics and the outcome of linear regressions on treatment status of the respondents' characteristics on rows. A unit of observation is a respondent, except in row *Willingness to Pay (UGX 1,000)*, where the unit of observation is a combination of one respondent and one table. Of the 897 respondents, 727 are individual consumers and 170 are industry insiders. Columns (1), (4), and (7) report the mean and the standard deviation for the variable in the respective row, across the respondents in the control group. Columns (2), (5), and (8) report, together with the standard errors, the differences in means for the respective row between the control and treatment group as resulting from the estimations on respondents in the column (that is, the β in $y_i = \beta\, Treatment_i + \varepsilon_i$, where y_i is the variable on the respective row for respondent i, and $Treatment_i$ their treatment assignment). Columns (3), (6), and (9) report the difference between the control and the treated respondents, computed as the difference in means between the treatment and control observations divided by the square root of the sum of their variances (Imbens and Wooldridge, 2009). All the normalized differences are smaller than one fourth of the combined sample variation. Most variables have self-explanatory labels. In Panel A, *Household Food Exp. > Med.* is a dummy for respondents whose household's expenditure on food in the previous 7 days was above median. *High Income Household* is a dummy for the 12.60% of respondents whose household's income in the previous 30 days was in the top 3 brackets (out of 8), i.e., above 2,000,000 UGX. In the same panel, the estimates for unemployment are omitted in columns (7)–(9) since all industry insiders are employed. In Panel B, *Complete Ranking Correct (0,1)*, *Easiest Ranking Correct (0,1)*, and *Hardest Ranking Correct (0,1)* are, respectively, a dummy for the respondent correctly ranking all tables at baseline, correctly ranking at baseline the tables for which the ratings of the master carpenters differ the most, and correctly ranking at baseline the tables for which the difference in the ratings of the master carpenters is the smallest. *Count Correct Rankings (#)* is the number of tables the respondent ranked correctly at baseline. While each respondent ranks five tables, if four are ranked correctly, the fifth one also is, hence *Count Correct Rankings (#)* spanning only the [0,4] interval. *Willingness to Pay (UGX 1,000)* is the baseline willingness to pay by the respondent for a table, i.e., the answer expressed before treatment to the question on what maximum price they would be willing to pay for a table in thousands of UGX.

Table 4: Treatment Effects on Quality Discernment

	Complete ranking correct (*1,0*)		Count correct rankings (#, *0-4*)		Easiest ranking correct (*1,0*)		Hardest ranking correct (*1,0*)	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Individual Consumer		-0.102***		-0.275***		-0.019**		-0.042
		(0.035)		(0.072)		(0.009)		(0.027)
Treated	0.229***	0.109**	0.499***	0.094	0.029***	0.006	0.140***	0.059
	(0.025)	(0.055)	(0.053)	(0.111)	(0.007)	(0.011)	(0.022)	(0.042)
Individual Consumer × Treated		0.148**		0.499***		0.028**		0.099**
		(0.062)		(0.127)		(0.013)		(0.050)
Total effect for treated		0.257***		0.593***		0.034***		0.159***
standard error		(0.029)		(0.060)		(0.008)		(0.026)
P-value treatment		0.004		0.000		0.024		0.037
Location FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Control	0.302	0.533	2.840	3.348	0.963	0.989	0.613	0.728
R^2	0.40	0.40	0.43	0.44	0.53	0.53	0.50	0.50
Obs.	897	897	897	897	897	897	897	897

Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table shows the results of linear regressions of the experiment outcomes after treatment on the respondent's type and treatment status, following equation (1). The regressions include location fixed effects (Kigo stall or Nateete stall). A unit of observation is a respondent. In columns (1) and (2), the dependent variable is a dummy for the respondent ranking all tables correctly. In columns (3) and (4), the dependent variable is the number of tables the respondent ranked correctly. While each respondent ranks five tables, the variable only takes values from 0 to 4, since if four tables are ranked correctly, the fifth is necessarily ranked correctly. In columns (5) and (6), the dependent variable is a dummy for the respondent correctly ranking the tables for which the ratings of the master carpenters differ most. In columns (7) and (8), the dependent variable is a dummy for the respondent correctly ranking the tables for which the difference in the ratings of the master carpenters is the smallest. The correctness of the respondent's rankings is based on the ratings of the master carpenters. The lower panel reports the *total treatment effect for treated* and the statistical significance of the difference between *Individual Consumer* and *Individual Consumer* × *Treatment*. *Individual Consumers* are respondents who are not carpenters, furniture middlemen, or re-sellers. Supplemental Appendix Figure A3 includes a depiction of the construction of the outcomes on quality discernment.

Table 5: Treatment Effect on Price for Quality

	Willingness to Pay (UGX 1,000)								
	ANCOVA			Before Treatment			After Treatment		
	All Customers (1)	Individual Consumers (2)	Industry Insiders (3)	All Customers (4)	Individual Consumers (5)	Industry Insiders (6)	All Customers (7)	Individual Consumers (8)	Industry Insiders (9)
Panel A: Baseline Specifications									
Quality (1-10)	0.953***	0.848***	1.390***	7.187***	6.309***	10.856***	7.592***	6.652***	11.529***
	(0.143)	(0.147)	(0.340)	(0.331)	(0.353)	(0.730)	(0.334)	(0.344)	(0.821)
Treated	-7.645***	-9.103***	-0.992	3.987	3.301	7.073	-3.961	-6.067**	5.614
	(1.396)	(1.563)	(3.000)	(2.857)	(3.067)	(6.967)	(2.708)	(2.814)	(7.483)
Treated × Quality (1-10)	1.380***	1.604***	0.387	0.163	0.378	-0.620	1.531***	1.952***	-0.192
	(0.242)	(0.272)	(0.516)	(0.512)	(0.535)	(1.314)	(0.525)	(0.541)	(1.434)
Location FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Control	61.67	59.15	72.45	61.28	58.78	71.99	61.67	59.15	72.45
R^2	0.91	0.90	0.92	0.15	0.13	0.28	0.19	0.16	0.31
Obs.	4484	3634	850	4484	3634	850	4484	3634	850
Panel B: Respondent's Own Quality Rating									
Quality (1-10)	0.815***	0.750***	1.076***	6.032***	5.664***	7.689***	6.207***	5.688***	8.404***
	(0.116)	(0.122)	(0.288)	(0.306)	(0.339)	(0.650)	(0.318)	(0.339)	(0.768)
Treated	-5.825***	-7.126***	-0.398	1.814	2.485	-0.519	-1.839	-3.259	3.154
	(0.993)	(1.092)	(2.297)	(2.901)	(3.204)	(6.418)	(2.692)	(2.939)	(6.431)
Treated × Quality (1-10)	1.057***	1.245***	0.290	0.467	0.471	0.521	1.015**	1.303***	0.161
	(0.173)	(0.190)	(0.410)	(0.487)	(0.538)	(1.066)	(0.466)	(0.496)	(1.182)
Location FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Control	61.67	59.15	72.45	61.28	58.78	71.99	61.67	59.15	72.45
R^2	0.91	0.90	0.92	0.18	0.16	0.27	0.20	0.19	0.30
Obs.	4484	3634	850	4484	3634	850	4484	3634	850
Panel C: Physical Quality Index									
Quality (1-10)	0.358***	0.303***	0.584***	2.972***	2.514***	4.859***	3.132***	2.639***	5.168***
	(0.060)	(0.059)	(0.170)	(0.157)	(0.167)	(0.376)	(0.163)	(0.166)	(0.440)
Treated	-1.807***	-2.405***	0.922	4.696**	4.326*	6.461*	2.576	1.614	7.017**
	(0.562)	(0.606)	(1.120)	(2.062)	(2.272)	(3.805)	(1.912)	(2.109)	(3.568)
Treated × Quality (1-10)	0.505***	0.599***	0.108	0.081	0.257	-0.445	0.581**	0.837***	-0.312
	(0.119)	(0.131)	(0.228)	(0.272)	(0.272)	(0.625)	(0.275)	(0.282)	(0.701)
Location FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Control	61.67	59.15	72.45	61.28	58.78	71.99	61.67	59.15	72.45
R^2	0.91	0.90	0.92	0.12	0.10	0.23	0.14	0.12	0.26
Obs.	4484	3634	850	4484	3634	850	4484	3634	850

Robust standard errors clustered at the respondent level are reported in parentheses in Panels A and B, while Panel C reports bootstrapped standard errors clustered at the respondent level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. In panel C, bootstrapping is used as a conservative approach to account for the additional sampling variability introduced by the prior estimation of the quality index. The table shows the results of linear regressions of the respondent's willingness to pay on treatment status and three different quality measures, following equation (2). In Panel A, *Quality (1-10)* is the median rating assigned to tables by the three master carpenters on a scale from 1 to 10, where 10 is the best quality produced in Uganda. In Panel B, *Quality (1-10)* is the rating assigned to the table either before or after the treatment by the respondent on a scale from 1 to 10, where 10 is the best produced in Uganda. In Panel C, *Quality (1-10)* is the recast quality score of tables in the experiment. This corresponds to deciles of the distribution of the quality score of tables in the market, such that a table in the experiment that is assigned to bin 1 corresponds to a table comparable to tables in the lowest decile of the quality distribution in the market. All regressions include location fixed effects (Kigo stall or Nateete stall). In all panels, a unit of observation is a combination of one respondent and one table, for a total of 897 respondents and 5 tables per respondent. In columns (1), (2), (3), (7), (8), and (9), the dependent variable is the respondent's willingness to pay after the treatment. In columns (4), (5), and (6), the dependent variable is the respondent's willingness to pay before the treatment. The outcome measured at baseline *Willingness to Pay Baseline* is included as a regressor in the ANCOVA specifications of columns (1), (2), and (3). The bottom panel reports the mean of the outcome in UGX 1,000 levels.

Quality Unobserved: Can Information Provision Unlock Demand-Side Incentives for Upgrading in Low-income Countries?

Julia Cajal-Grossi† Lore Vandewalle ‡ Christopher Woodruff §

Supplemental Materials (Not for Publication)

†Geneva Graduate Institute (IHEID), Switzerland. E-mail: julia.cajal@graduateinstitute.ch.

‡KU Leuven, Belgium, and Geneva Graduate Institute (IHEID), Switzerland. E-mail: lore.vandewalle@kuleuven.be.

§Oxford University, United Kingdom. E-mail: christopher.woodruff@economics.ox.ac.uk.

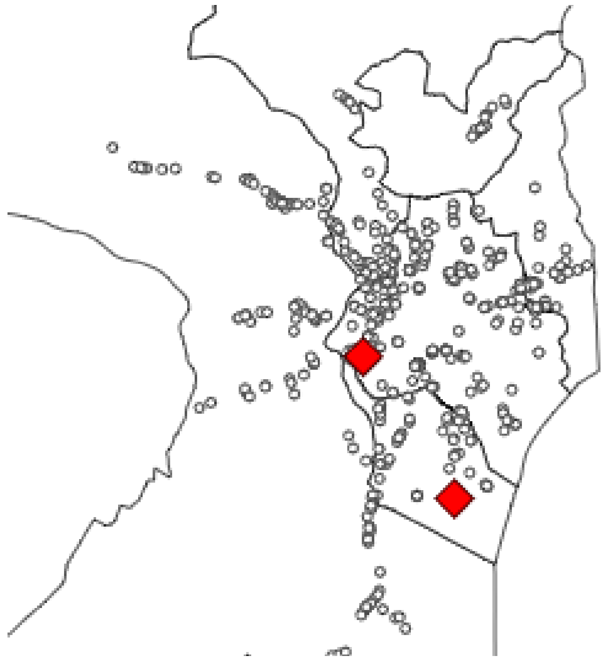
A Appendix Tables and Figures

Table A1: Characteristics of Carpentry Workshops

	Count (1)	Mean (2)	Std. Dev (3)	P10 (4)	P25 (5)	P50 (6)	P75 (7)	P90 (8)
Panel A: Business Profile								
Years of operation	748	9.73	7.06	3.00	4.00	8.00	15.00	20.00
Sales in last 12 months (1,000 UGX)	741	20,444	31,016	3,000	6,500	12,000	23,000	45,000
Number of product types sold	706	6.25	2.58	3.00	4.00	6.00	8.00	10.00
Share of top product in sales	748	0.65	0.18	0.40	0.50	0.60	0.80	0.95
Number of workers	656	4.61	3.06	2.00	3.00	4.00	6.00	8.00
Has carpentry certificate	682	0.24						
Has secondary education or higher	748	0.60						
Number of mechanized steps	748	5.66	2.06	3.00	4.00	6.00	7.00	8.00
Panel B: Expenditure Shares (Last 12 Months)								
Wood	745	0.41	0.17	0.21	0.31	0.42	0.52	0.62
Wages	745	0.15	0.10	0.04	0.08	0.13	0.20	0.28
All others	748	0.44	0.16	0.25	0.33	0.42	0.52	0.64
Panel C: Characterization of Demand								
Share of sales to individual consumers	746	0.86	0.25	0.40	0.80	1.00	1.00	1.00
Share of sales to industry insiders	746	0.08	0.21	0.00	0.00	0.00	0.00	0.30
Share of sales to others	746	0.06	0.16	0.00	0.00	0.00	0.00	0.20
New orders placed per week	707	1.23	0.91	0.38	0.63	1.00	1.50	2.25
Delivered orders per week	707	0.79	0.70	0.00	0.25	0.63	1.00	1.75
Items sold from stock per week	707	1.42	2.18	0.00	0.25	0.75	1.75	3.63
Panel D: Location of Competitors								
Distance to closest other carpenter (km)	707	0.21	0.43	0.00	0.01	0.06	0.24	0.57
Panel E: Carpenters' Quality								
Number of items	529	6.06	5.78	1.00	2.00	5.00	6.00	14.00
Quality (p50)	524	-0.01	0.88	-0.99	-0.50	-0.03	0.49	1.05

The table shows descriptive statistics of the characteristics of the 748 carpenters surveyed in Cajal-Grossi et al. (2025). The variation in the count of observations across rows reflects differences in coverage across variables. Most variables have self-explanatory labels. In Panel A, *Has secondary education or higher* is a dummy variable set to one for any of the following categories: Lower Secondary School (with Uganda Certificate of Education), Upper Secondary School (Uganda Advanced Certificate of Education), Post secondary Vocational Professional Training, and/or College (post upper secondary, two to three year, with Diploma). *Number of mechanized steps* is constructed from baseline survey questions on production technology. Carpenters report, for their main wood product sold, whether each of ten production steps (such as wood seasoning, cutting, planing, thickening, edging, sanding, mortising, etc.) is performed and carried out using machinery. Panel B breaks down carpenters' expenses into mutually exclusive categories. In Panel C, *Share of sales to individual consumers*, *Share of sales to industry insiders*, and *Share of sales to others* are mutually exclusive. In Panel D, *Distance to closest other carpenter (km)* reports the minimum distance to the closest competitor in the survey, computed from the geographical location of each carpenter. Panel E focuses on the total number of finished items for each carpenter, for which a physical quality assessment were performed by our field representatives. The quality score is computed on the following residualized individual components: *Joint Tightness*, *Coplanarity*, *No Nails*, *Squareness*, *Levelness*, *No Damage*, and *Sanding*. Each component is residualized using fixed effects for type and size of item, wood species, ornamentation, finishing type, and *assessor* (when applicable). *Quality (p50)* is the median of continuous quality score across all items produced by the carpenter. Supplemental Appendix B discusses the measurement of quality in full detail.

Figure A1: Map of Kampala and Wakiso with Carpenters and Outlets Locations

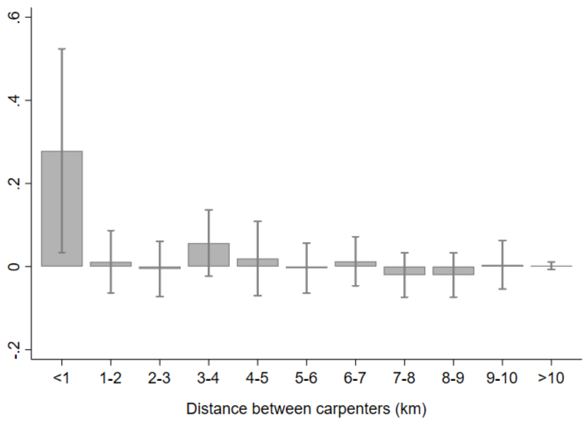


The figure illustrates the location of 563 of the carpenters surveyed in Cajal-Grossi et al. (2025) in the districts of Kampala and Wakiso. Carpenters outside these districts are not shown for visual clarity of the map. The map also shows the location of the two furniture outlets of Kigo and Nateete, where the kiosks were embedded for the experiment. The borders between districts and subdistricts are displayed in black. The carpenters are represented by black circles, while the location of Kigo and Nateete outlets is marked with red diamonds outlined in black.

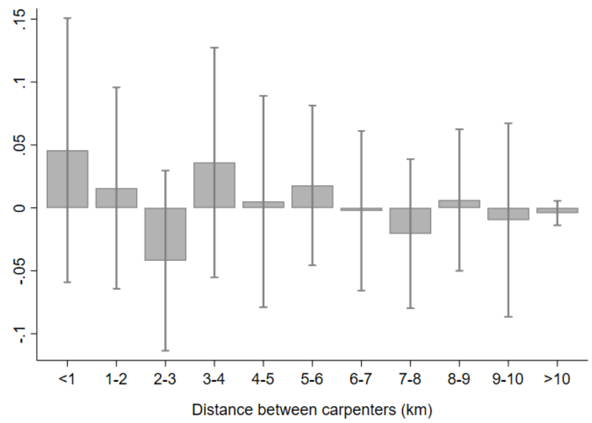
Figure A2: Spatial Autocorrelation of Carpenters' Quality and Prices

Prices

(a) Residualized Against Enumerator

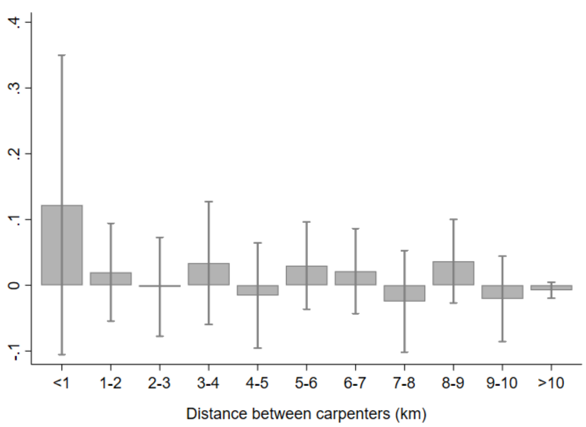


(b) Residualized Against Horizontal Characteristics

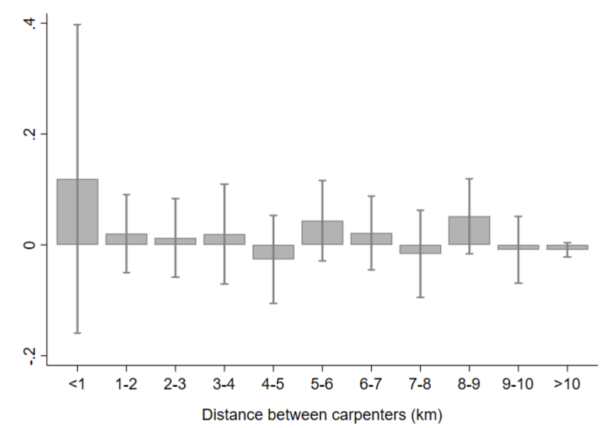


Quality

(c) Residualized Against Enumerator



(d) Residualized Against Horizontal Characteristics



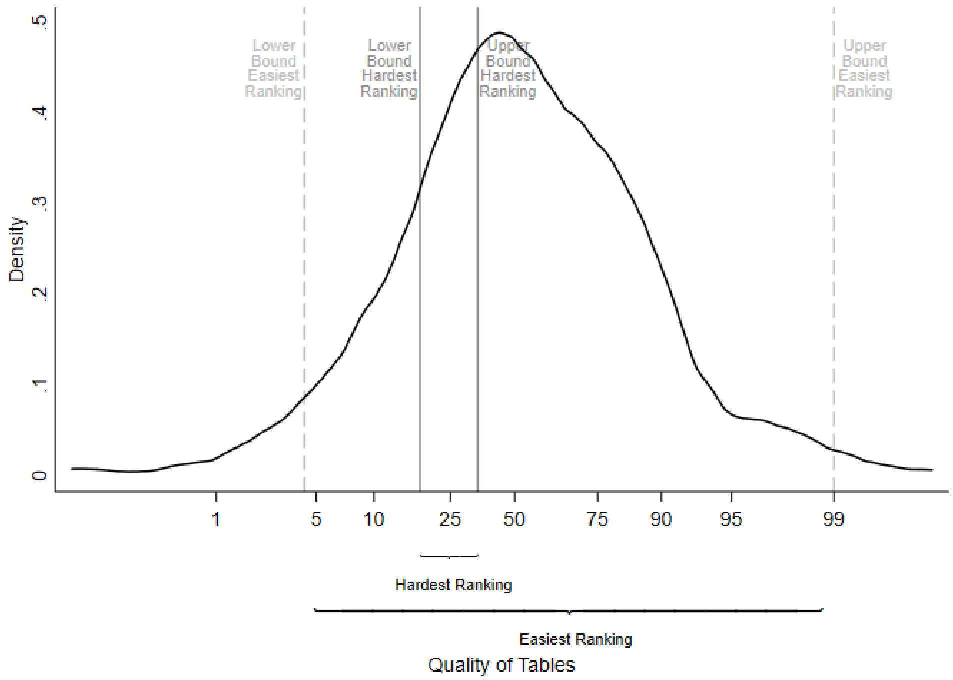
The figure shows the Moran's I measures of spatial autocorrelation between carpenters surveyed in Cajal-Grossi et al. (2025), based on their location and the respective values for *Quality* and *Price*. *Quality* is the quality score computed for products in the market (i.e., finished tables, beds, or door shutters) and *Price* is the price the carpenter expects to obtain once the customer bargains. The bars represent the Moran's I calculated for *Price* in the top panel and *Quality* in the bottom panel. Sub-figures (a) and (c) report spatial correlations of outcomes residualized only against enumerator (or quality assessor) fixed effects, while sub-figures (b) and (c) use outcomes residualized against fixed effects for varieties (product group, type, and size, type of finishing, ornamentation level, wood species) as well as enumerator. For each carpenter, *Quality* and *Price* are the means of the variables across their respective products. The horizontal axis represents distance bins, each defined by a specific weight matrix used to calculate Moran's I for the corresponding distance band. For example, in the 1 to 2 km distance bin, the weights for carpenters i and j in the Moran's I calculations take value > 0 only if the distance between the workshops falls inside that range. The Moran's I is calculated as follows, where N is the sample size, x_i corresponds to the value for carpenter i (either for *Quality* or *Price*), and W is the sum across all elements w_{ij} of the weight matrix, with elements $w_{ij} > 0$ if the distance between i and j is within a given range, and $w_{ij} = 0$ otherwise: $I = \frac{N}{W} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$. The number of carpenter pairs is 3,236 below 1 km, 2,446 between 1 and 2 km, 3,214 between 2 and 3 km, 3,120 between 3 and 4 km, 4,288 between 4 and 5 km, 4,874 between 5 and 6 km, 4,806 between 6 and 7 km, 5,364 between 7 and 8 km, 5,712 between 8 and 9 km, 6,518 between 9 and 10 km, and 129,062 beyond 10 km.

Table A2: Customer's Purchase Behavior

	Count (1)	Mean (2)	Std. Dev (3)	P10 (4)	P25 (5)	P50 (6)	P75 (7)	P90 (8)
Panel A: Search Behavior								
Search intensity	824	2.84	2.14	1.00	1.00	2.00	3.00	5.00
Ever bought furniture	897	0.92						
Bought furniture last year	897	0.65						
Amount spent on furniture last year (UGX 1,000)	581	679.74	1,383	97	200	400	650	1,200
Panel B: Criteria for Buying Wooden Furniture								
Horizontal attributes	897	0.96						
Wood	897	0.64						
Structual quality	897	0.47						
Price	897	0.31						
Sales service	897	0.07						
Panel C: Explanations of Quality Rankings (Highest Quality)								
Horizontal attributes	897	0.95						
Symmetry	897	0.54						
Joints	897	0.47						
Wood	897	0.36						
Panel D: Explanations of Quality Rankings (Lowest Quality)								
Horizontal attributes	897	0.94						
Joints	897	0.60						
Symmetry	897	0.52						
Wood	897	0.24						

The table shows summary statistics of customer purchase behavior at baseline for the 897 respondents in the experiment. Most variables have self-explanatory labels. Panel A shows information on furniture purchasing behavior. *Search intensity* corresponds to the number of outlets visited before purchasing the most recent piece of wooden furniture. This information is reported by 824 respondents (out of 897). Panel B shows the features listed as possible answers to the question *What are the three most important aspects you consider when buying furniture?*. *Horizontal attributes* include the size, ornamentation, and finishing of the item. *Wood* groups categories of wood species, seasoning, and weight. *Structural quality* refers to workmanship, symmetry, joints, and damage. *Price* refers to the item's price. *Sales service* groups delivery, location, and customer care of the item. Panels C and D show the baseline responses of all individuals in the experiment to the questions *Which features led you to rate this table as the highest quality?* and *Which features led you to rate this table as the lowest quality?*, respectively. In panels C and D, *Horizontal attributes* groups ornamentation, finishing, and size of the item; *Wood* groups wood species, seasoning, and weight; *Symmetry* includes sturdiness, legs, and the top of the item; whereas *Joints* includes squareness and the use of filler or nails.

Figure A3: Definition of Outcomes on Quality Discernment



The figure illustrates the definitions of the outcomes on quality discernment used in our experiment. The density depicted with a solid black line corresponds to the quality score of 416 tables in the market. The quality score is computed from the following residualized individual components: *Joint Tightness, Coplanarity, No Nails, Squareness, Levelness, No Damage*, and *Sanding*. Each component is residualized using fixed effects for type of item and size, wood species, ornamentation, finishing type, and assessor, when applicable. Vertical lines represent the quality scores for the highest-rated and lowest-rated tables in the *easiest* and *hardest* rankings in the experiment. In particular, in the experiment, respondents quality discernment was assessed on the basis of four outcomes measuring the correctness of the customers ranking, relative to the ranking produced by the master carpenters in the study. The four outcomes reflect whether the respondent gets the entire ranking correct, gets the *easiest ranking* correct, gets the *hardest ranking* correct, and the count of items the respondent orders correctly. The *easiest ranking* is the comparison of the two tables presented to the respondent for which the master carpenters ratings differ the most. Analogously, the *hardest ranking* is the comparison of the two tables presented to the respondent for which the difference in ratings is the smallest. Vertical lines illustrate the difficulty in quality discernment implied by the *easiest* and *hardest* rankings. In particular, for both the *easiest* and *hardest* comparisons, we identify the table with the highest rating from the master carpenters (upper bound) and the table with the lowest rating (lower bound). Gray lines are then placed at the quality scores of these tables. The x-axis is recast to present the percentiles of the quality score distribution in the market instead of the quality score itself. Not illustrated in this figure are the two other outcomes on quality discernment, whose definitions are more straightforward: a dummy variable that takes the value one if the ranking of all five tables is entirely correct (*Complete ranking correct*) and a count variable indicating the number of tables correctly allocated to their order in the ranking (*Count correct rankings*).

Table A3: Balance of Outcomes Conditioning on Set of Tables

	All Customers			Consumers (81.05%)			Industry Insiders (18.95%)		
	Mean (SD) (1)	Estimate (SE) (2)	Normalized Difference (3)	Mean (SD) (4)	Estimate (SE) (5)	Normalized Difference (6)	Mean (SD) (7)	Estimate (SE) (8)	Normalized Difference (9)
Complete Ranking Correct *(0,1)*	0.261 (0.440)	-0.073** (0.033)	-0.127	0.210 (0.408)	-0.018 (0.037)	-0.065	0.467 (0.502)	-0.205 (0.141)	-0.338
Count Correct Rankings *(#, 0-4)*	2.672 (1.145)	-0.221** (0.093)	-0.165	2.555 (1.150)	-0.147 (0.112)	-0.142	3.141 (1.001)	-0.693** (0.327)	-0.256
Easiest Ranking Correct *(0,1)*	0.961 (0.194)	0.002 (0.015)	0.007	0.957 (0.203)	0.000 (0.019)	0.003	0.978 (0.147)	-0.024 (0.061)	0.048
Hardest Ranking Correct *(0,1)*	0.590 (0.492)	-0.028 (0.037)	-0.088	0.561 (0.497)	-0.004 (0.043)	-0.070	0.707 (0.458)	-0.079 (0.140)	-0.155
Willingness to Pay (UGX 1,000)	58.707 (37.365)	2.949 (2.259)	0.096	55.878 (36.199)	4.804* (2.725)	0.108	70.113 (39.788)	-2.142 (6.511)	0.073
Obs.	463	858	897	371	672	727	92	77	170

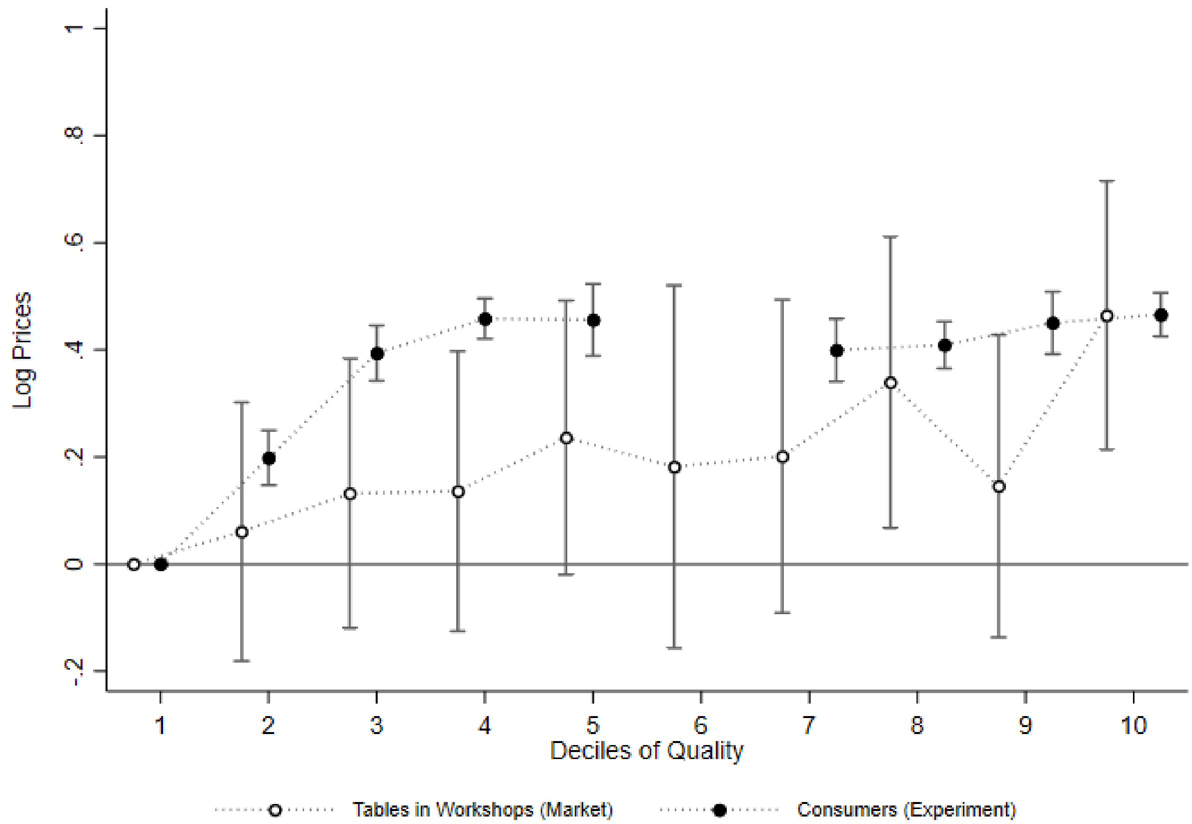
Standard deviations (columns (1), (4), and (7)) and robust standard errors (columns (2), (5), and (8)) are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table reproduces Table 3, with the only difference being that the regressions in columns (2), (5), and (8) include a set of fixed effects that capture the mix of tables presented to each respondent, irrespective of their display order. A unit of observation is a respondent, except in row *Willingness to Pay (UGX 1,000)*, where the unit of observation is a combination of one respondent and one table. Of the 897 respondents, 727 are individual consumers and 170 are industry insiders. Columns (1), (4), and (7) report the mean and the standard deviation for the variable in the respective row, across the respondents in the control group. Columns (2), (5), and (8) report, together with the standard errors, the differences in means for the respective row between the control and treatment group as resulting from the estimations on respondents in the column (that is, the β in $y_i = \phi_m + \beta\, Treatment_i + \varepsilon_i$, where y_i is the variable on the respective row for respondent i, and $Treatment_i$ their treatment assignment). All differences in means include menu fixed effects (ϕ_m). A menu is defined as the combination of tables among the five tables presented to the respondent, regardless of the order in which they were presented. Columns (3), (6), and (9) report the difference between the control and the treated respondents, computed as the difference in means between the treatment and control observations divided by the square root of the sum of their variances (Imbens and Wooldridge, 2009). All the normalized differences are smaller than one fourth of the combined sample variation. Most variables have self-explanatory labels. In Panel A, *Household Food Exp. > Med.* is a dummy for respondents whose household's expenditure on food in the previous 7 days was above median. *High Income Household* is a dummy for the 12.60% of respondents whose household's income in the previous 30 days was in the top 3 brackets (out of 8), i.e., above 2,000,000 UGX. In the same panel, the estimates for unemployment are omitted in columns (7)–(9) since all industry insiders are employed. In Panel B, *Complete Ranking Correct (0,1)*, *Easiest Ranking Correct (0,1)*, and *Hardest Ranking Correct (0,1)* are, respectively, a dummy for the respondent correctly ranking all tables at baseline, correctly ranking at baseline the tables for which the ratings of the master carpenters differ the most, and correctly ranking at baseline the tables for which the difference in the ratings of the master carpenters is the smallest. *Count Correct Rankings (#)* is the number of tables the respondent ranked correctly at baseline. While each respondent ranks five tables, if four are ranked correctly, the fifth one also is, hence *Count Correct Rankings (#)* spanning only the [0,4] interval. *Willingness to Pay (UGX 1,000)* is the baseline willingness to pay by the respondent for a table, i.e., the answer expressed before treatment to the question on what maximum price they would be willing to pay for a table in thousands of UGX.
tab:sumstat

Table A4: Robustness of Price-Quality Gradient Results using Different Measures of Quality

	Market for Tables in Workshops			Experiment		
				Pre-Treatment WTP		
	Unit Cost	Markup Factor	Unit Price	Consumers	Insiders	Master Carpenters
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Variations in the Index Construction						
1. Baseline (Table 2)	0.038***	0.001	0.039***	0.042***	0.064***	0.124***
	(0.014)	(0.006)	(0.013)	(0.002)	(0.005)	(0.011)
2. Continuous Measure	0.136***	-0.003	0.134***	0.113***	0.174***	0.330***
	(0.040)	(0.022)	(0.039)	(0.007)	(0.012)	(0.028)
3. No Residualization	0.045***	0.003	0.048***	0.050***	0.088***	0.160***
	(0.016)	(0.008)	(0.016)	(0.003)	(0.007)	(0.012)
4. Residualization Post-Aggregation	0.036***	0.001	0.037***	0.043***	0.067***	0.128***
	(0.014)	(0.006)	(0.013)	(0.003)	(0.005)	(0.011)
5. No Missing Components	0.036***	0.001	0.037***	0.042***	0.064***	0.124***
	(0.014)	(0.006)	(0.013)	(0.002)	(0.005)	(0.011)
6. Unweighted Standardization	0.036***	-0.001	0.036***	0.041***	0.059***	0.122***
	(0.014)	(0.007)	(0.013)	(0.003)	(0.004)	(0.010)
Panel B: Exclusion of Individual Quality Components						
7. Excluding Coplanarity	0.038***	-0.001	0.037***	0.042***	0.058***	0.121***
	(0.013)	(0.006)	(0.012)	(0.003)	(0.004)	(0.010)
8. Excluding Joint Tightness	0.036***	0.001	0.037***	0.041***	0.062***	0.122***
	(0.014)	(0.006)	(0.013)	(0.002)	(0.005)	(0.011)
9. Excluding No Nails	0.035***	0.002	0.037***	0.045***	0.065***	0.130***
	(0.012)	(0.007)	(0.013)	(0.002)	(0.005)	(0.011)
10. Excluding Squareness	0.038***	0.003	0.041***	0.049***	0.078***	0.144***
	(0.014)	(0.007)	(0.014)	(0.003)	(0.005)	(0.011)
11. Excluding Levelness	0.035**	0.000	0.035**	0.039***	0.060***	0.120***
	(0.015)	(0.006)	(0.014)	(0.003)	(0.005)	(0.010)
12. Excluding No Damage	0.030**	-0.001	0.029**	0.044***	0.067***	0.130***
	(0.014)	(0.007)	(0.014)	(0.003)	(0.005)	(0.011)
13. Excluding Sanding	0.034***	0.002	0.035***	0.039***	0.060***	0.120***
	(0.013)	(0.006)	(0.012)	(0.002)	(0.005)	(0.010)
Mean Outcome	1249.15	1.88	224.95	58.77	71.99	173.38
Obs.	315	315	315	3635	850	120

Bootstrapped standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Bootstrapped standard errors are clustered at the carpenter, at the respondent, and at the master carpenter level for the *Market*, the *Experiment*, and the *Master Carpenters* regressions, respectively. Bootstrapping is used as a conservative approach to account for the additional sampling variability introduced by the prior estimation of the quality index. The table reports the results of re-estimating the regressions in Table 2 under alternative definitions of the quality measure. In the *Market for Tables in Workshops*, column (1) reports the logarithmic unit production cost per table, column (2) the logarithmic markup factor, and column (3) the logarithmic price listed by the carpenter. The top and bottom 1% of the price distribution are excluded in columns (1), (2), and (3). In the *Experiment*, columns (4) and (5), the outcome variable is the baseline willingness to pay (WTP) by the respondent for the table, i.e., the answer expressed before treatment in logarithmic UGX from the 897 respondents to the question of what maximum price they would be willing to pay for each table. Column (4) studies *Consumers*, i.e., respondents that are not *Insiders*, and column (5) considers *Insiders*, i.e., the 170 respondents who are carpenters, furniture middlemen, or re-sellers. Column (6) considers the logarithmic price that master carpenters consider a suitable market price for the item at hand. A unit of observation is an item in the *Market for Tables in Workshops* and a combination of a table and a respondent in the *Experiment*, except for Column (6) under the header *Master Carpenters* whereby a unit of observation is a table and master carpenter combination. *1. Baseline (Table 2)* is the quality score like in Table 2. *2. Continuous Measure* is the continuous quality score without being recast into bins. *3. No Residualization* is the quality score computed and recast into bins analogously to *1. Baseline (Table 2)*, except that no residualization is applied. *4. Residualization Post-Aggregation* applies the same residualization procedure as *1. Baseline (Table 2)*, but ex post on each individual component. *5. No Missing Components* is computed and recast into bins analogously to *1. Baseline (Table 2)*, except for including dummy variables for the missing individual components from the fixed effects in the residualization procedure – we note that missing values in physical attribute variables are extremely rare, with only 11 tables in columns (1)–(3) having any missing index component and no respondents (columns (4)-(5)) exhibiting positive missingness. *6. Unweighted Standardization* is computed and recast into bins analogously to *1. Baseline (Table 2)*, except that the standardized individual components are aggregated through an unweighted sum rather than through a sum with inverse-covariance weights. In Panel B, each row, *Excluding Component* is the quality score like in Table 2, computed excluding the specific *Component*. The regressions include fixed effects for varieties: item type and size, wood species, ornamentation, finishing type, and enumerator (or quality assessor). The bottom panel reports the mean of the outcome in UGX 1,000 levels.

Figure A4: Price-Quality Gradient in the Market and in the Experiment at Baseline



Bootstrapped standard errors are clustered at the carpenter and the respondent level, in the *Market* and *Experiment* regressions, respectively. Bootstrapping is used as a conservative approach to account for the additional sampling variability introduced by the prior estimation of the quality index. The graph presents the estimates of the coefficient on each decile of quality in the regressions of the table-level and table-respondent-level logarithmic prices in the *Market* and the *Experiment*, respectively. In the *Market*, the specification follows $y_j = \sum_{q=2}^{10} \beta_q \times D_{j\in q} + FE + \varepsilon_j$ where y_j is the logarithmic price that the carpenters expect to obtain from table j once the customer bargains and $D_{j\in q}$ is a dummy equal to one if the quality of table j belongs to the q^{th} decile in the distribution of quality in the market. Point estimates for β_q in the *Market* regression are indicated with white, round scatter markers. For *Consumers (Experiment)*, the specification is $y_{ij} = \sum_{q=2}^{10} \beta_q \times D_{j\in q} + FE + \varepsilon_{ij}$ where the outcome y_{ij} is the logarithmic willingness to pay at baseline of individual consumer respondent i for table j – i.e., the answer expressed in logarithmic UGX from participants who are individual consumers (both assigned to the treatment and the control group) to the question of what maximum price they would be willing to pay for each table they are presented with. Point estimates for β_q in the *Experiment* regression are indicated with black, round scatter markers. The *Market* and *Experiment* regressions include fixed effects for varieties (product group, type, and size), type of finishing, ornamentation level, wood species, and enumerator. In addition, the *Experiment* regression further includes respondent fixed effects. We assign tables in the experiment to deciles of quality in the market, $D_{j\in q}$, such that a table in the experiment that is assigned to bin 1 corresponds to a table comparable to tables in the lowest decile of the quality distribution in the market. Analogously, a table in the experiment assigned to bin 10 has a quality comparable to the top decile in the market. None of the tables that were used in the experiment have a quality score that places them in bin 6 as defined in the market. These estimations are performed over a total number of 383 tables in the market, and 4,485 observations in the experiment corresponding to 897 respondents. A regression of the respective specification with the same structure, but with a linear restriction on the deciles of the quality measure gives a slope of 0.035 (S.E. 0.011) in the *Market*, 0.124 (S.E. 0.0108) in the *Experiment*.

Table A5: Randomization Inference Robustness

Equation	Column	Row	Original	Randomization Inference		
			P-Value	P-Value	Lower Bound	Upper Bound
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Table 4						
eqn. (1)	col. (1)	row 2	0.000	0.000	0.000	0.007
eqn. (1)	col. (2)	row 2	0.047	0.030	0.017	0.049
eqn. (1)	col. (2)	row 3	0.018	0.012	0.004	0.026
eqn. (1)	col. (3)	row 2	0.000	0.000	0.000	0.007
eqn. (1)	col. (4)	row 2	0.397	0.380	0.337	0.424
eqn. (1)	col. (4)	row 3	0.000	0.000	0.000	0.007
eqn. (1)	col. (5)	row 2	0.000	0.000	0.000	0.007
eqn. (1)	col. (6)	row 2	0.570	0.620	0.576	0.663
eqn. (1)	col. (6)	row 3	0.033	0.016	0.007	0.031
eqn. (1)	col. (7)	row 2	0.000	0.000	0.000	0.007
eqn. (1)	col. (8)	row 2	0.161	0.144	0.114	0.178
eqn. (1)	col. (8)	row 3	0.047	0.036	0.021	0.056
Panel B: Panel A of Table 5						
eqn. (2)	col. (1)	row 2	0.000	0.000	0.000	0.007
eqn. (2)	col. (1)	row 3	0.000	0.000	0.000	0.007
eqn. (2)	col. (2)	row 2	0.000	0.000	0.000	0.007
eqn. (2)	col. (2)	row 3	0.000	0.000	0.000	0.007
eqn. (2)	col. (3)	row 2	0.741	0.736	0.695	0.774
eqn. (2)	col. (3)	row 3	0.454	0.484	0.439	0.529
eqn. (2)	col. (4)	row 2	0.163	0.178	0.145	0.214
eqn. (2)	col. (4)	row 3	0.750	0.772	0.733	0.808
eqn. (2)	col. (5)	row 2	0.282	0.266	0.228	0.307
eqn. (2)	col. (5)	row 3	0.480	0.486	0.441	0.531
eqn. (2)	col. (6)	row 2	0.311	0.312	0.272	0.355
eqn. (2)	col. (6)	row 3	0.637	0.634	0.590	0.676
eqn. (2)	col. (7)	row 2	0.144	0.156	0.125	0.191
eqn. (2)	col. (7)	row 3	0.004	0.004	0.000	0.014
eqn. (2)	col. (8)	row 2	0.031	0.028	0.015	0.047
eqn. (2)	col. (8)	row 3	0.000	0.000	0.000	0.007
eqn. (2)	col. (9)	row 2	0.454	0.460	0.416	0.505
eqn. (2)	col. (9)	row 3	0.893	0.886	0.855	0.913

The table reports p-values for all treatment or treatment-interacted coefficients shown in Tables 4 and 5 (Panel A) under alternative inference methods. Panels separate different tables. Column (1) reports the estimating equation associated with each panel, using the equation numbering in the main paper. Columns (2) and (3) indicate the column and row location of the corresponding coefficient in the original table. Column (4) reports robust standard-error-based p-values, as obtained from the specifications of the main tables. Columns (5)–(7) report p-values computed using randomization inference. Randomization inference is based on 500 permutations of the treatment indicator. The reported randomization inference p-value equals the share of permutations in which the absolute value of the permuted coefficient is at least as large as the observed estimate. Columns (6) and (7) report the 95 percent confidence interval for this permutation-based p-value.

Table A6: Treatment Effects on Quality Discernment Conditioning on Set of Tables

	Complete ranking correct *(1,0)*		Count correct rankings *(#, 0-4)*		Easiest ranking correct *(1,0)*		Hardest ranking correct *(1,0)*	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Individual Consumer		-0.075		-0.240**		-0.018*		-0.055
		(0.048)		(0.099)		(0.009)		(0.040)
Treated	0.224***	0.082	0.453***	0.006	0.032***	0.015	0.142***	0.017
	(0.031)	(0.066)	(0.062)	(0.141)	(0.008)	(0.014)	(0.027)	(0.056)
Individual Consumer × Treated		0.175**		0.557***		0.021		0.156**
		(0.075)		(0.160)		(0.015)		(0.062)
Total effect for treated		0.258***		0.562***		0.037***		0.173***
standard error		(0.035)		(0.069)		(0.009)		(0.030)
P-value treatment		0.025		0.001		0.086		0.020
Location FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Control	0.301	0.517	2.844	3.326	0.964	0.989	0.622	0.719
R^2	0.57	0.58	0.60	0.61	0.63	0.63	0.65	0.65
Obs.	858	858	858	858	858	858	858	858

Robust standard errors are in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table reproduces Table 4, with the only difference being that all regressions include a set of fixed effects that capture the mix of tables presented to each respondent, irrespective of their display order. The regressions include location fixed effects (Kigo stall or Nateete stall). A unit of observation is a respondent. In columns (1) and (2), the dependent variable is a dummy for the respondent ranking all tables correctly. In columns (3) and (4), the dependent variable is the number of tables the respondent ranked correctly. While each respondent ranks five tables, the variable only takes values from 0 to 4, since if four tables are ranked correctly, the fifth is necessarily ranked correctly. In columns (5) and (6), the dependent variable is a dummy for the respondent correctly ranking the tables for which the ratings of the master carpenters differ most. In columns (7) and (8), the dependent variable is a dummy for the respondent correctly ranking the tables for which the difference in the ratings of the master carpenters is the smallest. The correctness of the respondent's rankings is based on the ratings of the master carpenters. The lower panel reports the *total treatment effect for treated* and the statistical significance of the difference between *Individual Consumer* and *Individual Consumer* × *Treatment*. *Individual Consumers* are respondents who are not carpenters, furniture middlemen, or re-sellers.

Table A7: Treatment Effect on Price for Quality Conditioning on Set of Tables

	Willingness to Pay (UGX 1,000)								
	ANCOVA			Before Treatment			After Treatment		
	All Customers (1)	Individual Customers (2)	Industry Insiders (3)	All Customers (4)	Individual Customers (5)	Industry Insiders (6)	All Customers (7)	Individual Customers (8)	Industry Insiders (9)
Quality (1-10)	1.141***	1.043***	1.896***	7.227***	6.323***	10.739***	7.734***	6.753***	11.533***
	(0.164)	(0.179)	(0.488)	(0.329)	(0.363)	(0.637)	(0.332)	(0.359)	(0.698)
Treated	-7.553***	-9.248***	-0.752	1.599	1.576	-1.031	-6.094**	-7.825**	-1.677
	(1.509)	(1.676)	(3.996)	(3.044)	(3.428)	(8.201)	(3.043)	(3.288)	(9.433)
Treated × Quality (1-10)	1.393***	1.641***	0.459	0.225	0.537	-0.187	1.599***	2.126***	0.291
	(0.258)	(0.295)	(0.602)	(0.516)	(0.548)	(1.305)	(0.539)	(0.564)	(1.459)
Location FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean Control	61.67	59.15	72.45	61.28	58.78	71.99	61.67	59.15	72.45
R^2	0.91	0.91	0.93	0.30	0.32	0.66	0.33	0.35	0.66
Obs.	4484	3634	850	4484	3634	850	4484	3634	850

Standard errors in parentheses, clustered at respondent level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. The table reproduces Panel A of Table 5, with the only difference being that all regressions include a set of fixed effects that capture the mix of tables presented to each respondent, irrespective of their display order. The regressions include location fixed effects (Kigo stall or Nateete stall). A unit of observation is a combination of one respondent and one table, for a total of 897 respondents and 5 tables per respondent. In columns (1), (2), (3), (7), (8), and (9), the dependent variable is the respondent's willingness to pay after the treatment. In columns (4), (5), and (6), the dependent variable is the respondent's willingness to pay before the treatment. The outcome measured at baseline *Willingness to Pay Baseline* is included as a regressor in the ANCOVA specifications of columns (1), (2), and (3). *Quality (1-10)* is the median rating assigned to tables by the three master carpenters on a scale from 1 to 10, where 10 is the best quality produced in Uganda. The bottom panel reports the mean of the outcome in UGX 1,000 levels.

B Measuring Quality

This appendix presents the diagnostic tool that we use for measuring the (vertical) quality of horizontally differentiated furniture items. It then discusses the sources of variability in observed quality in the furniture items used in our experiment, and presents a scalar metric that captures overall vertical quality, combining the most relevant dimensions of heterogeneity.

B.1 Quality Assessment Tool

We partnered with a team of master carpenters (MCs) and a researcher from the Department of Forestry of Makerere University to design a diagnostic tool that assesses the quality of wooden furniture. The tool collects information on horizontal and vertical physical attributes of the item. The horizontal attributes allow us to construct narrow *varieties* defined by the broad class of product, its type, dimensions, finishing, ornamentation level, variety-specific attributes, and the wood species used to produce it. For concreteness, when maximally disaggregated, an example of a variety in our nomenclature would read as *bed, single bed, 90 cm × 180 cm, unstained and varnished, plain ornamentation, with a headboard, no mosquito stands, made of musambya.*

The vertical attributes cover dimensions that impact the functionality and durability of the item, including structural aspects of the joinery, stability, symmetry, and levelness properties, the humidity of the lumber, the sealing of the wood pores, and the presence of cracks and damage, among others. Within a variety, observed variability in these dimensions responds almost exclusively to 'pure' workmanship quality, carpentry-specific managerial practices, capabilities in the organization of work, and the use of hand tools and/or machines.

This diagnostic tool was used to determine the quality of the tables presented to customers in the experiment discussed in the main paper. In Cajal-Grossi et al. (2025), the tool was used to diagnose the quality of virtually all furniture items produced or stored by a random sample of carpenters in five regions in Uganda, over the period between October 2022 and December 2022. This amounts to over 3,600 furniture items, which we leverage for constructing an index of quality, discussed in Section B.3, and for characterizing the relationship between quality and prices in the market, in the main text. Before turning to these, we study the sources of variation in quality across the items used in our experiment.

B.2 Experts' Quality Ratings

Verification of Quality Assessments. In the context of our experiment, by construction, there is no horizontal variation across furniture products – they are all side tables of the same size, design, and wood. This allows us to study quality differences within a variety, hence abstracting away from products' horizontal attributes.[1]

[1]There are two exceptions to this statement. The first is that a handful of the procured tables had minimal design differences that could be detected visually, responding to slight misinterpretations of the commissioned design on the part of the carpenters. Despite the almost negligible discrepancy with the rest

In addition to completing the diagnostic tool, three MCs enlisted to work in our experiment were asked to offer a score, from 1 (lowest) to 10 (highest), on the overall quality of each of the tables in our experiment. The lowest rating awarded by the MCs is 2 and the highest is 9, with an average (median) score across MCs and items of 6.3 (6.0) and a standard deviation of 1.95.

We perform three checks on the consistency and informativeness of these ratings. First, we observe that the correlation of ratings between the three MCs and across all tables is very high, despite the assessments being done independently by each MC. The lowest unconditional correlation between any two of the three MCs is 0.88 and the highest is 0.93. Second, MCs were asked to re-assess a sample of 8 tables for a second time, weeks after their first assessment and rating exercise. Re-assessments are also highly correlated with first assessments of the same item, within MC: the lowest unconditional within-master-carpenter correlation coefficient is 0.85 and the highest is 0.97. Third, we corroborate that there is meaningful variation in quality across items, and that this is well-captured by our measurement exercise. To establish this, we inspect close-up pictures of the items, to assess the correspondence between our observation, the physical measurements in the quality assessment tool, and the ratings offered by the MCs. An illustration of such examination is presented in Supplemental Appendix Figure B1, which contrasts close-up pictures of an item that received an average rating across MCs of 3 (Panels (a) and (b)), with an item that obtained a high average score of 8 (Panels (c) and (d)). In the physical assessment of the low-quality item MCs noted joints that were not tight, nor filled, and indicated the presence of cracks on the wood and poor finishing – all faults visible in the top pictures of Supplemental Appendix Figure B1. In contrast, the highly rated item, which is portrayed at the bottom of the figure, is assessed as having tight joints affixed with dowels and glue, and no imperfections or damage in the finishing are recorded.

Sources of Variability in Quality. As discussed above, the physical assessment of items using our diagnostic tool covers a large number of dimensions, from structural aspects of the joinery to defects in the wood surface. We use the overall quality ratings offered by the MCs to understand which of these physical aspects appear to drive experts' views on an item's quality. We do so by decomposing the variability in MCs' ratings over a set of key dimensions –joints, symmetry, defects, and finishing– each of which, in turn, covers specific tests and measurements.[2]

of the tables, we capture these departures as a distinct feature of ornamentation. The second exception is that, for the purpose of a series of auxiliary exercises, we commissioned 8 additional tables to be produced with a different species of wood (mvule instead of musambya). We account for the material accordingly in the discussions below. As such, applying the diagnostic tool to the items in the experiment renders only vertical quality differentiation.

[2]The joints dimension covers whether joints are tight and, if not, whether any gaps have been filled, whether the sides of the joints protrude over vertices, and whether nails (which are weaker and rust) have been used to affix the joint, instead of the best practice use of dowels and glue. The symmetry dimension collects whether a principal angle of the item (typically a critical joint) is square, whether opposing joints are symmetric, and whether the item is leveled relative to the ground. The presence of defects includes whether there are visible cracks in the wood, whether 'perverse' (as opposed to functionally innocuous) knots in the wood are present, and whether there is any other sign of damage (such as insects, mold, etc.). Finally, the

We perform a series of variance decomposition exercises that study the variability in MCs' ratings over the different dimensions of quality. For compactness, we leave a detailed exposition of this analysis outside the scope of this text and list here two important findings. First, we note that a rich model including all quality dimensions, as listed in footnote 2, accounts for nearly 90% of the variation in MCs' ratings. Second, the realization of item characteristics is highly correlated across quality dimensions, consistent with the diagnostic tool recovering workmanship quality and production and managerial practices: on average, carpenters who are able to assemble tight joints are also able to sand and seal the wood properly. An indication of this is that the loss of fit from removing any individual quality dimension while keeping all other dimensions fixed results in minimal losses of explanatory power.

B.3 An Index of Vertical Quality

Part of our analysis involves the characterization of the relationship between quality, customers' knowledge and prices. While the core of our analysis uses MCs' ratings as measures of quality, some validation exercises and descriptives require the use of a measure of quality that is available not only for the items in our experiment, but also for items produced by carpenters in the market (as surveyed by Cajal-Grossi et al. (2025)). To this end, we combine the quality measurements recovered by our assessment tool, to construct a scalar index increasing in vertical quality. To remain conservative, we exclude from the components of this index any physical characteristic of carpentry items that may also reflect horizontal differentiation across varieties.[3] Concretely, the quality index combines the following markers of quality: the presence (lack thereof) of nails, the tightness of the joints and the use of filler, the coplanarity of the joints (sometimes referred to as 'flushness'), the squareness of key angles, the levelness of horizontal pieces, the presence (lack thereof) of observable cracks, damage, etc., and the appropriate sanding of the piece.

As we construct this index not only for the tables of our experiment, but also for items truly sold by carpentry workshops in Uganda (as discussed in Section B.1), we residualize each of the components of the index against fixed effects that account for horizontal differences across items, as well as attributes that can affect the measurement process with our assessment tool. These include fixed effects for the type of product (such as single bed or two-panel door), its size, the wood species the item is made of, the extent of ornamentation of the item, the type of finishing, and the trained assessor performing the measurements.

Our baseline measure aggregates these residualized components, weighting each of them according to the inverse of their in-sample variance. In practice, this is computed using the *swindex* Stata command of Schwab et al. (2020), implementing the GLS-based index

finishing dimension identifies whether the item has been sealed (either with sanding sealer or an alternative), whether it has been properly sanded and whether there are spillages of glue, paint, varnish, filler, or any other material applied to the item. The measurements that underlie these dimensions are either categorical by construction or recast as such for the purpose of the analysis here.

[3] Examples of these attributes are the type of lumber the item is made of, the finishing of the item (i.e. whether an item is painted, varnished, stained, etc.), or characteristics of the ornamentation.

construction proposed by Anderson (2008). As a result, the quality index reflects a standardized inverse-covariance weighted average of components using all available data to assign less weight to highly correlated outcomes than to outcomes that are uncorrelated.

Supplemental Appendix Table B1 shows the results of regressing the MCs' ratings against each individual marker of quality, as well as the constructed index. The exercise shows that in general individual components are conditionally positively correlated with the ratings that MCs award, as is the constructed quality index.

To facilitate the interpretation of quality differences in the experiment in a meaningful way (i.e. market-relevant), we recast the quality index into bins, corresponding to deciles of quality in the broader market for tables (intended as the tables sold by the random sample of 748 carpenters surveyed in Cajal-Grossi et al. (2025)). As such, a table in the experiment that is assigned to bin 1 corresponds to a table comparable to tables in the lowest decile of the quality distribution in the market. Analogously, a table in the experiment assigned to bin 10 has a quality comparable to the top decile in the market. This recasting allows for a clearer interpretation of quality differences in tables in the experiment, as well as for a benchmarking of our findings with the patterns in the market. Supplemental Appendix Figure B2 shows that the ratings awarded by MCs are (weakly) increasing in the quality decile to which each experiment table belongs.

The results in the main paper that make use of this quality index (Table 2) are studied under a large array of alternative index constructions in Supplemental Appendix Table A4, always showing very robust results. In particular, we entertain variations of the index that exclude each individual component of the index in turn, others where we do not perform the residualization, an index where we allow for residualization only after the aggregation, and a version in which we perform the aggregation without using weights, thus producing an unweighted standardized sum of all components, and a version in which we do not recast the quality index onto quality bins.

Figure B1: Observed Quality and Master Carpenters' Ratings

(a) Low Rating: Joints

(b) Low Rating: Cracks and Finishing

(c) High Rating: Joints

(d) High Rating: Cracks and Finishing

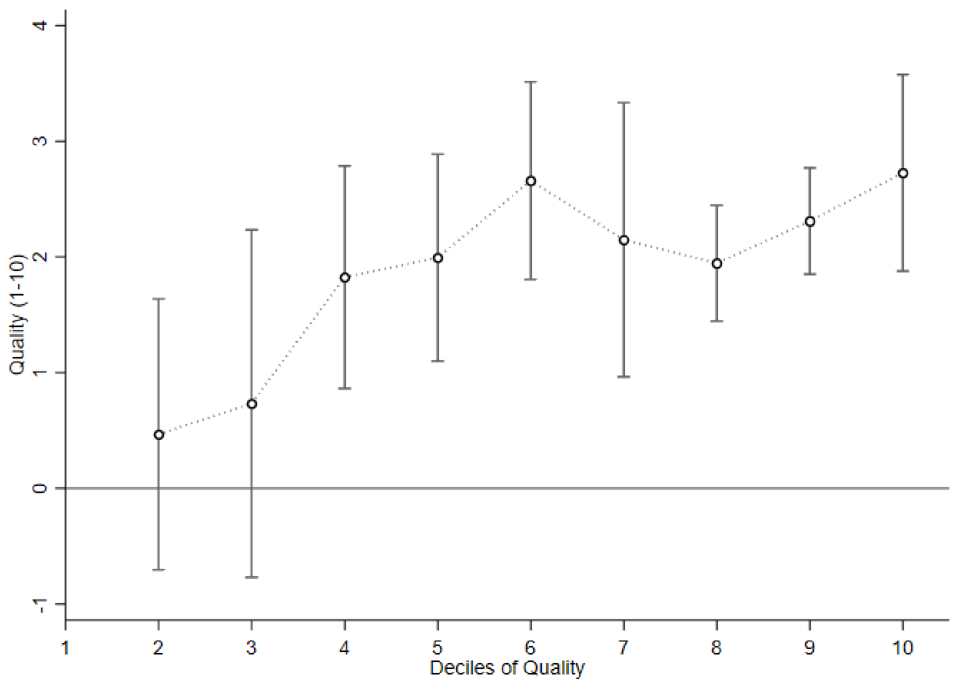
Panels (a) and (b) show close-up pictures of the item with study identification number 1993, whose average master carpenters' score is 3. The physical assessments of the item indicate that joints are not tight, nor are they sealed with filler, as shown in Panel (a), and indicate the presence of cracks and poor finishing, as depicted in Panel (b). Panels (c) and (d) show close-up pictures of the item with study identification number 9991, whose average master carpenters' score is 8. The physical assessments indicate that the joints are tight and affixed using dowels or glue, as illustrated in Panel (c), and that the finishing is smooth with no visible cracks or damage (Panel (d)).

Table B1: Informativeness of the Assessment Tool and Quality Measure

	Quality (1-10, Standardized)								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Coplanarity	-0.181							0.009	
	(0.114)							(0.067)	
Joint Tightness		0.747***						-0.062	
		(0.128)						(0.238)	
No Nails			1.065***					0.621*	
			(0.293)					(0.350)	
Squareness				0.514***				0.005	
				(0.083)				(0.117)	
Levelness					0.961***			0.632***	
					(0.183)			(0.163)	
No Damage						1.074***		0.276	
						(0.150)		(0.249)	
Sanding							0.727***	0.443***	
							(0.108)	(0.154)	
Quality									0.366***
									(0.055)
Mean Outcome	6.20	6.20	6.20	6.20	6.20	6.20	6.20	6.20	6.20
Mean Regressor	1.94	1.89	0.29	1.13	0.66	0.46	2.15		0.02
R^2	0.52	0.68	0.60	0.68	0.67	0.69	0.71	0.82	0.72
Obs.	189	189	189	189	189	189	189	189	189

Robust standard errors clustered at the table level are reported in parentheses for columns (1)–(8), while column (9) reports bootstrapped standard errors clustered at the table level. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$. Bootstrapping is used as a conservative approach to account for the additional sampling variability introduced by the prior estimation of the quality index. A unit of observation is a combination of one assessor and one table, for a total of 189 unique observations corresponding to 64 tables commissioned from carpenters, and assessed by three different master carpenters. We remind the reader that only 40 of these tables are used in the experiment (10 per kiosk). In all columns, the outcome variable (*Quality (1-10, Standardized)*) is the standardized rating assigned to tables by the three master carpenters on a scale from 1 to 10, where 10 is the best produced in Uganda (although the master carpenters only expressed ratings within a range from 2 to 9). In column (9), Quality is the continuous quality score computed on the following residualized individual components: *Joint Tightness*, *Coplanarity*, *No Nails*, *Squareness*, *Levelness*, *No Damage*, and *Sanding*. The estimation includes fixed effects for varieties, type of finishing, ornamentation level, and master carpenter. Each individual component of the quality score is considered (without residualization) separately in columns (1)–(7), and all individual components are pooled in column (8). The bottom panel reports the mean of the standardized outcome and the mean of the main regressor in each specification.

Figure B2: Informativeness of Quality Assessments



Bootstrapped standard errors clustered at the table level. Bootstrapping is used as a conservative approach to account for the additional sampling variability introduced by the prior estimation of the quality index. The figure shows the outcome of a linear regression of the rating expressed by each of the three master carpenters (*Quality (1-10)*) against deciles of the quality score. The quality score is computed as discussed in Supplemental Appendix B.3, combining the following residualized individual components: *Joint Tightness, Coplanarity, No Nails, Squareness, Levelness, No Damage*, and *Sanding*. Each component is residualized against fixed effects for type of item and size, wood species, ornamentation, finishing type, and assessor (when applicable). We note that while there may be no variation along most of these dimensions across the items commissioned for the experiment, there is substantial variation in them across the tables in the market, on the basis of which the quality distribution is partitioned into deciles. The deciles of the quality score were obtained by recasting the continuous quality score of tables in the experiment into ten bins. These correspond to deciles of the distribution of the quality score of tables in the market, such that a table in the experiment that is assigned to bin 1 corresponds to a table comparable to tables in the lowest decile of the quality distribution in the market. Analogously, a table in the experiment assigned to bin 10 has a quality comparable to the top decile in the market. A unit of observation is a combination of one assessor and one table, for a total number of 189 unique observations corresponding to 64 tables assessed by three different master carpenters (the 40 tables used in the experiment are obtained from the 64 commissioned tables studied in this figure). The specifications include master carpenter fixed effects, as well as fixed effects capturing the wood species and small horizontal differences across items, as explained in footnote 1. A regression of the respective specification with the same structure, but with a linear restriction on the deciles of the quality measure gives a slope of 0.283 (S.E. 0.036).