Journal of Development Economics 152 (2021) 102701

Contents lists available at ScienceDirect

Journal of Development Economics

journal homepage: www.elsevier.com/locate/devec

Short Communication

Testing willingness to pay elicitation mechanisms in the field: Evidence from Uganda

Konrad B. Burchardi [a,b,1], Jonathan de Quidt [a,c,*,1], Selim Gulesci [d,e,1], Benedetta Lerva [f,g,1], Stefano Tripodi [h,1]

[a] Institute for International Economic Studies, Stockholm University, SE-106 91 Stockholm, Sweden
[b] BREAD, CEPR, CESifo, ThReD
[c] CAGE, CEPR, CESifo, ThReD
[d] Department of Economics and TIME (Trinity Impact Evaluation Unit), Trinity College, Dublin, Ireland
[e] BREAD, CEPR, EUDN, J-PAL, LEAP
[f] Mistra Center for Sustainable Markets, Stockholm School of Economics, SE-113 83 Stockholm, Sweden
[g] The World Bank, Washington, DC, United States
[h] Department of Economics, Copenhagen Business School, Porcelaenshaven 16A, DK-2000 Frederiksberg, Denmark

ARTICLE INFO

JEL classification:
C90
C93
D44
O12

Keywords:
Willingness to pay
Becker–DeGroot–Marschak
Field experiment

ABSTRACT

Researchers frequently use variants of the Becker–DeGroot–Marschak (BDM) mechanism to elicit willingness to pay (WTP). These variants involve numerous incentive-irrelevant design choices, some of which carry advantages for implementation but may deteriorate participant comprehension or trust in the mechanism, which are well-known problems with the BDM. We highlight three such features and test them in the field in rural Uganda, a relevant population for many recent applications. Comprehension is very high, and 86 percent of participants bid optimally for an induced-value voucher, with little variation across treatments. This gives confidence for similar applications, and suggests the comprehension-expediency trade-off is mild.

1. Introduction

Measurement of willingness to pay (WTP) is an important part of the economist's toolkit. Conceptually, knowledge of consumers', households', or firms' WTP for goods or services is key to constructing non-parametric demand curves, which can be used to predict the effects of counterfactual policies such as price subsidies. Furthermore, standard elicitation mechanisms allow the researcher to conduct *selective trials* (Chassang et al., 2012), which measure marginal treatment effects as a function of WTP. Marginal treatment effects can be used to compute a variety of policy-relevant treatment effects (Heckman and Vytlacil, 2005).

In recent years development economists have increasingly adopted these techniques in the field. Researchers have studied demand for health products (Hoffmann, 2009; Hoffmann et al., 2009; Cohen and Dupas, 2010; Ashraf et al., 2010; Dupas, 2014; Fischer et al., 2016; Hoffmann, 2018; Grimm and Hartwig, 2018; Fischer et al., 2019; Berry et al., 2020), improved latrines (Yishay et al., 2017), fuel-efficient stoves (Berkouwer and Dean, 2020), solar electricity (Grimm et al., 2020), education programs (Berry and Mukherjee, 2019; Burchardi et al., 2020a), business training (Maffioli et al., 2020), rainfall insurance and agricultural information (Cole et al., 2020), farmer training (Lerva, 2020), and fertilizer (Burchardi et al., 2020b).

The dominant approach to WTP elicitation uses some variant of the classic Becker, DeGroot, and Marschak (1964) (BDM) mechanism. These variants share a common structure. First, participants report a WTP value, W. Second, a random price P is drawn. Third, if $W \geq P$, the participant purchases the good at price P, otherwise they do not purchase and pay nothing. This mechanism shares the incentive properties of a second-price sealed-bid auction: truthful reporting is a weakly dominant strategy.

That approach is persuasive and powerful in theory. This paper is concerned with its practical implementation. There are many incentive-equivalent ways to implement this structure, with different auxiliary

* Corresponding author at: Institute for International Economic Studies, Stockholm University, SE-106 91 Stockholm, Sweden.
E-mail addresses: konrad.burchardi@iies.su.se (K.B. Burchardi), jonathan.dequidt@iies.su.se (J. de Quidt), gulescis@tcd.ie (S. Gulesci), blerva@worldbank.org (B. Lerva), st.eco@cbs.dk (S. Tripodi).

[1] All authors contributed equally to the project.

https://doi.org/10.1016/j.jdeveco.2021.102701
Received 13 February 2021; Received in revised form 27 May 2021; Accepted 6 June 2021
Available online 10 June 2021
0304-3878/

properties. Some have statistical or practical advantages for the experimenter; they elicit richer data or give the experimenter more control over the randomization. But these may come at the cost of participant comprehension or trust in the mechanism. This is important, because it is widely recognized that even highly-educated participants can find second-price sealed-bid auctions difficult to understand (Kagel et al., 1987; Ausubel, 2004; Li, 2017), so equivalent incentives may not lead to equivalent behavior. These concerns may be particularly important for field research in low-income countries, where participants have limited education, are unfamiliar with intricate incentive structures and randomization, and may feel unable to verify or trust that the scheme will be honestly implemented (Akbarpour and Li, 2020).

We conduct an experiment in rural Uganda to compare the performance of four different implementations of the basic structure described above. These allow us to test the influence of three design features: the response mode (how participants state their WTP); whether prices are randomly assigned onsite after elicitation, or preassigned; and how much information we give participants about the price distribution. As we explain in the next section, these features capture key trade-offs between design expediency and ease of comprehension.

We measure performance in two ways. First, we implement an extensive set of comprehension checks, inspired by those used by Berry et al. (2020). Second, we measure optimal bidding performance, inspired by Cason and Plott (2014). We elicit WTP for an induced-value voucher that is immediately redeemable for its face value in cash: 1400 Ugandan Shillings (UGX, $1.27 at 2016 PPP, or around 30 percent of the daily agricultural wage in the region). It is a weakly dominant strategy to report WTP equal to the face value, so deviations from optimal behavior are easy to observe. As an auxiliary check we also examine bidding for a well-known soap product, a more natural transaction but for which optimal behavior is more difficult to define.

We find two main results. First, mechanism comprehension is very high. Participants answer our comprehension checks correctly 94 percent of the time, and approximately 86 percent of participants bid optimally for the voucher. This is reassuring news for other work using similar mechanisms to ours. Second, performance is very similar (and statistically indistinguishable) across our four treatment arms. This suggests that researchers can feel confident in following the approach that best suits their field setting, and readers can be confident in interpreting findings from implementations similar to those we test.

Besides the field applications cited above, we relate to a large experimental literature testing different auction and WTP elicitation mechanisms in lab settings (some relevant examples are Bohm et al. (1997), Rutström (1998), Andersen et al. (2006), Brebner and Sonnemans (2018)). Most relevant for our design is Cason and Plott (2014), who find that many participants report WTP less than $2 for a token worth $2. In a field setting, Berry et al. (2020) find that elicited WTP is higher under "take it or leave it" offers than BDM while Cole et al. (2020) find they are more similar. We also relate to the parallel literature on "contingent valuation," which uses hypothetical questions to elicit values of public goods, see e.g. Hanemann (1994).

2. Design

We begin this section by motivating our three design features of interest: response mode, price revelation, and information about the price distribution. Then, we describe the four treatments that we use to test them. Finally, we highlight the design features common to all of our treatments.

We selected the experimental and common features for pragmatic reasons: they invoke important tradeoffs that arise when eliciting WTP in the field (including in our own work). We summarize our view of those tradeoffs below. There are other important aspects that we do not address in our design. In particular, different designs might induce different *anchoring* effects, or differences in *default* behavior. We return to these in Section 5.2.

2.1. Design features of interest

Response mode

The classic approach to the BDM has participants simply state a WTP value on a continuous scale (for a recent implementation see Berry et al., 2020). A practical advantage of this approach is the richness of the data generated. But it relies on participants being readily able to retrieve or construct their own WTP value, and provides little guidance to help them evaluate if that is indeed their true (maximum) WTP.

An alternative is to use a *Multiple Price List* (MPL).[2] This approach asks, for each of a discrete set of prices, whether the participant would be willing to pay that price should it be drawn. Thus the procedure is broken down into a sequence of isolated "yes/no" questions. This loses the point identification of the classic BDM, giving only interval identification.[3] But it is arguably more transparent and easier to understand (Andersen et al., 2006).

When implementing MPL, we adopt a format that mimics an ascending auction. Starting at zero, we ask the participant if they are willing to pay each possible price in ascending order, stopping once they say "no." We do this because the ascending auction is obviously strategy-proof (Li, 2017), and by mimicking its format we may improve comprehension.[4]

To summarize, BDM has the statistical advantage of richer data, permitting point identification of WTP, but comprehension may be improved under MPL.

Price revelation

A standard approach is to draw the price using a transparent randomizer once WTP has been stated, e.g. a bingo cage or paper slips in a hat. Sometimes (including in our application) the participant draws the random price themselves. We refer to this as *Onsite* randomization. A key argument for doing so is trust: participants can see the process and verify that it is fair.

This approach has a statistical disadvantage in the context of running a selective trial (Chassang et al., 2012). A selective trial uses the price assignment to assign a treatment, so as to later evaluate its impact.[5] But *Onsite* randomization makes it impossible to stratify this randomization, increasing the risk of imbalances and reducing statistical power. Researchers might therefore prefer *Preassigned* random prices, which can be stratified on variables available prior to elicitation.[6] The cost is reduced transparency: it is harder for participants to verify that the randomization has been conducted as described.

[2] Sometimes BDM is used to describe the entire class of elicitation mechanisms that prompt respondents to state their WTP before revealing the price; under this terminology MPL is a subset of BDM. However it is also common to use "BDM" and "MPL" as we do, to specifically distinguish between variants with and without a price list.

[3] If a participant says "yes" to price P_t and "no" to price P_{t+1} we learn that her WTP is in the interval $[P_t, P_{t+1})$.

[4] Note that MPL is *not* obviously strategy-proof. The difference is that in an ascending auction, saying "No" guarantees that the participant will not buy. Under MPL, they might still end up buying at a lower price. Breitmoser and Schweighofer-Kodritsch (2021) present evidence that some of the gains from OSP mechanisms can be achieved in similar non-OSP designs.

[5] Conditional on WTP, treatment is random and depends only on the price draw. Thus it is possible to estimate marginal treatment effects and use them to reconstruct other effects of interest (Heckman and Vytlacil, 2005). Burchardi et al. (2020b) use a selective trial to measure returns to fertilizer.

[6] An intermediate approach that is easy to implement is to randomize the price after WTP has been elicited, but using a survey implement such as a tablet or mobile phone. We conjecture that this has more in common with *Preassigned* than *Onsite* randomization, because the participant has little insight into how the randomization was conducted. Another way to achieve stratification, which we do not explore, would be to randomize *without replacement*, for example through a public lottery draw, once WTP has been elicited from each member of a stratification group.

Our second feature test compares *Onsite* to *Preassigned* randomization. We implement this using scratchcards. Under *Onsite* draws, the participant is given a scratchcard with eleven panels, each concealing a different price. They choose and scratch one of these to determine the price they will face. They are told in advance that once the price has been recorded, they will be able to scratch the remaining panels to verify that the scratchcard was fair. Under *Preassigned* prices the participant is given a scratchcard with a single panel concealing a preassigned price. They are told that this price was chosen by a computer and the enumerator does not know what it is, but of course they cannot verify this information.

To summarize, *Onsite* randomization has the advantage of transparency, but cannot be stratified. Preassigned randomization allows for stratification, but may lower trust in the mechanism.

Information about the price distribution

The standard approach is to use a uniform price distribution, and inform the participant about this fact. This is simple to explain, even to participants unfamiliar with probabilities, and straightforward to implement transparently with familiar randomization devices.

There are good reasons for uniform prices. While in principle the incentive properties of the mechanism only depend on full support – that all prices are drawn with some positive probability – in practice, the distribution might matter. Mazar et al. (2013) find that shifting probability mass from the bottom to the top of the price distribution increases WTP in a BDM mechanism (see also Bohm et al., 1997). It may be that participants perceive salient features of the price distribution as anchors, or as informative about the true underlying value or as hints as to what to bid.

However, nonuniform prices can be very useful in practice.[7] The findings highlighted above suggest that if researchers take this route, they may not want to *inform* participants about it.

It is even possible that not reporting the price distribution improves comprehension. Going to great lengths to precisely explain the price distribution might be interpreted as saying this information is important for incentives, when in fact it is irrelevant.

Our final feature test compares a *Uniform* price distribution, where the participants are informed about both the support and the distribution, to an *Unstated* distribution, where we only tell them its support. To operationalize the process, under Uniform we tell them that all of the prices are equally likely, and ask comprehension questions probing whether they understand what this means.

To summarize, Uniform prices have the advantage of transparency. The Unstated price distribution gives the researcher more freedom to use alternative distributions. It also has the potential advantage of not drawing attention to irrelevant information.

2.2. Treatments

We have three design features of interest, but not all nine possible combinations are interesting or practical to test. We designed four experimental treatments, each of which changes one feature from the previous one. We can therefore test the effects of each feature independently by comparing pairs of treatments. Experimental instructions can be found in Appendix C.

Treatment 1: BDM, Onsite, Uniform. Our benchmark treatment is a classic BDM implementation with *Onsite*, *Uniform* price randomization, similar to Berry et al. (2020). Participants state their value for the good, then prices are drawn from a uniform distribution.

Treatment 2: MPL, Onsite, Uniform. Relative to Treatment 1, Treatment 2 changes the reporting mode, from BDM to MPL.

Treatment 3: MPL, Preassigned, Uniform. Relative to Treatment 2, Treatment 3 changes the price randomization, from Onsite to Preassigned.

Treatment 4: MPL, Preassigned, Unstated. Relative to Treatment 3, Treatment 4 changes price distribution information, from Uniform to Unstated.

Thus, comparing treatments 1 and 2 is informative about the effects of BDM versus MPL. Comparing treatments 2 and 3 is informative about *Onsite* versus *Preassigned* randomization. Comparing treatments 3 and 4 is informative about *Uniform* versus *Unstated* price distributions.

2.3. Common design features

Price distribution

In all treatments we kept the price distribution the same, a discrete uniform distribution over eleven prices. Participants were always informed of its support.[8]

Price is independent of wtp

If participants believe their WTP can somehow influence the price, incentive-compatibility is lost. This could be a concern when the price is drawn using a non-transparent randomizer.

To underline that price was independent of WTP, we always revealed prices using a scratchcard, produced using scratch-off stickers placed over pre-printed price cards.[9] The scratchcard had either one panel (*Preassigned* treatments) or eleven (*Onsite* treatments). In *Onsite* treatments, the participant was told they could first choose which panel to scratch to reveal their price, after which they would be allowed to scratch the remaining panels if they wanted (so that they could verify the card was fair).

We believe the pre-printed scratchcards make it very clear that the price could not be influenced by WTP. One possible exception would be if the participant believed they could collude with the enumerator to find a lower price on the eleven-panel scratchcard. Our next design feature addresses this concern.

Collusion-proofness

Our scratchcard implementation is also robust to collusion, between the enumerator and participant, to provide information about or distort the price draw. Besides lowering data quality, collusion is a potential concern in selective trials where the researcher wants treatment to be assigned correctly and fairly.

Collusion could come in two forms: a) the enumerator could tell the participant what the price will be prior to WTP elicitation; or b) the enumerator could distort the price by ensuring that a low price is drawn.

We rule these out by using pre-printed scratchcards assigned to the respondent's unique ID. The enumerator did not know what price or what random order of prices was printed on the card. The enumerator

[7] We provide two examples. Burchardi et al. (2020a) elicit WTP for an education intervention. There were in fact only two possible prices, due to the program design. Participants were simply (truthfully) informed that the price was random, but not of its distribution or support. Burchardi et al. (2020b) use a selective trial to study returns to fertilizer. A challenge in selective trials is that, with uniform prices, individuals with low WTP are unlikely to be treated, while individuals with high WTP are very likely to be treated, reducing statistical power. They instead use a bimodal price distribution with full support but 90 percent of its density split between zero and the maximum price, P_{max}. Thus the probability of treatment, that is, $P \leq WTP$, is close to 50 percent for anyone with $WTP < P_{max}$ (the probability is one if $WTP = P_{max}$).

[8] The standard BDM implementation uses a (near-)continuous price distribution, to maintain incentives to report any WTP value. We kept the discrete distribution throughout to maintain similarity across treatments. Thus there is no strict incentive to report WTP in-between price increments. It is unclear whether this makes the basic incentive properties of the BDM easier or harder to understand. Some participants do state intermediate values, particularly in the Soap elicitation (for Voucher the vast majority state 1400 UGX).

[9] Cason and Plott (2014) use a similar technique with prices concealed by opaque stickers.

Table 1
Selected villages' characteristics.

	BDM, Onsite, Uniform	MPL, Onsite, Uniform	MPL, Preassigned, Uniform	MPL, Preassigned, Unstated	Total	Female
Village 1	17	6	4	5	32	0.57
Village 2	9	19	0	3	31	0.64
Village 3	6	3	17	6	32	0.37
Village 4	2	5	5	1	13	0.43
Village 5	3	7	4	14	28	0.55
Village 6	2	5	14	11	32	0.38
Village 7	11	5	6	10	32	0.33
Total	50	50	50	50	200	0.47

was also required to write the final WTP on the card, and take a photograph of it, before any panels were scratched. This prevented them from revealing the price before WTP was elicited.

In the *Preassigned* treatments the enumerator cannot distort the price. In *Onsite* treatments they could do this by scratching several panels and reporting the lowest price as the true price. We rule this out by requiring them to photograph the card once again, after the first panel has been scratched.

Comprehension checks

Our elicitation procedure always included a battery of comprehension checks, described in detail in Section 4. These were slightly adapted according to the treatment. Participants could revise their WTP after the comprehension questions.

3. Methodology

3.1. Sample and randomization

We conducted the experiment in Mbale District (Eastern Subregion Uganda, see map in Appendix Figure B.1). We randomly selected seven villages, subject to a minimum population in the 2010 Census of 300 people or around 50 households, given an average household size of six. This criterion simplified finding participants. Basic information about the sample villages is given in Table 1. All villages were in rural areas, the closest being about 7 km from Mbale city, while the furthest was around 19 km away. Field activities took place from 18–24 November 2016. Each village was visited for one day by a team of four enumerators, targeting seven to eight participants each per day.[10]

Participants were mobilized on the day of the survey, by asking the village chairperson to gather around 35 people in the village that were willing to participate. We targeted 200 participants in total, and preassigned 50 participant IDs to each of our four experimental treatments by a simple (non-stratified) randomization. Prices, if preassigned, were randomized within treatment.

The only descriptive information we collected was participant gender, which is reassuringly close to 50% female. But since we did not target a representative sample we do not claim representativeness, and our results might look different in different subpopulations.

To benchmark our sample, Appendix Table A.1 reports basic summary statistics for adult respondents in both Eastern Subregion and the whole country. The data are taken from the 2013/2014 wave of the nationally representative Uganda National Panel Study (UNPS, part of the World Bank's *Living Standards Measurement Study* or LSMS, Uganda Bureau of Statistics (2016)). The statistics highlight that our experiment was conducted in a rural setting, with low levels of literacy and education, and large households. All of these characteristics are typical for Uganda as a whole, and the Eastern subregion in particular. Our participants therefore differ substantially from those in typical university-based lab studies, and might find the mechanisms we test less familiar and less easy to comprehend.

3.2. Goods and optimal bidding

We elicited participants' WTP for two goods. The first good was an induced-value voucher, redeemable for its face value in cash. We begin by demonstrating how the voucher works, and build trust, by giving the participant a voucher worth 100 UGX which they can immediately exchange for 100 UGX in cash. Then, we introduce the experimental voucher, which has a face value of 1400 UGX ($1.27 at 2016 PPP, or around 30 percent of the daily agricultural wage in the region).[11] We reproduce the voucher in Appendix Figure C.1.

The second good was a 600g bar of blue soap of a brand commonly available in local markets. The typical market price for this product was 2000 UGX.

There were eleven possible prices for each product. For the voucher, these were {0, 200, 400, …, 2000}. It is a weakly dominant strategy to report WTP equal to the voucher's face value.[12]

For the soap, possible prices were {0, 400, 800, …, 4000}. After conducting the WTP elicitation, we elicited participants' beliefs about the market price. Private values for the product might be heterogeneous, so any bid below market price could be optimal. We code bids strictly above the participant's perceived market price as non-optimal, but this is only suggestive (bidding above market price could be optimal if doing so saves on travel or effort costs).

4. Comprehension checks

We implemented several comprehension checks to test participants' understanding of the overall procedure and expose them to possible consequences of their choices. We conjectured that this would induce them to change their behavior if their first choice was a mistake.

The comprehension check questions were asked both in the voucher and the soap round, although with some differences described below. Appendix D provides question phrasing.

In the voucher round, we asked two set of comprehension questions before WTP elicitation (*Chart checks* and *Price checks*) and two afterwards (*Would-you-buy checks* and *Profit checks*). In the soap elicitation round, only the *Price checks* and the *Would-you-buy checks* were asked.[13]

After the checks were completed, the participants were asked whether they wanted to change their WTP. If they did, the elicitation was repeated and a new WTP was recorded. Five participants asked to

[10] The project was conducted in the piloting phase of a larger field experiment, which subsequently received IRB approval from MUREC (national IRB in Uganda). Informed consent was obtained from all participants. This is consistent with the institutional rules at the authors' institutions.

[11] The median adult agricultural daily wage in the 2013/14 wave of the UNPS is 4000 UGX, corresponding to around 4500 UGX at 2016 prices.

[12] Since prices in the experiment are discrete, reporting WTP of one price increment below 1400 UGX, i.e. 1200 UGX, is also weakly dominant and we code it as optimal. In the BDM treatment, participants are free to submit any bid, so we code any bid between 1200–1400 UGX as optimal. Furthermore, one participant did not have enough cash with them to pay 1400 UGX, we code them as bidding optimally if they do so given the liquidity constraint they faced.

[13] Each participant was supposed to be asked each check question at least once. If the answer to the question was correct, the enumerator was to move to the next question in the survey. In case of a wrong answer, the enumerator was supposed to re-explain the relevant feature of the procedure and to ask the check question again, for a maximum of three attempts. However, due to an implementation error, in some cases the enumerators skipped questions, sometimes on the first time of asking and sometimes on the repeats. We code unasked questions as missing, and focus on the share of correct answers for the first attempt of the check questions. The most notable cases were that 11 participants in the voucher and 36 in the soap treatments were not asked the price check questions.

Table 2
Proportion of correct answers to each comprehension check (first attempt).

	Overall	BDM, Onsite, Uniform	MPL, Onsite, Uniform	MPL, Preassigned, Uniform	MPL, Preassigned, Unstated
Panel A: Pre-elicitation checks, Voucher					
1st chart check	0.86	0.88	0.88	0.88	0.82
2nd chart check	0.93	0.88	1.00	0.94	0.90
Prices listed	0.94	0.96	0.89	0.94	0.96
1st price check	0.97	0.98	0.96	0.96	.
2nd price check	0.97	1.00	0.96	0.96	.
Panel B: Post-elicitation checks, Voucher					
Buy at $p >$ WTP	0.94	0.94	0.96	0.88	1.00
Buy at $p <$ WTP	0.99	0.98	1.00	0.98	1.00
Avg. voucher	0.93	0.92	0.94	0.92	0.93
Panel C: Pre-elicitation checks, Soap					
Prices listed	0.88	0.86	0.90	0.87	0.90
1st price check	0.97	1.00	1.00	0.92	.
2nd price check	0.98	1.00	1.00	0.95	.
Panel D: Post-elicitation checks, Soap					
Buy at $p >$ WTP	0.99	0.98	1.00	0.98	1.00
Buy at $p <$ WTP	0.99	0.98	1.00	1.00	1.00
Avg. soap	0.96	0.95	0.97	0.96	0.97
Avg. voucher + soap	0.94	0.93	0.95	0.93	0.95
N	200	50	50	50	50

The Table reports the share of correct answers to the first attempt of each check question, by Treatment assignment. In the *"1st chart check"*, participants were asked to choose a chart describing the situation with the highest likelihood of buying the voucher. In the *"2nd chart check"*, participants were asked to choose a chart describing the situation with the highest likelihood of buying the voucher, conditional on making no loss. In the *"Prices listed check"*, participants had to list each possible price for the voucher/soap. In the *"1st price check"*, participants were asked whether each price for the voucher/soap was equally likely to be drawn. In the *"2nd price check"*, participants were asked whether a price for the voucher/soap had no chance to be drawn. In the *"Buy at $p >$ WTP check"* and *"Buy at $p <$ WTP check"*, participants had to state whether they would have been able to buy the voucher/soap given a random price higher or lower than their initial WTP, respectively. Averages are calculated as the mean of correct answers across all comprehension checks, excluding *"1st price check"* and *"2nd price check"*. F-tests for equality of means in rows Avg. voucher and Avg. soap have p-values 0.63 and 0.39 respectively in our primary specification (full regression results reported in Table A.2).

change their initial WTP in the voucher round and five in the soap round. Each time, only one of them actually ended up providing a different WTP value.

4.1. Pre-elicitation comprehension checks

The first set of questions (*Chart checks*) was meant to test the participant's general understanding of the elicitation procedure, which had been explained by the enumerator during the introductory part of the survey. The participant was presented with three hypothetical scenarios, depicted in three charts. In each scenario a fictitious auction for a voucher worth 1600 UGX was presented (i.e., different to the real 1400 UGX voucher in the experiment), along with the profit or loss conditional on each possible random price draw. Participants had to indicate which chart corresponded to (i) the scenario with the highest likelihood of purchasing the voucher (*1st chart check*)[14] and (ii) the scenario with the highest likelihood of purchasing the voucher without suffering a monetary loss (*2nd chart check*).[15]

In the second set of checks (*Price checks*), we tested the participants' understanding of the price distribution (excluding those assigned to MPL, Preassigned, Unstated treatment condition, as they had not been told about the price distribution). Participants had to (i) list the full set of possible prices (*prices listed*), and (ii) indicate whether, in their understanding, one of the eleven random prices was more likely to be drawn (*1st price check*), or had no chance of being drawn (*2nd price check*).

As shown in Panel A and Panel C of Table 2, the share of correct answers for each check question is high for each treatment arm. Overall, in the voucher round, 86% and 93% of the participants answered the first and the second chart questions correctly, respectively. For the voucher price checks, 94% of participants correctly listed all possible prices, 97% correctly stated that no price was more likely than any other, and 97% that no price was guaranteed not to appear. For the soap, these figures are 88%, 97%, and 98% respectively.

4.2. Post-elicitation comprehension checks

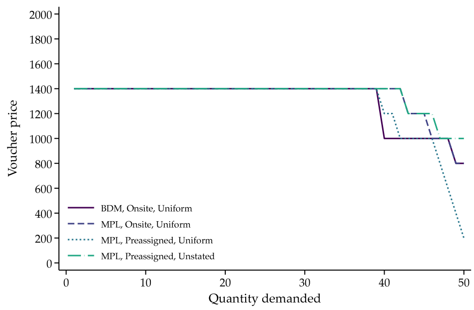
After eliciting WTP, we tested whether participants understood the possible consequences of their choices.

First, in the *would-you-buy checks*, we presented two hypothetical random price draws, one higher (*Buy at $p >$ WTP check*) and one lower (*Buy at $p <$ WTP check*) than their WTP, and asked if they would purchase in each case (the correct answers are no and yes respectively).

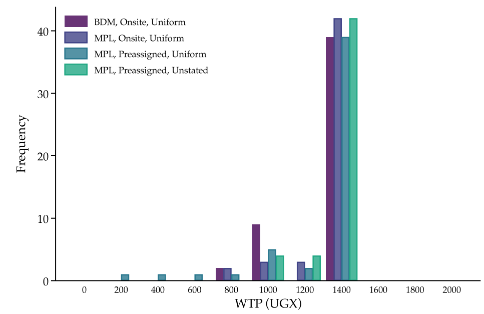
Next, in the voucher round only, we implemented the *profit checks* examining whether participants could calculate the monetary payoffs resulting from their choices. We presented them with some possible price draws and asked them to calculate their payoff conditional on their chosen WTP for the voucher, as well on alternative (hypothetical) WTP values. These questions were only asked if a participant's initial WTP was different than 1400 UGX. In retrospect, it was an oversight not to also ask when WTP equaled 1400 UGX. However, these checks do not appear to have affected our WTP data. All participants answered correctly, and only one participant changed their WTP after the post-elicitation checks.

[14] "Among these possible answers, what is the answer where you have the highest chance of purchasing the card?".

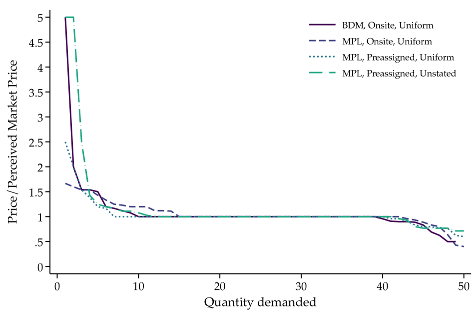
[15] "Among these possible answers, what is the answer where you have the highest chance of purchasing the card without suffering a loss?".



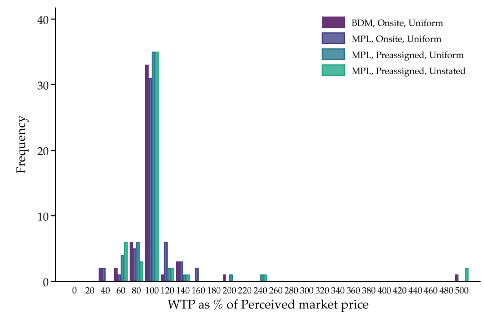
(a) Voucher demand



(b) Distribution of Voucher WTP



(c) Soap demand, scaled by perceived market price



(d) Distribution of Soap WTP/perceived market price

Fig. 1. Demand curves and WTP distributions. Note: in panels (b) and (d) WTP is rounded down to the nearest price increment when elicited via BDM.

Panel B of Table 2, shows the proportions of correct responses to the *would-you-buy checks*. For the first and second *would-you-buy* checks, comprehension averages 94% and 99% in the voucher round and 99% in both checks for the soap round.

4.3. Analysis of comprehension

In the Appendix we analyze comprehension in a regression framework. For voucher and soap separately, we compute average comprehension across the questions that were asked in all treatments (i.e., excluding the price checks which were not asked in the MPL, Preassigned, Unstated treatment). We then regress the average score on design features. Appendix Table A.2 presents the results. We find no evidence of systematic differences in comprehension between design features. All regression coefficients are close to zero and precisely estimated, and we cannot reject the null of no difference between design features in any specification.

We conclude that comprehension was high across treatments, and design features do not seem to substantially affect it.

5. Bidding behavior

Fig. 1 plots demand curves and the distribution of WTP for the voucher and the soap, by treatment.

For the voucher, it is easy to see from panels (a) and (b) that the vast majority of participants bid optimally, at 1400 UGX. Moreover, strikingly, nobody bids above this level, which is a common issue in second-price mechanisms (Kagel et al., 1987). There is little variation in demand between treatments.

For the soap, bidding is more heterogeneous. Some participants bid above the notional market price of 2000 UGX and others bid below. To code (potentially) optimal bidding, we express their bid as a fraction of their belief of the actual market price (34% of participants believe the price is above 2000 UGX). If this value is above 1 we consider the bid (potentially) suboptimal. Using this scaling, we plot demand and the WTP distribution in panels (c) and (d) (Appendix Figure B.2 provides the unscaled plots). The modal bid lies at 1, i.e., WTP equal to perceived market price. A small number of bids are substantially higher, going up to five times the market price.

While it is hard to fully characterize optimal bidding for the soap product, it is notable that as with the voucher, the demand curves and distributions of WTP are very similar between treatments. This suggests that the WTP elicited is very similar whichever mechanism is used.

Fig. 2 plots optimal bidding rates by treatment. Panel (a) shows these for the voucher round. Optimal bidding is high, averaging 86%, and variation across treatments is small relative to sampling error. Panel (b) shows (potentially) optimal bids for soap. These lie in the 70%–90% range with slightly more variation across treatments. Comparing voucher and soap bids, there are no obvious patterns in the variation across treatments.

Table 3 analyzes the three design features in a regression framework. We present linear probability regressions with optimal bids for vouchers or (potentially) optimal bids for soap as the dependent variable. We present estimates with no controls, village fixed effects, and village and enumerator fixed effects.

Our primary specification is column (3), which examines optimal voucher bidding by design feature, net of heterogeneity at the village or enumerator level. Optimal bidding in the BDM treatment is 78%, and the regression coefficients are small and precisely estimated. Our 95% confidence intervals rule out differences across features larger than around 10 percentage points. By the standard ex-post power rule, we have 80% power at the 5% level to detect differences in optimal bidding rates of 2.8 standard errors, or around 14 percentage points.

The specifications without enumerator fixed effects have larger point estimates but the differences across treatments are still small relative to standard errors, and not systematic, similar to what we saw in the figures. It is also notable that the enumerator fixed effects explain a large fraction of the outcome variation (R-squared in the voucher regression increases from 6.4 percent to 47.2 percent), suggesting that comprehension depends strongly on enumerator skill.

We conduct F-tests for the joint null of no differences in optimal bidding between design features. The p-value on this test in our primary

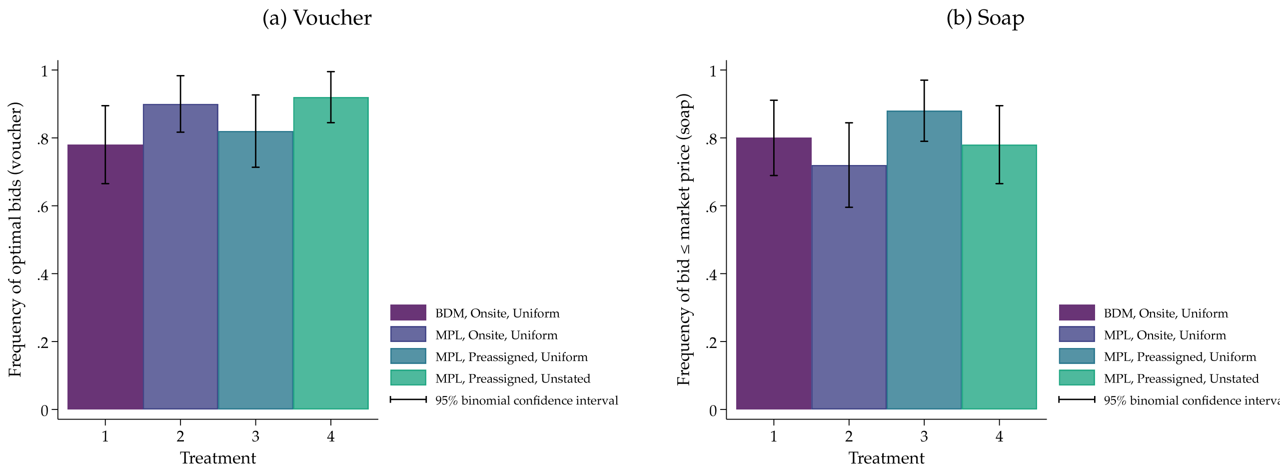


Fig. 2. Optimal bidding.

Table 3
Optimal bidding for voucher and soap by design feature.

	(1) Voucher	(2) Voucher	(3) Voucher	(4) Soap	(5) Soap	(6) Soap
MPL	0.120	0.128*	−0.007	−0.080	−0.018	−0.004
	(0.073)	(0.073)	(0.050)	(0.086)	(0.091)	(0.080)
Preassigned price	−0.080	−0.102	0.033	0.160**	0.131	0.071
	(0.070)	(0.068)	(0.052)	(0.079)	(0.093)	(0.086)
Unstated distribution	0.100	0.106*	−0.011	−0.100	−0.075	−0.073
	(0.067)	(0.061)	(0.047)	(0.075)	(0.080)	(0.080)
Village FE	No	Yes	Yes	No	Yes	Yes
Enumerator FE	No	No	Yes	No	No	Yes
F-test p-value	0.157	0.089	0.914	0.213	0.434	0.734
N	200	200	200	200	200	200
R-squared	0.026	0.064	0.472	0.020	0.065	0.216
BDM, Onsite, Uniform mean	0.780	0.780	0.780	0.800	0.800	0.800

Linear probability model. Robust standard errors in parentheses. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$. Outcome in columns (1)–(3) is a dummy for optimal bidding in the voucher WTP elicitation. Outcome in columns (4)–(6) is a dummy for bidding no more than the perceived market price in the soap WTP elicitation. "MPL" is a dummy for all MPL treatments, "Preassigned price" is a dummy for all preassigned price treatments, "Unstated distribution" is a dummy for the treatment with unstated price distribution. "F-test p-value" corresponds to a test of the null hypothesis that the coefficients on all three dummies are zero.

specification is 0.914, i.e. we find no evidence of significant differences between features. The smallest such p-value, 0.089, arises in the voucher specification with village fixed effects only.

We conclude that all four elicitation mechanisms, and hence all three design feature variations, resulted in both high and very similar levels of optimal bidding.

5.1. Heterogeneity

We only collect one demographic characteristic, which is participant gender. Appendix Table A.3 analyzes whether optimal bidding behavior differs by gender. Columns (1) and (3) pool all treatments and test for overall differences in optimal bidding for voucher and soap respectively. Columns (2) and (4) split out the features and interact with gender. Overall, women are slightly more likely than men to bid optimally for the voucher, but the difference is never more than 4.5 percentage points and never significant. The pattern is less consistent for soap and some point estimates are fairly large but again never significant. F-tests for the joint hypothesis of no differences across the eight feature × gender subgroups have p-values close to 1. Realistically we are underpowered to detect design feature-specific differences by gender. That said, our findings do not suggest major heterogeneity along this dimension.

5.2. Limitations

Default. A common implementation challenge in WTP studies is "default," or refusal to pay once the price has been realized (Maffioli et al., 2020). This seems to be a particular challenge when payment takes place some time after elicitation, for example when the good or service is offered on credit (Grimm et al., 2020), and when contract enforcement is weak.

Default may result from unanticipated liquidity or preference shocks that change the participant's WTP *ex post*. It may also reflect misunderstanding of the mechanism: if participants make mistakes, and end up committing to prices in excess of their true WTP, they may be more likely to default.

Default is a concern in part because if participants anticipate that they will be allowed to default, they may have less incentive to truthfully report their WTP.

In our study, payment took place immediately after WTP elicitation and enumerators were required to ensure participants had enough cash with them before recording their final WTP, so we did not observe any default. But our findings of high comprehension and optimal bidding suggest that, in implementations similar to ours, misunderstanding is not the most likely explanation of default behavior.

Anchoring. A common concern in WTP elicitation is *anchoring* or *framing effects* (see, e.g. Andersen et al., 2006). The range of prices under consideration guides responses away from values outside that range. Mass points or midpoints of the price distribution or price list might also serve as salient anchors. Prompting a particular set of responses (MPL) might anchor participants on different values than free response (BDM).

Our experiment is not designed to isolate anchoring effects, as the same support of the price distribution is always provided and the MPL procedure asks about different prices sequentially and verbally rather than presenting a list with a salient midpoint. We therefore conjecture, consistent with our overall null findings, that salient anchors are similar across our treatments.[16]

6. Conclusion

In this paper we experimentally test four variants of the Becker–DeGroot–Marschak mechanism for eliciting willingness to pay, with participants in rural Uganda. Our goal was to understand whether variation in design features led to differences in optimal bidding, in particular for a voucher with known induced value, following Cason and Plott (2014).

Comprehension and optimal bidding are high and similar across treatments. This is good news for practitioners, giving confidence that these procedures can elicit meaningful, truthful willingness to pay reports from participants. Design considerations can take the lead, rather than concerns about miscomprehension.

We note however, that our focus on optimal bidding led us to choose products (voucher and soap) with known value or familiar characteristics. Many WTP studies consider unfamiliar products, or those with higher or more uncertain value. We see our findings as good news for these settings too. But our findings do not obviate the need for careful piloting and comprehension checks in those settings.

Acknowledgments

We thank Apollo Tumusiime and Charles Angebault for outstanding field support. We thank the editor, Tom Vogl, two anonymous referees, Shengwu Li, David McKenzie, and Sebastian Schweighofer-Kodritsch for very helpful feedback. This research was supported by the Swedish Research Council, grant no. 2015-03426. Burchardi, de Quidt, and Lerva acknowledge financial support from Handelsbanken's Research Foundations. All errors and omissions are our own. Declarations of interest: none.

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jdeveco.2021.102701.

References

Akbarpour, M., Li, S., 2020. Credible auctions: A trilemma. Econometrica 88 (2), 425–467.
Andersen, S., Harrison, G.W., Lau, M.I., Rutström, E.E., 2006. Elicitation using multiple price list formats. Exp. Econom. 9 (4), 383–405.
Ashraf, N., Berry, J., Shapiro, J.M., 2010. Can higher prices stimulate product use? Evidence from a field experiment in Zambia. Amer. Econ. Rev. 100 (5), 2383–2413.
Ausubel, L.M., 2004. An efficient ascending-bid auction for multiple objects. Amer. Econ. Rev. 94 (5), 1452–1475.
Becker, G.M., DeGroot, M.H., Marschak, J., 1964. Measuring utility by a single-response sequential method. Behav. Sci. 9 (3), 226–232.
Berkouwer, S.B., Dean, J.T., 2020. Credit and attention in the adoption of profitable energy efficient technologies in Kenya. Berkeley Mimeo.
Berry, J., Fischer, G., Guiteras, R., 2020. Eliciting and utilizing willingness to pay: Evidence from field trials in northern Ghana. J. Polit. Economy 128 (4), 1436–1473.
Berry, J., Mukherjee, P., 2019. Pricing private education in urban India: Demand, use, and impact. Working paper.
Bohm, P., Lindén, J., Sonnegård, J., 1997. Eliciting reservation prices: Becker–DeGroot–Marschak mechanisms vs. markets. Econ. J. 107 (443), 1079–1089.
Brebner, S., Sonnemans, J., 2018. Does the elicitation method impact the WTA/WTP disparity? J. Behav. Exp. Econ. 73, 40–45.
Breitmoser, Y., Schweighofer-Kodritsch, S., 2021. Obviousness around the clock. Exp. Econom.
Burchardi, K., de Quidt, J., Gulesci, S., Sulaiman, M., 2020a. Credit constraints and demand for remedial education: Evidence from Tanzania. Working Paper.
Burchardi, K., de Quidt, J., Lerva, B., Tripodi, S., 2020b. Credit constraints and capital allocation in agriculture: Theory and evidence from Uganda. Working Paper.
Cason, T.N., Plott, C.R., 2014. Misconceptions and game form recognition: Challenges to theories of revealed preference and framing. J. Polit. Economy 122 (6), 1235–1270.
Chassang, S., Miquel, G.P.i., Snowberg, E., 2012. Selective trials: A principal-agent approach to randomized controlled experiments. Amer. Econ. Rev. 102 (4), 1279–1309.
Cohen, J., Dupas, P., 2010. Free distribution or cost-sharing? Evidence from a randomized malaria prevention experiment. Q. J. Econ. 125 (1), 1–45.
Cole, S., Fernando, A.N., Stein, D., Tobacman, J., 2020. Field comparisons of incentive-compatible preference elicitation techniques. J. Econ. Behav. Organ. 172, 33–56.
Dupas, P., 2014. Short-run subsidies and long-run adoption of new health products: Evidence from a field experiment. Econometrica 82 (1), 197–228.
Fischer, G., Karlan, D., McConnell, M., Raffler, P., 2016. To charge or not to charge: Evidence from a health products experiment in Uganda. Working paper.
Fischer, G., Karlan, D., McConnell, M., Raffler, P., 2019. Short-term subsidies and seller type: A health products experiment in Uganda. J. Dev. Econ. 137, 110–124.
Grimm, M., Hartwig, R., 2018. Unblurring the market for vision correction: A willingness to pay experiment in rural Burkina Faso. IZA DP No. 11929.
Grimm, M., Lenz, L., Peters, J., Sievert, M., 2020. Demand for off-grid solar electricity: Experimental evidence from Rwanda. J. Assoc. Environ. Resour. Econ. 7 (3), 417–454.
Hanemann, W.M., 1994. Valuing the environment through contingent valuation. J. Econ. Perspect. 8 (4), 19–43.
Heckman, J.J., Vytlacil, E., 2005. Structural equations, treatment effects, and econometric policy evaluation. Econometrica 73, 669–738.
Hoffmann, V., 2009. Intrahousehold allocation of free and purchased mosquito nets. Amer. Econ. Rev.: Pap. Proc. 99 (2), 236–241.
Hoffmann, B., 2018. Do non-monetary prices target the poor? Evidence from a field experiment in India. J. Dev. Econ. 133, 15–32.
Hoffmann, V., Barrett, C.B., Just, D.R., 2009. Do free goods stick to poor households? Experimental evidence on insecticide treated bednets. World Dev. 37 (3), 607–617.
Kagel, J.H., Harstad, R.M., Levin, D., 1987. Information impact and allocation rules in auctions with affiliated private values: A laboratory study. Econometrica 55 (6), 1275.
Lerva, B., 2020. Quantifying externalities in technology adoption: Experimental evidence from Ugandan farmers. Working paper.
Li, S., 2017. Obviously strategy-proof mechanisms. Amer. Econ. Rev. 107 (11), 3257–3287.
Maffioli, A., McKenzie, D., Ubfal, D., 2020. Estimating the demand for business training: Evidence from Jamaica. World bank policy research working paper no. 9415.
Mazar, N., Koszegi, B., Ariely, D., 2013. True context-dependent preferences? The causes of market-dependent valuations. J. Behav. Decis. Mak. 27 (3), 200–208.
Rutström, E.E., 1998. Home-grown values and incentive compatible auction design. Internat. J. Game Theory 27 (3), 427–441.
Uganda Bureau of Statistics, 2016. Uganda National Panel Survey (UNPS) 2013–2014 (UGA_2013_UNPS_v01_M). Government of Uganda, Available at: https://microdata.worldbank.org/index.php/catalog/2663. (Accessed 18 May 2021).
Yishay, A.B., Fraker, A., Guiteras, R., Palloni, G., Shah, N.B., Shirrell, S., Wang, P., 2017. Microcredit and willingness to pay for environmental quality: Evidence from a randomized-controlled trial of finance for sanitation in rural Cambodia. J. Environ. Econ. Manage. 86, 121–140.

[16] It is possible that different, untested combinations of features might induce different anchoring effects (a referee suggests BDM with Unstated price distribution might differ more from MPL).