REVIEW OF CRIME, PEACE AND SOCIETY
ISSN (O): 3079-6105, ISSN (P): 3079-6091
Vol. 3, No. 6, 2026, 397-409

Fielding Without a Frame: Design, Paradata and Data Quality in a Mobile Survey of Lahore Students

Ali Aman[1], Rana Eijaz Ahmad[2]

Abstract

Survey research on political attitudes outside high-income democracies usually proceeds without a sampling frame and without platform data against which self-reports can be checked. What remains is the instrument and the record it keeps of its own administration. This article reports the design and the measured data quality of a web survey of university students in Lahore, Pakistan (analytic N = 779, fielded 23 February to 15 March 2026), covering institutional trust, political conflict online and support for extra-constitutional alternatives. A pilot on a general-purpose form platform failed in four ways, and each failure was answered by a feature of a purpose-built application: one experimental list per respondent, section-level timing, completion enforced at the section level while refusal remained available, and local caching for interrupted mobile sessions. The paradata show a sample arriving in institution-led bursts, with 32.7% of responses on a single day; a median completion time of 11.4 minutes against the 10 to 15 announced at consent, with only 35.3% of respondents inside that window; and 98.8% mobile completion. Refusal followed the international pattern for financial items, at 20.9% on a household-asset question against 1.7% on direct support for a military takeover. Consent refusals, breakoff points and duplicate submissions were not recorded at all, and the article specifies the fields that would have captured them. Where a design cannot warrant its sample, the administration is the evidence, and it should be reported to the standard usually reserved for results.

Article History
Received 20 August 2026
Accepted 25 August 2026
Published 07 September 2026

OPEN ACCESS

Keywords
web surveys, paradata, non-probability samples, item non-response, sensitive questions, Pakistan

1. Introduction

Two questions in the same instrument, answered by the same 779 students, produced opposite patterns of refusal. Asked which of five assets their household owned, 163 respondents (20.9%) chose "prefer not to answer". Asked whether they supported the military taking over the government to ensure national stability, 13 (1.7%) declined to answer. On the face of it, a question about the army taking power drew a small fraction of the refusal drawn by a question about owning a motorcycle or an air conditioner.

The two items did not, however, offer respondents the same ways out, and that matters from the outset: the asset question was a checklist on which a respondent either disclosed or declined, while the military question was a five-point scale whose neutral midpoint 297 respondents (38.1%) selected. Section 5.4 sets both formats out and gives the comparison its bounds.

Against the survey-methods literature, only one half of that contrast is surprising. Income and financial-asset questions are associated with missing-data rates between 20% and 40%, ordinary survey questions with rates between 1% and 4% (Tourangeau & Yan, 2007; Yan et al., 2010). The asset result is therefore unremarkable. What is notable is the behaviour of an extra-constitutional intervention item in a country where political expression carries documented legal and institutional risk (Ahmed et al., 2025; Anwar et al., 2024): its non-

[1] Ali Aman (Corresponding Author): PhD Scholar, Department of Political Science, University of the Punjab, Lahore (aliaman000@gmail.com).
[2] Prof. Dr. Rana Eijaz Ahmad: Chairperson, Department of Political Science, University of the Punjab, Lahore.

disclosure rate is an ordinary question's rate. Interpretation of such a result depends on documentation of its production conditions, that is, on respondent recruitment, the assurances communicated, the refusal provision, and the screening applied afterwards.

Those conditions are what this article reports, and it reports them largely through paradata: the record a web instrument keeps of its own administration — how long each section took, on what device, in what order, and what was left unanswered — collected passively while respondents answer. It is a fieldwork paper. It carries no substantive result about trust, polarisation or misinformation; those are reported elsewhere. Its subject is the administration of a politically sensitive survey in a setting that offers neither of the two resources on which survey inference is usually built.

The first absent resource is a sampling frame. No accessible list of university students in Lahore exists. The Higher Education Commission publishes the recognised institutions (Higher Education Commission, n.d.), but enrolment registers are held by universities and are not released for research. Recruitment therefore proceeds through networks, and the sample is a non-probability sample whose composition, within whichever institutions a study selects, is decided by who passes the link on. The methodological consequences are well established: self-selection and under-coverage produce estimates whose bias cannot be quantified from the sample itself, and claims of representativeness are not available (Baker et al., 2013; Bethlehem, 2010).

The second absent resource is platform data. Studies of digital political behaviour in high-income settings increasingly validate self-reports against trace data. In Pakistan no such access exists, so every measure of the information environment is a self-report, and the instrument that collects it bears the full weight of measurement.

When both resources are missing, the design of the instrument and the record of its administration become the evidence a reader has for judging what the data can support. Yet fieldwork is typically compressed into a paragraph. This article takes the opposite approach and reports the administration in full — the setting, the pilot and what it exposed, recruitment and consent, the quality metadata, the limitations, and the lessons — including the parts that did not work.

The intended reader is specific. Anyone fielding a survey on contested political questions where no frame exists faces the same four decisions this study faced: which platform to field on, what to require of respondents, what to record passively while they answer, and what to do with records that look careless. Sections 3 to 5 give a worked answer to each, with the numbers that answer produced, and Section 6 lists the fields this design should have written and did not. Both halves are meant to be reused, and the second is the cheaper lesson. Treating the administration of a survey as an empirical object, rather than as a paragraph of preamble, is what this article offers a field in which access to frames and platform data is the exception rather than the rule.

2. Setting and constraints

The study population was students enrolled at Higher Education Commission-recognised universities in Lahore. Twelve institutions appear in the analytic sample. The public and private sectors are close to evenly represented, at 404 and 374 respondents (51.9% and 48.0%), as are men and women, at 397 and 376 (51.0% and 48.3%); 96.3% were undergraduates.

University students are a defensible population for a study of digitally mediated political conflict, and a convenient one, and the two reasons should not be confused. Political

information reaches them through the same peer networks the study asks about, and campus life gives their political talk a shared institutional setting; they are also, unlike most of the country, reachable online at all. Neither reason makes them representative of Pakistani youth, and nothing here treats them as such. Their institutional composition is given in **Table 1**, and Section 5.1 shows it to be a product of how the link travelled and not of any sampling decision.

Table 1

Institutional Composition of the Analytic Sample (N = 779)

Institution	n	%
Government College University (GCU)	258	33.1
Forman Christian College (FCCU)	201	25.8
University of Central Punjab (UCP)	101	13.0
University of the Punjab (PU)	53	6.8
Lahore College for Women University (LCWU)	46	5.9
University of Engineering & Technology (UET)	34	4.4
Lahore University of Management Sciences (LUMS)	21	2.7
Beaconhouse National University (BNU)	19	2.4
National College of Arts (NCA)	13	1.7
Lahore School of Economics (LSE)	11	1.4
University of Lahore (UOL)	11	1.4
Kinnaird College for Women (KCW)	10	1.3
Could not be classified	1	0.1

Note. The instrument listed eleven institutions and an "Other (HEC-recognised)" option with a free-text box. Forty-seven respondents chose Other; forty-six of them named Lahore College for Women University, which is therefore the twelfth institution, and one entry could not be classified. Public sector 404 (51.9%), private 374 (48.0%), one case unclassified on sector.

Three setting features governed every design decision. The first is topic sensitivity. Items on confidence in the military establishment and the intelligence services, on partisan hostility, and on support for extra-constitutional intervention were administered to respondents in a country where digital political expression is subjected to documented legal and institutional pressure (Ahmed et al., 2025; Anwar et al., 2024). Under such conditions, self-administration and anonymity constitute measurement decisions and not merely ethical commitments, since the alternative is systematic misreporting (Tourangeau & Yan, 2007).

The second is the device. The population is reached through phones. Pakistan had 116 million internet users and 190 million active cellular connections in early 2025 (Kemp, 2025). In the achieved sample, 98.8% of respondents completed on a mobile device. Research in comparable settings has reached populations the same way, and for the same reason — landlines are few and mobile coverage is rising — while noting how little is known about the quality of the samples such methods produce (L'Engle et al., 2018). An instrument that assumed a desktop, a stable connection or an uninterrupted sitting would have been answering a different population.

The third is the channel. Recruitment had to run through the same semi-private routes through which political content circulates among students, above all messaging groups. That is a practical necessity, and it parallels what has been documented next door in India, where

WhatsApp groups carry political claims into everyday circulation (Banaji et al., 2019). As Section 5 shows, the choice leaves a clear signature in the data.

3. From a pilot platform to a purpose-built instrument

The instrument was piloted before it was fielded, for the reasons the trials literature gives for piloting anything — to test feasibility and recruitment before the main effort is committed (Thabane et al., 2010) — and, as it turned out, to expose problems while they could still be fixed. Pilot work ran on a general-purpose form platform with students enrolled at universities in Gujranwala, a city near Lahore that stands in for the target population, between 23 December 2025 and 10 January 2026. Piloting in a different city kept the networks that would later carry the main study free of students who had already seen the questionnaire.

The pilot exposed four failures, all of them properties of the platform rather than of the questionnaire.

Routing leaked on the experimental module. The survey contained a list-count module in which one group saw a list of four non-sensitive items and the other saw the same list plus a sensitive fifth item, with respondents reporting only how many items they supported. The method depends on a respondent seeing one list and not the other (Blair & Imai, 2012). Section placement on the form platform let essentially all pilot respondents see both, which destroyed the comparison.

Timing was not recorded usably. Without completion times, no defensible screen for careless responding is possible, and completion time is the single most reliable non-reactive indicator of low-quality records (Leiner, 2019).

Enforced completion could not coexist with refusal. The platform's required-question setting was binary: either an item could be skipped, or it had to be answered. There was no way to require that a respondent engage with a sensitive item while still offering an explicit way to decline. That distinction matters, because forcing answers raises dropout and lowers answer quality, whereas an explicit non-disclosure option preserves the respondent's autonomy without producing silent missing data (Décieux et al., 2015).

Branching was too shallow to collect paired ratings of a respondent's own party and its rival, which meant a planned measure of partisan affect could not be constructed.

A purpose-built application, written in React with a Firebase backend, was fielded instead. Its features map one-to-one onto the failures above and onto the risks the methods literature identifies for long self-administered instruments.

List-module allocation is performed once, within the application, at a fixed transition, and the version not presented is discarded from the stored record, so that contribution to both arms is prevented. No re-allocation is produced by back-navigation. Section-level timings and total duration are recorded passively at submission, together with the device string and a mobile indicator; direct identifiers and IP addresses are not retained. The device string, a model and browser version shared across identical handset and browser combinations, is the file’s only quasi-identifier, and its weakness is demonstrated in Section 6. Completion validation is enforced section by section, and every sensitive item carries an explicit "prefer not to say" or "prefer not to answer" option, the combination the pilot platform could not provide. Responses in progress are cached on the device, a warning is triggered before tab closure, and cache clearance follows submission success. These provisions were designed for a population answering on phones over intermittent connections.

Three smaller choices follow the same logic. The questionnaire advances section by section, so a respondent meets a bounded task at each step; a progress indicator and an estimate

of the minutes remaining address the point at which respondents decide whether to continue (Galesic, 2006); and back-navigation is allowed without re-drawing the arm assignment, so no one can be moved into the other arm or shown a second list.

Two burden decisions were taken at the same time. A thirteen-item social-desirability battery (Reynolds, 1982) that had been carried in the pilot was removed, because questionnaire length is associated with lower participation and poorer answer quality (Galesic & Bosnjak, 2009), and because long instruments invite the shortcuts that satisficing theory describes (Krosnick, 1991). The veracity-judgement module was reformatted to binary true/false items, in line with validated practice for measuring discernment as performance (Maertens et al., 2024). The instrument was then locked before fielding. As locked, it presented approximately 70 items, the exact number depending on a respondent's routing.

4. Recruitment and consent

Twelve universities in Lahore were selected for the study. Recruitment then proceeded through three channels, that is, departmental mailing lists, link forwarding by faculty members to their own students, and student WhatsApp groups. No panel provider or sample supplier was involved in distribution. **Table 2** summarises the study as fielded.

Table 2

Fieldwork at a Glance

Item	Detail
Design	Self-administered web survey, single wave
Population	Students at HEC-recognised universities in Lahore
Field period	23 February – 15 March 2026
Platform	Purpose-built React application, Firebase backend, anonymous authentication
Recruitment	Departmental mailing lists; faculty forwarding; student WhatsApp groups
Sampling frame	None available
Sessions past the consent screen	1,071
Completed submissions	786
Completed / sessions	73.4%
Exclusions	7 (named an institution outside Lahore)
Analytic N	779 — twelve institutions, plus one case that could not be classified
Identifiers collected	None; no IP addresses stored
Passive metadata	Total duration, section timings, device string, mobile flag

Because invitations travelled through networks and not from a list, the number of individuals reached by the invitation is unknown, and no response rate can be computed. There is nothing peculiar to this study in that limitation; it is a general property of open web surveys, and the reporting literature recommends avoidance of the term "response rate" altogether, with publication instead of whichever ratios a design can support (Eysenbach, 2004). This design supports the ratio of completed submissions to consent-screen sessions: **786 of 1,071, or 73.4%**. Seven further respondents were excluded after identification of a Gujranwala institution, the first author's own university, and not a Lahore one; circulation of the link had extended beyond the target city. Their responses fall outside the population definition, and their exclusion leaves an analytic sample of 779. Neither author had taught any of them.

Entry was gated twice. A consent screen stated that participation was voluntary and anonymous, that no name, student identification number or IP address would be collected, that

responses would be analysed in aggregate, that the respondent could withdraw at any point, and that the survey would take 10 to 15 minutes. Respondents then confirmed consent and current enrolment; failing either routed them out before any substantive item. The study was approved by the Advanced Studies and Research Board of the University of the Punjab at its meeting of 7 November 2024, with notification dated 18 November 2024. That approval covers the doctoral project within which this survey was fielded, and it runs to the thesis submission deadline the same notification sets.

5. What the quality metadata showed

5.1 The sample arrived in bursts, and the bursts were institutionally concentrated

Completions occurred on 16 of the 21 days of fielding, but they were not spread across them. A single day, 5 March, produced 255 completions — 32.7% of the analytic sample — and the three busiest days produced 522, or 67.0%.

Each burst was dominated by one institution, though never composed of it alone. On the four days that produced more than 80 completions, the leading institution supplied 65.3% (23 February), 56.6% (25 February), 48.3% (4 March) and 47.1% (5 March) of that day's responses — and the college that led 5 March contributed 59.7% of its own total on that single day. This is what recruitment by forwarding looks like in the data: a faculty member or a group administrator passes the link to a cohort, and a day's worth of the sample arrives from one institution. Whether any single day corresponds to one forwarding event cannot be established, because the instrument recorded no channel indicator — a gap Section 6 returns to.

The tail matters as much as the peaks. After 5 March, the remaining ten days of fielding produced 47 completions in total, and five days produced none at all (Table 3). A field period of three weeks was not, in practice, three weeks of data collection; it was four or five productive days and a long wait for stragglers. A study that closed during the first quiet stretch would have lost roughly a third of its sample, and one that budgeted effort evenly across the calendar would have misjudged where the effort goes, which is into securing the next forwarding event.

Table 3

Completions by Field Day (N = 779)

Date (2026)	Completions	Date (2026)	Completions
23 February	124	4 March	89
24 February	19	**5 March**	**255**
25 February	143	6 March	3
26 February	20	9 March	21
27 February	56	10 March	11
1 March	1	11 March	7
2 March	13	14 March	1
3 March	12	15 March	4

Note. Pakistan Standard Time. Five days within the field period produced no completions and are omitted. The three busiest days account for 67.0% of the analytic sample.

Two consequences follow. The first concerns composition. Institution selection was deliberate, but respondent distribution across those institutions was not, being a by-product of link forwarding, and its description should reflect that origin and not a presentation as though within-institution sampling had occurred. The second is inferential: any institution-level characteristic inherits a clustered structure, and ordinary standard errors will understate it. Both

consequences are the applied form of non-probability literature expectations where recruitment travels through networks (Baker et al., 2013; Bethlehem, 2010).

5.2 Completion time, against the time the consent screen promised

The consent screen announced 10 to 15 minutes. Median completion required 11.4 minutes (interquartile range 9.1 to 16.1), so the announcement was accurate at the distribution centre and inaccurate for most individuals: only 35.3% of respondents finished inside the announced window. A slightly smaller proportion, 31.1%, achieved completion in less time, and 24.1% required between 15 and 30 minutes. Completion by a small group required considerably longer, 3.6% between 30 and 60 minutes and 1.8% more than an hour, a reflection of sessions left open and not of sustained application, and a reminder that elapsed time on a self-administered instrument is not time on task.

Time allocation is equally informative. The longest block was not the veracity battery but the digital-habits and network-composition block, at a median of 178 seconds, approximately 27% of the summed section medians, against 124 seconds for the battery and 23 seconds for the consent screen. The full distribution is reported in Table 4.

Table 4

Completion Time Against the Announced 10–15 Minutes (N = 779)

Completion time	n	%
Under 5 minutes (speeder flag)	32	4.1
5 to under 10 minutes	242	31.1
10 to under 15 minutes (announced window)	**275**	**35.3**
15 to under 30 minutes	188	24.1
30 to under 60 minutes	28	3.6
60 minutes or more	14	1.8

Note. Computed from summed section timings. Median 11.4 minutes, interquartile range 9.1 to 16.1. Cases of 30 minutes or more reflect sessions left open rather than time on task.

5.3 Devices

Of 779 respondents, 770 (98.8%) achieved completion on a mobile device; 589 (75.6%) were recorded on Android and 181 (23.2%) on iOS. Desktop computers were used by nine respondents. Instrument caching and resumption were therefore not a marginal convenience but the collection condition for most of the data.

One incidental observation is relevant to a later section: across 779 respondents only 54 distinct device strings were recorded. Device-string identification extends to a model and a browser version, not to a person, and in a population with few popular handsets the collision rate is high.

5.4 What respondents actually declined to answer

Four items carried an explicit non-disclosure option, and **Table 5** reports how often each of them was used.

Table 5

Use of the Non-Disclosure Option, by Item (N = 779)

Item	Non-disclosure n	%
Household assets owned	163	20.9
Direct support for a military takeover	13	1.7
Gender	6	0.8
Ethnicity / mother tongue	3	0.4

Note. These four items carried an explicit "prefer not" option. On the military-takeover item a further 297 respondents (38.1%) selected the neutral midpoint, which this design cannot distinguish from reticence.

The distribution follows the financial-item literature and not intuitions about political risk. Household-asset non-disclosure reached 20.9%, a value inside the 20% to 40% band associated with income and asset questions (Tourangeau & Yan, 2007; Yan et al., 2010). Almost no non-disclosure was recorded on the demographic items. Non-disclosure on direct military-takeover support reached 1.7%, a value within the 1% to 4% band associated with ordinary questions.

That comparison requires bounding, because construction of the two items was not identical, and their differences are of four kinds simultaneously. The asset item listed five household possessions for selection, with "prefer not to answer" as an exclusive alternative, so that disclosure and refusal were the only available positions and no intermediate position existed. The military item comprised five ordered options from strongly oppose to strongly support, with "prefer not to answer" alongside them, so a neutral midpoint was available, and selection of it was made by 297 respondents (38.1%). Beyond format, the asset item is a factual request and the military item an attitudinal one; the asset item concerns the household, so disclosure extends to a family's standing and not merely to an individual position, while the military item concerns the respondent's personal conviction; and box selection and scale placement are different tasks with different costs. Any of the four differences could carry part of the discrepancy.

The consequence is reticence measurement over different ranges. Military-item refusal is 1.7%; refusal together with the midpoint is 39.8%. The first treats no midpoint answer as reticence, the second treats every midpoint answer as reticence, and neither assumption is independently credible, since a neutral answer may represent indifference, genuine ambivalence, issue ignorance, or a costless refusal. **This design cannot separate them**, and the honest interpretation of the asset comparison is therefore a location between those boundaries and not at either.

What the distribution establishes is narrower and still useful: explicit refusal was used heavily on household wealth and almost not at all on the army, and only the military item offered a midpoint with the capacity for absorption of what refusal might otherwise have carried. A low refusal rate is not candour evidence, and nothing here should be read as establishing frankness in the military-item responses. The consideration is not hypothetical: Chinese list experiments place self-censorship of regime support at 24.5 to 26.5 percentage points, far above the impression produced by direct questioning (Robinson & Tannenberg, 2019). Where concealment on that scale is possible, a 1.7% refusal rate provides a researcher with almost no information about respondents' actual convictions.

5.5 Careless responding: two screens that caught different people

Two pre-specified screens were available. The first was speed. A threshold of 300 seconds across summed section timings was fixed before fielding, on the pilot's timing distribution and a reading-time floor for the instrument's demanding blocks, and it flagged 32 respondents (4.1%). Because a threshold set in advance can still be a poor one, it is worth benchmarking after the fact: 300 seconds is 44.0% of this study's median completion time, and cut-offs at 30% and 50% of that median would have flagged 12 and 51 respondents. The pre-set value sits between the two, which is the reassurance a reader is entitled to ask for. Speed is

in any case the screen the literature supports most strongly, since completion time identifies careless records more reliably than the alternatives, and Leiner (2019) proposes a relative completion-speed index as a refinement of it.

The second screen was straight-lining, that is, identical answers across every item in a grid. It flagged 32 respondents (4.1%) on the seven-item institutional-trust grid and 51 (6.5%) on the four-item intervention grid. Where those 32 sat on the scale matters for how the flag should be read: 17 of them answered "none at all" to all seven institutions, 7 "a fair amount", 6 "not very much" and 2 "a great deal". A respondent with no confidence in any of the seven has an honest reason to give seven identical answers, so on this grid the screen flags a pattern that is often a substantive position. A three-item conspiracy grid produced 173 (22.2%), which is reported here for completeness and not used as a quality flag: across three closely related statements, agreement throughout is an ordinary attitude pattern rather than a sign of inattention. Patterns of this kind are response styles, and response styles are largely consistent within a respondent over the course of a questionnaire (Weijters et al., 2010), which is a further reason not to read a short grid as a quality flag. No respondent straight-lined all three grids.

The two screens flagged non-overlapping sets of respondents: none of the 32 trust-grid straight-liners was among the 32 speeders. It is tempting to read that as a local counter-example to the association between speeding and straight-lining reported in web surveys (Zhang & Conrad, 2014). The temptation should be resisted. With 32 flagged on each screen out of 779, the overlap expected if the two were unrelated is 1.3 respondents; observing none gives Fisher's exact $p = .64$. Across the whole sample, straight-lining is unrelated to how long people took (point-biserial $r = -.010$, $p = .785$; median 667 seconds for straight-liners against 684 for everyone else). These data are simply too thin at the tails to speak to the association either way, and treating the absence of an overlap as a finding would be the error Gelman and Stern (2006) describe. What does follow is operational: the two screens identified different people here, so applying only one of them would have left the other group unexamined.

Flagged cases were retained, and the question a fieldwork paper can answer is whether they answered differently from everyone else. Largely they did not. Speeders declined the military item at 3.1% against 1.6% (Fisher $p = .423$), took its midpoint at 40.6% against 38.0% ($p = .853$), straight-lined the trust grid at 0.0% against 4.3% ($p = .637$) and the intervention grid at 6.2% against 6.6% ($p = 1.000$). The one difference of any size is on the asset item, which speeders declined at 34.4% against 20.3%, and even that does not reach conventional significance ($p = .073$) on 32 cases. That last figure is the right place to state the limit of this comparison: with 32 flagged cases against 747, only differences of roughly 20 percentage points or more are detectable at conventional power, so these are weak nulls rather than demonstrations of sameness. Inferential models fitted to these data are reported in companion articles, and each refits its own models with and without flagged cases; what belongs here is the finding that the flagged records are not a behaviourally distinct group.

5.6 Missingness

Enforced completion operated as intended. Across the 60 non-routed substantive columns no empty cells are present in the analytic dataset, and exactly one list-module version was answered by every respondent. Refusal, where utilised, is recorded as an explicit category and not as absence, and preservation of that distinction was the design objective. In accordance with the consent-screen undertaking, all reporting in this article is aggregate; the Section 6

duplicate check is a record comparison for data integrity, with no presentation of any individual's answers.

6. Limitations: what the design could not see

A fieldwork report with a success account alone is of limited utility. Four gaps are stated precisely below, since an inexpensive remedy exists for each.

Consent refusals leave no trace. Respondents declining consent were routed out, and no record was written. Session logging begins at consent-screen departure, so the denominator for any participation rate, whether invitation recipients or first-page visitors, does not exist. No CHERRIES-style participation rate can be computed for this study (Eysenbach, 2004). Refusal logging as a bare, anonymous event would have cost nothing and would have rendered the consent stage measurable.

Breakoff cannot be located. The session log records only that a session started and whether it completed. Of 1,071 sessions, 285 did not finish, and it is not possible to say where they stopped. Where a respondent stops is what the burden literature is about: the position and the duration of a block of questions predict how interested and how burdened a respondent feels at that moment (Galesic, 2006). This design cannot test any of that on its own data, even though it records section timings for everyone who finished. Writing the last section reached into the session record, which is one additional field, would have converted 285 unusable rows into a usable breakoff curve.

Duplicate submissions cannot be ruled out. Three response pairs are identical across all 60 non-routed items; two pairs arrived from an identical device string less than a minute apart, and the third six days apart from different devices. A further six pairs are separated by exactly one item of 60. Without IP retention and with only 54 distinct device strings, no available screen can separate a re-submission from two students with similar handsets and similar convictions. A plausible technical explanation exists for the sub-minute pairs. Anonymous authentication was performed for every respondent, so a stable per-session identifier existed throughout; no application use was made of it. Each submission was written as a new document under a server-generated key, no rule restricted a respondent identifier to a single write, and after write failure a further Submit press was invited by the interface, so server success combined with handset timeout would produce exactly this pattern. That mechanism fits the observation without constituting proof. Submission under the identifier already held would have rendered a second write impossible.

Deliberate misreporting is out of reach. The screens above detect inattention, not intent; non-reactive measures of this kind cannot identify deliberate faking (Leiner, 2019), and the instrument carried no attention check and no consistency pair. For a survey about what people will and will not say about powerful institutions, that is a real gap, though not one a technical screen closes.

Assignment was not blocked. The two versions of the list module were assigned by an unblocked draw inside the application. Simple randomisation of that kind is unbiased in expectation but does not guarantee balance in any one sample, and blocking on a few observed characteristics costs little at the design stage.

Beyond these four, the standing limitations of the design apply. The sample is non-probability, drawn from one city and overwhelmingly undergraduate, and supports no population inference. The study is cross-sectional, so every relationship it reports is an association. And recruitment through messaging groups is a plausible source of selection on

precisely the network characteristics some of the measures concern; it is disclosed here rather than claimed to have been ruled out.

7. Lessons

The decisions this study took, the risk each was meant to address, and what the data say about them are set out in **Table 6**.

Table 6

Design Decisions and What the Data Show

Decision	Risk addressed	What the data show	Recommendation
Purpose-built app replacing a form platform	Routing leakage, no timing, forced disclosure	Each respondent answered exactly one list version; timings available for all	Keep
Section validation + explicit refusal option	Silent missing data vs forced disclosure	0 empty cells; refusal visible and concentrated on one item	Keep
Passive timing and device metadata	Undetectable careless responding	32 speeders flagged; they differ from the rest on none of five answering behaviours	Keep; add a per-question timer on the longest block
Local caching and resume	Mobile session loss	98.8% mobile completion	Keep
Session log with start/complete only	Unmeasured attrition	285 unfinished sessions, location unknown	Log consent refusals and last section reached
New document per submission	Duplicates	3 exact pairs, 2 of them sub-minute; 6 near pairs	Client-generated id, written once
Unblocked assignment in the app	Chance imbalance	—	Block on a small number of observed characteristics

Five points carry beyond this study. Pair enforced completion with an explicit refusal option: an answer requirement without one raises dropout and degrades quality (Décieux et al., 2015), while the combination produced a dataset without missing cells in which refusal remains visible as a selection. Record the administration, not merely the answers, since timings and a device indicator are passive and carry the whole of Section 5. Log what does not happen: consent refusals and the final section reached constitute one field each, and their absence is the largest deficiency in this study's record. Make submission idempotent, which eliminates the duplicate category these data cannot resolve. And expect the sample in bursts, with a record of the channel behind each; institutional composition here was determined by forwarding behaviour, and its recovery afterwards is possible only as a date-by-institution pattern.

8. Conclusion

A survey fielded without a sampling frame and without platform data cannot borrow authority from its sampling design, because it has none to borrow. What it can do is make its administration inspectable: state how people were reached, what they were promised, what they were allowed to refuse, how long they took, on what devices, and which records were screened and why. This study did that for 779 students in Lahore, and the record is mixed in a way that is more useful than a clean one: complete data with refusal preserved as a visible category; a sample assembled in institutional bursts; a promised completion time that fitted a third of respondents; two quality screens that caught disjoint groups; and three parts of the administration — consent refusals, breakoff location and duplicate submissions — not recorded at all, each for want of a field that would have cost nothing to write.

The refusal contrast with which this article opened is, in the end, a case for the same argument. Household assets drew twelve times the non-disclosure of a question about military

takeover. Neither number interprets itself. Both become interpretable only against the record of how the instrument was built, what alternatives it offered a reluctant respondent, and what it failed to observe.

References

Ahmed, Z. S., Yilmaz, I., Akbarzadeh, S., & Bashirov, G. (2025). Contestations of internet governance and digital authoritarianism in Pakistan. *International Journal of Politics, Culture, and Society, 38*, 499–526. https://doi.org/10.1007/s10767-024-09493-2

Anwar, A., Hassan, A., & Yousaf, S. (2024). Digital authoritarianism and journalistic dissent in Pakistan: An empirical investigation of PECA. *Journal of Social & Organizational Matters, 3*(4), 641–651. https://doi.org/10.56976/jsom.v3i4.170

Baker, R., Brick, J. M., Bates, N. A., Battaglia, M., Couper, M. P., Dever, J. A., Gile, K. J., & Tourangeau, R. (2013). Summary report of the AAPOR task force on non-probability sampling. *Journal of Survey Statistics and Methodology, 1*(2), 90–143. https://doi.org/10.1093/jssam/smt008

Banaji, S., Bhat, R., Agarwal, A., Passanha, N., & Pravin, M. S. (2019). *WhatsApp vigilantes: An exploration of citizen reception and circulation of WhatsApp misinformation linked to mob violence in India*. Department of Media and Communications, London School of Economics and Political Science.

Bethlehem, J. (2010). Selection bias in web surveys. *International Statistical Review, 78*(2), 161–188. https://doi.org/10.1111/j.1751-5823.2010.00112.x

Blair, G., & Imai, K. (2012). Statistical analysis of list experiments. *Political Analysis, 20*(1), 47–77. https://doi.org/10.1093/pan/mpr048

Décieux, J. P., Mergener, A., Neufang, K. M., & Sischka, P. (2015). Implementation of the forced answering option within online surveys: Do higher item response rates come at the expense of participation and answer quality? *Psihologija, 48*(4), 311–326. https://doi.org/10.2298/PSI1504311D

Eysenbach, G. (2004). Improving the quality of web surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). *Journal of Medical Internet Research, 6*(3), Article e34. https://doi.org/10.2196/jmir.6.3.e34

Galesic, M. (2006). Dropouts on the web: Effects of interest and burden experienced during an online survey. *Journal of Official Statistics, 22*(2), 313–328.

Galesic, M., & Bosnjak, M. (2009). Effects of questionnaire length on participation and indicators of response quality in a web survey. *Public Opinion Quarterly, 73*(2), 349–360. https://doi.org/10.1093/poq/nfp031

Gelman, A., & Stern, H. (2006). The difference between "significant" and "not significant" is not itself statistically significant. *The American Statistician, 60*(4), 328–331. https://doi.org/10.1198/000313006X152649

Higher Education Commission. (n.d.). *HEC recognized campuses*. Higher Education Commission, Pakistan. https://www.hec.gov.pk/english/universities/pages/recognised.aspx

Kemp, S. (2025, March 3). *Digital 2025: Pakistan*. DataReportal. https://datareportal.com/reports/digital-2025-pakistan

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology, 5*(3), 213–236. https://doi.org/10.1002/acp.2350050305

Leiner, D. J. (2019). Too fast, too straight, too weird: Non-reactive indicators for meaningless data in internet surveys. *Survey Research Methods, 13*(3), 229–248. https://doi.org/10.18148/srm/2019.v13i3.7403

L'Engle, K., Sefa, E., Adimazoya, E. A., Yartey, E., Lenzi, R., Tarpo, C., Heward-Mills, N. L., Lew, K., & Ampeh, Y. (2018). Survey research with a random digit dial national mobile phone sample in Ghana: Methods and sample quality. *PLoS ONE, 13*(1), Article e0190902. https://doi.org/10.1371/journal.pone.0190902

Maertens, R., Götz, F. M., Golino, H. F., Roozenbeek, J., Schneider, C. R., Kyrychenko, Y., Kerr, J. R., Stieger, S., McClanahan, W. P., Drabot, K., He, J., & van der Linden, S. (2024). The Misinformation Susceptibility Test (MIST): A psychometrically validated measure of news veracity discernment. *Behavior Research Methods, 56*, 1863–1899. https://doi.org/10.3758/s13428-023-02124-2

Reynolds, W. M. (1982). Development of reliable and valid short forms of the Marlowe-Crowne Social Desirability Scale. *Journal of Clinical Psychology, 38*(1), 119–125. https://doi.org/10.1002/1097-4679(198201)38:1<119::AID-JCLP2270380118>3.0.CO;2-I

Robinson, D., & Tannenberg, M. (2019). Self-censorship of regime support in authoritarian states: Evidence from list experiments in China. *Research & Politics, 6*(3), 1–9. https://doi.org/10.1177/2053168019856449

Thabane, L., Ma, J., Chu, R., Cheng, J., Ismaila, A., Rios, L. P., Robson, R., Thabane, M., Giangregorio, L., & Goldsmith, C. H. (2010). A tutorial on pilot studies: The what, why and how. *BMC Medical Research Methodology, 10*, Article 1. https://doi.org/10.1186/1471-2288-10-1

Tourangeau, R., & Yan, T. (2007). Sensitive questions in surveys. *Psychological Bulletin, 133*(5), 859–883. https://doi.org/10.1037/0033-2909.133.5.859

Weijters, B., Geuens, M., & Schillewaert, N. (2010). The individual consistency of acquiescence and extreme response style in self-report questionnaires. *Applied Psychological Measurement, 34*(2), 105–121. https://doi.org/10.1177/0146621609338593

Yan, T., Curtin, R., & Jans, M. (2010). Trends in income nonresponse over two decades. *Journal of Official Statistics, 26*(1), 145–164.

Zhang, C., & Conrad, F. G. (2014). Speeding in web surveys: The tendency to answer very fast and its association with straightlining. *Survey Research Methods, 8*(2), 127–135. https://doi.org/10.18148/srm/2014.v8i2.5453