Designing Integrated Multi-Role Survey Instruments: A Converge–Diverge–Reconverge Framework for Complex Organisations

Jennifer R. Ayres[1a], Rosmina Bustami[2a], Jethro H. Adam[a]

[a]UNIMAS Water Centre (UWC), Faculty of Engineering, Universiti Malaysia Sarawak, Kota Samarahan, 94300, Malaysia.

ORCID IDs: [1] 0000-0002-4538-6512 | [2] 0000-0002-8438-8932

Abstract

Multi-role organisational surveys must trade role-specific depth, which cannot be compared across roles, against role-agnostic breadth, which is comparable but shallow. This paper documents a converge–diverge–reconverge architecture addressing this through deterministic role-based routing, and a protocol for verifying that a branched instrument behaved as designed. It was deployed in an engineering workplace survey (N = 335, 22 countries). Respondents completed a shared entry block, were routed by role to one of five role-specific modules or a single-item fallback and reconverged on 35 common items. Routing operated without error across seven pathways, and response quality on the universal block did not degrade with pathway length. Construct scores varied with the amount of preceding role-specific content, but because exposure and role are confounded, sequencing effects cannot be fully separated from role differences without randomising block order within role. The paper contributes documented design decisions and a post-deployment verification protocol for branched instruments.

Keywords: engineering survey, systemic risk, survey methodology, branching logic, Converge-Diverge-Reconverge Design, multi-role workplace research, survey approaches, integrated multi-role survey

1. Introduction

1.1 A persistent tension in multi-role organisational research

Researchers studying organisations in which people perform fundamentally different functions face a persistent measurement problem. Universal survey instruments — those that treat the entire workforce as a homogeneous population — enable cross-group comparison but obscure how shared pressures are experienced and acted upon differently across roles. Role-specific instruments capture functional depth but fragment the dataset, preventing system-level analysis. Choosing between comparability and specificity is, for many researchers, not a design preference but a forced compromise.

This paper documents an architecture developed to address that tension, and a protocol for verifying that a branched instrument behaved as designed. The architecture is demonstrated through a multi-country engineering workplace survey, a context characterised by high role diversity and interdependent, safety-critical functions, and is transferable to other multi-role settings where role diversity creates measurement difficulty.

This paper's contribution is explicitly methodological. It documents how to design the instrument, not what the instrument found.

1.2 Why existing approaches fall short

Survey-based research in multi-role organisations has tended to resolve the comparability–specificity tension by defaulting to one end or the other. Role-agnostic instruments are common in large-scale workforce studies: they enable broad comparison and are straightforward to administer, but they implicitly assume that all respondents interpret and experience survey constructs in the same way. In contexts where responsibility, authority, and proximity to risk vary substantially by function, this assumption is difficult to defend. A question about schedule pressure means something different to an engineer verifying a safety-critical system than to a project controls analyst preparing a milestone forecast — yet a role-agnostic instrument treats these responses as equivalent.

Role-specific instruments address this by targeting individual functions, but they are typically designed in isolation. Each instrument captures what happens within a function; none captures how pressures originate, propagate, and manifest across functions. The result is a fragmented picture of organisational dynamics that cannot support system-level analysis.

Adaptive survey designs and branching logic offer a technical route out of this impasse and are well established in health and clinical research. Their application within workplace and organisational research, however, remains limited and inconsistently reported. Where branching is used, it is typically treated as an administrative convenience — a way to skip irrelevant questions — rather than as a deliberate methodological strategy for preserving analytical integrity across a fragmented population. What is rarely documented is how such instruments should be designed when cross-role comparison is the analytical objective: which content to branch, how to sequence it, and how to verify afterwards that the instrument behaved as intended.

1.3 The practical guidance gap

The absence of a documented framework is not only a gap in the research literature. It is a practical obstacle for graduate students, early-career researchers, and applied evaluators who need to study multi-role organisations but lack access to methodological models for doing so.

Existing survey design texts provide strong foundations for general instrument construction (Bradburn et al., 2004). They offer limited guidance, however, on how to manage role diversity at the design stage — how to decide what belongs in a universal module versus a role-specific branch, how sequencing decisions bear on the comparability the design is meant to deliver, and how to verify routing integrity and structured missingness once the instrument has been fielded.

1.4 Purpose and contribution of this paper

The paper addresses three methodological questions.

RQ1. Can deterministic role-based branching deliver role-specific content without degrading response quality on the common comparative measures?

RQ2. Can routing integrity and structured missingness be verified from deployed survey data rather than from pre-deployment testing?

RQ3. Does differential exposure to role-specific content appear to influence subsequent responses to the common measures?

These are not formal hypotheses. RQ1 and RQ2 are answerable from the deployed sample; RQ3 is addressed exploratorily and, for reasons set out in Section 4.5, cannot be answered definitively by this design.

The paper makes two contributions.

First, it provides a replicable design framework for researchers seeking to study systemic phenomena across diverse functional roles. The architecture is demonstrated in engineering but is explicitly framed for transfer to other multi-role organisational contexts. Replication materials permitting independent adoption are listed under Data and supplementary materials.

Second, it documents the practical design decisions — including sequencing, branching logic and burden constraints — together with a post-deployment verification protocol for branched instruments, and shows that deterministic routing confounds the quantity of preceding content with role. These decisions are rarely made explicit in published work; this paper surfaces them as methodological choices with identifiable rationale and trade-offs.

2. Background

2.1 The importance of role in engineering workplace research

Role is a critical determinant of how organisational pressures are perceived, negotiated, and acted upon in engineering workplaces (Argyris, 1990; Argyris and Schön, 1978). Differences in decision authority, temporal exposure to risk, and accountability structures shape not only what pressures are experienced, but also how those pressures translate into practice (Hollnagel, 2018; Weick and Sutcliffe, 2007).

For example, schedule pressure may lead to accelerated verification and acceptance within commissioning roles, deferred maintenance decisions in operations, constrained risk tolerances in health and safety functions, or optimistic forecasting within planning and project controls (Woods, 2006). Each response reflects a localised adaptation to shared

organisational demands, yet traditional survey approaches rarely capture these differentiated pathways within a single instrument.

2.2 Comparison with existing approaches

Table 1 compares the main approaches to multi-role measurement by design feature: generic single-path instruments correspond to the role-agnostic surveys, and parallel role surveys to the role-specific instruments, described in Section 1.2. Entries record whether a feature is characteristic of an approach as described in the methodological literature, not how well each approach performs.

Table 1. Design-feature comparison of multi-role survey approaches

Design feature	Generic single-path	Parallel role surveys	Split questionnaire	Present architecture
Content tailored to respondent characteristic	No	Yes	No	Yes
Module allocation randomised	n/a	No	Yes	No
Module allocation deterministic on observed variable	n/a	Yes	No	Yes
Missingness by design	No	n/a	Yes	Yes
Missingness requires imputation	n/a	n/a	Yes (typically)	No
Universal block completed by all respondents	Yes (all items)	No	variable	Yes
Cross-group comparison supported within one dataset	Yes	No	Yes	Yes
Differential content precedes universal block	No	n/a	variable	Yes

Notes: Yes = characteristic of the approach; No = not characteristic; variable = present in some implementations and absent in others; n/a = not applicable. Entries are descriptive, not evaluative rankings.

2.3 Modular and split questionnaire designs

Dividing a questionnaire so that no respondent answers every item is an established response to instrument length. The formal treatment originates with Raghunathan and Grizzle (1995), whose split questionnaire design partitions an instrument into components, administers varying subsets to different respondents, and recovers full-sample estimates through multiple imputation of the unadministered parts. It has since been extended to large commercial surveys (Adıgüzel and Wedel, 2008), to official statistics (Chipperfield and Steel, 2009) and to panel designs (Imbriano and Raghunathan, 2020). Modularisation can raise start rates in web administration, but it has not been shown to improve overall completion: breakoff rises with the number of parts, and completion across all parts is no higher than for a single questionnaire (Andreadis and Kartsounidou, 2020; Toepoel and Lugtig, 2022). How modules are composed also matters. Placing highly correlated items in the same module degrades the imputation the design depends on, and spreading them across modules is recommended (Axenfeld et al., 2022).

A recurring variant places a core block, administered to every respondent, alongside the rotating modules. This preserves a common basis for comparison and anchors the imputation model and is the structural feature most closely resembling the architecture reported here.

The resemblance is partial and the difference consequential. In split questionnaire designs, module allocation is randomised, and it is that randomisation which makes the resulting missingness ignorable and licenses imputation, because the missingness is then missing completely at random (Graham et al., 2006; Rhemtulla and Little, 2012). The design reported here allocates modules deterministically, on a characteristic the respondent reports within the instrument. The resulting missingness is not random and not ignorable in the technical sense. Nor is it imputed. It is fully explained by an observed variable, and it confines role-specific items to within-branch description.

The two designs therefore solve different problems. Split questionnaire designs shorten an instrument whose content applies to everyone but is too long to administer in full. Deterministic branching delivers content that does not apply to everyone, to the subset for whom it is meaningful. The modular literature is largely silent on the second case, because random allocation removes the questions deterministic allocation raises.

2.4 Comparability as a measurement condition

Multi-role designs rest on an assumption rarely stated explicitly: that constructs administered to different groups in identical wording are measured equivalently across those groups. Identical items are necessary but not sufficient. Where a construct is measured differently across groups, observed differences in group means may reflect measurement non-equivalence rather than genuine differences in the underlying construct (Meredith, 1993).

The measurement invariance literature sets out the conditions under which cross-group comparison is defensible, and they are tested in sequence. Configural invariance establishes that the same factor structure holds in each group. Metric invariance establishes that items relate to the latent construct with equal strength. Scalar invariance establishes equal item intercepts, the condition required before latent means may be compared (Vandenberg and Lance, 2000; Putnick and Bornstein, 2016). Invariance is nonetheless rarely tested in organisational research, despite being a prerequisite for the group comparisons that research reports routinely.

This condition bears directly on branched instruments in a way the modular survey literature has not addressed. A design that routes respondents to different content before administering a shared construct block introduces a group-specific difference in the conditions under which that block is completed. If those conditions affect responding, the assumption of equivalent measurement is threatened not by the wording of the items, which is identical, but by the sequence in which respondents encounter them. Invariance testing addresses whether a construct is measured equivalently; it does not by itself identify sequencing as the mechanism where equivalence fails. The diagnostics reported in Section 4.5 address that gap by examining whether responses vary with the quantity of preceding content.

2.5 Question order and the response process

That preceding content can shape subsequent answers is well established. Tourangeau and Rasinski (1988) model the survey response as a four-stage process of interpretation, retrieval, judgement and response. Context effects arise because retrieval is neither exhaustive nor stable. Respondents draw on a sample of relevant beliefs, and that sample over-represents whatever is most accessible when the question is asked (Tourangeau et al., 1989; Tourangeau et al., 2000). Two consequences follow for branched designs. First, context effects are strongest where preceding items and the target item share subject matter, because topical overlap determines which beliefs are made accessible, and are weaker where preceding

content addresses different material. Second, prior items can supply an interpretive frame, so that respondents answer a general item in terms established by the specific items preceding it (Schuman and Presser, 1981). Both mechanisms predict that any sequencing effect in a branched instrument should be construct-specific rather than uniform. Such effects need not be assimilative. Where specific items precede a general one and the two are presented as belonging together, respondents may exclude information they have already reported from the general judgement, producing contrast rather than assimilation (Schwarz et al., 1991; Schwarz and Bless, 1992). The direction of any sequencing effect is therefore not fixed in advance.

A second body of work addresses what respondents do when questions become burdensome. Krosnick (1991) proposes that respondents facing high cognitive demand may satisfice, providing an adequate rather than an optimal answer through incomplete retrieval or, in stronger forms, through none at all. Satisficing manifests in identifiable patterns, including failure to differentiate among items in a battery, endorsement of the first acceptable option, and non-substantive responding. Because it increases with fatigue, it is a specific concern for content positioned late in a long instrument (Krosnick and Presser, 2010; Smyth, 2016). In the present architecture, that is where the comparative constructs sit. In web surveys, a longer stated length reduces participation, and items positioned later receive faster, shorter and more uniform answers (Galesic and Bosnjak, 2009).

Topical overlap can be specified from the instrument itself. Of the four multi-item constructs in the universal block, workplace pressure shares subject matter most directly with the role-specific modules, each of which asks mainly about experienced pressure to depart from intended practice. The other three constructs overlap less, but not to zero. The modules also ask about peer or team pressure (Q23, Q45), recognition by management (Q35), team competence (Q36) and senior stakeholders bypassing safety processes (Q63). The boundaries within the universal block are not clean either: the pressure scale includes peer judgement (Q68) and harsh treatment of mistakes (Q69), and the psychological outcomes scale includes the effect of peer pressure on wellbeing (Q76). On the accessibility account, any sequencing effect should therefore be strongest for workplace pressure and weaker, though not absent, for the other three. This expectation follows from the response-process literature; the analysis that examines it in Section 4.5 is exploratory and was conducted after data collection.

The two accounts are separable in the data. Assimilation predicts movement in construct means with exposure, concentrated by topical overlap; satisficing predicts reduced differentiation and increased non-substantive responding, rising with cumulative burden and largely indifferent to topic. Response variation, non-differentiation and construct means are therefore examined jointly in Section 4.5.

2.6 Positioning the present study

Three bodies of work bear on the design reported here without jointly addressing it: the modular survey literature, which establishes that partitioning an instrument is viable and specifies how to recover estimates when allocation is random; the measurement invariance literature, which specifies the conditions under which cross-group comparison is defensible; and the question order literature, which establishes that preceding content shapes responses and predicts where such effects should concentrate. We identified no guidance integrating the three features found here: allocation is deterministic, comparison across allocation groups is the analytical objective, and the allocated content itself precedes the comparative measures. That combination is the subject of this paper.

3. Instrument Design Methodology

3.1 The framework: five design decisions

The framework comprises an architecture — common entry, role-conditional divergence, common reconvergence — and five decisions required to implement it. Three levels should be distinguished throughout: the framework itself; the case implementation, the engineering instrument deployed here (N = 335), which demonstrates the framework but is not an ideal instance of it; and the exploratory diagnostic in Section 4.5. Claims made at one level do not transfer automatically to another. Each decision below carries consequences that the verification protocol in Section 5.3 is designed to surface.

The routing variable. Routing must depend on a characteristic respondents can report accurately about themselves, that is stable within the reporting period, and that genuinely predicts which specialised content is answerable. A routing variable respondents interpret inconsistently will produce apparently clean routing while distributing content incorrectly, and no post-deployment check will detect this, because the audit verifies that routing followed the recorded answer, not that the answer was correct. Where role definitions are contested, an explicit definition should accompany the routing item. Routing can also shape what respondents report. Where respondents learn that an answer leads to additional questions, some choose answers that shorten the instrument (Eckman et al., 2014; see also Kreuter et al., 2011). A routing variable that sends one group to a substantially longer module creates that incentive, and the risk should be weighed when module lengths differ.

The number of categories. Each additional branch reduces the sample available for within-branch analysis while adding design and testing burden. Researchers should estimate expected cell sizes from the anticipated sampling frame before finalising categories and set minimum branch sizes in advance from the intended analysis and the effect sizes it must detect. No universal threshold applies. Supplementary Material 2 sets out conditions this study does not itself meet — thirty respondents per role and demonstrated completion time parity — offered as the standard a prospective implementation should aim at rather than a description of this deployment.

Module depth. Depth should be allocated according to the analytical weight each branch is expected to carry, not according to the researcher's familiarity with the domain; where it must vary, the variation should be stated and justified rather than left for readers to infer from module tables.

Composition of the universal block. This is the only content on which cross-role comparison rests, and it should be selected for comparability rather than coverage: constructs must be answerable by every respondent irrespective of routing, without reliance on role-specific vocabulary or assumed familiarity with processes.

Sequencing. Terminal reconvergence grounds general judgements in concrete recall but exposes respondents to unequal preceding content; initial convergence removes that differential but asks for generalised judgements before any specific material has been brought to mind. A third option, unavailable on the platform used here but supported by several commercial survey systems, is to randomise the order of the role-specific and universal blocks within each role group, which preserves grounding for half the sample while producing a within-role comparison that separates sequencing effects from role characteristics directly. A decision rule follows: where a universal construct overlaps substantially in subject matter with the branch content, administer it before branching or

randomise block order within role; where contextual grounding is the priority and overlap is limited, terminal reconvergence is justified.

3.2 Conceptual framing

The instrument is grounded in a systemic view of workplace pressure within engineering organisations (Hollnagel, 2018; Weick and Sutcliffe, 2007). Rather than treating pressure as an individual-level attribute, the design conceptualises it as an organisational condition that manifests differently across roles depending on proximity to risk, authority, and operational consequence (Rasmussen, 1997). The substantive relationships among the constructs are reported elsewhere and are not examined here.

3.3 Instrument structure

3.3.1 Overall sequence and rationale

The instrument follows a converge–diverge–reconverge sequence: a shared entry block completed by every respondent, role-conditional routing, and a universal block completed by all respondents after routing. Here, universal means administered to all respondents; it does not imply that the constructs are interpreted equivalently across roles, which is a measurement question addressed in Section 5.5.

Convergence (Q1-Q19). All respondents completed four eligibility and consent items (Q1-Q4) followed by fifteen demographic and work-context items (Q5-Q19), covering age, gender, education, geographic location, discipline, industry sector, work setting, employment type, career choice, intention to leave the profession, job satisfaction, work-life balance, working hours, team size and intention to leave the current organisation. This block establishes the shared sampling and contextual frame against which all subsequent responses are interpreted; it does not contain the study's comparative constructs.

Routing (Q20) and divergence (Q21-Q65). The routing item, Q20, was answered by all respondents, recorded primary role and served as the sole routing variable. Respondents were then directed to at most one role-specific module. Branching was adopted at the design stage rather than retrofitted. Several phenomena central to the study - pressure to begin testing prematurely, downgrading of defect severity, bypassing of permit systems, adjustment of schedule forecasts - are meaningful only to practitioners who encounter them directly. Presenting such items to respondents with no working exposure would have produced uninformative responses while consuming limited attention, increasing both fatigue and attrition. Branching was therefore a relevance and burden decision: each respondent received only the specialised content their role made answerable.

Reconvergence (Q66-Q100). All respondents, irrespective of routing, completed the same thirty-five universal items measuring workplace pressure (Q66-Q70), delay and cost attribution (Q71-Q72), psychological outcomes (Q73-Q76), peer influence and team culture (Q77-Q80), peer pressure impacts on quality (Q81-Q86) and on project delivery (Q87-Q96; a section the form described as optional, in which every item offered a Not applicable option, although all items were set as required), and organisational culture (Q97-Q100; headed "Organisational Context" on the form). A single optional open-text item (Q101) closed the instrument. Once the universal block had been added (Section 5.5), placing it after the role-specific content was a deliberate choice made on substantive grounds: workplace pressure, peer influence and organisational climate are diffuse constructs that respondents interpret against their own working experience, and asking respondents first about concrete, situated

events was judged more likely to elicit grounded rather than abstract judgements. The sequence carries a known cost. Respondents arrive at the universal block having been exposed to differing amounts of role-specific content, which may influence subsequent responses; this trade-off is examined in Section 4.5.

Cross-role comparative analysis draws exclusively on Q5-Q19 and Q66-Q100; role-specific items are not comparable across roles by construction and are used only for within-role description.

3.3.2 Role-to-module mapping

Sixteen role options were available at Q20, together with an "Other" free-text field; Appendix A sets out the full instrument structure. The sixteen listed options were chosen to cover the great majority of occupations expected in the sampling frame. The free-text field was retained deliberately rather than as an afterthought: a respondent whose role is absent from a fixed list may reasonably conclude the survey is not for them and abandon it, and the intention was that anyone working in the sector could see themselves represented and complete the instrument. Respondents using that field receive no role-specific module and pass directly to the universal block, which is a cost the design accepts in exchange for not excluding them. Four respondents used the field. Designers should decide in advance whether unlisted roles are excluded, routed to a catch-all module written for none of them, or admitted without role-specific content, and should report the number affected. Table 2 sets out the mapping to modules.

Table 2. Role-to-module mapping (N = 335).

Module	Items	Role options routed	n
Commissioning and start-up (Q21-Q40)	20	Commissioning / start-up	81
Operations and maintenance (Q41-Q47)	7	Operations, plant operations and production; maintenance and reliability engineering	33
Planning and scheduling (Q48-Q52)	5	Planning, scheduling and project controls	22
Commercial and contracts (Q53-Q57)	5	Commercial, contracts and procurement	14
HSE/WHS (Q58-Q64)	7	Health, safety and environment; quality assurance and quality control	39
Unbranched pathway (Q65)	1	Design engineering; project management; construction; trades and technicians; leadership and management; academia and research; graduate and early career; technical specialist; controls, automation and instrumentation	142
Unrouted free-text entries	0	None. "Other" entries matching no listed option went directly to Q66	4

Two mappings warrant note. Maintenance and reliability respondents were routed to the operations module, whose content addresses deferred maintenance, operating outside limits and equipment defect tolerance; quality assurance and control respondents were routed to the HSE/WHS module, whose content addresses premature sign-off, rushed assessment and pushback on raised concerns. Both groupings were made on content grounds and are reported here so that role-specific results are not read as pertaining to a single occupational label.

3.3.3 Item load by pathway

Item load varies between pathways only within the role-specific block (Table 3).

Table 3. Item load by pathway, excluding the optional open-text item (Q101). The common core comprises Q1–Q20 and Q66–Q100.

Pathway	Common core	Role-specific	Total
Commissioning	55	20	75
Operations and maintenance	55	7	62
HSE/WHS and QA/QC	55	7	62
Planning and scheduling	55	5	60
Commercial and contracts	55	5	60
Unbranched pathway	55	1	56
No role-specific module	55	0	55

The commissioning pathway is therefore the longest at seventy-five items and the shortest routed pathway fifty-six. This asymmetry follows from the study's development history and is addressed in Section 5.5.

3.4 Branching logic implementation

The instrument was implemented using Google Forms, selected for its stable skip-logic functionality, device accessibility, and suitability for anonymous data collection in a global context. Figure 1 presents the full survey structure. The platform did not record completion time and could not randomise the order of sections, two capabilities whose absence limits what this deployment can establish (Sections 4.2 and 4.5). Role selection served as the sole branching variable, directing respondents to relevant role-specific modules, which precede the universal block.

All branching paths were tested before deployment to confirm that each role selection reached its assigned module and that all paths terminated in the universal block. This architecture avoided presenting respondents with role-specific questions outside their reported area of work and was intended to reduce unnecessary burden relative to a flat, role-agnostic design (Dillman et al., 2014; Krosnick, 1991). No flat-survey comparison was administered, so the size of any burden reduction is not established.

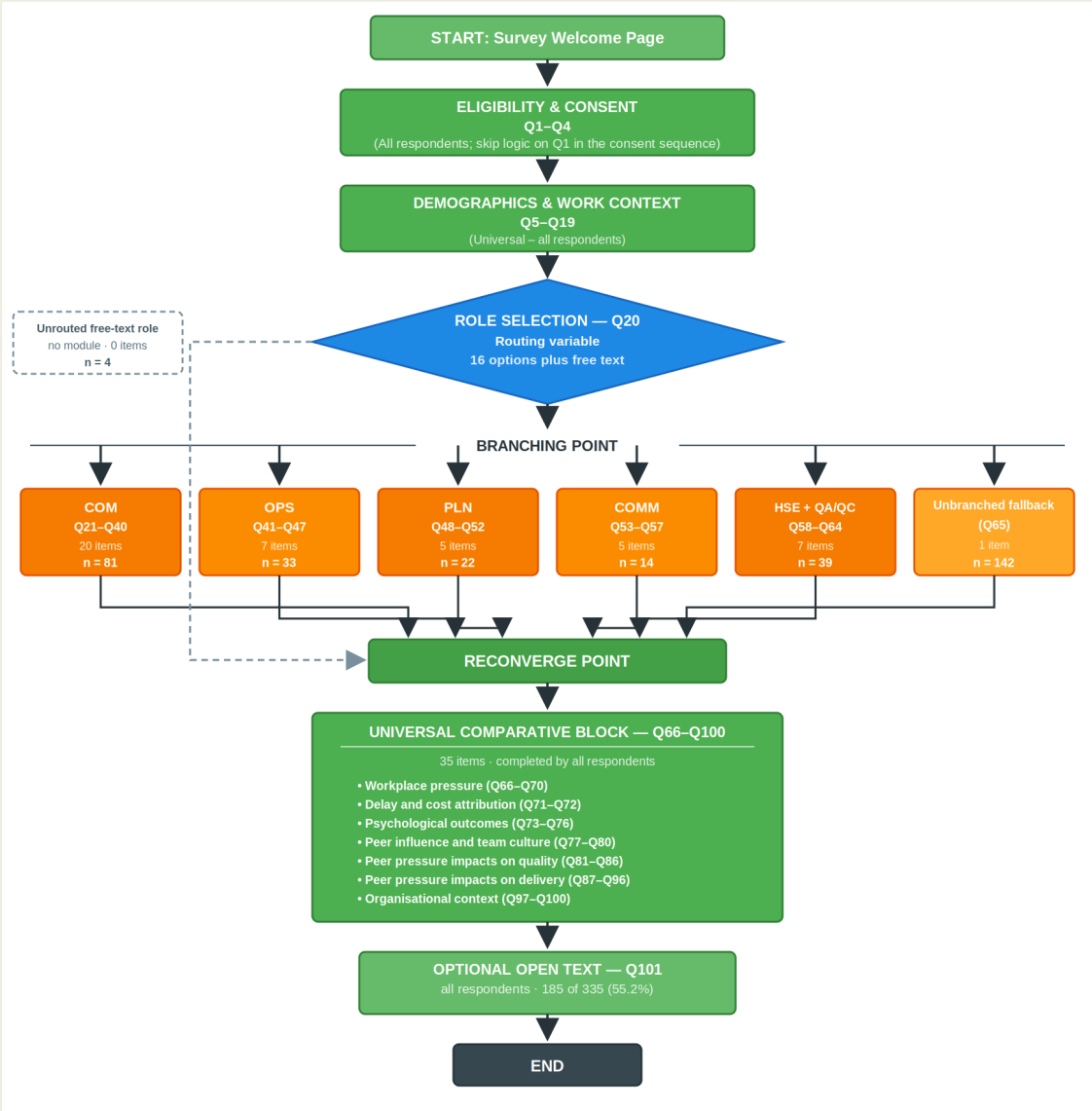


Figure 1. Converge–diverge–reconverge architecture as deployed (N = 335). Branch counts sum to 331; four respondents whose free-text role entry matched no listed option received no role-specific module and passed directly to Q66.

Table 4 summarises the focus areas and representative themes of each role-specific module.

Table 4. Role-specific module content summary

Role-specific module	Primary focus areas	No. of items	Example themes
Commissioning & start-up	Schedule-driven verification pressure, defect handling, readiness risk, permit and documentation completeness	20	Premature testing, defect severity downgrading, out-of-sequence work, risk acceptance
Operations & maintenance	Asset reliability under pressure, maintenance deferral, operational risk trade-offs	7	Running with known defects, operating outside limits, fast-tracked work
Planning & scheduling / project controls	Forecast realism, reporting pressure, milestone dominance	5	Schedule forecast adjustment, delay suppression, incomplete inputs
Commercial, contracts & procurement	Contractual pressure on technical decisions, milestone incentives	5	Incomplete documentation acceptance, variation suppression
Health, Safety & Environment (HSE/WHS)	Safety governance under delivery pressure, risk escalation	7	Premature sign-off, superficial risk assessment, safety pushback
Unbranched fallback (Q65)	No role-specific module. Q65 records supervisory scope for nine role categories: design engineering; project management; construction; trades and technicians; leadership and management; academia and research; graduate and early career; technical specialist; controls, automation and instrumentation	1	Number of people directly supervised or led (n = 142)
Unrouted free-text entries	—	0	Role descriptions not matching a listed option; proceed directly to Q66 (n = 4)

Notes: Role-specific modules are delivered only to respondents whose Q20 selection routes them to that module. Module lengths vary by role; see Table 3 for item load by pathway.

3.5 Question development and scaling

Survey items were developed through an iterative process that combined a review of existing workplace and safety literature with practitioner-informed insights drawn from the research team's professional experience across engineering, commissioning, operations, and project delivery contexts.

Where possible, items were framed to reflect observable practices and experiences rather than abstract attitudes, reducing ambiguity and supporting consistent interpretation across roles. Multi-item scales were used for latent constructs and single-item measures for descriptive or concrete variables, with frequency scales for behavioural and experiential items, where the object of measurement is occurrence, and agreement scales for attitudinal items, where it is a stance toward a condition. Formats were not standardised across modules, a choice intended to reduce mechanical response patterns of the kind that uniform anchoring across a long instrument may encourage. This is a design rationale, not an established mechanism: no counterfactual was administered, and varying anchors may equally increase the effort of reorienting to each scale.

Role-specific items addressing workplace pressures and practices were framed to capture experiences across respondents' entire engineering careers rather than their current role or employer alone, with section headers framing the role-specific modules as covering the respondent's "WHOLE career past & present", the peer-impact items (Q81–Q96) as covering "CURRENT & PAST" roles, and the four multi-item scales as referring to the "CURRENT role". The whole-career framing enhanced anonymity by preventing attribution to specific organisations or projects, and captured respondents who had moved between functional roles.

The broader recall window increases experiential coverage but introduces recall and period effects, which limit interpretation of any prevalence estimate derived from these items.

3.6 Ethical considerations

Participation in the survey was voluntary and limited to individuals aged 18 or older who were currently working, or had worked in the past five years, in engineering, technical, industrial, manufacturing, construction or infrastructure environments, as stated on the welcome page; the eligibility item itself (Q1) asked only about current work, and offered a separate option for engineering-related academic or research roles. Eligible roles included direct engineering and technical positions, as well as operational, project support, commercial, supply chain, and administrative functions within these organisational contexts. The study received approval from the Human Research Ethics Committee (Non-Medical), Universiti Malaysia Sarawak (HREC(NM)/2023 (3)/45).

Anonymity was protected by instrument design rather than by post-collection de-identification. No names, email addresses, organisational identifiers or IP addresses were collected. Login requirements and response tracking were disabled on the survey platform, and location was requested only as country or region rather than as site or city. Responses were exported from the platform to secure storage accessible only to the research team.

Informed consent was obtained through a structured multi-step process. The survey welcome page served as the Participant Information Sheet, providing study details including purpose, eligibility, data handling, anonymity protections and researcher contact information. Respondents were then asked to confirm that they had read the information sheet (Q2), confirmed age eligibility (Q3), and recorded voluntary consent (Q4), a required item with the single option “I agree to participate”. Four respondents answered “No” at Q2, an item whose text stated that a “No” answer would return the respondent to the welcome page, although no such routing was configured; all four then gave consent at Q4. Within the consent sequence, eligibility screening at Q1 was the only item carrying skip logic; the consent items themselves were not conditionally routed. Supplementary Material 1 documents the consent and eligibility process in full.

3.7 Sample, recruitment and analysis

Data were collected from an international cross-sectional sample of engineering professionals. The survey was open from January to March 2026. Respondents were recruited through professional social media, engineering online communities, personal and professional networks, university alumni networks and engineering association outreach. A total of 336 responses were received; one was excluded at eligibility screening, giving an analytical sample of N = 335 across 22 countries and six continents. Recruitment through self-selecting online channels may over-represent professionals engaged with workplace conditions, and observed proportions describe the responding sample rather than population prevalence.

All analyses were conducted in JASP. Universal-block items are ordinal five-point scales and exposure groups are markedly unequal in size (n = 36 to n = 142), so non-parametric tests were used throughout: Kruskal-Wallis for group comparisons, with eta-squared based on the H statistic (η^2H) as the accompanying effect size (Tomczak and Tomczak, 2014), and Spearman’s rho for monotonic association across ordered exposure levels. Rank-biserial correlation, the effect size for two-group comparisons, does not apply to the four-group exposure tests reported here. Because η^2H adjusts H for the number of groups compared,

epsilon-squared ($\varepsilon^2 = H/(n - 1)$) is also reported where that adjustment affects interpretation (Section 4.5.4). Cronbach's alpha is reported descriptively by exposure group. Where four construct-level tests are compared, a Bonferroni correction is applied and stated.

4. Diagnostic evidence

4.1 Content validity

Content validity was supported through structured review and testing by practising engineers, with at least one individual representing each role-specific module working through the full survey before public deployment. Review participants were drawn from commissioning, operations, maintenance, planning, project delivery, commercial, and health and safety roles, ensuring coverage across all branched pathways within the instrument.

Reviewers were asked to assess the relevance of items to their role, the clarity of wording, the completeness of role-specific content, the appropriateness of response options, and the overall survey flow.

Feedback from this practitioner review led to targeted refinements in question wording, ordering, and response options to improve clarity and reduce ambiguity. Several items were rephrased to reflect better operational language used in practice, and explanatory text was added where reviewers identified potential for misinterpretation. No items were removed or added following this process; instead, refinements focused on improving alignment between constructs and real-world engineering experience.

This review supports content validity in the conventional sense — expert judgement that items are relevant, intelligible and adequately cover the domain — and no stronger claim is made for it.

Any responses submitted during this review were deleted with the other testing data and were not analysed. Because the instrument was fully anonymous, it cannot be determined whether any reviewer subsequently completed the deployed survey.

Practitioner review covered the branched modules. The nine role categories routed to the unbranched pathway receive no role-specific content, and content validity was therefore not separately assessed for them; the universal block was reviewed by practitioners across the branched roles and is common to all pathways.

4.2 Instrument testing prior to deployment

Completion burden was treated as a design constraint from the outset, on the reasoning that fatigue-driven abandonment would compromise the universal block disproportionately, since it sits at the end of the instrument. Pathway lengths were checked through self-timed completion during instrument construction, although role-specific content ranged from one to twenty items for reasons of development history (Section 5.5). The welcome page described the survey as taking 10 to 12 minutes; no timing data were recorded against which to check this.

A pilot deployment with 52 testers was then conducted to verify instrument behaviour prior to formal data collection. Its objectives were to confirm that branching routed respondents correctly and that no structural modification was required. Routing operated as designed and no modifications were made. All pilot and testing responses were permanently deleted before

data collection began; none are included in the 336 responses received. As with the practitioner review, it cannot be determined whether any pilot participant subsequently completed the deployed survey.

4.3 Routing integrity in deployment

Routing was verified against the full deployed sample rather than inferred from pilot behaviour. Each role-specific block was cross-tabulated against the routing variable (Q20) for all 335 respondents.

No cross-contamination was observed. Every role-specific block was answered only by respondents whose Q20 selection routed them to it, and by no others (Table 5).

Table 5. Routing audit against the full deployed sample (N = 335).

Block	Respondents	Composition	Contamination
Q21-Q40 Commissioning	81	Commissioning (81)	0
Q41-Q47 Operations	33	Operations (24); maintenance and reliability (9)	0
Q48-Q52 Planning	22	Planning and project controls (22)	0
Q53-Q57 Commercial	14	Commercial and procurement (14)	0
Q58-Q64 HSE/WHS	39	HSE (19); quality assurance and control (20)	0
Q65 Unbranched pathway	142	Nine unbranched role categories	0
No module	4	Free-text role entries	n/a

Because the platform executes skip logic deterministically, a clean audit is the expected result rather than a finding about the design. Its value is as a verification step: it checks routing under live conditions across the role selections actually made, and it would detect a misconfigured branch that pre-deployment testing had missed.

4.4 Structured missingness and response completion

All items in the universal block were set as requiring a response on the deployed form, and the platform retained only submitted responses. Across the seventeen Likert-scaled universal items and 335 respondents, comprising 5,695 individual responses, no item was left unanswered. That completeness is a consequence of the field configuration rather than evidence about respondent behaviour and is reported here as a property of the dataset. Apart from the optional final item, missingness in the instrument is therefore confined to the role-specific block, where it is deterministic and fully explained by the routing variable, with one exception: a single non-response on an optional commissioning item (Q39, n = 80 of 81). That structured missingness is a property of the architecture and does not depend on the field configuration.

The universal block contains 35 items across seven content areas, of which 17 form the four multi-item agreement-scaled constructs analysed in Section 4.5: workplace pressure (Q66-Q70), psychological outcomes (Q73-Q76), peer influence and team culture (Q77-Q80), and organisational culture (Q97-Q100). The remaining 18 items were excluded from the sequencing diagnostic on format grounds rather than by selection: Q71-Q72 are multi-select attribution items with no ordinal scale, and Q81-Q96 use yes-gradient, impact, count and

frequency formats that are not commensurable with the agreement scale. No item meeting the format criterion was excluded.

The optional open-text item (Q101), the final question on every pathway, drew text from 185 of 335 respondents (55.2 per cent), two of which were non-substantive ("Nil" and "Prefer not to say"). A further two entries reading "NA" and "N/A" were not counted as text, so 187 entries were non-empty in total. Because only completed submissions are stored, this figure describes respondents who reached the end and cannot speak to those who did not. A companion analysis reports the content of these responses; for present purposes the response rate is reported as evidence of willingness to answer an optional item at the end of every pathway. Open-text response is not itself a measure of engagement, since respondents may decline for reasons unrelated to burden, and the figure is offered as a descriptive indicator rather than as evidence of sustained attention.

4.5 Exploratory sequencing diagnostic

The architecture exposes respondents to differing quantities of role-specific content before they reach the universal block. Commissioning respondents completed twenty preceding items; operations, maintenance, HSE and quality respondents seven; planning and commercial respondents five; and respondents in nine unbranched roles a single item. This section reports a post hoc examination of whether universal-block responses varied with that exposure.

Preceding-item exposure is defined throughout as the number of role-specific items a respondent completed before reaching Q66. Four exposure levels arise from the routing structure: 1 item (n = 142), 5 items (n = 36), 7 items (n = 72) and 20 items (n = 81), totalling 331. The four respondents entering free-text role descriptions received no role-specific module and are excluded from all exposure-level analyses, since they have no exposure level. Figure 2 sets out which roles contribute to each level and makes the resulting confound explicit. Exposure level is a deterministic function of role and is not independent of it. The term denotes quantity of preceding content and should not be read as implying that role has been controlled. The single item at the lowest exposure level (Q65) records supervisory scope rather than pressure content, so exposure levels differ in kind as well as in quantity. The analysis that follows is a diagnostic of the deployed architecture, not an estimate of an exposure effect.

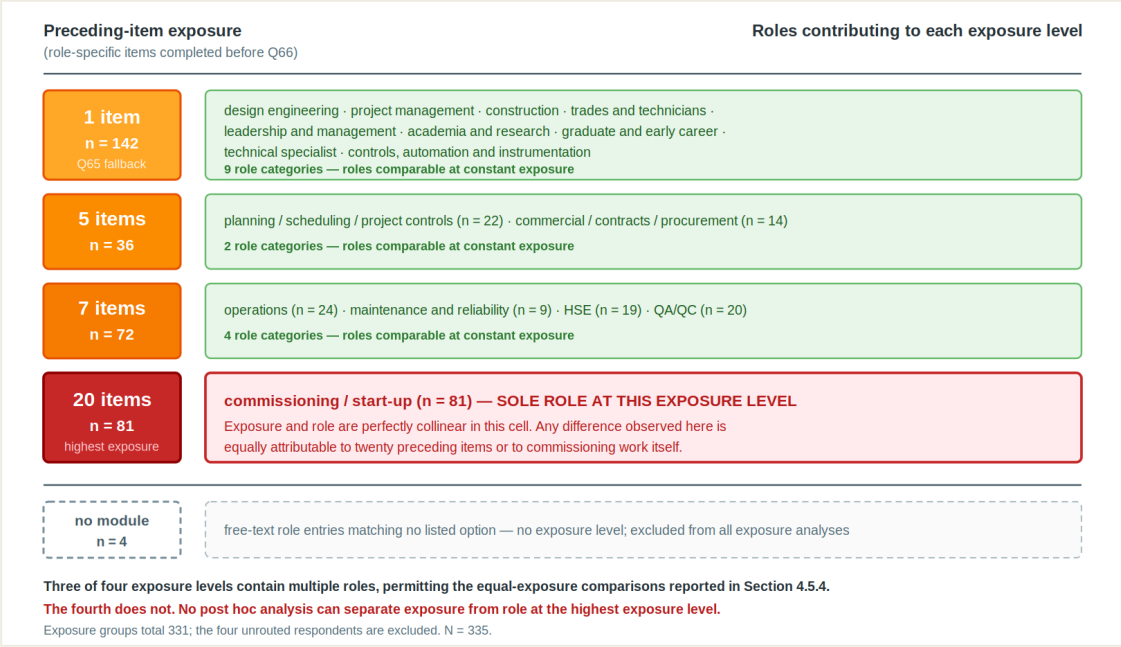


Figure 2. Preceding-item exposure by role (N = 335). Three exposure levels contain multiple roles and permit the equal-exposure comparisons in Section 4.5.4. The 20-item level contains commissioning alone, so exposure and role cannot be separated there by any post hoc analysis.

4.5.1 Response variation

Within-respondent standard deviation across the seventeen universal Likert items was lower at every higher exposure level than at the lowest: 1.02 at one preceding item (n = 142), 0.89 at five (n = 36), 0.91 at seven (n = 72) and 0.91 at twenty (n = 81). The difference was significant (Kruskal-Wallis H = 14.59, p = 0.002), with a weak negative monotonic association (Spearman rho = -0.18, p = 0.001). The pattern is a step down from the one-item group rather than a steady decline: the three higher exposure groups are effectively indistinguishable from one another.

Non-differentiation did not follow the same pattern. Straight-lining, defined as two or fewer distinct response values across all seventeen items, occurred in 5.6 per cent of respondents at one preceding item, none at five, 4.2 per cent at seven and 2.5 per cent at twenty. Response variation therefore narrowed with exposure without a corresponding rise in disengaged responding, a pattern that, on these response-quality indicators, is more consistent with more uniform judgements across items than with satisficing. Satisficing takes several forms, not all of which are detectable through non-differentiation.

4.5.2 Scale reliability

Internal consistency was acceptable throughout and did not degrade systematically with exposure (Table 6). Mild attenuation appears at the highest exposure for pressure and culture. This is not explained by the response variation reported above, which is identical at seven and twenty preceding items, and no mechanism is proposed for it here; with n = 81 in that group and no confidence intervals, the difference may not be meaningful. No scale fell below conventional acceptability.

Table 6. Cronbach's alpha by preceding item count.

Scale	k	Overall	1 (n=142)	5 (n=36)	7 (n=72)	20 (n=81)
Workplace pressure	5	0.78	0.75	0.87	0.82	0.71
Psychological outcomes	4	0.86	0.82	0.88	0.89	0.85
Peer influence	4	0.81	0.85	0.79	0.74	0.79
Organisational culture	4	0.91	0.92	0.91	0.91	0.85

Notes: The Overall column is calculated on all 335 respondents; the exposure-level columns exclude the four respondents without a role-specific module (n = 331).

4.5.3 Construct scores

Construct means varied with preceding item count, but not uniformly (Table 7). Four construct-level tests were conducted. Applying a Bonferroni correction across those four tests (alpha = 0.0125), pressure and culture remain significant (both $p < 0.001$) while psychological outcomes ($p = 0.014$) and peer influence ($p = 0.018$) do not. Peer influence additionally shows no monotonic association despite a nominally significant omnibus test. The equal-exposure comparisons reported in Section 4.5.4 were not separately corrected and are treated as descriptive. Effect sizes are small to moderate ($\eta^2H = 0.022$ to 0.094).

Table 7. Universal construct means (1-5) by preceding item count, with Kruskal-Wallis tests and Spearman correlations.

Construct	1 item (n = 142)	5 items (n = 36)	7 items (n = 72)	20 items (n = 81)	H	p	η^2H	rho
Workplace pressure	3.14	3.14	3.45	3.62	25.46	<0.001	0.069	0.263
Organisational culture	3.56	3.42	3.26	2.90	33.62	<0.001	0.094	-0.309
Psychological outcomes	3.02	3.15	3.35	3.28	10.62	0.014	0.023	0.151
Peer influence	2.96	3.32	3.25	3.01	10.06	0.018	0.022	0.022 ns

Notes: Spearman correlations between preceding item count and construct score: workplace pressure $p < 0.001$; organisational culture $p < 0.001$; psychological outcomes $p = 0.006$; peer influence $p = 0.687$.

4.5.4 Role differences at constant exposure

Because several roles share an identical preceding item count, exposure and role can be partially separated: if differences are driven by exposure, roles at equal exposure should not differ from one another.

Table 8 reports role differences within each exposure level.

Table 8. Role differences at constant exposure, compared with the between-exposure gradient.

Exposure level	Roles compared	Culture	Pressure
1 preceding item	9 categories (n = 142)	$H = 27.92$, $p < 0.001$, $\eta^2H = 0.150$	$H = 9.38$, $p = 0.311$, $\eta^2H = 0.010$
5 preceding items	2 categories (n = 36)	$H = 0.59$, $p = 0.442$, ns	$H = 0.05$, $p = 0.818$, ns
7 preceding items	4 categories (n = 72)	$H = 10.90$, $p = 0.012$, $\eta^2H = 0.116$	$H = 7.19$, $p = 0.066$, $\eta^2H = 0.062$
Between-exposure (reference)	4 exposure groups (n = 331)	$\eta^2H = 0.094$	$\eta^2H = 0.069$

Read with η^2H, the pattern appears asymmetric. For organisational culture, role variation at constant exposure ($\eta^2H = 0.116$ and 0.150) exceeds the exposure gradient (0.094), so culture differences appear better explained by role than by sequencing. For workplace pressure, role variation is small at one preceding item (0.010) and effectively zero at five, while the exposure gradient is 0.069; at seven preceding items, however, role variation in pressure reaches 0.062 ($p = 0.066$). The asymmetry depends on the effect-size measure. η^2H subtracts the number of groups minus one from H, which penalises the nine-group comparison at one preceding item most heavily. With epsilon-squared ($\varepsilon^2 = H/(n - 1)$), which carries no such adjustment, role variation in pressure at constant exposure is 0.067 at one preceding item and 0.101 at seven, against an exposure gradient of 0.077; the corresponding values for culture are 0.198 and 0.154 against 0.102. On that measure, role differences in pressure are as large as the exposure gradient. For peer influence, no exposure gradient exists to explain.

Two cautions attach to these comparisons. The effect sizes being set against one another come from tests of different designs — nine groups within n = 142, four within n = 72, four exposure groups across n = 331 — and both η^2H and ε^2 are sensitive to group count and sample size, so the comparison is offered as a qualitative pattern rather than as a numerical ranking. Confidence intervals are not available at these subgroup sizes. The comparison does not separate role from sequencing; it is not a decomposition of variance between them, and should not be read as one.

The observed pattern is therefore compatible with a sequencing effect on workplace pressure, the construct sharing most subject matter with the role-specific modules, but it does not establish one. Role differences at constant exposure are of similar size on one effect-size measure, culture moved more than pressure across exposure groups, and psychological outcomes showed a weak monotonic trend ($\rho = 0.151$) although its omnibus test did not survive correction.

The 20-item exposure group comprises a single role, commissioning. That group cannot be distinguished from a genuine characteristic of commissioning work, and no design-based remedy is available post hoc. Randomising block order among respondents sharing a role would separate the two; this was not done and is recommended for future implementations.

4.6 Scope of validation

To be explicit about scope: content validity was addressed through practitioner review; internal consistency was examined descriptively; routing integrity and structured missingness were verified against the deployed sample. Factorial and construct validation, criterion validity and measurement invariance remain unestablished and are future work; test–retest reliability could not be assessed because the survey carried no respondent identifiers. The instrument should not be described as validated on the evidence presented here.

5. Discussion

5.1 What the architecture achieved

Deterministic routing delivered role-relevant content without evidence of response-quality degradation in the deployed sample. Completeness across the universal block follows from its required-response configuration and is not evidence about respondent behaviour. Two observations are not determined by that configuration: the optional final free-text item drew text from 55.2 per cent of the sample, and non-differentiation remained low on every pathway. Structured missingness is fully explained by an observed variable recorded within the instrument and requires no imputation.

5.2 What the architecture cost

Terminal reconvergence exposes respondents to unequal quantities of preceding content. The diagnostics reported in Section 4.5 are compatible with a cost for the construct sharing most subject matter with the role-specific modules, though the design cannot separate it from role. Researchers adopting the architecture should treat sequencing effects as likely where role modules and universal constructs overlap in subject matter, and should design accordingly, either by randomising block order within role or by selecting universal constructs semantically distant from role-module content.

5.3 A verification protocol for branched instruments

The diagnostics applied here generalise to any deterministically branched instrument. Table 9 sets out the protocol.

Table 9. Post-deployment verification protocol for deterministically branched instruments.

Check	Procedure	Interpretation
Routing integrity	Cross-tabulate each role-specific block against the routing variable	Any response outside an assigned pathway indicates a logic fault
Structured missingness	Confirm missingness is fully explained by routing	Unexplained missingness is consistent with burden or comprehension problems
Response behaviour	Compare within-respondent variation and non-differentiation across exposure groups	Narrowing with rising non-differentiation is consistent with burden-related satisficing
Reliability	Compute alpha by exposure group	Systematic degradation would suggest exposure-related measurement instability
Sequencing	Compare construct means across exposure groups; use roles at equal exposure to separate exposure from role	Gradients surviving the equal-exposure control support a sequencing interpretation

5.4 Transferability and boundary conditions

The architecture suits research settings with three properties. Respondents must be divisible into a modest number of groups whose work differs substantively enough that shared items would be uninformative for some of them. There must be constructs of genuine interest that apply across all groups, since without these the design reduces to a set of parallel single-group surveys with a common demographic preamble. And the routing characteristic must be

recordable within the instrument rather than inferred from sampling frame membership, since it is the recorded answer that drives both routing and subsequent analysis.

Potential applications include healthcare, education, emergency services and logistics, in each of which workforces divide into functionally distinct groups with shared organisational conditions. Role-adapted instruments with common cores have been reported in health settings; the transferability of the architecture to the others is proposed rather than demonstrated. The architecture is less suited to settings where the population is functionally homogeneous, where specialised content is genuinely universal, or where the analytical interest lies entirely within a single group; in the last case a single-role instrument is simpler and avoids the sequencing question altogether.

Two boundary conditions follow from the design itself. Deterministic routing confines role-specific analysis to within-branch description, since no comparable data exist outside the branch. And because respondents in different branches encounter different content before the universal block, cross-role comparison carries a measurement risk that role-agnostic instruments do not. This is the cost of the depth the design provides.

5.5 Limitations

Eleven limitations bear on the interpretation of this work.

Design origin. The instrument began as a study of commissioning practice and was extended to comparison roles during development. This history is visible in the architecture: the commissioning module contains twenty items against a maximum of seven elsewhere, and commissioning is the largest single role group. The universal block was introduced to enable the comparison the extended design required. The depth asymmetry is therefore a consequence of development sequence rather than a considered allocation, and a framework designed for multi-role comparison from the outset would likely distribute module depth more evenly. Against this, all pathways shared a common core of fifty-five items, and the variation is confined to a single block.

Confounding of exposure with role. Preceding item count is determined by role. The equal-exposure comparisons in Section 4.5.4 provide partial separation but cannot resolve the highest exposure category, which contains a single role.

Post hoc analysis. The diagnostics in Section 4.5 were conducted after data collection and were not pre-specified. Two of the four constructs examined would not survive correction for multiple comparisons.

Construct structure and invariance not established. Two distinct limitations apply. The dimensionality of the universal block was not examined at the full-sample level, so the four scales are specified rather than demonstrated. Separately, group sizes ranging from seven to twenty-four in several role categories preclude multi-group confirmatory factor analysis, so measurement invariance cannot be tested even were the structure established. Comparability across roles is therefore assumed, not demonstrated, and researchers using this architecture for substantive cross-role hypotheses should establish structure at the full-sample level and configural, metric and scalar invariance across groups before comparing group means.

Constrained construct coverage. Content was deliberately restricted to limit both instrument length and the risk of framing responses. Topics with plausible bearing on the phenomena studied, including leave entitlements, remuneration and retention drivers, were excluded on

those grounds. Their omission limits the explanatory scope of any analysis drawing on this instrument.

Completion duration unrecorded. The survey platform captured no timing data, so burden parity is a design target established through self-timed construction rather than a verified field outcome.

Completion could not be observed. The universal items were configured to require a response, and the survey platform retained only completed submissions. Partial responses were not recorded, so the proportion of respondents who began the instrument but did not submit it cannot be determined. Claims about completion in this paper are restricted to the retained sample, and the effect of the architecture on abandonment remains untested.

Common-method effects. The universal constructs were collected through a single self-report instrument in one administration context, so common-method variance cannot be excluded from any association observed among them.

Sample and self-report. Participation was voluntary and recruitment self-selecting, so respondents experiencing higher workplace pressure may be over-represented. Responses are self-reported and cross-sectional, so recall effects cannot be excluded and temporal stability cannot be assessed. The sample is not representative of any national engineering workforce, and the English-language instrument may have limited participation in some regions.

Instrument wording. Several items in the deployed instrument have weaknesses that cannot be corrected after fielding. Two age categories overlap (65–74 and 74+), two items pose a question answered on an agreement or frequency scale (Q80 and Q36), and several items combine more than one object in a single question (for example Q21, Q32, Q42 and Q85). The instrument is deposited as deployed (Supplementary Material 3) so that these can be assessed directly.

Routing incentive. The commissioning pathway was substantially longer than the others. Respondents who discovered this could have selected a different role to shorten the instrument, and the design cannot detect such misreporting.

5.6 Future research directions

The limitations above define a specified programme rather than an open list, and it has two strands.

The first concerns the instrument. Dimensionality of the universal block should be established at the full-sample level before its scales are used substantively, using estimators appropriate to ordinal indicators; measurement invariance across role groups should then be tested once per-group samples are large enough to support it; and test–retest reliability requires an identifier design that the anonymity commitment here precluded. Until that work is done, the universal block should be treated as a set of a priori content groupings.

The second concerns the architecture, and is separable from the first. Randomising block order within role groups would separate sequencing effects from role characteristics, which the present design cannot. Recording completion time would convert burden parity from a design target into a measured outcome. Administering a flat counterfactual alongside the branched instrument would establish whether branching and format variation affect response quality, which the present single-arm design cannot. And applying the verification protocol in

Section 5.3 across multiple branched instruments would establish whether the pattern reported here generalises.

The distinction matters for readers adopting the architecture: the first strand governs whether these particular scales measure what they claim, while the second governs whether the design behaves as intended. The routing and missingness results in Sections 4.3 and 4.4 depend on neither.

6. Conclusion

The architecture reported here addresses a design tension rather than resolving it. Deterministic role-based routing delivers specialised content to the respondents who can answer it, and produces structured missingness that is fully known and requires no imputation. The deployed instrument provides evidence consistent with this: routing operated without error across seven pathways and 335 responses, with no cross-contamination between blocks, and response quality on the universal block did not degrade with pathway length.

What the diagnostics also show is that the design is not free of consequence. Construct scores varied with the quantity of role-specific content preceding them, but in this design exposure and role cannot be separated, and whether the gradients reflect sequencing or role depends on how the comparison is made. That uncertainty is itself the lesson for designers: the comparability such instruments are built to deliver cannot be assumed when differential content precedes the universal block, and it can only be fully tested by randomising block order within role.

The practical implication is that branched instruments require verification after fielding, not only testing before it. Pre-deployment testing establishes that routing logic executes; it cannot establish what differential exposure did to the responses that logic produced. The protocol set out in Table 9 is offered for that purpose, and can be applied to any deterministically branched instrument using data the instrument already generates.

Data and supplementary materials

Appendix A presents the survey instrument in tabular form, with representative item wording from each module. Figure 1 shows the branching architecture. Supplementary Material 1 documents the consent and eligibility process in full. Supplementary Material 2 provides implementation guidance for researchers adopting the architecture, covering sequencing options, module sizing and common design faults. Supplementary Material 3 contains the complete instrument as deployed. Together, these materials permit direct replication of the instrument.

Data availability. The survey data are not publicly available because the ethics approval for the study does not permit individual responses to be shared. The deployed instrument is provided as Supplementary Material 3.

During the preparation of this work, the authors used Grammarly to edit the manuscript. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

Adıgüzel, F., Wedel, M., 2008. Split questionnaire design for massive surveys. J. Mark. Res. 45, 608-617. https://doi.org/10.1509/jmkr.45.5.608

Andreadis, I., Kartsounidou, E., 2020. The impact of splitting a long online questionnaire on data quality. Surv. Res. Methods 14, 31-42. https://doi.org/10.18148/srm/2020.v14i1.7294

Argyris, C., 1990. Overcoming organizational defenses: facilitating organizational learning. Allyn and Bacon, Boston.

Argyris, C., Schön, D.A., 1978. Organizational learning, Addison-Wesley OD series. Addison-Wesley Pub. Co, Reading, Mass.

Axenfeld, J.B., Blom, A.G., Bruch, C., Wolf, C., 2022. Split questionnaire designs for online surveys: the impact of module construction on imputation quality. J. Surv. Stat. Methodol. 10, 1236-1262. https://doi.org/10.1093/jssam/smab055

Bradburn, N.M., Sudman, S., Wansink, B., 2004. Asking questions: the definitive guide to questionnaire design - for market research, political polls, and social and health questionnaires, Revised ed. Jossey-Bass, San Francisco, CA.

Chipperfield, J.O., Steel, D.G., 2009. Design and estimation for split questionnaire surveys. J. Off. Stat. 25, 227-244.

Dillman, D.A., Smyth, J.D., Christian, L.M., 2014. Internet, Phone, Mail, and Mixed-Mode Surveys: The Tailored Design Method, 4th ed. Wiley, Hoboken, NJ.

Eckman, S., Kreuter, F., Kirchner, A., Jäckle, A., Tourangeau, R., Presser, S., 2014. Assessing the mechanisms of misreporting to filter questions in surveys. Public Opin. Q. 78, 721-733. https://doi.org/10.1093/poq/nfu030

Galesic, M., Bosnjak, M., 2009. Effects of questionnaire length on participation and indicators of response quality in a web survey. Public Opin. Q. 73, 349-360. https://doi.org/10.1093/poq/nfp031

Graham, J.W., Taylor, B.J., Olchowski, A.E., Cumsille, P.E., 2006. Planned missing data designs in psychological research. Psychol. Methods 11, 323-343. https://doi.org/10.1037/1082-989X.11.4.323

Hollnagel, E., 2018. Safety-I and Safety-II: The Past and Future of Safety Management, 1st ed. CRC Press, Boca Raton, FL. https://doi.org/10.1201/9781315607511

Imbriano, P.M., Raghunathan, T.E., 2020. Three-form split questionnaire design for panel surveys. J. Off. Stat. 36, 827-854. https://doi.org/10.2478/jos-2020-0040

Kreuter, F., McCulloch, S., Presser, S., Tourangeau, R., 2011. The effects of asking filter questions in interleafed versus grouped format. Sociol. Methods Res. 40, 88-104. https://doi.org/10.1177/0049124110392342

Krosnick, J.A., 1991. Response strategies for coping with the cognitive demands of attitude measures in surveys. Appl. Cogn. Psychol. 5, 213-236. https://doi.org/10.1002/acp.2350050305

Krosnick, J.A., Presser, S., 2010. Question and questionnaire design, in: Marsden, P.V., Wright, J.D. (Eds.), Handbook of Survey Research, 2nd ed. Emerald, Bingley, pp. 263-314.

Meredith, W., 1993. Measurement invariance, factor analysis and factorial invariance. Psychometrika 58, 525-543. https://doi.org/10.1007/BF02294825

Putnick, D.L., Bornstein, M.H., 2016. Measurement invariance conventions and reporting: the state of the art and future directions for psychological research. Dev. Rev. 41, 71-90. https://doi.org/10.1016/j.dr.2016.06.004

Raghunathan, T.E., Grizzle, J.E., 1995. A split questionnaire survey design. J. Am. Stat. Assoc. 90, 54-63. https://doi.org/10.1080/01621459.1995.10476488

Rasmussen, J., 1997. Risk management in a dynamic society: a modelling problem. Saf. Sci. 27, 183-213. https://doi.org/10.1016/S0925-7535(97)00052-0

Rhemtulla, M., Little, T.D., 2012. Planned missing data designs for research in cognitive development. J. Cogn. Dev. 13, 425-438. https://doi.org/10.1080/15248372.2012.717340

Schuman, H., Presser, S., 1981. Questions and Answers in Attitude Surveys: Experiments on Question Form, Wording, and Context. Academic Press, New York.

Schwarz, N., Bless, H., 1992. Constructing reality and its alternatives: an inclusion/exclusion model of assimilation and contrast effects in social judgment, in: Martin, L.L., Tesser, A. (Eds.), The Construction of Social Judgments. Lawrence Erlbaum Associates, Hillsdale, NJ, pp. 217-245.

Schwarz, N., Strack, F., Mai, H.-P., 1991. Assimilation and contrast effects in part-whole question sequences: a conversational logic analysis. Public Opin. Q. 55, 3-23. https://doi.org/10.1086/269239

Smyth, J.D., 2016. Designing questions and questionnaires, in: Wolf, C., Joye, D., Smith, T.W., Fu, Y.-c. (Eds.), The SAGE Handbook of Survey Methodology. SAGE, Los Angeles, pp. 218-235. https://doi.org/10.4135/9781473957893.n16

Toepoel, V., Lugtig, P., 2022. Modularization in an era of mobile web: investigating the effects of cutting a survey into smaller pieces on data quality. Soc. Sci. Comput. Rev. 40, 150-164. https://doi.org/10.1177/0894439318784882

Tomczak, M., Tomczak, E., 2014. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. Trends Sport Sci. 21, 19-25.

Tourangeau, R., Rasinski, K.A., 1988. Cognitive processes underlying context effects in attitude measurement. Psychol. Bull. 103, 299-314. https://doi.org/10.1037/0033-2909.103.3.299

Tourangeau, R., Rasinski, K.A., Bradburn, N., D'Andrade, R., 1989. Belief accessibility and context effects in attitude measurement. J. Exp. Soc. Psychol. 25, 401-421. https://doi.org/10.1016/0022-1031(89)90030-9

Tourangeau, R., Rips, L.J., Rasinski, K., 2000. The Psychology of Survey Response. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511819322

Vandenberg, R.J., Lance, C.E., 2000. A review and synthesis of the measurement invariance literature: suggestions, practices, and recommendations for organizational research. Organ. Res. Methods 3, 4-70. https://doi.org/10.1177/109442810031002

Weick, K.E., Sutcliffe, K.M., 2007. Managing the Unexpected: Resilient Performance in an Age of Uncertainty, 2nd ed. Jossey-Bass, San Francisco, CA.

Woods, D.D., 2006. Essential characteristics of resilience, in: Hollnagel, E., Woods, D.D., Leveson, N. (Eds.), Resilience Engineering: Concepts and Precepts. Ashgate, Aldershot, pp. 21-34.