Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Accounting for the insider researcher's propositions in interview data: an evidential classification from a military education design study

Azem Aminuddin

azem821@gmail.com

National Defense University Malaysia

Zahimi Zainol Abidin

National Defense University Malaysia

Research Article

Keywords: insider research, researcher influence, semi-structured interviews, endorsement, military education, reflexivity

Posted Date: November 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11650188/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Insider researchers often voice their own views during interviews, and participants' agreement then becomes evidence for those same views. Realist and deliberative interviewing manage this exchange when it is planned, but no analysis-stage procedure exists for weighting agreement when the researcher's views entered unplanned. This article proposes and tests such a procedure. Using a methodological instrumental case study and directed qualitative content analysis, it re-examines ten semi-structured interviews and two written responses from an insider study of a small-state military academy in Southeast Asia, where the first author is a serving officer. Each of 138 exchanges between a researcher proposition and a participant's reply was coded as independent testimony, endorsement (bare, elaborated or qualified), corrective challenge or divergent position. The study's themes were then re-weighted under three increasingly strict evidential rules. Endorsement made up 109 of the 138 exchanges. Challenge did not rise steadily with seniority, but every factual correction came from participants with senior rank or authority over the matter. Three structural themes survived the strictest rule; a central design proposition did not. The article contributes an auditable way to separate what participants testified from what they accepted, and to show which feedback reached the insider researcher. Future work should test the scheme with independent coders, larger samples and other hierarchical organisations.

1 Introduction

A researcher who studies his own institution rarely arrives at an interview with nothing to say. He has a history in the setting, a view of its problems and, in design research, a framework he hopes to build. When he voices these views in the interview, the participant's reply becomes part of the evidence later used to judge them. A participant who says "yes, I agree" to the researcher's diagnosis has produced something, but it is not the same thing as a participant who raises that diagnosis unprompted. Qualitative reports seldom keep the difference visible.

The problem is well recognised at the interview stage. Realist interviewing treats the researcher's theory as something to be shared and tested with participants (Manzano 2016), and deliberative interviewing asks the interviewer to put forward positions so that participants can contest them (Berner-Rodoreda et al. 2020). Remedies for acquiescence are likewise built into the interview itself (Mukumbang et al. 2020). These approaches assume that the sharing of propositions was planned. They say little about what an analyst should do afterwards with transcripts in which a researcher voiced his propositions without design, and in which agreement, elaboration and disagreement now sit side by side with testimony the researcher never prompted.

Three gaps follow. First, no analysis-stage rule is available for weighting endorsement of researcher-voiced propositions against independent testimony; member checking comes closest but is a planned, post-interview step that asks participants to review results (Birt et al. 2016), and post hoc ratings of interviewer questions grade the question rather than the participant's uptake (Cairns-Lee et al. 2022).

Second, existing methods presume that theory-sharing was intended, so they offer no means of locating where unplanned propositions entered. Third, power in the interview is usually assumed to run from the interviewer to the participant. Insider research in hierarchical organisations complicates this, since participants may be senior, equal or junior to the researcher (Mercer 2007; Holian and Coghlan 2013), yet the one empirical study of interviewer challenge known to the authors compared interview styles rather than participants' standing relative to the researcher (Berner-Rodoreda et al. 2021).

This article addresses these gaps through a methodological instrumental case. It asks three questions:

- RQ1. Where and how did researcher propositions enter interviews that were not designed for theory-testing?
- RQ2. How did participants respond, and what patterns of corrective and divergent challenge are observable across participants of different seniority relative to the researcher?
- RQ3. How far does each study finding remain supportable as proposition-responsive testimony is progressively excluded?

The guiding principle is simple. Endorsement of a researcher's proposition is not treated as independent validation of it. The analysis asks instead how much of a study's account survives as different forms of proposition-responsive testimony are set aside, and it records which participants were in a position to correct the researcher and did so. For a journal concerned with systemic practice, the second point matters as much as the first. A researcher who is part of the system he studies is also a channel through which the system's feedback about itself is filtered. Classifying responses to his propositions is one way of showing which feedback reached him, which he took up, and which he passed over.

The setting sharpens this point. A defence academy studying its own education through one of its officers is, in effect, an institution observing itself. The researcher's propositions were not neutral hypotheses: several of them restated the reform agenda he had helped to carry while in post. When colleagues responded to those propositions, the interview became one of the few occasions on which the institution's account of itself was put to its members and could be confirmed, qualified or corrected. Who corrected it, and on what matters, is therefore a question about how feedback moves through a hierarchy as well as a question of method.

The article proceeds in six further sections. Section 2 reviews how theory-driven, deliberative and insider interviewing treat the researcher's own views, and where each stops short. Section 3 sets out the case, the data, the coding scheme and the three re-weighting rules. Section 4 reports where the researcher's propositions entered the interviews, how participants of different seniority responded, and how the study's themes changed under the rules. Section 5 discusses what the findings mean for treating endorsement as evidence, for the role of rank in correcting an insider researcher, and for the insider researcher as part of the system studied. Sections 6 and 7 state the limitations and conclude.

2 Researcher propositions in qualitative interviews

2.1 Theory-driven and epistemic interviewing

Several traditions invite the interviewer to say what he thinks. In realist evaluation the interviewer presents programme theory and asks participants to confirm, refine or refute it, a cycle Manzano (2016) describes as a teacher-learner exchange. Deliberative interviewing goes further and treats the interview as a space for argument in which the interviewer states positions and participants are encouraged to disagree (Berner-Rodoreda et al. 2020). Curato (2012) frames respondents as interlocutors whose views are tested and defended rather than merely collected, and notes that military officers in particular may be socialised into an official discourse that such questioning can unsettle.

These approaches share two features. They plan the exchange in advance, and they locate its rigour in the conduct of the interview. Potter and Hepburn (2005) press a related point from discourse analysis: the interviewer's talk is part of the data, and reports that delete it misrepresent how the participant's words were produced. Their argument implies an analytic duty that the planned approaches leave open. If the interviewer's propositions are part of the data, the analyst needs a way of distinguishing what the participant contributed from what the participant accepted.

The closest empirical evidence comes from a test of the deliberative style. Berner-Rodoreda et al. (2021) compared deliberative and conventional interviews (20 in all) and found that participants disagreed with or challenged the interviewer in six of the ten deliberative interviews. They also observed that participants who were unsure tended to agree once the interviewer shared more of his own view, and that the interviewer felt like a junior "PhD student" in conversation with senior participants. The study shows that challenge happens and that it is shaped by who is speaking to whom. It does not grade the agreement it observed, nor does it ask how the study's conclusions would change if agreement were discounted.

2.2 Acquiescence and interview-stage remedies

Agreement offered to please the interviewer is a familiar threat. Hsieh and Shannon (2005, p. 1283) warn that in directed content analysis participants may answer in the direction the researcher's framework suggests, and they recommend an audit trail and attention to non-supporting evidence. Mukumbang et al. (2020) met acquiescence while piloting realist interviews and countered it by offering participants worked and failed examples so that agreement had to be earned. Holian and Coghlan (2013) warn insider action researchers that colleagues may say what the researcher wants to hear, and that apparent consensus can echo the dominant voices in the room.

Conversation analysis adds a finer distinction. Heritage and Raymond (2005) show that speakers who agree can mark their agreement as dependent on the first speaker's view or as independent of it, for example by confirming before agreeing or by recasting the claim in their own terms. Agreement, in other words, comes in grades. The present study does not attempt conversation analysis; its transcripts were not prepared for it. It borrows the insight that a bare "yes" and an elaborated reply that adds the

participant's own example are different evidential objects, and it treats the participant's own example as elaborated uptake rather than as independent corroboration.

Member checking is the analysis-stage tool most often used to settle such questions. Birt et al. (2016) show its value and its limits: it is planned, it occurs after the interviews, and it asks participants to react to the researcher's synthesis. It cannot recover what happened inside interviews where the synthesis was already being voiced.

2.3 Insider research, rank and the direction of agreement

Insider researchers face these problems in sharper form. Mercer (2007) describes the "double-edged sword" of researching one's own educational institution. Colleagues already know the researcher's views and reciprocity in the interview is hard to refuse (pp. 8 to 11); in her own study a colleague later judged that she had led participants (p. 11). She doubts that insider positions can be reduced to a tidy taxonomy (p. 14) and calls for more insider research by academic managers across levels of hierarchy (pp. 14 to 15). Holian and Coghlan (2013) describe the role duality of practitioner-researchers and note that the researcher's level of authority shapes what colleagues will say. Berglund (2025) writes of the "true insider" whose identity as practitioner and researcher cannot be separated.

Within systemic and action research, Coghlan and Shani (2021) distinguish first-person inquiry (the researcher's own reflection), second-person inquiry (inquiry with others) and third-person inquiry (claims offered to a wider audience), and argue that abductive reasoning links the three. Casey and Coghlan (2026, p. 4) caution that participants' accounts should not be taken as the explanation of events. Read together, these arguments place the present problem precisely. A researcher's propositions voiced in an interview are first-person hunches entering a second-person exchange. When the transcript is analysed, the exchange becomes third-person evidence. The classification developed here is a device for keeping that passage visible rather than a claim that endorsement can be made independent after the fact.

Rank adds a further layer. The usual assumption is that the interviewer holds more power than the participant. In a military academy, participants may outrank the researcher, share his rank, or sit below him, and some hold posts that give them authority over the very matters he raises. It is plausible that seniors feel freer to challenge. It is equally plausible, on Heritage and Raymond's (2005) account, that seniors confirm a junior's claims in ways that assert their own authority without contesting them. The direction of the effect is an empirical question that the existing literature has not examined.

3 Method

3.1 Context and the researcher's position

The study is a methodological instrumental case (Stake 1995) nested in doctoral educational design research (McKenney and Reeves 2019) on a pedagogical framework for the defence academy of a small Southeast Asian state. The first author is a full insider: a serving officer of the rank of major at the time

of the interviews, who had previously held a senior academic post at the academy and led the establishment of a faculty of military studies. Participants knew his history and many knew his views. Because the journal uses single-blind review, this position is declared here rather than concealed.

The interview protocol was written to explore the state of professional military education, not to test the researcher's framework. During the interviews, however, the first author voiced his diagnoses and his emerging framework conversationally. The eight themes re-weighted here come from the substantive analysis in the first author's doctoral thesis (in preparation). The present article uses the same corpus to examine how the researcher's propositions entered it.

3.2 Data

The corpus comprises ten semi-structured interviews of one and a half to two and a half hours, conducted in person or online between March and November 2025, and two written-response sets, one from a participant who answered only in writing and one from a participant who also gave a follow-up interview. Eleven officers and staff of the academy took part. Relative to the researcher, five held senior military rank, four held equivalent rank, one held junior rank, and one held no military rank but a senior post. Bands were assigned by substantive rank at the time of interview, compared with the researcher's rank of major and using equivalent ranks across services (a lieutenant commander counts as a major, a commander as a lieutenant colonel). Rank was taken from participants' own introductions and confirmed by the first author. It does not capture seniority within a rank, length of service or authority of post, which are considered as competing explanations in Section 4.3. Interviews were conducted in English with some Brunei Malay; transcripts were produced by automatic transcription and then cleaned. Four of the cleaned files are question-by-question summaries that retain few interviewer turns; the remainder retain the turn-by-turn record.

Ethical approval

was granted by the National Defence University of Malaysia Ethics Committee. All participants gave informed consent. One participant asked not to be identified; in this article participants are referred to only by code and band, and details of post, service and nationality are withheld.

3.3 Unit of analysis and coding scheme

The unit of analysis is the proposition-response exchange: a researcher proposition together with the participant's following turn or turns, up to the point where the topic changes or a new proposition is voiced. Keeping the interviewer's turn inside the unit follows Potter and Hepburn's (2005) argument that the interviewer's talk is data. A proposition was defined as an utterance or protocol question that asserts or presupposes a claim about the setting, or proposes a remedy, which the participant can accept or reject. Information questions and topic prompts were not counted.

Two origins were distinguished. Planned propositions were embedded in the protocol wording, for example a question asking whether the curriculum used "too much foreign case studies … or is it well-

balanced”, and were put to every participant in the same words. Unplanned propositions were voiced in interviewer turns. For protocol questions that presuppose change (“If we redesigned PME … what should be improved?”), simply answering within the presupposition was not treated as endorsement; a response was coded only when the participant explicitly affirmed, qualified or rejected the presupposition.

Each exchange received one response code (Table 1). The codes are mutually exclusive at the exchange level. Annotations were added separately for responses that rested on what policy requires rather than on what happens, for irony, for agreement later contradicted in the same interview, for proposition type (diagnostic or design) and for strength (assertive or tentative).

The last column of Table 1 refers to three evidential rules, used later to re-weight the study’s themes and set out in full in Section 3.4. Each rule is stricter than the one before. Rule 1 counts all independent testimony and all endorsement. Rule 2 removes bare endorsement. Rule 3 keeps only testimony given before the researcher voiced any view of his own in conversation, the written responses and, for claims about formal arrangements, documents.

Table 1
Response codes

Code	Label	Decision rule	Evidential treatment
A-pre	Independent testimony before exposure	Participant raises the matter before the researcher has voiced any view of his own in conversation (beyond the protocol questions) in that interview, or in a written set	Counts under all three rules
A-post	Independent testimony after exposure	Participant raises a new matter after the researcher has begun voicing views on other matters, but before he has voiced a view on this one	Counts under Rules 1 and 2
A (absence)	Testimony of absence	Participant reports not knowing or not having encountered something; not knowing is not rejecting	Treated as A
B1	Bare endorsement	Agreement without the participant's own content	Excluded under Rules 2 and 3
B2	Elaborated endorsement	Agreement with the participant's own example or reasoning; annotated as illustrating or extending the proposition	Counts under Rules 1 and 2 as uptake, not as corroboration
B3	Qualified endorsement	Agreement that narrows, conditions or partly rejects the proposition	Treated as B2 for the accepted part
C1	Corrective challenge	Participant shows that the researcher misread a fact about the setting	The analysis must be corrected
C2	Divergent position	Participant holds a different view or experience	Retained as a contested finding
E1	Non-substantive response	Acknowledgement, deflection or silence	Excluded from all denominators

Three boundary rules were fixed before coding. An answer to an open question about a matter the researcher had already voiced in that interview was coded as delayed uptake (B2 or B3), not as A-post, because it fails the "not yet voiced" test. When an immediate response was missing and a later one existed, the later one was coded. C1 and C2 were distinguished by function rather than form: C1 required a claim about a fact of the setting that would oblige a correction, while C2 covered disagreement about interpretation, priority or design. A C1 code records the corrective function of the response, not the truth of the correction.

3.4 Procedure and re-weighting rules

The procedure followed directed qualitative content analysis (Hsieh and Shannon 2005), with the response codes as the initial category set and the substantive study's eight themes as the objects of re-weighting. Each transcript was read in full, exchange by exchange, and every interviewer turn was checked against its speaker label before coding. Notes were kept in a running log that records, for each

exchange, the proposition's wording, origin, type and strength, the response code and the theme concerned.

For RQ2, codes were aggregated at the participant level, which remains the unit of comparison, since exchanges within one interview are not independent observations. For each participant, the challenge share was calculated as the number of C1 and C2 exchanges over all eligible exchanges (A, B1, B2, B3, C1, C2), with E1 and unanswered propositions excluded. Absolute counts are reported beside every proportion. The directional proposition that challenge rises with seniority was read on a graded scale: consistent, partly consistent or non-linear, not supported, contrary, or no clear band-level pattern. A competing-explanation matrix was used to consider closeness to the researcher, proposition type and strength, topic, mode (oral or written) and the number of propositions each participant met.

For RQ3, each theme was re-examined under three rules applied in turn. Rule 1 counts independent testimony and all endorsement. Rule 2 removes bare endorsement. Rule 3 admits only independent testimony given before the researcher's first unplanned proposition, written responses and, where a theme concerns formal arrangements, documents. Themes were then labelled robust (supported under Rule 3), robust after exposure (supported by A-post testimony from three or more participants but thin under Rule 3), co-produced (the phenomenon is attested independently but its label or mechanism rests on endorsement) or design conjecture (resting mainly on endorsement with little or no independent testimony). Contrary testimony and C2 responses were recorded against each theme and flagged as contests.

3.5 Second coding, trustworthiness and AI assistance

The second author, who is the first author's doctoral supervisor, acted as second coder, coding the transcripts with the rules in Table 1. Participants' bands were not disclosed to him, so that knowledge of rank could not influence the coding. The two codings agreed on 125 of the 138 exchanges (90.6%). Part of the difference arose because the second coder treated some adjacent exchanges as a single unit; the remaining disagreements were settled by discussion, and the original codes were retained. Because the second author is a co-author and not independent of the study, the second coding shows that the rules were applied consistently; it cannot show that the researcher's interpretation is free of insider bias. For that reason a reflexive log records the points at which the researcher's interpretation was challenged and whether it was revised or retained (Section 4.2).

Documents were used only to corroborate claims about formal arrangements, such as whether a policy or criterion exists, and not to corroborate participants' experience. A policy may establish what an institution prescribes; it does not establish what its members routinely encounter.

The first-pass coding was carried out with the assistance of a large language model (Claude, Anthropic) working from the full transcripts under the first author's written coding rules. The first author reviewed the coding and verified every quotation against the source transcript; where his reading diverged from

the first pass, his reading was adopted. The model did not generate themes or interpret participants' meaning independently of these rules.

4 Findings

4.1 Where propositions entered the interviews

Researcher propositions entered by four routes. The first was the protocol itself. Questions on culture and religion asked how important it was for education "to reflect Brunei's culture and Islamic values", whether Islamic teaching principles should shape "how subjects are taught", and how important it was "to embed Islamic pedagogical principles into our formal teaching doctrine". Another asked whether there were "too much foreign case studies". Because these questions were identical for every participant, they hold the proposition constant across bands, and the written sets were exposed to them as fully as the interviews were.

The second route was the interviewer's own turns. In the turn-by-turn transcripts the first author voiced diagnoses ("we don't have that … framework at all"), causal claims ("the format … is not a very effective way of learning") and design proposals ("one of the simplest solution would be run the train the trainer program"). The number of such turns varied by an order of magnitude: one participant met a single recorded unplanned proposition, another met thirty-six. Some of this variation is real and some is an artefact of transcription. The four summary-style files kept few interviewer turns, and in one of them the interviewer's propositions could be recovered only through the participant's uptake ("as you said", "you're right", "like you said"). Cleaning a transcript can erase the evidence that a proposition was voiced.

The third route was reformulation. The first author repeatedly restated a participant's point in his own terms and then built on it. One senior participant had proposed bridging the gap between classroom and operations through "Simulation, that is as real as possible"; the interviewer recast this as "experiential learning" and placed it in "a new pedagogical framework". The participant later named "experiential learning" as the most effective method and, near the end, listed it as a framework component "as we discussed earlier". The phenomenon was the participant's; the label and its place in the framework were the researcher's.

The fourth route was the read-back. When the first author summarised his thesis aim near the end of an interview, one participant replied by listing the researcher's earlier suggestions back to him as recommendations: "I agree with your suggestion and observation earlier when you mentioned about involving more local case studies … you highlighted about importance of ensuring that that instructors, directing staff are equipped … you mentioned about erm, embracing the latest technology, such AI … yes I agree with that". The same participant had earlier said that their own course was "quite balance[d]" in local content. In a later recap the interviewer described the earlier discussion as agreeing "that we are very pretty much Western influence", overwriting the participant's divergent answer.

4.2 How participants responded

Across the eleven participants, 138 exchanges received a substantive response (Table 2). Endorsement dominated: 23 bare, 66 elaborated and 20 qualified, or 109 of 138. Independent testimony in direct response to a proposition accounted for 2 exchanges, both of them testimony of absence. Challenge accounted for 27: 8 corrective and 19 divergent. Most independent testimony in the corpus did not occur inside proposition exchanges at all but in answers to neutral questions, which is why it appears in the theme analysis rather than in Table 2.

Bare endorsement often closed down an exchange that had been heading elsewhere. One participant answered the researcher's question about training instructors by suggesting that the academy "look to the civilian sector … ubd … Ibte", since the usual view was taken "through a very military lens"; when the researcher then proposed that "one of the simplest solution would be run the train the trainer program", the reply was "Yeah, absolutely." Another participant, asked whether short postings made it pointless to train instructors because of "the culture and the mindset", said only "Yes I agree, it's the culture." These replies are compatible with agreement and with politeness, and the transcript cannot distinguish them.

Elaborated endorsement was the commonest response and the most ambiguous. Some elaborations illustrated the proposition. Others extended it with matter the researcher had not raised, as when a participant agreed that continuing education was poorly supported and added that their own request to study for a doctorate had been refused because the service "can't afford to lose people". Collaborative completion was frequent in the longest interview: the researcher began a sentence about the gap between courses and units and the participant finished it ("What use would that be in practical terms when they go back to the units"). Such completions show alignment but leave authorship shared.

Corrective challenge was rare but consequential. A senior participant, told that a topic had been added mid-course "not because you wanted to align it back with … national strategic interests", replied that alignment had "already been done before the start of the course because the syllabus has to be approve[d] by high headquarters", and that the mid-course addition had filled a gap left by a cancelled visit. Another senior participant corrected the claim that professional military education had begun with the academy ("not during training institute been in existence since the 1990s, huh?") and, more substantively, the premise that the academy owned military education: under the systems approach to training the academy is "the training delivery authority" and "directionally, it doesn't come from Defense Academy itself". When the researcher described a briefing that placed cadets at lower levels of Bloom's taxonomy, this participant supplied the reason the researcher's account had omitted: the mapping "was designed for people who has not obtained their degree. Like, people like me", before agreeing that it no longer fitted graduate entrants.

Divergent positions took several forms. Some rejected a design proposition outright. Told that the framework would embed Islamic pedagogical concepts in place of an approach that was "superficial … ritualistic", one participant replied: "All well and good. But … Actually, it's just a label … how do you measure it?", and argued that the aim should be Islamic concepts of governance rather than spirituality. A written respondent judged the curriculum's local content "well-balanced" and the inclusion of Islamic

teaching principles “not necessary … but it would be a bonus”, adding that it “needs justification”. Others rejected a presupposition. One participant, asked what a redesign should contain, said it was “hard to tell if we need to redesign it wholly” and that the academy should “wait until there’s enough evidence”. Another, asked how instruction inculcated critical thinking, replied that “critical thinking cant be taught”, only its habit inculcated. A participant answering the opening question declined the deficit framing altogether: “Pme works … it isn’t broken. And I think to say so would be … fundamentally disingenuous.”

The reflexive log records what the researcher did with these challenges. Some were taken up within the interview. After a senior participant argued that a usable framework already existed in the systems approach to training but was not used, the researcher restated the point accurately (“it’s not because there is a lack of guidance, it’s because there is guidance, but we’re not … following it”) and, at the end of the interview, said that he had “really overlooked” this point. Others were absorbed. When the same participant rejected the framing of Islamic integration as spirituality, the researcher replied that he was “really glad that … you understood where I’m going with this one, which is … not about spirituality”, and later called the participant “one of the … few interviewees that understand what I’m talking about”. The challenge was recast as agreement, and the participant then turned it on the researcher’s own formation: “your worldview might be impacted by that”.

4.3 Challenge and relative seniority

Table 2 reports the participant-level distribution.

Table 2
Responses to researcher propositions by participant

ID	Band	Mode	Props met	A	B1	B2	B3	C1	C2	n	Challenge
SN1	Senior	Oral	16	1	4	2	1	2	3	13	5/13 (38%)
SN2	Senior	Oral (summary)	1	0	1	1	1	0	1	4	1/4 (25%)
SN3	Senior	Oral	19	0	3	6	2	3	5	19	8/19 (42%)
SN4	Senior	Oral and written	12	0	1	13	1	0	1	16	1/16 (6%)
SN5	Senior	Oral	7	0	2	2	3	1	0	8	1/8 (12%)
PR1	Equivalent	Oral (summary)	5	0	2	4	2	0	0	8	0/8 (0%)
PR2	Equivalent	Oral (summary)	2	0	0	4	0	0	1	5	1/5 (20%)
PR3	Equivalent	Written only	0	0	0	1	1	0	4	6	4/6 (67%)
PR4	Equivalent	Oral (summary)	7	1	2	4	0	0	0	7	0/7 (0%)
JR1	Junior	Oral	14	0	2	8	4	0	3	17	3/17 (18%)
UR1	Civilian (senior post)	Oral	36	0	6	21	5	2	1	35	3/35 (9%)
Total			119	2	23	66	20	8	19	138	27/138 (20%)

Note: Props met = unplanned propositions met, including those that drew no response; n = eligible exchanges (A, B1, B2, B3, C1, C2); Challenge = (C1 + C2)/n. Protocol propositions are included in the response counts. Two divergent positions (PR3, JR1) were responses to the framing of a protocol question.

Aggregated by band, the senior participants challenged in 16 of 60 eligible exchanges (27%), those of equivalent rank in 5 of 26 (19%) and the junior participant in 3 of 17 (18%). Read on the graded scale, the pattern is only partly consistent with the proposition that challenge rises with seniority. The senior band's share is higher, but it is carried by two participants (SN1 and SN3), who account for 13 of the band's 16 challenges, while a third senior participant (SN4) endorsed almost everything put to them. The highest individual share in the corpus belongs to a participant of equivalent rank answering in writing (PR3), and the junior participant challenged about as often as the equivalent band. Within-band variation is larger

than the difference between bands, so the band-level comparison is better described as showing no clear pattern than as supporting the proposition.

Corrective challenge shows a sharper pattern than challenge in general. All eight C1 exchanges came from four participants (SN1, SN3, SN5, UR1), three of senior rank and one in a senior civilian post, each of whom held knowledge of, or authority over, the matter corrected: the curriculum of a school they had led, the doctrine and training system in which they were the recognised specialist, the academic programme they advised, or the course on which they taught. Participants of equivalent and junior rank disagreed with the researcher, but none of them corrected him on a matter of fact.

The competing explanations qualify these readings. Exposure differed sharply: the participants with the highest challenge counts also met many propositions, while SN2 met one. Mode mattered: the written set with the highest share answered only planned propositions, at leisure and without the researcher present. Closeness mattered: the senior participant who endorsed almost everything had institutional responsibility for the reforms the researcher had helped to implement and was invited to react to a framework presented as their continuation. Proposition type mattered too. Design propositions about Islamic pedagogy drew most of the divergent positions, from senior, equivalent and written respondents alike, while diagnostic propositions about instructor training drew mostly elaborated endorsement.

4.4 Re-weighting the study’s findings

Table 3 reports the effect of the three rules on the substantive study’s eight themes. One theme (T7) is split into its diagnostic and design claims because they behave differently.

Table 3
Themes re-weighted under three evidential rules (participants)

Theme	Rule 1: participants with independent testimony	Rule 3: participants with A-pre or written testimony	Contrary testimony or divergent positions	Label
T1 Teaching without a pedagogy	11	5	Three participants reported improvement in part of the provision	Robust
T2 Critical thinking demanded, structurally withheld	8	2	Two participants judged current methods sufficient; one judged one course balanced	Robust after exposure; contested
T3 Governance without a governor	10	7	One senior participant: the governing system exists but is not used	Robust; refined
T4 Teaching as posting, not profession	11	4	One senior participant: selection “not such a big issue”	Robust
T5 Measured as training, not as education	9	3 (general weakness of evaluation only)	One senior participant: the training system fits if applied, and education is too academic; two reported useful feedback	Core robust; mechanism robust after exposure; contested
T6 Responsibility without discretion	5	1	None	Robust after exposure; thin
T7a Islamic values compartmentalised (diagnostic)	6	2	None	Robust after exposure
T7b Embed Islamic pedagogy in doctrine (design)	0	1 for, 1 against	Three divergent positions; four qualified endorsements	Design conjecture; contested
T7c Local content insufficient	0 (five elaborated endorsements)	1 for, 1 against	Three divergent positions, one correction, one report of heavy local content	Contested; partly corrected

Theme	Rule 1: participants with independent testimony	Rule 3: participants with A-pre or written testimony	Contrary testimony or divergent positions	Label
T8 A framework the field has already begun to write	4 participant-originated components	2	One: the framework already exists; one: wait for evidence	Co-produced

The three structural themes held. Teaching without a pedagogy (T1) and teaching as posting (T4) were each attested by eleven participants independently and by four or five before any unplanned proposition or in writing. One participant described the commonest method as “a non-method, which is reading off the slides”; another, before the researcher had voiced any view, said that instructors on one course had “just been told they’re going to be an instructor”. Governance without a governor (T3) survived Rule 3 with seven participants, and the senior participant’s challenge refined rather than weakened it: the point was that the governing system existed and was not used, which locates the gap in authority rather than in design.

Four themes were robust only on testimony given after the researcher had begun voicing his views. The claim that education was measured with training tools (T5) was raised by six participants before the researcher had mentioned it in their interviews, but always after he had begun voicing other views, so under Rule 3 only the broader claim that evaluation was weak and procedural survives, in the two written sets and one early interview remark distinguishing training “for certainties” from education “for uncertainties”. The withholding of critical thinking (T2), responsibility without discretion (T6) and the compartmentalisation of Islamic content (T7a) rest similarly on independent but post-exposure testimony. For T7a the strongest evidence is testimony of absence: three participants said they were not familiar with Islamic pedagogical principles, and a fourth recalled that the education policy said nothing about them (“I don’t remember reading anything about Islamic principles or Islamic pedagogy in the pme”).

The design claim that Islamic pedagogy should be embedded in teaching doctrine (T7b) did not survive. No participant raised it independently; it entered through the protocol and the researcher’s turns. It drew one strong elaborated endorsement, in a written set that adopted the protocol’s own terms, four qualified endorsements and three divergent positions. Under Rule 2 it rests on one participant; under Rule 3 it is supported by one written set and opposed by another. It is better reported as a contested design conjecture than as a finding. The claim that local content was insufficient (T7c) fared worse: it was corrected in part by a participant who reported that local security architecture had “just [been] introduced”, contradicted by a written respondent and by a report that one junior course was “heavy on Brunei security studies”, and endorsed mainly in reply to the protocol’s wording.

The final theme (T8) changes character under the rules. Several framework components were participant-originated: engaging the services as stakeholders in course design (“It’s not a Defense

Academy course. It’s an Rbaf. Course”), feedback on whether education is meeting its goals, an instructor exchange programme, a second edition of the education policy. Others were the researcher’s own suggestions read back to him. Removing bare endorsement removes the read-back but leaves the participant-originated components, so the theme survives as a co-produced claim: the field has begun to write parts of a framework, and the parts it wrote can be distinguished from those it accepted.

5 Discussion

5.1 Endorsement as second-person evidence

The first contribution is a way of keeping the passage from second-person exchange to third-person claim visible (Coghlan and Shani 2021). In this corpus endorsement was the default response to researcher propositions, and much of it was elaborated. Treating elaborated endorsement as corroboration would have strengthened every theme the researcher favoured. Treating it as uptake, and asking what survives without it, separated the structural diagnoses, which participants offered unprompted and in writing, from the design proposition the researcher brought to the field, which participants mostly accepted with conditions or contested.

The re-weighting is not a test of truth. A theme labelled robust is one that does not depend on the researcher’s prompting for its support; a design conjecture may still be a good design. The point is that the two should be reported differently. Hsieh and Shannon (2005) recommend reporting non-supporting evidence in directed content analysis, and Birt et al. (2016) show the value of returning claims to participants. The procedure here works earlier and from the other direction: it grades the support the researcher’s propositions received inside the interviews before any claim is returned.

The routes by which propositions entered also matter for practice. Protocol wording carried the researcher’s design commitments into every interview and into the written sets, which are therefore free of unplanned propositions but not of planned ones. Reformulation and read-back produced some of the most persuasive-looking evidence in the corpus, in which participants appeared to recommend the researcher’s framework. None of these is visible in a report that quotes only the participant’s words, which is Potter and Hepburn’s (2005) concern in a concrete form. Transcription can hide them too: in the summary-style files the researcher’s turns survived only as traces in the participant’s replies.

5.2 Rank, role and the direction of correction

The second contribution concerns power. The common assumption that the interviewer holds the advantage did not describe these interviews. Two senior participants challenged the researcher often and directly, one of them questioning the researcher’s own formation. Yet seniority alone did not predict challenge: the most senior participant endorsed almost everything, and the highest challenge share came from an equivalent-ranked respondent writing alone. What did travel with seniority, or more exactly with authority over the matter in question, was correction. Every corrective challenge came from someone who held senior rank or whose post made the disputed fact their own territory.

This fits Heritage and Raymond's (2005) account of epistemic authority better than a simple model of deference. Participants corrected the researcher where they had primary rights to know, and disagreed with him where the question was one of value or design. One senior participant put the methodological point directly: "Don't use rank. And not seniority of the person. Look at the role in the bigger scheme of things." Rank was treated here as an observable relational characteristic, not as a full measure of the power at work in each interview. Comparing by role as well as rank would be the natural refinement.

For systemic practice, the distinction between C1 and C2 describes two kinds of feedback. Corrective challenge changes the observer's picture of the system; divergent positions preserve accounts that the observer's framework cannot absorb. The reflexive log shows both being handled. One correction was acknowledged within the interview; one divergent position was recast as agreement. An institution, or a researcher, may tolerate disagreement while resisting correction, or accept correction while absorbing dissent into its own terms. The data cannot show how the academy learns. They do show the conversational conditions under which feedback reached the researcher, and what he did with it.

5.3 The insider researcher as part of the system

The researcher in this case was not outside the system he described. He had helped to build the faculty whose provision he was now diagnosing, and his propositions carried that history into the interviews. Holian and Coghlan (2013) describe this as role duality, and Berglund (2025) argues that for the true insider the two roles cannot be separated. The data show what this means in practice. The participant who endorsed almost everything was the one with institutional responsibility for the same reforms, so their agreement was in part the system confirming its own reform narrative. The participants who corrected the researcher were those whose posts held the relevant knowledge, which suggests that the institution's knowledge about itself was distributed by role and reached the researcher only where a holder of that role chose to speak.

The researcher's own handling of feedback belongs to the same system. Casey and Coghlan (2026, p. 4) warn against treating participants' accounts as the explanation of events; the reflexive log adds that the researcher's accounts of those accounts need the same caution. One correction, that a governing system already existed but was not used, changed his view within the interview and reframed a theme. One divergent position was absorbed into his framework as agreement. In Coghlan and Shani's (2021) terms, the first produced the abductive revision that links second-person exchange to third-person knowledge; the second short-circuited it.

For action researchers this suggests a practical check within the action cycle. Before acting on agreement gathered in conversation with colleagues, an insider can ask whether the claim would survive the strictest of the three rules: would it stand on what colleagues said before hearing the researcher's view, or wrote without him present? A claim that survives can carry action. A claim that rests on endorsement is better treated as a proposal still to be tested, which is how the design proposition in this study should now be read.

5.4 Implications for insider and design research

Two practical implications follow. First, insider researchers who find that they voiced propositions in interviews should not discard those passages or quote around them. Coding the exchange, reporting the counts and re-weighting the findings is more honest than either deletion or silent inclusion. Second, transcripts prepared for thematic analysis should retain interviewer turns, including reformulations and summaries, since these are where co-production happens.

The procedure is transferable to other settings where researchers study their own organisations: health professionals interviewing colleagues, academic managers interviewing staff, practitioner-doctorate candidates in any hierarchy. Its demands are modest. It requires a transcript that keeps the researcher's turns, a codebook fixed before coding, and the willingness to report a finding as co-produced or conjectural when the evidence says so.

6 Limitations

The case is single and small, and the comparison across bands rests on eleven participants, one of them alone in the junior band. Exchanges are nested in interviews and are not independent; the proportions are descriptive. Exposure differed greatly between participants, partly because of transcription, and the four summary-style files certainly under-record the researcher's turns. Two passages in one interview had uncertain speaker attribution, and in another the cleaned file omitted the participant's turns between four interviewer propositions; those exchanges were coded conservatively or left uncoded.

The classification was applied by the researcher whose propositions it evaluates, with machine assistance under his rules. The second coding, by the second author (the first author's supervisor), addresses consistency, not insider bias, and the reflexive log addresses bias only partially. Proposition strength was coded from the transcript and cannot recover tone. Documentary corroboration was limited to claims about formal arrangements.

One ethical limitation requires disclosure. In a small institution the combination of band and mode may identify some participants to colleagues; details of post and service have been withheld, and participants confirmed the band assignments.

7 Conclusion

An insider researcher who voices his views in interviews is not thereby disqualified from using what follows. He is obliged to show what he contributed and what his participants contributed. The classification developed here does this by grading responses to the researcher's propositions, distinguishing correction from divergence, and re-weighting the findings under progressively stricter rules. In this case the structural diagnoses survived the strictest rule, a central design proposition did not, and corrective feedback came from those whose rank or role gave them authority over the facts in question. The study's account of its institution is more modest for the exercise, and more defensible.

Declarations

Author contributions Wan Md Azem Hj Md Aminuddin: conceptualisation, data collection, analysis, writing (original draft). Zahimi Zainol Abidin: supervision, second coding, writing (review and editing). Both authors read and approved the final manuscript.

Funding The authors received no specific funding for this work.

Competing interests The first author is a serving officer of the armed forces whose academy is studied and previously held a senior academic post there. The second author is the first author's doctoral supervisor and acted as second coder. The authors declare no other competing interests.

Ethics approval and consent The study was approved by the National Defence University of Malaysia Ethics Committee. All participants gave written informed consent.

Use of AI A large language model (Claude, Anthropic) assisted first-pass coding under the first author's written rules, as described in Section 3.5. The first author verified all coding decisions and quotations.

Data availability Transcripts cannot be shared because of participant confidentiality and the security classification of some content. The codebook and an anonymised exchange log are available from the corresponding author on reasonable request.

References

1. Berglund M (2025) Going in or going out: practicing embedded research as a true insider. Systemic Pract Action Res 38(1):1–24. https://doi.org/10.1007/s11213-025-09713-6
2. Berner-Rodoreda A, Bärnighausen T, Kennedy C, Brinkmann S, Sarker M, Wikler D, Eyal N, McMahon SA (2020) From doxastic to epistemic: a typology and critique of qualitative interview styles. Qualitative Inq 26(3–4):291–305. https://doi.org/10.1177/1077800418810724
3. Berner-Rodoreda A, Bärnighausen T, Eyal N, Sarker M, Hossain P, Leshabari M, Metta E, Mmbaga E, Wikler D, McMahon SA (2021) Thought provoking, interactive, and more like a peer talk: testing the deliberative interview style in Germany. SSM - Qualitative Res Health 1:100007. https://doi.org/10.1016/j.ssmqr.2021.100007
4. Birt L, Scott S, Cavers D, Campbell C, Walter F (2016) Member checking: a tool to enhance trustworthiness or merely a nod to validation? Qual Health Res 26(13):1802–1811. https://doi.org/10.1177/1049732316654870
5. Cairns-Lee H, Lawley J, Tosey P (2022) Enhancing researcher reflexivity about the influence of leading questions in interviews. J Appl Behav Sci 58(1):164–188. https://doi.org/10.1177/00218863211037446
6. Casey M, Coghlan D (2026) Beyond the cycles: rethinking data collection, analysis and interpretation in action research in organisational setting: implications for QuARC. Systemic Pract Action Res

39:18. https://doi.org/10.1007/s11213-026-09766-1

7. Coghlan D, Shani AB (2021) Abductive reasoning as the integrating mechanism between first-, second- and third-person practice in action research. Systemic Pract Action Res 34(4):463–474. https://doi.org/10.1007/s11213-020-09542-9
8. Curato N (2012) Respondents as interlocutors: translating deliberative democratic principles to qualitative interviewing ethics. Qualitative Inq 18(7):571–582. https://doi.org/10.1177/1077800412450154
9. Heritage J, Raymond G (2005) The terms of agreement: indexing epistemic authority and subordination in talk-in-interaction. Social Psychol Q 68(1):15–38. https://doi.org/10.1177/019027250506800103
10. Holian R, Coghlan D (2013) Ethical issues and role duality in insider action research: challenges for action research degree programmes. Systemic Pract Action Res 26(5):399–415. https://doi.org/10.1007/s11213-012-9256-6
11. Hsieh H-F, Shannon SE (2005) Three approaches to qualitative content analysis. Qual Health Res 15(9):1277–1288. https://doi.org/10.1177/1049732305276687
12. Manzano A (2016) The craft of interviewing in realist evaluation. Evaluation 22(3):342–360. https://doi.org/10.1177/1356389016638615
13. McKenney S, Reeves TC (2019) Conducting educational design research, 2nd edn. Routledge, London
14. Mercer J (2007) The challenges of insider research in educational institutions: wielding a double-edged sword and resolving delicate dilemmas. Oxf Rev Educ 33(1):1–17. https://doi.org/10.1080/03054980601094651
15. Mukumbang FC, Marchal B, Van Belle S, van Wyk B (2020) Using the realist interview approach to maintain theoretical awareness in realist studies. Qualitative Res 20(4):485–515. https://doi.org/10.1177/1468794119881985
16. Potter J, Hepburn A (2005) Qualitative interviews in psychology: problems and possibilities. Qualitative Res Psychol 2(4):281–307. https://doi.org/10.1191/1478088705qp045oa
17. Stake RE (1995) The art of case study research. Sage, Thousand Oaks