Response to Reviewers and Academic Editor

Manuscript ID: behavsci-4169206

Title: Assessing Military Professionals' Endorsement of Decision-Making Assumptions: An Exploratory Factor Analysis

Thanks to the Academic Editor and both Reviewer 1 and Reviewer 2 for constructive insights, which have significantly strengthened the manuscript. The revisions and clarifications provided here should adequately address all concerns.

Response to Academic Editor

Comment AE1:

"Please provide factor loadings also for the items that have been omitted (Table 4)."

Response:

I fully agree with the intent of increasing transparency regarding item performance in the early stages of the factor analysis. As described in the Results section, the initial extraction with all 11 items produced a three-factor solution, but the third factor was defined by a single item (Q1), rendering the model psychometrically uninterpretable. The scree plot clearly indicated that only two factors were substantively meaningful, and I therefore proceeded with a two-factor solution.

In this two-factor extraction, Q1 remained in the model long enough to generate interpretable loadings, although they were weak (.33 on Factor 1), accompanied by low communalities, weak inter-item correlations, and a negative effect on internal consistency. Consistent with recommended criteria for item retention (primary loadings ≥ .40), Q1 was removed from the final model. Because Q1 was part of the two-factor extraction prior to exclusion, I have now added a row for Q1 in Table 4, showing its actual loadings as requested.

However, Q2 and Q7 cannot be added to Table 4, because they did not enter the two-factor extraction at all. Both items were removed before the two-factor model was estimated due to critically low communalities (< .10) and unstable or negligible loadings in the initial unstructured extraction. As a result, they never produced meaningful or stable loadings within the two-factor framework shown in Table 4. Including them would therefore be statistically misleading, as the values available from the preliminary, unrotated solution do not correspond to the factor structure presented in Table 4 and would not be interpretable within the retained two-factor model.

For transparency, in the revised manuscript I have added the requested Q1 information to Table 4. I hope this addresses your concern and provides readers with a complete and accurate understanding of the item-retention decisions and the development of the final factor structure. Location in revised manuscript: Table 4, p. 9.

Response to Reviewer 1

Reviewer 1 – Comment 1:

Some points are unclear to a reader unfamiliar with Klein's idea. For instance, you start with the proposition that Klein claims/structure is well known, but it is not (of course, it depends on the public (lines 35-39). So, I recommend more explicitly defining these for a broader audience.

- Response:

I agree that the original version of the manuscript did not provide sufficient conceptual grounding for readers unfamiliar with naturalistic decision-making or Klein's work. In the revised manuscript, I have substantially expanded the Introduction to offer a clearer and more detailed explanation of the Recognition-Primed Decision (RPD) model and its relevance to decision-making in military VUCA environments. This includes a stronger description of the cognitive processes underlying RPD—situation assessment and mental simulation—and how these mechanisms contrast with classical analytical models. I also clarify how Klein's 11 decision-making claims emerge from this theoretical foundation and why these claims may reflect multiple, distinct belief dimensions rather than a single construct. Location in revised manuscript: p. 1-2, lines 30-72.

Reviewer 1 – Comment 2:

Why Klein's decision-making claim could be a single dimension (or a multiple dimension)? I mean, a factor analysis is based on a latent construct. In other words, there appears to be a conceptual leap in treating "decisions" as a latent construct for factor analysis. Since decision-making is typically viewed as a process rather than a static latent trait, the theoretical justification for your factor structure remains unclear.

What is not clear is the relationship between the 11 claims, the dual-capacity, and the 3 dimensions described: "sensemaking frames what options become salient; decisions change the environment practitioners must interpret; and adaptation updates both understanding and action". You may suggest, if I understood correctly, that those 11 claims could be translated into competencies, which could then be organized into 2 (heuristic vs systematic) or 3 dimensions (sensemaking, adaptation and decision). However, based on Table 1, what you used is how people evaluate whether they agree with a sentence, not how they decide. If you would like to test the dual-process as Klein named, dual-capacity (dual-process has already been suggested and tested since 1980 with Richard Petty and Alice Eagly; although, as you said, Klein may not have been tested, but his idea is not new and there are several pieces of evidence with different methodologies…).

What you describe based on military demands, in fact, is more related to institutional logics that operate behind the military structure. This military institutional logic is more than 2 (e.g., bureaucratic, legalist, vigilant etc.). But I understand it is not the main point.

- Response:

To address all aspects of your critique, I have:

I. I clarified that the latent construct measured is attitudinal endorsement of decision-making beliefs, not decision-making ability, and the revised Methods section now states this explicitly. Specifically, I added text explaining that participants rated their agreement with the 11 claims using evaluative, attitude-based statements, and that the resulting factor analysis therefore examines belief structures, not decision-making performance or cognitive processes. Location in revised manuscript: p. 2, lines 64-69 and p. 5, lines 224-228

II. Added a detailed explanation of why multiple dimensions are theoretically expected. In the revised manuscript, the theoretical rationale for expecting multiple dimensions within the 11 decision-making claims is strengthened. Specifically, the Introduction (p. 2, lines 52–62) is expanded to incorporate Klein's own distinction, outlined in Streetlights and Shadows (Klein, 2009).

III. Explained why the items are attitudinal and not behavioral, and how that shapes interpretation. In the updated manuscript, I clearly explain that our results on Klein's 11 claims reflect attitudes. They show how respondents feel about general advice or beliefs on making decisions, rather than how decisions play out in real-life situations. Location in revised manuscript: p. 2, lines 57-66 and p. 5, lines 224-228.

IV. Removed any suggestions of testing dual-process theories. I now clarify that the factor analysis does not test Klein's cognitive processes directly p. 2, lines 76-83. I have also added a limitation stating that the study examines only attitudinal endorsement of Klein's claims and does not measure the cognitive processes underlying recognition-based or analytic reasoning. Location in revised manuscript: p. 13, lines 535-542.

V. Improved the justification for exploratory factor analysis. I have now added a paragraph describing the feasibility of our exploratory factor analysis. This clarification appears on p. 6, lines 267–273, where I explain that Klein's 11 claims stem from heterogeneous decision-making traditions and therefore lack latent structures, making EFA the appropriate methodological choice

Lastly, I thank reviewer 1 for raising the point about military institutional logics. I agree that military contexts are shaped by multiple, co-existing logics (e.g., bureaucratic, legalist, vigilant), and these extend well beyond the two attitudinal dimensions identified in our analysis. However, the aim of the present study is narrower: to examine how respondents organize their attitudinal endorsement of Klein's decision-making claims, not to map the broader institutional logic structure of the military.

Reviewer 1 – Comment 3:

Your sample is quite small (N=225), which falls below any psychometric sample cutoff/suggestion, and you should defend your point. Furthermore, it would be helpful to clarify why you expected a different factor structure, specifically within a military population (although you don't have any other sample to compare).

- Response:

I appreciate the feedback from both reviewers regarding the sample size. Reviewer 1 expressed concerns that the sample (N = 225) is small for psychometric analysis, while Reviewer 2 believes it is adequate based on the guideline of 20 participants per item. Thus, I acknowledge the concern but maintain that the sample of 225 is sufficient for exploratory factor analysis (EFA), since the sample size exceeds the recommended range for our 11-item instrument. Thus, no changes were made to the revised manuscript based on this rationale.

In response to Reviewer 1's question about why a different factor structure was expected in a military population, I note that the theoretical rationale is already described in the manuscript: military decision-making in VUCA environments emphasizes both structured, goal-oriented practices and analytic, evidence-based reasoning, which aligns with Klein's distinctions and supports expecting more than one latent belief dimension (p.3-4 lines 134-188). Regarding the request for a comparison sample, I agree that additional validation would be valuable. As stated in the manuscript under

Methodological Considerations (p.6 lines 268-279), we plan for replication and CFA in future studies. I have now made this point more explicit.

Reviewer 1 – Comment 4:

"Additionally, evidence of validity requires several steps and comparisons. Those steps are not presented in the present work."

- Response:

Thank you for raising this important point. I have expanded the Limitations section to clarify that the present study represents only the first step in a broader validation program. The revised manuscript now explicitly states that additional forms of validity evidence – such as confirmatory factor analysis, convergent and discriminant validity, criterion-related validity, and replication in independent samples – are required to fully establish the robustness of the factor structure (p. 13, lines 535–545).

In addition, I now clarify earlier in the manuscript that exploratory factor analysis is intentionally limited to identifying potential latent structures and does not provide the full range of validity evidence expected in later validation stages. As standard psychometric guidance notes, EFA precedes and informs subsequent validation methods rather than replacing them. For this reason, a single exploratory study cannot present all forms of validity evidence within one design. This clarification has been added to the methodological discussion on p. 6, lines 276–279.

Reviewer 1 – Comment 5:

"Klein's model often hinges on the distinction between experts and 'juniors'. Your data showed no significant differences based on age (a common proxy for seniority). This raises a critical question: is this due to a lack of variance in the sample's experience levels, or is the instrument perhaps not sensitive enough to capture these distinctions or the construct/items/factors are not well measured?"

- Response:

I agree that Klein's framework distinguishes between expert and novice decisionmakers. In our data, however, chronological age did not differentiate factor scores. As clarified in the revised manuscript (pp. 12, lines 497), this null effect should be interpreted cautiously. Chronological age is an imprecise indicator of expertise, as domain-specific skill develops through accumulated knowledge, exposure, and sustained deliberate practice rather than through age alone (Charness et al., 2008). The expertise literature further shows that self-assessed expertise often displays no systematic relationship with age, reflecting substantial inter-individual variability at all ages (Van Der Heijden, 2001).

Accordingly, the absence of age effects in the present study likely reflects restricted variance in operational experience within the sample, as well as the inherent limitations of using age as a proxy for expertise. To address this, the manuscript now explicitly notes that age-based comparisons should not be interpreted as tests of Klein's expert–junior distinctions.

I also highlight in the Limitations section (p. 14, lines 550–555) that future research should incorporate direct indicators of expertise (e.g., rank, years in operational roles, mission-specific experience) and behavioral or process-tracing decision tasks to more rigorously examine expert–novice distinctions central to Klein's model.

Reviewer 1 – Comment 6:

“The manuscript would benefit from a more detailed description of how the instrument was constructed and administered. Currently, the items seem to lack a unified ‘target’; some refer to leaders, some to ‘our plans,’ or ‘we can’ and others to generic decision-making biases. This makes it difficult to capture a cohesive and integrative latent construct.”

- Response:

This comment aligns with broader concerns you raised about what the instrument measures, and I have now strengthened the relevant sections of the manuscript accordingly.

This comment relates directly to your earlier concerns about conceptual coherence, and I addressed it by clarifying in Section 2.2.1 (p.5 lines 217-228), that the 11 items were drawn directly from Klein’s (2009) published formulations and translated into Norwegian with minimal adaptation to preserve their original meaning. Because these claims were not designed as psychometric indicators, they differ naturally in grammatical form and target focus (e.g., some referencing “leaders,” others referring to “we,” or expressing general prescriptive advice). I now explicitly state that this variation reflects the structure of the source material rather than inconsistencies introduced during instrument development.

Reviewer 1 – Comment 7:

“I noted that the mean score for the item regarding ‘teaching people procedures helps them make decisions and perform tasks more skillfully’ was 1.8 on a 5-point scale. This seems surprisingly low for a military cohort. It might be worth double-checking this data point and/or providing a contextual explanation.”

- Response:

I have double-checked the data and confirm that the mean score of 1.8 is correct. The confusion stemmed from an earlier inconsistency in scale direction, which has now been corrected in Table 1 (p.7). As noted in the revised manuscript (p. 5, lines 224–224), our study used a scale where 1 = Strongly Agree and 5 = Strongly Disagree, meaning that a score of 1.8 indicates strong endorsement, not low agreement. This finding is consistent with military training culture, where procedural instruction is highly valued. It also aligns with Klein’s (2009) data, where all retained claims scored above 5.0 on a 7-point “high = agree” scale; when reversed to match our scale direction, his values fall between 2.0 and 2.8, comparable to the endorsement observed in our sample.

Reviewer 1 – Comment 8:

“Please ensure you include internal consistency measures (such as Cronbach’s alpha or, ideally, McDonald’s omega) for your factors to support the evidence of validity of your findings (at least an alpha, as you used SPSS, I guess SPSS does not offer omega yet).”

- Response:

I have revised the Limitations section to explicitly address the issue of internal consistency metrics (p. 14, lines 556–558). As noted, IBM SPSS Statistics does not provide McDonald’s omega within its native EFA or reliability analysis functions, which limits the direct calculation of ω using the standard SPSS interface. For this reason, and consistent with the software capabilities available during the analysis, I reported Cronbach’s alpha, which remains the most used internal consistency metric in SPSS-based exploratory studies. The revised text now also states that future validation work will

employ software capable of estimating omega to provide a more comprehensive assessment of reliability.

Reviewer 1 – Comment 9:

"Given the many problems with your instrument, sample, and dual-process, I would consider your study a starting point and would suggest a review in the sentences based on the psychometric literature."

- Response:

I agree that the present study represents an initial exploratory step, and I have already positioned it accordingly in the manuscript. Following your suggestion, I carefully reviewed whether additional revisions were required. However, the concerns you raised regarding the instrument, sample, and theoretical scope are already comprehensively addressed in the revised *Limitations and Future Directions* section (p. 13, lines 500–528), where I explicitly note that (a) Klein's items were not designed as psychometric indicators, (b) the study provides only preliminary structural evidence, and (c) further validation—using purpose-built items, behavioral methods, and independent samples—is necessary. As these revisions fully incorporate the psychometric considerations you highlighted, no further changes were made in response to this comment.

Reviewer 1 – Comment 10:

"I believe the following sources may offer valuable perspectives on scale development as you revise your work (they are not mine):

http://journals.sagepub.com/doi/10.1177/109442819800100106

https://doi.org/10.1016/j.leaqua.2018.07.001

Doi: 10.1590/1982-7849rac2022210085.en

Doi: 10.3389/fpubh.2018.00149"

- Response:

Thank you for suggesting these useful methodological resources. I have reviewed your recommended papers (Hinkin, 1998; Rogers, 2021; Crawford & Kelder, 2019, Boateng et al. (2018)) and confirm that the revisions already implemented fully align with the best-practice guidance they outline. Because our study is positioned as an initial exploratory examination of belief structures rather than a full scale-development project, I did not incorporate these sources directly into the manuscript. However, the manuscript now clearly reflects their core recommendations. For example, transparent reporting of item origins, explicit acknowledgment of conceptual heterogeneity, clarity that the study is exploratory, justification of the EFA approach, and a strengthened Limitations section outlining the need for future confirmatory and validity-oriented work. I therefore believe the manuscript addresses the intent of your suggestion without requiring additional in-text citations of these methodological sources.

Response to Reviewer 2

Reviewer 2 – Comment 1:

"The paper examines 11 core claims of Gary Klein as part of his natural decision making and Recognition-Based Decision (RBD) theories on a sample of Norwegian Armed Forces personnel (N=225). The writers believe that these claims represent a two-factor structure, i.e. 'Planning/Structure' and 'Analytical/Evidence-Based Applications,' rather than a single-dimension one. The subject matter is very important to the understanding of decision-making mechanisms under VUCA conditions. Nonetheless, the introduction and literature review are not giving enough attention to the theoretical aspects of why Klein's claims are contradictory or why they should be multidimensional. Besides, the use of factor analysis as an exploratory tool results in a methodological inconsistency with the way hypotheses (H1) are presented."

- Response:

This comment aligns closely with concerns raised by Reviewer 1, and I have already addressed it by removing the hypothesis (H1) entirely. The study is now framed solely around a research question, ensuring that the analysis is presented as fully exploratory. This adjustment is reflected on p. 2, lines 64-72. I also revised terminology throughout the manuscript to avoid any implication of model validation, explicitly stating that EFA cannot be used to validate hypotheses.

Reviewer 2 – Comment 2:

"The paper is promoted as an 'Exploratory Factor Analysis' (EFA). Generally, EFA seeks to 'identify' the number of factors a certain structure is composed of. Nevertheless, by saying in H1 that 'a multi-factor model will fit better than a single-factor model,' the authors have basically given the research a semi-confirmatory character. The testing of a particular hypothesis (H1) would require Confirmatory Factor Analysis (CFA) as the analysis method. On the other hand, if the dataset is new and the structure has to be uncovered, the H1 should be an 'expectation', or the research question (RQ1) should be put forward."

- Response:

I agree that presenting a formal hypothesis (H1) suggesting that a multi-factor model would "fit better" than a single-factor model introduced a semi-confirmatory tone that is inconsistent with the exploratory purpose of EFA. In line with your recommendation, I have therefore removed H1 entirely from the manuscript and reframed the study as fully exploratory.

In the revised manuscript, the inquiry is now presented as a research question (RQ1), reflecting that the dataset is new and that the latent structure needed to be identified rather than tested. I also added explicit clarification that EFA cannot validate or confirm predetermined measurement models, and that confirmatory procedures such as CFA would be required to evaluate any hypothesized factor structure.

These changes are reflected in the Introduction (p. 2, lines 70–73) and Limitations (p.13, lines 530–534), where the manuscript now explicitly states that the goal is to uncover underlying structures of participants' attitudinal responses. I also note that Reviewer 1 raised a related concern about the conceptual justification for treating the items as indicators of a latent construct. By removing H1, reframing the study around an exploratory research question, and clarifying that the instrument assesses attitudinal beliefs rather than cognitive processes or decision-making performance, I have addressed this issue for both reviewers.

Reviewer 2 – Comment 3:

"11 items derived from Klein's (2009) book do not represent a standard 'scale,' but rather the authors' theoretical claims. The authors haven't indicated how they took these statements and turned them into a Likert-type questionnaire (the choice of words, the scaling technique). Adding a paragraph of 'Item Development' can help describe the transition from the original items to the survey format. If experts were consulted to assess the face validity of the items, this should also be mentioned."

- Response:

I have expanded the Instrument section to clarify how the 11 survey items were derived directly from Klein's (2009) published decision-making claims. This is described in the revised manuscript (p. 5 lines 217-223). Because these claims were formulated by a leading scholar in naturalistic decision-making and represent established conceptual propositions rather than psychometric indicators, I retained their original grammatical structure and referent focus to ensure fidelity to the source material. As noted in the revised manuscript (p. 14 lines 548-553), this study represents an initial exploratory assessment of how Klein's published claims are organized in respondents' belief structures.

Reviewer 2 – Comment 4:

"After the analysis, the authors concluded that two dimensions, 'Planning' and 'Analytical,' were identified (Abstract section). Nevertheless, the manuscript doesn't provide sufficient theoretical rationale for the differentiation of these two dimensions as evidenced by the literature review. More emphasis should be rather put on the discussion about the opposition of rational/analytical decision-making with intuitive/structural decision-making (for instance, with reference to Dual-Process Theory). Doing so will support your finding of a two-factor structure."

- Response:

I appreciate this point. In response to Reviewer 1 comment 2, I clarified why multiple dimensions are theoretically expected when practitioners evaluate Klein's claims. The revised Introduction explains how the 11 claims reflect distinct components of the RPD model and how military decision-making inherently involves both structured preparation and analytic evidence-based reasoning. This expanded theoretical grounding supports the emergence of two distinct but related factors. These revisions appear on pp. 1–2, lines 40–72.

Reviewer 2 – Comment 5:

"225 participants constitute a fair sample size for EFA according to the 20 participants per item criterion. However, 'dropping two items because of low commonality' (Line 12) might have resulted in the weakening of the scale from the content perspective. The Introduction or Methodology sections should mention which items were dropped, and how the changes weakened the content validity of the 'decision-making' domain."

- Response:

This issue was also raised by the editor, and the manuscript now includes a detailed description of the item-removal process. As explained in our response to the editor, two items (Q2 and Q7) were removed due to critically low communalities and unstable or negligible loadings in the initial extraction, while Q1 was removed for failing to meet minimum loading strength within the two-factor structure. Importantly, the manuscript now provides transparent justification grounded in best-practice

psychometric criteria, with specific loadings reported for Q1 in Table 4. See revised text in the Results section (pp. 7–9).

Reviewer 2 – Comment 6:

"The description of the VUCA concept is very superficial. What the Norwegian Armed Forces (2024–2025) are specifically facing (e.g., Arctic security, hybrid threats, etc.) should be tightly tied to Klein's claims. Referencing briefly Norwegian military doctrines on decision-making paradigms in the 'Military-specific studies' (Line 125) section would definitely add value to the study both locally and globally."

- Response:

In accordance with your recommendation, I have substantially expanded the description of the VUCA concept to make it more specific to the Norwegian Armed Forces' operational realities. The revised text (pp. 2–3, lines 76–104) now incorporates literature and concrete examples from both the High North and Norway's recent expeditionary operations. Together, these revisions directly address your comment by grounding the VUCA discussion in Norwegian strategic and operational experience and by linking these real-world conditions more tightly to Klein's theoretical framework.

Reviewer 2 – Comment 7:

"The selection of 'Principal Axis Factoring with Oblimin rotation' by the authors is a methodologically correct choice but a reasoning should be given why the factors are believed to be correlated (oblique). Hence, it would be best to refer to literature when explaining the expected correlation between the factors."

- Response:

In the revised Methods section, I explain that Klein's model conceptualizes sensemaking, decision-making, and adaptation as interdependent cognitive processes. Because these belief structures are theoretically expected to correlate, an oblique rotation is appropriate (Sass & Schmitt, 2010). These clarifications appear in the updated methodological explanation (pp. 3 and 6, lines 134-138/ 267–282).

Reviewer 2 – Comment 8:

"It seems that Klein's original statements were not a 'scale.' The translation process of these sentences into Norwegian (did you carry out a back-translation?) and the expert review stages (content validity) are not clear. How was the psychometric suitability of the items for a survey ensured?"

- Response:

I note that this issue was also raised earlier by both Reviewer 1 and Reviewer 2, and the manuscript has now been revised accordingly in response to those comments.

As clarified in the revised Instrument section (Section 2.2.1), the 11 items were taken directly from Klein's (2009) published decision-making claims and translated into Norwegian with minimal adaptation to preserve their meaning. Because the purpose of the study is to examine attitudinal endorsement of Klein's conceptual propositions, rather than to construct a new psychometric scale, fidelity to the original formulations was prioritized. For this reason, and as now explained in the

manuscript, external expert-panel review was not employed, as such procedures would have altered the conceptual integrity of Klein's statements.

These revisions were made earlier in the process based on overlapping feedback from Reviewer 1 (regarding item construction and conceptual heterogeneity) and from your earlier comments on methodological clarity. The updated Limitations section (p. 13, lines 529–545) now explicitly acknowledges these methodological boundaries. In combination, these revisions address concerns about translation, expert review, and psychometric suitability while remaining consistent with the exploratory aims of the present study.

- **Reviewer 2 – Comment 9:**

"The 5-point Likert scale was characterized as '1=Strongly Disagree, 5=Strongly Agree' (Line 176). Nevertheless, in Table 1, the means (e.g., Q11 = 1.64) reveal that almost all participants have disagreed with the statements. Hence, these items seem to possess very little discriminative power and that 'social desirability' bias might also take place. It is necessary to discuss this situation."

- Response:

There was indeed an inconsistency between the manuscript text and the caption of Table 1. The text (line 177) correctly stated the response scale as "1 = Strongly Agree, 5 = Strongly Disagree," which reflects the scoring used in the survey instrument and all statistical analyses. However, the caption of Table 1 (p. 7 line 287) mistakenly displayed the opposite labels. This created the appearance of a contradiction when comparing the described scale with the observed means.

I have corrected the Table 1 caption to display the correct scale direction: 1 = Strongly Agree, 5 = Strongly Disagree. The low means (e.g., 1.64 for Q11) therefore represent high endorsement, not disagreement.

In addition, as you recommended, I have expanded the Discussion section to address two issues raised by this pattern of findings (p. 12 lines 480-492):

- the implications of restricted variance and potential response tendencies within a relatively homogeneous military cohort, and
- why the discriminative power of certain items may be limited, particularly for statements that reflect widely shared doctrinal or procedural expectations.

These revisions clarify the endorsement patterns and strengthen the interpretive transparency of the results..

Reviewer 2 – Comment 10:

"Convenience sampling was chosen as the sampling method. 225 persons are an almost adequate number for EFA. Demographic variables that have a direct influence on decision-making styles, such as participants' rank and years of experience (seniority), should be included."

- Response:

In addressing Reviewer 1, I already clarified that age showed no relationship to factor scores and that age is an imprecise proxy for expertise. I also noted that future research should include more direct metrics of experience such as rank or years in operational roles. This point is reiterated in the revised discussion section (p. 13, lines 500–506).

Reviewer 2 – Comment 11:

“That a two-factor solution accounts only 29.24% of the total variance is, by academic standards, a very low number (usually one expects from 50 up to 60%). It may be interpreted as a lack of congruence of the model with Klein's theory or, alternatively, the items are insufficiently related. The low rate should not be neglected or regarded as ‘modest’ but rather as a serious limitation.”

- Response:

This aligns with Reviewer 1’s concern about missing evidence of validity. In response, I expanded the Limitations section to clarify that EFA represents only a first analytical step, and that additional validation (CFA, discriminant and convergent validity, criterion evidence) will be necessary. I also emphasize that the modest variance explained reflects the conceptual heterogeneity of Klein’s claims. These additions appear on p. 14, lines 548–555.

Reviewer 2 – Comment 12:

“The Cronbach's Alpha for the Factor 1 was .65. In social sciences, a .70 level is considered as a generally accepted lower limit. Thus, a value of .65 is perceived as ‘weak/borderline’ and consequently, using this factor as a ‘scale’ is a risky decision.”

- Response:

I agree that an α value of .65 falls below the conventional .70 guideline. In the revised manuscript, I now explicitly caution that the internal consistency estimate for Factor 1 should be interpreted as *acceptable but borderline (*p. 10, line 359), particularly given that the factor consists of only four items derived from Klein’s conceptually heterogeneous claims. As noted in literature, Cronbach’s alpha is known to underestimate reliability when scales are short or when items vary in conceptual breadth, due to restricted inter-item covariance and violations of tau-equivalence assumptions. This downward bias in α for short, heterogeneous attitudinal measures is well documented (Green & Yang, 2009).

Reviewer 2 – Comment 13:

“Some of the questions, namely Q2, Q7, and Q1, have been removed. However, Q1 and Q5 also have very low communality values of .17 and .18 respectively. A stronger rationale should come up technically for decisions on retention or removal of these items.”

- Response:

This issue overlapped directly with the editor’s and Reviewer 1’s question about conceptual alignment and item performance. I provide a clear explanation for the removal of Q1 and the retention of Q5 based on multiple indicators (loadings, communality, inter-item correlations, and internal consistency effects), consistent with recommended multifactor decision rules. These details are found in the Results section (pp. 7–9)

Reviewer 2 – Comment 14:

“After demonstrating the necessity of greater clarity the authors mention that they have carried out an ‘exploratory’ analysis and ‘validated’ H1 (Line 329). I cannot validate hypotheses with EFA; instead, I can only identify the structures. This error in nomenclature needs to be fixed everywhere in the text.

An update of the analysis to Confirmatory Factor Analysis (CFA) or the avoidance of the term ‘validation’ is also a way out.”

- Response:

As noted above in response to your Comment 2, I have already removed H1 entirely and revised all terminology throughout the manuscript to ensure that EFA is presented purely as exploratory. I also clarified that EFA does not provide hypothesis validation. This is located on p. 6, lines 277–279.

Reviewer 2 – Comment 15:

“Referring to the authors' statement that these outcomes will be implemented in training programs (AARs, etc.) (Line 421). A prudent military training session would perhaps not consider a ‘diagnostic tool’ whose inherent property appears to explain the 29% of the variance and be reliable up to .65. It is advised that this part be softened.”

- Response:

In addressing Reviewer 1, I expanded the Limitations section to explicitly state that the present study represents only an initial step and that the instrument does not measure decision-making performance. The revised text clarifies that further validation work is required before practical applications can be considered. The discussion of applied implications has therefore been made more cautious in tone. These revisions are included in pp. 13–14.

Reviewer 2 – Comment 16:

“It is a misconception to consider the limitation of Klein items not being written for psychometric purposes (Line 425) simply a ‘limitation.’ This is, in fact, the very essence of the problem of construct validity of the study. The impact of this situation on the results should be candidly revealed.”

- Response:

This point overlaps with Reviewer 1’s concern about limited validity evidence and measurement boundaries. In response, the Limitations section has been expanded to emphasize that the heterogeneous origin of Klein’s claims was the central reason for the modest variance explained and for needing to remove three items. I now explicitly state that future studies must employ purpose-built indicators to strengthen construct validity. These revisions are included on p. 13 lines 532-534.

References:

Charness, N., Krampe, R. T., Alwin, D. F., & Hofer, S. M. (2008). EXPERTISE AND KNOWLEDGE. In D. F. Alwin & S. M. Hofer (Eds.), *Handbook of Cognitive Aging* (pp. 244-258). SAGE Publications, Incorporated. https://doi.org/10.4135/9781412976589.n15

Green, S. B., & Yang, Y. (2009). Commentary on Coefficient Alpha: A Cautionary Tale. *Psychometrika, 74*(1), 121-135. https://doi.org/10.1007/s11336-008-9098-4

Klein, G. A. (2009). *Streetlights and shadows : searching for the keys to adaptive decision making* (1st ed. ed.). MIT Press.

Sass, D. A., & Schmitt, T. A. (2010). A Comparative Investigation of Rotation Criteria Within Exploratory Factor Analysis. *Multivariate behavioral research, 45*(1), 73-103. https://doi.org/10.1080/00273170903504810

Van Der Heijden, B. (2001). Age and Assessments of Professional Expertise: The Relationship between Higher Level Employees' Age and Self-assessments or Supervisor Ratings of Professional Expertise. *International Journal of Selection and Assessment, 9*(4), 309-324. https://doi.org/10.1111/1468-2389.00183