

DEFENSIBILITY OF GATEKEEPING DECISIONS IN CLINICAL SUPERVISION OF MALAYSIAN COUNSELLOR TRAINEES: A CONCEPTUAL FRAMEWORK

Hadijah Johari¹

Fatahyah Yahya^{1*}

Mohd Suhadi Mohamed Sidik¹

Masyhuri Ahmad¹

Amnah binti Maktar²

Mohd Fahrizal Asmy bin Mohd Ali³

¹ Kulliyah of Counselling, Education and Social Sciences, Universiti Islam Antarabangsa Sultan Abdul Halim Mu'adzam Shah (UniSHAMS), 09300 Kuala Ketil, Kedah, Malaysia. Emails: hadijah@unishams.edu.my; fatahyah.yahya@unishams.edu.my; suhadisidik@unishams.edu.my; masyhuriahmad@unishams.edu.my

² Bandar Tasek Mutiara Health Clinic, 14120 Simpang Ampat, Pulau Pinang, Malaysia.

Email: amnahmaktar@moh.gov.my

³ Department of Human Sciences, Universiti Sultan Azlan Shah (USAS), 33000 Kuala Kangsar, Perak, Malaysia.

Email: fahrizalasmy@usas.edu.my

*Corresponding author: fatahyah.yahya@unishams.edu.my

Article history

Received date : 15-8-2026

Revised date : 16-8-2026

Accepted date : 15-10-2026

Published date : 20-10-2026

To cite this document:

Johari, H., Yahya, F., Mohamed Sidik, M. S., Ahmad, M., Maktar, A., & Mohd Ali, M. F. A. (2026). Defensibility of gatekeeping decisions in clinical supervision of Malaysian counsellor trainees: A conceptual framework. *Journal of Islamic, Social, Economics and Development (JISED)*, 11 (87), 258 - 285

Abstract: *When making gatekeeping decisions, clinical supervision must safeguard clients while being fair towards counsellor trainees. The score for competence and written progression conditions only provide part of the explanation and cannot be the only basis for changing the training decision when evidence is contested. This critical conceptual synthesis serves to compare 48 purposively selected references by their evidential function, context and allowed level of inference. It distinguishes developmental support, measurement and decision justification. There are five conditional vulnerabilities that direct the analysis: unclear indicators, inadequate evidence, unexamined assessor disagreement, remediation without change criteria, reasons not linked to decisions. Malaysian institutional documents qualify blanket claims that safeguards are absent, because they already document ethical pass conditions and conditional independent reassessment. The proposed framework is based on the interlinkage of Standards, Evidence, Verification, Action, and Decision with return pathways for further evidence and review. A fictional case illustrates the possibility of similar overall scores having different evidential states, each requiring different next steps. The contribution is a case-transition account connecting current evidence, warranted inference, authorised response and conditions for reconsideration. The following three propositions concern reviewability, handling of discrepancies and demonstrated improvement following remediation. Due to incomplete search histories and uneven distribution of access to sources, coverage is limited. What the framework offers is an inspectable foundation to ground an*

empirical evaluation in and a foundation for discussing the subject with a supervisor, but not a validated algorithm or evidence of improved decision accuracy.

Keywords: *clinical supervision; counsellor trainees; gatekeeping; decision defensibility; conceptual framework*

Introduction

Clinical supervision is a process of professional education in which learning, assessment and responsibility for the welfare of clients are linked. Supervisors assist trainees in developing their skills. Meanwhile, they check whether the trainee's level of performance is suitable for further training without danger. Some core functions such as competence development, feedback, and the supervisory relationship are highlighted in the literature (Association for Counselor Education and Supervision [ACES], 2011; Bernard & Goodyear, 2019). Additionally, performance monitoring and public protection per professional standards are required (Council for Accreditation of Counseling and Related Educational Programs [CACREP], 2024).

More than a satisfactory total score or confirmation that supervision took place is necessary if it is a consequential training decision. The challenges are how assessors justify a response with some observations missing, inconsistent rating and modification of performance following upon intervention. An unsupported approval as well as an unsupported adverse restriction can matter to clients and trainees. It is a decision justification issue; the present analysis does not estimate the frequency in Malaysia (ACES, 2011; Cook et al., 2015).

In this paper, decision defensibility is the extent to which one can trace a gatekeeping decision, explain, evaluate, and review it on the basis of evidence and reasonable procedures. This definition is organised in five provisional design requirements which are not a validated scale, nor an exhaustive account of defensibility.

The proposed requirements are explicit standards and behavioural indicators; sufficient and relevant evidence; investigation of assessor disagreement; action in proportion to the concern and available evidence; and written reasons and opportunity for review. Their design rationale draws on supervision guidance, validity arguments, and programmatic assessment (ACES, 2011; Cook et al., 2015; van der Vleuten et al., 2012). Lembaga Kaunselor Malaysia (LKM, 2025) identifies the Malaysian ethical reference context. This five-part structure is the proposal made here, not a structure set out or endorsed by any one source.

Defensibility here is not synonymous with whether all assessors agree and is instead whether the inference and response can be examined. But a well-documented judgement might be an unsound one; reviewability, accuracy of the decisions, and fairness are therefore viewed as separate targets for evaluation.

Table 1 clarifies the conceptual definition and operational meaning of decision defensibility, competence, gatekeeping, remediation, non-compensatory rules and due process and review. Operational meanings describe proposed functions; they are not outcome measures that have been validated.

Table 1: Conceptual Definitions and Operational Meanings of the Core Constructs

Construct	Conceptual Definition in This Paper	Operational Meaning within the Framework	Source Basis
Decision defensibility	The extent to which a decision can be traced, explained, evaluated, and reviewed on the basis of evidence and reasonable procedures.	Clear standards; sufficient evidence; investigated disagreement; proportionate action; written reasons and a review mechanism.	Cook et al. (2015); ACES (2011); conceptual synthesis
Competence	Multidimensional professional capability, rather than a single overall score.	Domains and performance indicators are defined and assessed across time and sources.	Chen et al. (2025); Lambie et al. (2018); Arifin et al. (2022)
Gatekeeping	The process of identifying, assessing, and responding to competence concerns that affect trainee progression or client welfare.	Actions are documented, decisions are synthesised collectively, and a review pathway is provided.	Ziomek-Daigle and Christensen (2010); ACES (2011); Motroni Banik et al. (2024)
Remediation	Planned intervention to address specific, potentially remediable deficits.	Targets, time frames, support, evidence of success, and reassessment are specified.	Gaubatz and Vera (2002); ACES (2011)
Non-compensatory rules	Critical deficits are not automatically offset by high performance in other domains.	Considered only after critical indicators, evidence verification, and the consequences of misclassification have been specified and tested.	Meyers (2018); Cook et al. (2015); synthesis in this paper
Due process and review	Transparent and fair procedures for stating reasons, providing an opportunity to respond, and reviewing decisions.	Written reasons, links to evidence, an opportunity to respond, and a review or appeal mechanism.	ACES (2011); CACREP (2024). Local application requires programme-specific confirmation.

Source: The authors' synthesis based on the sources listed in the Source Basis column.

Developmental models are used to explain changes in the needs of trainees for autonomy and support (Stoltenberg & McNeill, 2010), and Gönültaş et al. (2024) discuss supervision and professional development. These sources provide information to developmental support but not gatekeeping decision rules.

The Counseling Competencies Scale-Revised (CCS-R) organises counselling skills and professional dispositions (Lambie et al., 2018). A score provides information for assessment but does not by itself specify a progression response. Accordingly, developmental support, measurement and decision justification remain as separate functions (Cook et al., 2015).

Role conflict, role ambiguity, and supervision anxiety are associated with competence in Malaysian studies (Arifin et al., 2019; Arifin, Abdullah, et al., 2023). Arifin, Zakaria, et al. (2023) explore supervision frequency and competence among trainees and supervisors from

public universities. Such designs do not identify causal links or effectiveness of gatekeeping decision rules.

A comprehensive description of the journey from concern to a reviewed progression decision for a contested Malaysian case is not present in the cited source set, whereas studies of competence and supervision associations are identifiable. This is a bounded observation, which prompts a case level synthesis. It is not a finding that all relevant research in Malaysia has been found or there are no institutional safeguards.

The argument is bounded by three connected gaps. Put together, they refer to the relationship between what is measured, what an institution requires and how a specific decision is justified.

- i. The evidence integration gap is a problem of how developmental support, competence evidence and gatekeeping actions are connected in an explicit account of a supervision decision.
- ii. The decision procedure gap relates to the case-level connection among conflicting assessments, authorised decisions, remediation and review. It is not an absence-of-policy claim, as relevant pass conditions and conditional reassessment provisions already exist in the documents supplied by both Universiti Teknologi Malaysia (UTM, 2019, p. 14) and Universiti Putra Malaysia (UPM, 2025, pp. 8–9).
- iii. The empirical evaluation gap relates to the failure of the mentioned Malaysian research to trace a complete case from concern identification to evidence review to remediation to reassessment to final decision. Written institutional safeguards narrow this gap to the case level application and effect they have, rather than their mere existence.

The legal, qualifications, and ethical context stated in this paper is the Counsellors Act 1998, Malaysian Qualifications Framework (MQF), and code of ethics in Malaysia (LKM, 2025; Malaysia, 1998; Malaysian Qualifications Agency [MQA], 2024). These sources serve different normative functions and none are evidence of uniform programme implementation.

What is key to understand is the difference between a score, an eligibility condition, and an authorised progression judgement. This analysis asks what additional evidence justifies shifting from uncertainty to remediation or reassessment, or to a final judgement. The difference is pertinent due to recent competence and remediation research, but despite this, deterioration in Malaysian practice is not established (Chen et al., 2025; Salpietro et al., 2025, 2026).

The purpose of this paper is to construct a conceptual synthesis and a testable framework to appraise the defensibility of gatekeeping decisions in the clinical supervision of counsellor trainees in Malaysia.

Four research objectives are established:

1. To analyse the conceptual relationships among trainee development, competence assessment, and the gatekeeping function.
2. To determine the limits of inference when international evidence and evidence from related professions are applied to the context of counsellor education in Malaysia.
3. To identify points of vulnerability in the chain connecting standards, evidence, assessor judgement, remediation, and the final decision.
4. To formulate a layered decision framework and propositions that can be tested through empirical research.

Based on these objectives, the paper addresses four research questions:

- i. RQ1: How are trainee development, competence assessment, and gatekeeping connected in terms of their decision-making functions?
- ii. RQ2: Which inferences from international literature and related professions can, and cannot, be transferred to the context of counsellor education in Malaysia?
- iii. RQ3: Which points in the supervision decision chain are conceptually vulnerable to insufficient evidence, assessor disagreement, unclear remediation, or untraceable decisions, based on the reviewed corpus?
- iv. RQ4: Which layers and testable propositions can be formulated for a gatekeeping decision framework intended to support transparency and reviewability?

The contribution is a case-transition account of decision defensibility. Each transition accounts for the evidence currently available, for the inference that the evidence supports, for the authorised response, and for the further evidence/review that could warrant a shift of response. It does not originate validity arguments, several assessment sources or gatekeeping. An analytical value of this is to differentiate cases with similar totals but different evidential and procedural states.

The Malaysian application integrates the programme requirements, academic observation, site observation, remediation responsibilities and review without assuming identical institutional procedures. Table 5 locates the proposal relative to existing approaches, and the following fictional case below Figure 1 illustrates the proposed distinction. Such comparisons delimit a conceptual contribution, not a statement of historical priority or established superiority.

Conceptual Analysis Approach and Evidential Boundaries

It is a critical conceptual synthesis with a focus on explaining relationships, not prevalence or the impact of interventions. The analytical unit is a claim supported by one or more sources and relates to assessment, supervision or progression. The unit of application is intended to be a "case episode," which consists of an evidential state, a proposed response, and the rationale for keeping (or changing) the proposed response. Neither is a nationally sampled trainee or institution. The design rationale is given in the form of explicit source roles, and an inspectable line of reasoning (Jaakkola, 2020).

This paper does not meet the requirements for a systematic review or meta-analysis. The purposive source set allows for conceptual comparison, but not for exhaustive retrieval, saturation, pooled effect or a representative account of the Malaysian programmes. The source-use record for this manuscript version is dated 22 September 2026, and this date was not retroactively assigned for the original searches.

Source Set and Selection Rationale

There are 48 references spanning empirical studies, conceptual works, reviews, books, standards, legislation and institutional documents included in the purposive source set. Sources are chosen for a specified analytical role, not as 48 independent empirical studies. Conceptual design is informed by Jaakkola (2020); additional counselling-model comparisons are provided by Letourneau (2016) and Grunhaus and Lyons (2024).

Previously, source discovery was carried out using nine Elicit exports grouped into four file groups. The previous manuscript reports 203 rows and 171 different exact title strings. These are discovery-record counts, not a verified screening flow or study denominator. The detailed

historical totals and metadata issues will remain as reported at the Supplementary File S1, Part C and will not be re-presented as a new row-level audit in this revision.

The documentation record is incomplete and does not include all original queries, search dates, filters, inclusion decisions, or the order of revisions to categories. The publication year, file names, and workbook dates cannot fill that void. Therefore, sources were kept for their current, explicit conceptual contribution, not expressed as the end point of a repeatable systematic search.

A source is used when its identifiable content informs any one or more of the following: developmental support; rating competence; gatekeeping or remediation; providing justification for use of the assessment; or programme requirements in the context of the Malaysian setting. A methodological source is used only as a justification for conceptual design. Evidence is not transferred across professions, countries, instrument versions or levels of decision making without an explicit boundary. These use rules are set out in Table 2.

A version-specific map of the 48 references, their analytical use, available locator and verification status is available in Supplementary File S1 (V3). It distinguishes checks which were finished in this revision from claims and document anchors which are passed forward from V2. It is not the missing original export workbook or the re-creation of screening decisions that were not recorded.

Table 2: Source Functions and Rules for Use in the Conceptual Synthesis

Source function	Illustrative sources	Use in this article	Inference not permitted
Development and measurement	Stoltenberg & McNeill (2010); Lambie et al. (2018); Arifin et al. (2022)	Distinguish support needs, indicators, and measurement evidence.	A score or stage is not an automatic progression rule.
Empirical experience and association	DeCino et al. (2020); Gonsalvez et al. (2023); Arifin, Zakaria, et al. (2023)	Identify actors, constructs, and possible decision concerns.	No causal effect or Malaysian prevalence inferred across designs.
Conceptual design and comparison	Cook et al. (2015); Jaakkola (2020); Letourneau (2016); Grunhaus & Lyons (2024)	Explain inferential links and delimit the proposed contribution.	A design rationale is not evidence of better decisions.
Normative and institutional material	ACES (2011); CACREP (2024); MQA (2024); UTM (2019); UPM (2025)	Identify written requirements with programme/edition locators.	A written provision is not uniform or effective implementation.
Identity-only and unresolved material	LKM (2025) edition listing; unresolved export records	Identify a reference context; retain provenance limitations in S1.	No unread ethical clause, guessed DOI, or screenshot attribution.

Note: These categories describe evidential roles. A source may inform more than one role; the categories are not a study-count denominator.

Source: Selection and interpretation rules applied to this conceptual article; reference-level use is mapped in Supplementary File S1 (V3).

Source Identity and Inferential Control

Bibliographic identity was considered as a distinct control and a support for a claim was considered separately. A work can be identified by a DOI, export label, or even by a publisher product page, but this does not establish the content of a specific passage. The access actually obtained is recorded on the V3 register. A prior verification note is labelled as inherited rather than counted as an independent recheck.

The earlier export audit reported a mismatch between the title and the DOI for Client Welfare as an Anchor and a record with an unknown status called Narrating Supervision. The export identities are not used as additional corroborating studies, if they have not been resolved. The exported wording is corrupted and is different from the cited Motroni Banik et al. (2024). The final bibliography is not used to infer original search and matching histories.

Works accessible only via an abstract or via catalogue entry are restricted to information available at that level. For instance, the official listing shows the second edition of the LKM code (2025), without giving the clauses. As a result, a specific remediation, appeal or veto rule is not assigned to that listing. Institutional provisions are read as an indication of the programme, edition and page at which it was stated, but not as evidence of current implementation.

Potentially relevant mechanisms or hypotheses were identified through evidence from health professions, social work and other countries. It has not been used to estimate prevalence or effectiveness in Malaysian counsellor education.

It is possible that there may be some selection bias as the original discovery history and access to the full-text are not available. This limits the completeness of the comparison, not the logical distinction between a score and a decision. The inference trail is clearly stated, and readers are able to question a specific premise, a specific source use, or a transition instead of accepting the framework by its labels.

Synthesis Procedures and Analytical Boundaries

There are three operations that are used in the analysis. First, every cited source is given an evidential function, and its current use in the argument (Table 2; S1). This operation differentiates measurement evidence, reports/descriptions of practice, conceptual explanations, and written requirements. It does not treat guidance or a handbook as an empirical outcome study.

Second, a claim is divorced from the interpretation it allows and the design response developed in this work. Worked source to interpretation links are provided in Table 4. The principal checks are whether the construct, actor, context and outcome correspond to the proposed inference, whether a source offers a rationale or an observed result, and whether it provides contrary evidence.

Thirdly, the argument structures the links into Standards, Evidence, Verification, Action, and Decision. Standards indicate the requirement; Evidence provides observations; Verification is an investigation of validity for an inference; Action indicates support, or a justified interim response; and Decision records the authorised progression judgement. These are not themes that have been inferred empirically but theory guided functions. Their partitioning is not

designed to imply that exactly five layers are required or sufficient for revisiting a case, but is intended to expose a variety of different reasons why this might be done.

Each objective and question in Table 3 is linked to the analytical output produced. The reasoning given for functions is recorded in Table 4; institutional evidence that challenges claims of absence of policy is included in Table 6. Case illustration tests internal consistency of the proposed distinctions when the overall score is kept constant and the evidential state changed. It is not participant data, a completed simulation study or a historical coding log but is a fictional analytic demonstration.

These are the operations that characterise the current structure of the argument, a structure that can be reproduced. They do not restore the sequence of source choice or prove that the categories were predetermined. Independent dual screening, formal risk-of-bias scoring, empirical validation or consensus of experts is not claimed. With S1, the unresolved provenance is maintained to be distinct from the current explanation.

The source set is unable to establish Malaysian rates of failure to fail, intervention effectiveness or classification performance. The three proposed tests, which test selected functions of the framework, would not by themselves establish the validity of the entire decision system, even with favourable outcomes.

Table 3: Alignment of Objectives, Questions, Sources, Analysis, and Outputs

Objective / Question	Source Basis	Analytical Operation	Output and Location
RO1 / RQ1 Conceptual relationships	Developmental, competence, and gatekeeping sources: Stoltenberg and McNeill (2010); Lambie et al. (2018); Ziomek-Daigle and Christensen (2010).	Compare the function of developmental support, measurement, and progression decisions; retain distinctions between them.	An account of complementary functions. Literature Review; Table 5; RQ1 discussion and conclusion.
RO2 / RQ2 Limits of transfer	Malaysian studies (Arifin et al., 2022; Arifin, Zakaria, et al., 2023), international gatekeeping research, and related-profession sources.	Contrast source design, country, profession, and outcome before applying an inference to Malaysia.	A bounded evidence comparison. Table 7; Malaysian context section; RQ2 discussion and conclusion.
RO3 / RQ3 Vulnerability points	Supervision guidance, assessment arguments, and discrepancy/remediation literature mapped in Tables 4 and 7.	Compare source-supported inferences; identify conditional failures of a link, not measured frequencies.	Five unranked vulnerabilities. Conceptual Synthesis; Table 8; Figure 1; RQ3 discussion and conclusion.
RO4 / RQ4 Framework and propositions	The relationships identified under RO1–RO3 and the existing approaches compared in Table 5.	Arrange provisional functions, specify return pathways, and formulate observable tests and alternatives.	Five-function case-transition account, fictional illustration, and three propositions. Table 8; Figure 1; Table 9; RQ4 discussion.

Note: RO = research objective; RQ = research question. This matrix documents the argument, not participant data or completed empirical tests. Source-use and access status are recorded in Supplementary File S1 (V3).

Table 4: Illustrative Source-to-Interpretation-to-Framework Reasoning

Source and Locator	Source-Derived Point	Interpretation and Boundary	Proposed Link
Standards S06: ACES (2011), §1.a.i, p. 2.	The supervision contract specifies evaluation criteria and consequences of underperformance.	Translate a requirement into explicit assessment criteria. This is normative guidance, not evidence of implementation.	Standards; vulnerability 1.
Evidence S40: van der Vleuten et al. (2012), abstract/model rationale.	The model links cycles of assessment and support to intermediate and final evaluation.	Use evidence across occasions to inform progression. This transfer does not validate a Malaysian counselling protocol.	Evidence; vulnerability 2.
Verification S34: Tweed and Wilkinson (2019), “Aggregating conflicting information”.	The conceptual discussion considers how conflicting information is synthesised in progression decisions.	Investigate a discrepancy before treating it as resolved. The specific return pathway is an adaptation proposed here.	Verification → Evidence; vulnerability 3.
Action S06: ACES (2011), §9.d.ii, p. 13.	A written remediation plan specifies objectives, requirements, timing, and consequences.	Distinguish a completed plan from new evidence of the target performance. The return link makes reassessment explicit here; it is not claimed as a newly invented practice.	Action → Evidence / Verification; vulnerability 4.
Decision S34: Tweed and Wilkinson (2019), practice and procedure discussion.	Progression deliberation requires sufficient information and attention to the decision process.	Link reasons, evidence, and review in a case record. This supports a design rationale, not classification accuracy.	Decision → Review; vulnerability 5.

Note: Paraphrases are separated from the interpretation proposed here. Source IDs are stable identifiers in S1 (V3). The mapping explains the current argument, not a recovered sequence of original coding.

Literature Review

Theoretical Positioning and Assumptions

The main lens, following Cook et al. (2015), is Kane’s perspective of validity argument, considering that the problem is related to the justification of assessment use. Its scoring, generalisation, extrapolation, and implications inferences find various assumptions between the observation and action. Programmatic assessment brings in the logic of the accumulation of evidence and of support over time (van der Vleuten et al., 2012). Neither point of view establishes local thresholds or the effectiveness of the current scheme.

In this application an observation-to-indicator link concerns scoring; a repeated observation, to represent performance, concerns generalisation; extending that judgement to practice concerns extrapolation; and using it to justify progression or restriction, concerns implications. All of these connections may be subject to Verification. Therefore, the five functions in Figure 1 are not new incarnations of Kane’s four inferences. They structure the work and documentation that allows the investigation of those inferences (Cook et al., 2015; Kinnear et al., 2021).

The level of support is informed by the Integrative Developmental Model (IDM; Stoltenberg & McNeill, 2010). CCS-R offers an assessment language, rather than a theory of progression decisions (Lambie et al., 2018). As a result of this separation developmental need is not treated as evidence of failure, nor is the total on a scale considered to provide authority for decision. Performance is examinable and sufficiency, risk, and remediability are context sensitive judgements. Normative guidance provides reasons to do a procedure; its empirical effects need to be shown separately.

Table 5: Existing Approaches and the Delimited Contribution of This Synthesis

Approach and Sources	Existing Contribution	Organisation Proposed Here	Boundary of Contribution
Validity arguments and validity mapping Cook et al. (2015); Kinnear et al. (2021).	Examine evidence, interpretations, and the justification of consequential uses.	Locate which inference needs support and record what evidence could justify revisiting the response.	Does not replace a validity argument or establish a new theory of validity.
Programmatic assessment van der Vleuten et al. (2012); Schut et al. (2021).	Connect multiple assessment occasions, learning support, and progression decisions.	Distinguish evidence collection, review, an interim response, and final progression in the same case episode.	No claim that multiple evidence sources or collective decisions are new, or automatically accurate.
Development and competence Stoltenberg and McNeill (2010); Lambie et al. (2018).	Explain support needs and provide an organised language for assessing competence.	Keep developmental support and competence evidence distinct while specifying their place in a decision.	No new competence scale, developmental stage, or validated threshold.
Counselling gatekeeping and supervision guidance Ziomek-Daigle and Christensen (2010); ACES (2011); Letourneau (2016); Grunhaus and Lyons (2024).	Address concerns, ethical perspectives, support pathways, and progression or remediation.	Specify the current evidence–inference–response link and the condition for reconsideration, demonstrated in the fictional illustration.	Does not claim earlier models lack review. Added benefit must be tested against the programme’s actual procedure.

Source: Comparison of the stated scope of the cited approaches. Letourneau (2016) and Grunhaus and Lyons (2024) are compared at the level supported by their accessible publisher abstracts; an unreported procedural detail is not coded as absent. The table does not establish priority or superiority.

Letourneau (2016) integrates ethical perspectives in responding to problematic behaviours. Grunhaus and Lyons (2024) make a distinction between recognising, resource connecting and concern referrals in online programmes. The examples demonstrate that combining support with gatekeeping is not itself new. It is not an assertion that previous contributions do not include ethical judgement, assistance or re-evaluation; the proposed contribution is the explicit case-transition record and its testable implications.

From Development to Competence Assessment

Developmental support, feedback and assessment need to be balanced in effective clinical supervision. Developmental models help supervisors adapt the level of structure and autonomy to trainees’ needs (Stoltenberg & McNeill, 2010). Clear goals, observation and feedback are

also central to the practice of professional supervision (ACES, 2011; Bernard & Goodyear, 2019).

Development should not, however, be an excuse for delaying the assessment of significant competence concerns. Supervision research has revealed a range of supervisory goal setting, feedback and assessment practices (Gonsalvez et al., 2017). A review of workplace evidence also indicates the impact of both individual and organisational factors on supervision quality (Rothwell et al., 2021).

Competence is not one undifferentiated total, but rather multiple domains (Chen et al., 2025). Performance assessment is structured by the CCS-R (Lambie et al., 2018), but evidence for an instrument must be distinguished from evidence for a particular decision made using it.

Validity is needed to support the inferences that may be made from the data, not simply to describe the properties of a measurement instrument, say Cook et al. (2015). Validity maps can be used to trace back relationships between evidence, inferences and decisions (Kinnear et al., 2021). Programmatic assessment, in turn, prioritises the triangulation of various forms of evidence over time, recognising the difficulties of implementation and combining judgements (Schut et al., 2021; van der Vleuten et al., 2012).

The extent to which the basis of observation is adequate must be justified in relation to the intended inference in order to make a final competence-based progression judgement. This does not preclude consideration of the need for immediate protection after a single credible incident under current procedures. There are differences between the claims and evidential requirements of a provisional risk response and those for a final judgement of competence.

Assessor Variation, Role Conflict, and Gatekeeping

Gonsalvez et al. (2023) compare the supervisor ratings to trainee self-assessments and compare the discrepancies at different developmental stages in different competence areas. It is not direct evidence of disagreement between the academic and the site supervisors, but is nonetheless useful for a limited discussion of rating differences. The possible behaviour of those supervisors watching another task or situation is a conceptual explanation that needs to be explored, not a finding established by that study.

In the proposed framework, when a mismatch arises, the task, time, assessor point of view, and evidence are all examined prior to reaching a consequential judgement. Aggregation may remain appropriate following contextual review, but the issue is unexamined averaging, not all uses of a mean.

Programmatic assessment and validity mapping offer a justification for taking into account multiple sources and the inferences drawn from them (Kinnear et al., 2021; Schut et al., 2021; van der Vleuten et al., 2012). The suggested protection is to seek behavioural evidence before interpreting anxiety or interpersonal style as incompetence. This is a precautionary design measure, not proven by the present paper, and a decrease in construct substitution is not demonstrated.

In the field of health professions education, Jamieson et al. (2023) discuss positioning the supervisor as teacher, gatekeeper or team member. This points to a potential role-related explanation, but it cannot be said there is any similar pattern in Malaysian counselling.

Counselling gatekeeping studies provide the more direct professional context as it relates to the identification of concern, collecting information, deciding on intervention and progression (Teixeira, 2017; Ziomek-Daigle & Christensen, 2010).

Formal procedures are linked to programme follow-up of trainees who are flagged as presenting concerns (Gaubatz & Vera, 2002). However, the decisions still place an emotional and ethical burden on counsellor educators (DeCino et al., 2020). Motroni Banik et al. (2024) have reinforced the importance of client welfare as an anchor for judgement. Furthermore, the literature of progression decisions strongly suggests that there should be sufficient information and collective deliberation prior to making a decision (Tweed & Wilkinson, 2019).

Having a positive supervisory relationship is thus key. But it cannot take the place of the responsibility to assess.

Failure to Fail, Remediation, and Non-Compensatory Decisions

There is concern in the international literature about the tendency to allow underperforming trainees to continue training. But the strength of the evidence varies across disciplines and countries.

Brear and Dorrian (2010) studied gatekeeping during counselling education in Australia. Yepes-Rios et al. (2016) performed a systematic review which synthesised mechanisms of *failure to fail* in health professions education.

This evidence can be used to make hypotheses related to discomfort of assessors, documentation or organisational pressures. It does not, however, provide prevalence estimates for Malaysian counsellor trainees.

Supervision guidance sets out recommended remediation goals, relevant interventions, a time frame, and documented consequences (ACES, 2011). Hylton et al. (2017) offer a related-profession behaviourally specific remediation policies example (social work); this is a transferable design rationale, not evidence of effectiveness in counselling. Due-process concerns are discussed by Isom (2021) for virtual counselling programmes.

A total of 54 doctoral counsellor education handbooks from the United States that include remediation policies have been analysed by Salpietro et al. (2025). Just one included all ten of the domains examined. Results are about documentary coverage, not programme effectiveness. The set analysed in this study (54 handbooks) excludes handbooks that do not include remediation policies, which further limits possibilities for generalisation.

In a study featuring educator responses to 80 closed cases in remediation within CACREP-accredited Counsellor Education programmes, Salpietro et al. (2026) analysed these responses. The educators indicated in 61.3% of these cases that the outcomes were successful. The percentage is not a causal estimate of the effectiveness of an intervention and the retrospective case reports are not direct prospective observations, and it is not a Malaysian remediation rate. Where there are ethical or safety concerns, the additional question is raised whether a high mark in other domains should make up for a critical deficiency.

Competence scoring models may be compensatory or non-compensatory (Meyers, 2018). But the existence of a mathematical rationale does not establish which domains should carry veto power.

If the stakes for the decisions to be made are significant, the consequences of score use need also to be part of the validity argument (Cook et al., 2015). Non-compensatory rules should thus be regarded as design hypotheses that must be subject to a domain validation, risk indicators and a fair review process. They are not an automatic solution.

The Malaysian Context: Evidence, Standards, and Limits of Inference

What Is Already Known

The cited Malaysian literature discussed is related to supervision development and experience, competence measurement, and supervision related factors. Developmental context is provided by Mohd Jaladin et al. (2012), supervisors' experiences by Ahmad (2016), and psychometric evidence by Arifin et al. (2022). The written training requirements are discussed separately (institutional documents).

The CCS label in Arifin et al. (2022) is carried over and not presumed to refer to the same version as the CCS-R; in Arifin et al. (2019) and Arifin, Abdullah, et al. (2023), role-related or anxiety variables are linked with competence but gatekeeping outcomes are not established.

Arifin, Zakaria, et al. (2023) report associations between supervision frequency and two competence subscales in 204 trainees and 62 lecturer supervisors from nine public universities in Malaysia. These are associations, not evidence that frequency of supervision causes improved gatekeeping decisions. The corresponding sample sizes reported in other Arifin publications must be examined, before they can be considered as independent replications, and sample totals are not added across publications.

Review of Malaysian studies collectively substantiates that supervision and competence have been studied. With these conceptual, correlational and psychometric designs, however, there is not a resolution for how to deal with contradictory evidence, who is to make the decisions, the type of remediation that is provided and how decisions play out over time.

There are various contexts from the legal and qualifications, and ethical sources, such as Malaysia (1998), MQA (2024), and LKM (2025). An ethical reference is identified in the listing of the LKM edition and is not used to evidence an unread clause, nor as if this framework were endorsed.

The institutional comparison identifies the programme, the edition/session and the location associated with each claim, and retains the document-specific findings captured in V2 (Table 6). The associated source map clearly shows which files or passages in the original were not separately reopened for V3. The Universiti Malaya (UM, 2022) placement-registration item is not an entire assessment manual, but a contextual reference. No institutional origin or edition has been authenticated for the unbranded screen shot labelled as USAS.jpg, thus it is excluded from the institutional comparison.

According to the supplied handbooks of the Universiti Utara Malaysia (UUM) undergraduate and post-graduate students, 80% of the weightage is given to the academic assessment and 20% towards the agency, with the combined weightage for competence and ethical-compliance being 15 and 5 percentage points, respectively (UUM, n.d.-a, p. 10; UUM, n.d.-b, §11.0). These are separate programme documents and not indicative of a decrease in ethics requirements over time. The ratio of 90:10, as stated in the two available handbooks in the two sessions (UNIMAS,

n.d., p. 59; 2024, p. 88) is specified by the Universiti Malaysia Sarawak (UNIMAS). These allocations identify scoring contributions, not complete gatekeeping rules.

Trainees must be satisfactorily endorsed by the supervisor and demonstrate appropriate conduct that aligns with counselling ethics as specified in UTM (2019, Part A, §7.0, p. 14). It also specifies that failure can occur due to an unsatisfactory placement endorsement. These conditions coexist with a 100-point domain table, 10 of which are awarded for ethics/professionalism (p. 18). Hence, the document directly qualifies the inference that aggregation necessarily permits ethical shortcomings to be offset. No validated automatic veto is established by the conditional wording.

Under UPM (2025, §§3.4, 3.6–3.8, pp. 7–9), reflection assessment is allocated to a university supervisor and a conditional independent-reassessment option is available. It has a school contribution related to issues outside of counselling practice. The teaching-ethics wording in the shared handbook is not interpreted as a counselling specific rule. Universiti Pendidikan Sultan Idris (UPSI, 2025, §5.4.1, p. 19) refers readers to a separate counselling internship manual, so a missing weighting in the provided guide does not represent a zero site contribution. As a whole, the audit clearly separates between the written scoring, progression conditions, and the decision authority from their unobserved application.

Table 6: Document-Specific Assessment Structures and Decision Provisions

Document and location	Assessment structure	Written decision provision / scope	Limit of inference
UUM undergraduate, n.d.-a D01: pp. 10, 50–51; D05: p. 1	Academic 80%; agency 20%. Combined competence/ethics: 15.	Separate agency form confirms a 20% contribution, not additional marks.	Not an ethics-only weight or a tested veto.
UUM postgraduate, n.d.-b D08; §11.0; academic assessment form (source code BPPA/AK)	Academic 80%; agency 20%. Combined competence/ethics: 5.	Rubric conversion: (raw score / 15) × 5.	Different programme level; filename 2026 is not an authenticated edition.
UNIMAS D02, 2024/2025, p. 88; D06, 2020/2021, p. 59	Academic 90%; site 10% in both supplied sessions.	The counselling section specifies the allocation.	Same written split does not establish unchanged implementation.
UPM (2025) D04; §§3.4, 3.6–3.8; pp. 7–9	Supervisor 60; principal 10; mentor 20; reflection 10. A 70:30 grouping is derived.	School assessment excludes counselling practice; §3.7(a) describes conditional independent reassessment.	Shared teaching- practice/internship (LMI) guide. Teaching- ethics clauses are not assumed to apply identically to counselling.
UTM (2019) D07; Part A §7.0, p. 14; Part D §1.1, p. 18	Seven domains total 100; ethics/professionali sm 10. No assessor split established.	Endorsement and ethical conduct are pass conditions; an unsatisfactory placement endorsement may lead to failure.	Written safeguard, not an automatic validated veto or implementation finding.

Document and location	Assessment structure	Written decision provision / scope	Limit of inference
UPSI (2025) D03; pp. 9, 19, 33	Counselling internship KKP3108 identified; assessor split not established.	§5.4.1 refers to a separate counselling manual; §9.0 refers to assessment materials.	Missing allocation is not 100:0; teaching-specific rules are not transferred.

Note: Allocations are percentage points of the course total. Printed-page and section locators are retained from V2 and mapped in S1, Part B. D01–D08 are document identifiers, not institution counts. LMI = latihan mengajar dan internship (teaching practice and internship); BPPA/AK is retained as the source's form code, not expanded speculatively; KKP3108 is a course code. D09 is excluded as unauthenticated evidence. The table does not establish current policy or implementation.

Limits of Inference and a Defensible Research Gap

The research cited here did not find a complete Malaysian case-decision study done longitudinally. UTM and UPM do however record relevant safeguards (Table 6). The issue that still remains is how programme-specific provision is applied if there is a conflict of evidence, when remediation is considered, or if a decision is disputed. This distinction is important to prevent a limited literature search resulting in the unsupported claim that institutions have no procedures.

Thus there are two claims which cannot be adequately defended. The first is that in Malaysia, there is a particular *failure to fail* rate. The second is that existing scoring systems cause trainees who are unsafe to practise to be passed.

The lack of evidence is not evidence for the lack of a phenomenon. But equally it is not a licence to assume a phenomenon is occurring.

The comparisons between the Malaysian and international material cited are as shown in Table 7. The institutional counterevidence that qualifies claims of policy-absence is reported in Table 6. Neither table establishes implementation fidelity, accuracy of decisions made or national prevalence.

Table 7: Critical Comparison and Evidence Trail within the Cited Source Set

Dimension	Key Supporting Sources	International Evidence	Malaysian Evidence in This Corpus	Implications
Competence	Chen et al. (2025); Lambie et al. (2018); Kinnear et al. (2021); Arifin et al. (2022); Arifin, Zakaria, et al. (2023)	A multidimensional construct; assessment instruments and models are available.	Psychometric and supervision-frequency–competence evidence is cited. No cited Malaysian study tests the proposed gatekeeping decision chain.	Test the function of scores in decisions, not only the instrument.

Dimension	Key Supporting Sources	International Evidence	Malaysian Evidence in This Corpus	Implications
Assessors	Gonsalvez et al. (2017); Gonsalvez et al. (2023); Jamieson et al. (2023); Schut et al. (2021); Tweed and Wilkinson (2019); UPM (2025)	Supervisor–self rating differences and supervisor positioning are discussed. Related-profession arguments support contextual review, not a demonstrated Malaysian academic/site disagreement pattern.	UPM documents conditional independent reassessment in its shared LMI guide. Actual use and outcomes of academic/site discrepancy handling are not established here.	Examine reliability and evidence review protocols.
Gatekeeping	Brear and Dorrian (2010); DeCino et al. (2020); Salpietro et al. (2025); Salpietro et al. (2026); Yepes-Rios et al. (2016)	Processes, decision-making pressures, variation in remediation policies, and case outcomes have been reported outside Malaysia.	The corpus provides no basis for estimating local prevalence.	Track actual cases and outcomes across multiple institutions.
Remediation / ethics	ACES (2011); Hylton et al. (2017); Isom (2021); Motroni Banik et al. (2024); UTM (2019)	Guidance supports explicit goals, behavioural evidence, proportionate action, and review. These are design rationales, not demonstrated intervention effects.	UTM specifies ethical and endorsement-related pass conditions. These written safeguards qualify absence claims but do not establish implementation or rule performance.	Test remediation criteria and critical-domain rules before using them as vetoes.

Source: The authors' synthesis based on the sources identified in the table.

Conceptual Synthesis: From Scores to a Decision Chain

Five conditional vulnerabilities are suggested: vague indicators, evidence insufficient for the intended inference, lack of examination of any disagreement, lack of change criteria in remediation, and decisions detached from reasons. These are not measured failures or independently induced themes. Table 6 qualifies their application: a vulnerability is not an assumed lack of institutional protections, but a vulnerability concerns a link in a specific case.

Table 8: Specifications and Source Trail of the Proposed Layered Decision Framework

Layer	Supporting Sources	Key Question	Minimum Evidence / Action	Intended Safeguard
1. Standards	MQA (2024); CACREP (2024); Chen et al. (2025). LKM (2025) identifies ethical reference context only.	Which competence is being assessed?	Defined domains, indicators, and performance levels.	Avoid vague labels and construct substitution.

Layer	Supporting Sources	Key Question	Minimum Evidence / Action	Intended Safeguard
2. Evidence	Lambie et al. (2018); Cook et al. (2015); Kinnear et al. (2021); Schut et al. (2021)	Is the evidence sufficient?	Observations relevant to the specific inference; repeated or additional evidence where needed.	Avoid an insufficient observation base for the proposed judgement; distinguish interim protection from final progression.
3. Verification	Kinnear et al. (2021); Schut et al. (2021); Tweed and Wilkinson (2019). Rating-discrepancy context: Gonsalvez et al. (2023). Documentary example: UPM (2025, §3.7(a)).	Have discrepancies been investigated, and is the evidence sufficient for the proposed inference?	Investigate context and contradictory evidence; return to Evidence when further observations are needed.	Treat disagreement as a trigger for review.
4. Action Interim response	ACES (2011); Hylton et al. (2017); Isom (2021); Motroni Banik et al. (2024)	What justified interim response, support, or reassessment is required?	Record grounds, authoriser, scope, support, and review timing. Obtain new evidence after remediation; do not prejudge final progression.	Avoid automatically offsetting critical risks.
5. Decision Final progression judgement	ACES (2011); Ziomek-Daigle and Christensen (2010); Tweed and Wilkinson (2019). Written pass conditions: UTM (2019, Part A, §7.0).	What progression judgement is supported now, and what could warrant review?	Document evidence, reasons, unresolved disagreement, decision authority, and review; re-open earlier functions if warranted.	Make reasoning open to scrutiny; do not equate consensus with accuracy.

Source: The authors' synthesis based on the sources identified in the table.

EXISTING FOUNDATIONS: Act 580 | Malaysian Qualifications Framework (MQF) | Code of Ethics | Professional Standards | Developmental Theory

1 STANDARDS

Which competence is being assessed? Domains, indicators, and the consequences of decisions must be clear.

VULNERABILITY

Vague labels or construct substitution

2 EVIDENCE

Is the observation base sufficient for this inference? Identify what is known, uncertain, and still needed.

VULNERABILITY

Insufficient evidence for the proposed inference

3 VERIFICATION

Have discrepancies been investigated, and is the evidence sufficient for the proposed inference?

VULNERABILITY

Disagreement averaged out rather than investigated

↓ ↺ Return to 2: obtain further evidence where gaps remain.

4 ACTION — INTERIM RESPONSE

Record the grounds and authority for support, remediation, or proportionate temporary protection; set reassessment and review.

VULNERABILITY

Remediation without targets or criteria for change

↓ ↺ Return to 2–3: reassess performance after remediation.

5 DECISION — FINAL PROGRESSION

Document the authorised judgement, evidence, reasons, unresolved issues, and conditions for reconsideration.

VULNERABILITY

Decisions cannot be traced to evidence

↓ Review may reopen any material earlier function; document the reason.

ANTICIPATED OUTCOMES TO TEST: More traceable, explainable, and reviewable decisions; no demonstrated accuracy gain.

Figure 1: An Integrated Conceptual Framework for the Defensibility of Gatekeeping Decisions

Note: Grey identifies reference contexts; blue identifies proposed functions; orange identifies conditional vulnerabilities. The outcome box contains expectations for testing. Downward arrows show functional dependencies; return paths identify where further observations, reassessment, or review can change the reasoning. A justified interim protective response can arise before final progression adjudication. No arrow represents a demonstrated causal effect.

Source: Conceptual synthesis of the sources in Table 8. Malaysia (1998), MQA (2024), and LKM (2025) provide reference contexts, not endorsement. Return paths and the separation of interim action from final progression are design proposals.

Action refers to an interim response, such as support provision, a remediation plan or provision within the programme's existing procedures for temporary protection in a proportionate way. It is itself a reasoned, authorised option. Decision represents the actual final progression judgement made for the case episode (that is written down). There is no option for automated sanction for either function and interim action does not prejudge the final outcome.

If information is missing or inconsistent, Verification can go back to Evidence. New performance evidence is then evaluated following Action against the relevant requirement before the final progression judgement. In the case of any material inference, consideration will be given to reopening the matter for review. A programme must specify how an authoriser will be identified, what the evidential bases are, who has rights of response and how reviews will be arranged; none of this can be assumed from an arrow or a colour.

Fictional Case Illustration: A Similar Score, a Different Evidential State

The subsequent illustration is a fiction invented only to try out the internal logic of the framework. It is not a participant event, institutional event or outcome of a simulation study. Suppose a trainee achieves a satisfactory overall course score and receives two discordant ratings for one specified competence indicator and has an incomplete observation record. There is no critical ethical violation proven. It is not possible from the total alone to differentiate between lack of information and a demonstrated competence deficit.

At the first transition, Standards states the indicator that is in question and the progression condition of the programme that applies. Evidence documents observations by each assessor and knowledge gaps. Verification is checking task, timing, context and conflicting observations. The justified first-step response should thus be a written request for clarification or further evidence, rather than a pronouncement of competence or culpability. Any interim protective action requires its own recorded reasons and authorisation under the existing programme procedure.

Imagine that they gained enough supplementary observations to determine a specific shortcoming which can be remediated. The evidential state has changed but the overall course score remains the same. Action now includes the ability to set a behavioural target, support and review date, and evidence for reassessment. The attainment of a later progression judgement will be based on reassessed indicator, plus the programme's other requirements, and not just on attendance in remedial activities. When evidence remains insufficient there remains an element of uncertainty; if evidence is adequate and demonstrates that the target is not being achieved, then there is support for a different, informed re-consideration of the next step.

At every moment of change, four elements are recorded: the existing evidence; the inference that it allows; the authorised response; and the evidence or review that would justify a change to the response. This gives different next steps a justification without varying the overall score. This reflects the analytical distinction of the proposal; it does not demonstrate greater accuracy than an existing programme procedure. Here a comparator may already preserve the same links, which means that no benefits can simply be assumed, but must be demonstrated.

Three propositions make selected expected benefits empirically challengeable (Table 9). P1 deals with reconstruction of reasons, P2 deals with conflicting evidence and P3 deals with targeted performance change. Together they do not encompass all of fairness, authority, appeals, risk classification or client outcomes; these need to be assessed separately.

- i. Proposition 1 — Reviewability: For the same case evidence, records that explicitly link standards, observations, and reasons are expected to enable independent reviewers to reconstruct the evidence-to-decision reasoning more accurately than equally detailed records without explicit links.

- ii. Proposition 2 — Discrepancy handling: For the same discordant ratings and case evidence, adding an explicit evidence-linked discrepancy record to the programme's documented existing procedure is expected to improve detection and justified handling of decision-relevant contradictions. An averaging-only condition may test a narrower mechanism in simulation, but is not assumed to represent usual practice.
- iii. Proposition 3 — Remediation: For comparable remediable deficits, targeted plans with behavioural goals, support, a time frame, and reassessment are expected to produce greater independently observed improvement in the targeted performance than usual support of comparable duration and intensity.

Independent criteria must be able to distinguish between a supported inference, an unresolved uncertainty and acceptable alternative responses. They should be identified prior to the outcome of the framework, and reviewed by those not involved in its design. Completing the five labels is implementation fidelity; reviewer performance, handling of contrary evidence and observed change are outcomes. Reproducing the framework in a criterion key would be circular.

Table 9: Proposed Empirical Tests, Outcomes, and Disconfirming Conditions

Proposition	Illustrative Comparison	Observable Outcome	What Would Challenge the Proposition
P1: Review accuracy	Use matched case records with identical factual content and explicit versus unlinked reasoning; independent reviewers do not receive the study hypothesis.	Accuracy in reconstructing the evidential basis and identifying unsupported inferences against a prespecified case key; time and reviewer agreement as secondary measures.	No prespecified meaningful benefit, or more unsupported inferences, challenges the expectation. The key must not define correctness as using the framework's labels.
P2: Discrepancy handling	Primary: the programme's documented existing procedure versus that procedure plus an evidence-linked discrepancy record. Keep case facts and resources comparable. Optional simulation arm: averaging only, to isolate a mechanism.	Detection and justified handling of contradictions against an independently specified evidence key. Preserve valid minority evidence and reasonable alternative interpretations.	No meaningful gain over the real procedure, greater burden without benefit, or suppression of sound contrary evidence challenges added value. Agreement alone is insufficient.
P3: Remediation	Compare targeted plans with usual support, matching support exposure and accounting for baseline severity and opportunity to practise.	Change in observed targeted performance at prespecified reassessment; independent ratings and sustained improvement where follow-up permits.	No meaningful performance improvement, or an explanation based solely on unequal support or baseline severity, would challenge the proposition. Plan quality is not the outcome.

Note: Designs and measures are proposed, not implemented or validated. Predefine meaningful differences, precision, sample size, analysis, and masking where feasible. Match case information, exposure, resources, and time; document any differences. Retain reasonable alternative judgements in the reference key. Nonsignificance alone does not demonstrate equivalence. P1–P3 assess selected functions, not the entire framework.

Non-compensatory rules are a separate design question, distinct from the outcome of Proposition 3. Initial evaluation should use simulated cases with critical indicators that are explicitly defined. Prior to high-stakes application of its decisions, any rule must be tested for

its false-positive and false-negative decisions, its remediability, its proportionality, and fairness of review. A stricter rule is not presumed to be the more valid rule (Cook et al., 2015; Meyers, 2018).

Matched case records and simulations can be used to explore reasoning while not altering the trainees' progression in initial studies. An averaging-only arm isolates a mechanism, but not general effectiveness; so P2 should use the documented procedure of the participant programme as its main comparator. Review, burden, decision and performance can be examined in later longitudinal work across institutions. It is important to differentiate between interrater agreement and agreement on the correct action (Tweed & Wilkinson, 2019).

The prospective case record should include the requirement, the observations and dates, current evidential state, alternative interpretations, the authoriser and reasons for authorising the response, the evidence or review conditions that would allow this to change. This record is offered as the proposed practical expression of the case-transition account.

Discussion and Implications

For RQ1, the synthesis makes a distinction between the three (often discussed together) tasks of adapting support, assessing competence and authorising progression. Their relationship is significant because the same observation can inform instruction while still not being enough to warrant a final determination of failure. The functional layout locates that inferential boundary; it does not propose that Malaysian programmes today have independent systems that break up those tasks.

The theoretical interpretation is based on the validity-argument perspective. The observations need to be defensibly interpreted, have a sufficient basis for generalisation, and be justified for consequences assigned to them (Cook et al., 2015). IDM informs support and the CCS-R informs observation, but neither fills an unsupported progression inference merely by being present (Stoltenberg & McNeill, 2010; Lambie et al., 2018).

A case-transition record is the unit of application, not a new measurement score. To make a change of response accountable to change in evidence, interpretation or authorised review is its analytical gain. The fictional illustration demonstrates why an unchanged total does not establish an unchanged basis for the next step.

International and related-profession findings for RQ2 indicate mechanisms to investigate, rather than Malaysian rates or a ready-made local decision rule (Brear & Dorrian, 2010; DeCino et al., 2020; Yepes-Rios et al., 2016). Their impact is greatest when the actor, purpose of the assessment and outcome are retained rather than collapsed into a generic claim about competence.

Supervision development, experience, psychometric properties, and associated factors have been studied in Malaysian research (Ahmad, 2016; Arifin et al., 2019, 2022; Arifin, Abdullah, et al., 2023; Arifin, Zakaria, et al., 2023; Mohd Jaladin et al., 2012). The remaining question in this source set has to do with how contested case episodes are handled, and what the consequences are, not with whether supervision research exists.

The Malaysian question is thus not about the presence or not of supervision or written safeguards. It is about how evidence, institutional conditions, review and consequences are

related in real cases. It is a less general formulation that is required by the pass conditions of the UTM and by UPM's independent-reassessment provision (Table 6), and their effectiveness cannot be assumed from the language used.

The five vulnerabilities for RQ3 can be interpreted as failures to justify different links rather than as evidence of a lack of policy. This is important because the proper action alternative is not the same, as obtaining an unobserved performance is not the same as addressing disagreements or reassessing a remediable deficit. The distinction is explained in Table 4 and Table 8, but their frequency and impact have not been measured.

For RQ4, the five-function arrangement and P1–P3 convert this account into contestable design expectations. Agreement with existing validity and programmatic-assessment approaches is deliberate. The contribution is less general than a new general theory, namely, explicit connections between the current evidential state, a justified response and the condition for revisiting this response.

An unresolved inference could be identified through the record for programmes to use, without replacing their procedures of carrying out an authorised assessment system. Letourneau (2016) emphasises a multitude of ethical viewpoints and Grunhaus and Lyons (2024) differentiate between pathways of support and concern. The proposed account does not imply that there is no review in these models, nor that the present approach is better; it simply adds visibility of the evidence to change a response in a case record.

The possibility of erroneous approval and unjustified restriction, opportunity for the trainee to respond, impact of delay, and apparent strictness must be taken into account during evaluation. A documented reason is necessary for scrutiny, but even if it is documented, it may still be incorrect. This is not to say that there is evidence to support a particular critical-domain veto; it is a design and validity consideration (Cook et al., 2015; Meyers, 2018).

In the Malaysian application, the programme-level decision procedures, professional ethics, and qualification expectations are distinguished from each other, and institutions, MQA, and LKM could not be used interchangeably (LKM, 2025; MQA, 2024). It is not said that either body endorses this framework. Each institution would have to check its applicable requirements, decision making authority and review arrangements.

Its purpose is not to remove judgement; rather, the intent of the use is to make the professional judgement open to scrutiny, but not to equate documentation with sound practice.

Alternative Interpretations and Implementation Risks

There are three alternatives that constrain the argument. First, disagreement could simply show that the competence of their performance is different with varying observation contexts, rather than their weak competence. Second, an institutional safeguard could be in place already that covers the proposed vulnerability (UTM and UPM are provided as documentary examples, see Table 6). This framework needs therefore to be tested not against an assumed lack of safeguards, but as a supplement to, or clarification of, existing procedures. Third, collective review can have the potential to increase bias (Tweed & Wilkinson, 2019). There should be no concealment of the contradictory evidence or unresolved disagreement.

Another possibility is that there is a superficial lift which could be the appearance of defensibility through clearer records without commensurate improvement in judgement. Then, there is the issue of administrative burden that may supplant observation or feedback. This paper does not generate negative results; these are plausible alternative explanations to be tested. Reviewers' performance, decision consequences and workload, in addition to document completeness, should be studied in future. This is a limitation of the bounded selection process, as the source register cannot demonstrate that all disconfirming literature was located.

Three outputs specify the contribution: the difference between assessment weightings, eligibility conditions and decision authority; a case-transition record connecting evidence, inference, response and reconsideration; and separate tests of reviewability, discrepancy handling and performance change. Their proposed value may lie in finding out why a decision should be kept provisional, and/or altered, and not for counting boxes on a framework.

One of the virtues of this writing is the disciplined way of dealing with inferential boundaries. Psychometric evidence is not equated with decision evidence. Correlations are not stated as causal relationships. Public documents are also not assumed to be typical of the practice nationwide.

The source-use register will allow for inspection of the present argument, but it will not be able to recover any lost original searches or establish uniform full-text appraisal. Institutional findings are constrained by programme and version, and evidence based on screenshots without authentication is not utilised. Coverage and source verification are limited by these boundaries. These are separate from the untested effectiveness of the conceptual proposal.

Evaluation should start with document-review and/or simulated-case studies which are feasible and should move on to longitudinal, multi-institutional work if applicable. In addition to usability studies, burden of implementation, assessor reliability, trainee fairness, and consequences of decision must also be assessed. This present proposal has not undergone expert validation or empirical effectiveness testing.

Conclusion

The analysis addresses RQ1 by making a distinction between developmental support, competence assessment and authorised progression. It addresses RQ2 by disentangling transferable conceptual observations from Malaysian prevalence and/or implementation claims. The institutional material qualifies the problem, in the sense that written safeguards exist, and their usage in cases of conflict remains an empirical question.

For RQ3, five conditional vulnerabilities are stated that signal potential areas where an observation, interpretation, intervention or progression judgement may lack adequate support. Return pathways and three propositions that test selected anticipated benefits are connected to these functions by the framework for RQ4. It does not validate all the functions nor create a pass/fail rule.

The central contribution is the case-transition account, in which the proposed response is traceable to the available evidence and comes with the condition that would justify its review. This goes beyond naming five functions to specify what should be tied when the evidential state changes. Effectiveness or historical originality is not shown in the fictional illustration; it shows only internal logic.

However, there are some constraints when it comes to the comparison and these are, purposive selection, incomplete discovery histories, uneven access to sources, and document-version variations. Independent source verification and prospective testing are still required. Prior to use for high-stakes purposes, there should be simulations using programme procedures already established as comparators and fairness, burden and consequences examined alongside reviewer performance.

The final claim is thus constrained: there must be an examinable connection between evidence, inference, authorised response and reconsideration in order to make a defensible decision. The relationship is organised in the framework for criticism and testing, but the decision is not made by the framework on behalf of the programme.

Acknowledgements

The authors would like to express their gratefulness for the academic and professional assistance obtained while this conceptual paper was being developed and critically reviewed.

Author Contributions

Hadijah Johari took the lead as author and conceptualised the article, developed the conceptual argument and framework, performed literature synthesis, wrote the original paper, performed the integration of revisions, and led in coordination of the manuscript. The conceptual argument was reviewed by Fatahyah Yahya and Mohd Suhadi Mohamed Sidik who critically examined the manuscript and verified its academic content and overall scholarly coherence. Masyhuri Ahmad, Amnah binti Maktar and Mohd Fahrizal Asmy bin Mohd Ali contributed to reviewing and refining the manuscript in terms of its structure, format, language and grammatical presentation.

Transparency and Research Integrity

This is a conceptual paper that does not incorporate any newly gathered participant data, or identifiable client records. Its case illustration is fictional and only serves the purpose of analysing the internal logic of the proposed framework. Documentary provisions, standards of practice and published research are analysed, but the analysis is not an institutional compliance audit. The manuscript does not assert empirical validation of the proposed framework, improved decision accuracy, or effectiveness. As a conceptual paper no new human-participant research was conducted and no identifiable personal data were collected and thus participant consent was not applicable.

ChatGPT was used to help translate, edit for language and structure, and give suggestions on source-checking. Elicit was used to facilitate earlier source discovery, while available Consensus and SciSpace outputs supported targeted source identification. AI outputs were not considered to be participant data, expert independent review, or empirical evidence or validation. The authors bear the responsibility for source verification, evidence interpretation, accuracy of citations, disclosure of the real tools used and approval of the final version of the text.

Funding and Conflict of Interest

There was no specific grant or other type of sponsorship from any funding agency in the public, commercial or not-for-profit sector for this research. The authors have no financial or non-financial conflict of interest to disclose that might be construed to influence the preparation, interpretation, or publication of this manuscript.

Author Approval and Publication Declaration

All six authors have reviewed and approved the final version of the manuscript and agree to be accountable for their respective contributions. The authors confirm that the manuscript is original, has not been published previously, and is not currently under consideration for publication by another journal.

Similarity Screening

Similarity screening of the manuscript was performed before its submission. The similarity index is less than 20%, in line with the journal's submission requirement.

Source-Use Record

A retrospective source-use/revision record is provided as Supplementary File S1 (V3). It provides a mapping of the 48 references cited in the conceptual synthesis, their respective analytical roles, the references' current verification status and available locators. To make it clear, historical export counts are not presented as reconstructed raw export data, but as information carried over from V2. The supplementary record is not intended to be a reproduction of the original discovery history, screening log or entire source archive. Unresolved provenance and verification issues are preserved transparently, and not inferred or reconstructed retrospectively.

References

- Ahmad, N. (2016). *Meneroka penggunaan model penyeliaan kaunseling dalam kalangan penyelia latihan profesional di universiti awam* [Exploring the use of counselling supervision models among professional training supervisors at public universities] [Doctoral dissertation, Universiti Malaya]. https://knova.um.edu.my/student_works_2010s/3663/
- Arifin, S., Abdullah, S. S., Mat Min, R., Ramley, F., Harun, M. M., & Md Nor, S. B. (2022). Counseling competencies scale (CCS) psychometric properties: Application for trainee counselors in Malaysia. *International Journal of Advanced and Applied Sciences*, 9(11), 144–152. <https://doi.org/10.21833/ijaas.2022.11.018>
- Arifin, S., Abdullah, S. S., Ramley, F., Husain, Z., Mohd Yusof, R. A. A., & Ahmad, R. (2023). The influence of anxiety in supervision toward trainee counselors' competencies in Malaysia. *International Journal of Advanced and Applied Sciences*, 10(4), 88–95. <https://doi.org/10.21833/ijaas.2023.04.010>
- Arifin, S., Mohd Noah, S., Wan Jaafar, W. M., & Zakaria, N. S. (2019). The influence of the role conflict and role ambiguity towards counseling competency in supervision among trainee counselors in Malaysia. *International Journal of Academic Research in Progressive Education and Development*, 8(4), 77–90. <https://doi.org/10.6007/IJARPED/v8-i4/6437>
- Arifin, S., Zakaria, N. S., Wan Jaafar, W. M., Ghazali, N. M., & Sakari, N. S. (2023). The impact of supervision process toward counselling competencies among Malaysian trainee counsellors. *IAFOR Journal of Education*, 11(3), 93–111. <https://doi.org/10.22492/ije.11.3.05>
- Association for Counselor Education and Supervision. (2011). *Best practices in clinical supervision*. <https://acesonline.net/wp-content/uploads/2018/11/ACES-Best-Practices-in-Clinical-Supervision-2011.pdf>
- Bernard, J. M., & Goodyear, R. K. (2019). *Fundamentals of clinical supervision* (6th ed.). Pearson.

- Brear, P., & Dorrian, J. (2010). Gatekeeping or gate slippage? A national survey of counseling educators in Australian undergraduate and postgraduate academic training programs. *Training and Education in Professional Psychology*, 4(4), 264–273. <https://doi.org/10.1037/a0020714>
- Chen, C., Zhang, Y., Guo, Q., Wang, X., & Chen, S. (2025). Core competencies for psychological counselors: A scoping review. *Behavioral Sciences*, 15(2), 147. <https://doi.org/10.3390/bs15020147>
- Cook, D. A., Brydges, R., Ginsburg, S., & Hatala, R. (2015). A contemporary approach to validity arguments: A practical guide to Kane's framework. *Medical Education*, 49(6), 560–575. <https://doi.org/10.1111/medu.12678>
- Council for Accreditation of Counseling and Related Educational Programs. (2024). 2024 CACREP standards. <https://www.cacrep.org/for-programs/2024-cacrep-standards/>
- DeCino, D. A., Waalkes, P. L., & Dalbey, A. (2020). "They stay with you": Counselor educators' emotionally intense gatekeeping experiences. *The Professional Counselor*, 10(4), 548–561. <https://doi.org/10.15241/dad.10.4.548>
- Gaubatz, M. D., & Vera, E. M. (2002). Do formalized gatekeeping procedures increase programs' follow-up with deficient trainees? *Counselor Education and Supervision*, 41(4), 294–305. <https://doi.org/10.1002/j.1556-6978.2002.tb01292.x>
- Gonsalvez, C. J., Riebel, T., Nolan, L. J., Pohlman, S., & Bartik, W. (2023). Supervisor versus self-assessment of trainee competence: Differences across developmental stages and competency domains. *Journal of Clinical Psychology*, 79(12), 2959–2973. <https://doi.org/10.1002/jclp.23590>
- Gonsalvez, C. J., Wahnon, T., & Deane, F. P. (2017). Goal-setting, feedback, and assessment practices reported by Australian clinical supervisors. *Australian Psychologist*, 52(1), 21–30. <https://doi.org/10.1111/ap.12175>
- Gönültaş, M., Meydan, B., & Sağkal, A. S. (2024). The examination of the counseling supervision effect on personal and professional development of first-time supervisees. *International Journal of Psychology and Educational Studies*, 11(4), 342–356. <https://doi.org/10.52380/ijpes.2024.11.4.1375>
- Grunhaus, C. M., & Lyons, M. (2024). A model of retention and remediation for online counselor education. *Teaching and Supervision in Counseling*, 6(1), Article 5. <https://doi.org/10.7290/tsc06qkja>
- Hylton, M. E., Manit, J., & Messick-Svare, G. (2017). Gatekeeping and competency-based education: Developing behaviorally specific remediation policies. *Journal of Teaching in Social Work*, 37(3), 249–259. <https://doi.org/10.1080/08841233.2017.1313359>
- Isom, E. (2021). Remediation practices and considerations for virtual counseling programs: An emphasis on due process. *The Advocate*, 26(2). <https://doi.org/10.4148/2637-4552.1157>
- Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, 10(1–2), 18–26. <https://doi.org/10.1007/s13162-020-00161-0>
- Jamieson, J., Gibson, S., Hay, M., & Palermo, C. (2023). Teacher, gatekeeper, or team member: Supervisor positioning in programmatic assessment. *Advances in Health Sciences Education*, 28(3), 827–845. <https://doi.org/10.1007/s10459-022-10193-9>
- Kinnear, B., Kelleher, M., May, B., Sall, D., Schauer, D. P., Schumacher, D. J., & Warm, E. J. (2021). Constructing a validity map for a workplace-based assessment system: Cross-walking Messick and Kane. *Academic Medicine*, 96(7S), S64–S69. <https://doi.org/10.1097/ACM.00000000000004112>

- Lambie, G. W., Mullen, P. R., Swank, J. M., & Blount, A. (2018). The Counseling Competencies Scale: Validation and refinement. *Measurement and Evaluation in Counseling and Development*, 51(1), 1–15. <https://doi.org/10.1080/07481756.2017.1358964>
- Lembaga Kaunselor Malaysia. (2025). *Kod Etika Kaunselor* [Counsellor code of ethics] (2nd ed.). <https://lembagakaunselor.kpwkm.gov.my/product/kod-etika-kaunselor/>
- Letourneau, J. L. H. (2016). A decision-making model for addressing problematic behaviors in counseling students. *Counseling and Values*, 61(2), 206–222. <https://doi.org/10.1002/cvj.12038>
- Malaysia. (1998). *Akta Kaunselor 1998 (Akta 580)* [Counsellors Act 1998 (Act 580)]. <https://www.kpwkm.gov.my/portal-main/documents?type=akta>
- Malaysian Qualifications Agency. (2024). *Malaysian Qualifications Framework* (2nd ed.). <https://www.mqa.gov.my/new/mqf.cfm>
- Meyers, J. L. (2018). Scoring models in competency-based educational assessment. *The Journal of Competency-Based Education*, 3(3), e01173. <https://doi.org/10.1002/cbe2.1173>
- Mohd Jaladin, R. A., Amit, N., & Sumari, M. (2012). Counseling supervision for counselor development in Malaysia. *IAMURE International Journal of Social Sciences*, 3(1), 1–21. <https://doi.org/10.7718/ijss.v3i1.26>
- Motroni Banik, J. L., Coule, K. F., Bradley, N., & Rizzo, J. M. (2024). Client welfare as an anchor: Lived experiences of gatekeepers in counselor education. *Journal of Counselor Preparation and Supervision*, 18(2). <https://doi.org/10.70013/pzfx2vw9>
- Rothwell, C., Kehoe, A., Farook, S. F., & Illing, J. (2021). Enablers and barriers to effective clinical supervision in the workplace: A rapid evidence review. *BMJ Open*, 11(9), e052929. <https://doi.org/10.1136/bmjopen-2021-052929>
- Salpietro, L., Clark, M., & Frederick, J. (2026). Problems of professional competence, remediation interventions, and outcomes in counselor education: A case-level study. *Counselor Education and Supervision*. Advance online publication. <https://doi.org/10.1002/ceas.70035>
- Salpietro, L., Clark, M., & Zacarias, R. (2025). A content analysis of doctoral-level counselor education remediation policies. *Teaching and Supervision in Counseling*, 7(3), Article 1. <https://doi.org/10.7290/tsc0740fd>
- Schut, S., Maggio, L. A., Heeneman, S., van Tartwijk, J., van der Vleuten, C. P. M., & Driessen, E. (2021). Where the rubber meets the road: An integrative review of programmatic assessment in health care professions education. *Perspectives on Medical Education*, 10, 6–13. <https://doi.org/10.1007/s40037-020-00625-w>
- Stoltenberg, C. D., & McNeill, B. W. (2010). *IDM supervision: An integrative developmental model for supervising counselors and therapists* (3rd ed.). Routledge. <https://doi.org/10.4324/9780203893388>
- Teixeira, V. B. (2017). Gatekeeping practices of counselor educators in master's level counseling programs. *Journal of Counselor Preparation and Supervision*, 9(1). <https://doi.org/10.7729/91.1177>
- Tweed, M., & Wilkinson, T. (2019). Student progress decision-making in programmatic assessment: Can we extrapolate from clinical decision-making and jury decision-making? *BMC Medical Education*, 19, Article 176. <https://doi.org/10.1186/s12909-019-1583-1>

- Universiti Malaya. (2022, December 20). *Garis panduan pendaftaran tempat PLA4010 – Latihan profesional* [Guidelines for placement registration for PIA4010 – Professional training]. <https://education.um.edu.my/garis-panduan-pendaftaran-tempat-pia4010-latihan-profesional>
- Universiti Malaysia Sarawak. (n.d.). Buku panduan pelajar Fakulti Sains Kognitif dan Pembangunan Manusia sesi 2020/2021 [Faculty of Cognitive Sciences and Human Development student handbook, 2020/2021 session].
- Universiti Malaysia Sarawak. (2024). *Buku panduan pelajar FSKPM sesi 2024/2025* [FSKPM student handbook for the 2024/2025 academic session]. <https://www.fcshd.unimas.my/index.php/my-unimas/student>
- Universiti Pendidikan Sultan Idris. (2025). *Garis panduan latihan mengajar / internship kaunseling UPSI edisi 2025* [UPSI guidelines for teaching practice / counselling internship, 2025 edition]. Pusat Latihan Mengajar dan Industri.
- Universiti Putra Malaysia. (2025). *Buku panduan penyelia & guru pembimbing: Latihan mengajar dan internship* [Handbook for supervisors and mentor teachers: Teaching practice and internship]. Fakulti Pengajian Pendidikan. <https://educ.upm.edu.my/documentfile?L=bm>
- Universiti Teknologi Malaysia. (2019). *Garis Panduan Latihan Praktikum dan Internship Kaunseling: Program Sarjana Pendidikan (Bimbingan dan Kaunseling)* [Guidelines for counselling practicum and internship training: Master of Education (Guidance and Counselling) programme] (1st ed.).
- Universiti Utara Malaysia. (n.d.-a). *Buku Panduan Internship Program Sarjana Muda Kaunseling* [Internship handbook for the Bachelor of Counselling programme].
- Universiti Utara Malaysia. (n.d.-b). *Buku panduan internship Program Sarjana Sains Kaunseling (SSKX5996)* [Internship handbook for the Master of Science in Counselling programme]. Pusat Pengajian Psikologi Gunaan, Dasar dan Kerja Sosial.
- van der Vleuten, C. P. M., Schuwirth, L. W. T., Driessen, E. W., Dijkstra, J., Tigelaar, D., Baartman, L. K. J., & van Tartwijk, J. (2012). A model for programmatic assessment fit for purpose. *Medical Teacher*, 34(3), 205–214. <https://doi.org/10.3109/0142159X.2012.652239>
- Yepes-Rios, M., Dudek, N., Duboyce, R., Curtis, J., Allard, R. J., & Varpio, L. (2016). The failure to fail underperforming trainees in health professions education: A BEME systematic review: BEME Guide No. 42. *Medical Teacher*, 38(11), 1092–1099. <https://doi.org/10.1080/0142159X.2016.1215414>
- Ziomek-Daigle, J., & Christensen, T. M. (2010). An emergent theory of gatekeeping practices in counselor education. *Journal of Counseling & Development*, 88(4), 407–415. <https://doi.org/10.1002/j.1556-6678.2010.tb00040.x>