Tạp chí Khoa học Công nghệ Xây dựng, ĐHXDHN, 2026, 20 (4V): 86–99

XÂY DỰNG KHUNG CHỈ SỐ ĐÁNH GIÁ KẾT QUẢ THỰC HIỆN AN TOÀN TRÊN CÔNG TRƯỜNG XÂY DỰNG BẰNG KỸ THUẬT DELPHI MỜ

Nguyễn Tuấn Ngọc Tú[a,*], Trần Quang Dũng[b], Lê Hồng Hà[b], Hồ Ngọc Khoa[b]

[a]*Cục Giám định nhà nước về chất lượng công trình xây dựng, Bộ Xây dựng, số 80 Trần Hưng Đạo, Phường Cửa Nam, Quận Hoàn Kiếm, Hà Nội, Việt Nam*
[b]*Khoa Xây dựng Dân dụng và Công nghiệp, Trường Đại học Xây dựng Hà Nội, 55 Giải Phóng, phường Bạch Mai, Hà Nội, Việt Nam*

Nhận ngày 16/8/2026, Sửa xong 2/10/2026, Chấp nhận đăng 8/10/2026

Tóm tắt

Đo lường là điều kiện để chuyển quản lý an toàn công trường từ ứng phó sang phòng ngừa. Tại Việt Nam, đánh giá thực hiện an toàn lao động (ATLĐ) chủ yếu dựa vào chỉ báo muộn và chưa được chuẩn hóa; dẫn đến kết quả không so sánh được giữa các công trường. Nghiên cứu này đề xuất một khung chỉ số ba cấp gồm 9 khía cạnh, 28 tiêu chí và 95 chỉ báo, xây dựng trên lý thuyết quản lý rủi ro an toàn, lý thuyết nguyên nhân tai nạn và hướng dẫn OECD. Dựa trên tiêu chí sàng lọc về tính trùng lặp nội dung, tính khả thi và sự phù hợp với bối cảnh thi công dân dụng và công nghiệp tại Việt Nam, thảo luận nhóm tập trung (focus group) đã rút gọn tập chỉ báo xuống còn 67 chỉ báo. Tiếp theo, kỹ thuật Delphi mờ với 22 chuyên gia thuộc năm nhóm chủ thể dự án đã được triển khai. Kết quả nhận diện được 26 chỉ báo quan trọng và 41 chỉ báo có mức quan trọng thấp với độ nhạy ổn định. Bài báo cung cấp bằng chứng định lượng về mức ưu tiên của chuyên gia đối với chỉ báo an toàn trong bối cảnh Việt Nam.

Từ khoá: an toàn lao động xây dựng; khung chỉ số; chỉ báo kết quả thực hiện an toàn; chỉ báo sớm; chỉ báo muộn; Delphi mờ.

DEVELOPMENT OF A SAFETY PERFORMANCE ASSESSMENT INDICATOR FRAMEWORK ON CONSTRUCTION SITES USING FUZZY DELPHI METHOD

Abstract

Measuring safety performance is a prerequisite for shifting construction site safety management from reactive response to proactive prevention. In Vietnam, occupational safety performance evaluation relies primarily on lagging indicators and lacks standardization, leading to non-comparable results across project sites. This study proposes a three-tier framework comprising 9 dimensions, 28 criteria, and 95 indicators, grounded in safety risk management theory, accident causation theory, and OECD guidelines. Based on screening criteria covering content overlap, feasibility, and contextual relevance to civil and industrial construction in Vietnam, a focus group study refined the set to 67 indicators. Subsequently, a Fuzzy Delphi method involving 22 experts across five project stakeholder groups was conducted. The results identified 26 critical indicators and 41 indicators of low importance with stable sensitivity. This paper provides quantitative evidence regarding expert prioritization of safety indicators within the Vietnamese context.

Keywords: construction safety; multi-criteria framework; safety performance indicators; leading indicators; lagging indicators; fuzzy Delphi.

https://doi.org/10.31814/stce.huce2026-20(4V)-07 © 2026 Trường Đại học Xây dựng Hà Nội (ĐHXDHN)

1. Giới thiệu

Đo lường là cấu phần kiểm tra trong chu trình quản lý an toàn theo mô hình PDCA (Plan - Do - Check - Act): không có đo lường thì không có căn cứ khẳng định rủi ro mất an toàn trên công trường đã được kiểm soát, cũng không nhận diện được dấu hiệu cảnh báo sớm. Bối cảnh Việt Nam đặt ra hai vấn đề song song. Thứ nhất là mức độ nghiêm trọng và tính dai dẳng của tai nạn lao động. Trong sáu

*Tác giả đại diện. *Địa chỉ e-mail:* tuantucgd@gmail.com (Tú, N. T. N.)

năm 2020–2025, xây dựng luôn thuộc nhóm ba lĩnh vực có số vụ tai nạn lao động chết người cao nhất, chiếm tỷ trọng từ 9,73% đến 18,27% tổng số vụ [1, 2]; năm 2025 số vụ giảm 15,47% nhưng số người bị thương nặng tăng 5,27% và số người chết ở khu vực lao động không theo hợp đồng tăng 4,9%, cho thấy mức độ nghiêm trọng của từng vụ chưa giảm tương ứng. Thứ hai là năng lực đo lường: tỷ lệ doanh nghiệp gửi báo cáo tai nạn lao động định kỳ tăng từ 5,47% năm 2020 lên 17,52% năm 2025, tức 82,48% không báo cáo [1, 2], nên tần suất tai nạn ở phạm vi toàn quốc hoặc địa phương không thể tính toán một cách tin cậy, khiến các chỉ báo muộn dựa trên thống kê tai nạn thiếu độ tin cậy.

Trên thế giới, hướng nghiên cứu chủ đạo trong hai thập kỷ gần đây là chuyển trọng tâm từ chỉ báo muộn sang chỉ báo sớm [3, 4]. Việc áp dụng nguyên trạng các khung chỉ báo quốc tế vào Việt Nam gặp hai hạn chế. Các yếu tố tác động đến kết quả an toàn thay đổi theo môi trường pháp lý, kinh tế, xã hội và trình độ công nghệ nơi hoạt động thi công diễn ra [5]. Bên cạnh đó, chỉ báo an toàn có thể bị lạm dụng khi trở thành mục tiêu báo cáo thay vì công cụ cải tiến [6].

Từ đó hình thành khoảng trống nghiên cứu: chưa có khung chỉ số được chuẩn hóa, có cấu trúc phân cấp rõ ràng và được kiểm chứng bởi các chủ thể trực tiếp quản lý an toàn lao động tại công trường. Nghiên cứu hướng tới ba mục tiêu: xác lập cơ sở lý luận và đối sánh các khung hiện có; đề xuất khung chỉ số ba cấp phù hợp điều kiện thi công trong nước; và sàng lọc tập chỉ báo bằng khảo sát chuyên gia. Đóng góp gồm cơ chế gắn mỗi chỉ báo với một cặp vai trò nhập liệu và phê duyệt, và bằng chứng định lượng về mức ưu tiên cùng mức đồng thuận của chuyên gia trong bối cảnh Việt Nam.

Đối tượng nghiên cứu chủ đạo là kết quả thực hiện an toàn ở cấp tổ chức và cấp dự án, tức mức độ mà hệ thống quản lý an toàn của công trường đạt được các đầu ra dự kiến, đo bằng tổ hợp chỉ báo sớm và chỉ báo muộn. Phạm vi bài báo giới hạn ở vòng khảo sát Delphi thứ nhất; việc xác định trọng số và kiểm định khung chỉ số sẽ được thực hiện trong nghiên cứu tiếp theo.

2. Cơ sở khoa học

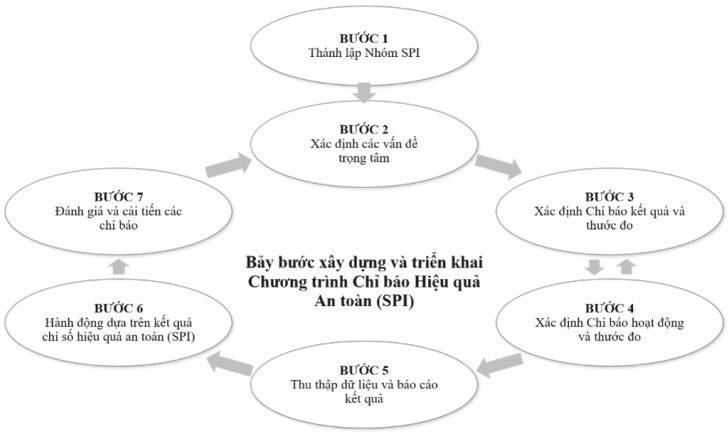
2.1. Khái niệm chỉ báo sớm, chỉ báo muộn

Chỉ báo kết quả thực hiện an toàn được phân thành hai nhóm theo quan hệ thời gian với sự kiện tai nạn. Chỉ báo muộn đo hậu quả đã xảy ra, điển hình là tần suất tai nạn có mất ngày công. Chỉ báo sớm đo các hoạt động phòng ngừa thực hiện trước khi sự cố phát sinh, ví dụ tỷ lệ hoàn thành kế hoạch huấn luyện hoặc số báo cáo sự cố cận nguy [3, 4].

Bài báo dùng ba thuật ngữ theo quy ước sau. Chỉ số là giá trị tổng hợp phản ánh tổng thể kết quả thực hiện an toàn lao động trên công trường. Tiêu chí là nội dung cần đo lường, đánh giá thông qua một hay một nhóm chỉ báo. Chỉ báo sớm và chỉ báo muộn là hai loại chỉ báo phân theo quan hệ thời gian với sự kiện tai nạn, không phải giá trị tổng hợp. Hạn chế của chỉ báo muộn gồm bốn khía cạnh. Thứ nhất, về thống kê, tai nạn nghiêm trọng là sự kiện ít phổ biến nên tần suất trên một dự án đơn lẻ có phương sai lớn, không đủ độ nhạy để phát hiện thay đổi trong hệ thống quản lý [7]. Thứ hai, về hành vi, gắn chỉ báo muộn với đánh giá thành tích tạo động cơ giấu báo cáo, làm sai lệch dữ liệu quản lý [6]. Thứ ba, về lý thuyết, sự cố là kết quả của chuỗi sai lệch tích lũy trong hệ thống tổ chức và bởi vậy, năng lực duy trì hoạt động an toàn mới là yếu tố cần đo lường, đánh giá hơn là đếm số lần thất bại [8]; và thứ tư, quan hệ giữa chỉ báo sớm với kết quả an toàn có độ trễ, không tuyến tính, nên bộ chỉ báo cần được đo ở tần suất đủ dày [9].

2.2. Khung chỉ báo của OECD

Hướng dẫn của Tổ chức Hợp tác và Phát triển Kinh tế (OECD) về chỉ báo kết quả thực hiện an toàn dành cho doanh nghiệp [10] là cơ sở phương pháp luận cho khung đề xuất. Về quy trình triển khai, OECD đề xuất một chu trình bảy bước có tính lặp, tóm lược ở Hình 1 [10]; trình tự này cho thấy chỉ báo không phải danh mục cố định mà là sản phẩm của vòng rà soát định kỳ.



Hình 1. Bảy bước xây dựng và triển khai chương trình chỉ báo hiệu quả an toàn [10]

2.3. Các khung chỉ báo khác và thực tiễn doanh nghiệp

Ngoài khung OECD, nghiên cứu đối sánh thêm ba tài liệu hướng dẫn quốc tế: hướng dẫn của Cơ quan An toàn và Sức khỏe Vương quốc Anh (HSE) với khái niệm cặp chỉ báo kiểm tra chéo cho mỗi rào chắn rủi ro quan trọng [11], tổng quan của Viện Y tế Công cộng và Môi trường Hà Lan (RIVM) với bộ tiêu chí đánh giá chất lượng chỉ báo [12], và bộ thước đo của Trung tâm An toàn Quy trình Hóa chất (CCPS) phân tầng chỉ báo theo bốn cấp độ nghiêm trọng của sự kiện [13]. Cả bốn khung đều không quy định chủ thể nhập liệu và phê duyệt cho từng chỉ báo [6, 14]. Khung đề xuất khác biệt ở: mỗi chỉ báo gắn một cặp vai trò nhập liệu - phê duyệt, tách biệt bắt buộc với chỉ báo rủi ro cao; và cấu trúc ba cấp cho phép đối chiếu đánh giá cấp nhóm với cấp chỉ báo.

Bảng 1 so sánh khung đề xuất với bốn khung tiền lệ theo bốn trục cấu trúc.

Bảng 1. So sánh cấu trúc khung đề xuất với các khung tiền lệ

Khung	Số cấp phân tầng	Phạm vi khía cạnh	Cơ chế trọng số	Phân công trách nhiệm dữ liệu
OECD No. 19 (2008) [10]	2 cấp (chỉ báo kết quả / chỉ báo hoạt động)	An toàn quy trình công nghiệp; không bao phủ chuỗi nhà thầu phụ và an toàn trong thiết kế	Không quy định; tổ chức tự đặt ưu tiên	Không quy định
HSE (2006) [11], RIVM [12], CCPS (2022) [13]	2 cấp, phân tầng theo mức nghiêm trọng sự kiện	Rào chắn rủi ro kỹ thuật cố định	Không quy định	Không quy định
Tổng quan chỉ báo sớm trong xây dựng [4]	Danh mục phẳng, không phân cấp	Xây dựng, chủ yếu chỉ báo cấp công trường	Xếp hạng theo tần suất xuất hiện	Không quy định

Khung	Số cấp phân tầng	Phạm vi khía cạnh	Cơ chế trọng số	Phân công trách nhiệm dữ liệu
Khung đánh giá quản lý ATLĐ trong nước [15]	2 cấp (nhóm yếu tố / yếu tố)	Quản lý ATLĐ cấp công trường, phạm vi một địa phương	Trọng số từ khảo sát, không xét tính nhất quán	Không quy định
Khung đề xuất của nghiên cứu này	3 cấp (9 khía cạnh / 28 tiêu chí / 95 chỉ báo)	Bao phủ cả chuỗi cung ứng và an toàn trong thiết kế, tiền xây dựng	Delphi mờ để sàng lọc	Mỗi chỉ báo gắn một cặp vai trò nhập liệu-phê duyệt

Hai nhánh nghiên cứu gần đây bổ trợ cho đối sánh này là nhận thức của người đứng đầu dự án đối với kiểm soát rủi ro nghề nghiệp [16] và khung đánh giá quản lý an toàn lao động tại các công trường trong nước [15].

2.4. Đặc thù điều kiện thi công và quản lý an toàn lao động tại Việt Nam

Tại Việt Nam, dữ liệu thường phân mảnh, chưa tồn tại cơ sở dữ liệu quốc gia về tai nạn và hành vi không an toàn ở cấp công trường. Ban an toàn lao động đặt dưới Ban chỉ huy công trường nên thiếu vị thế độc lập khi báo cáo vi phạm liên quan đến tiến độ. Tỷ trọng lao động không có hợp đồng ở mức cao đòi hỏi một chỉ báo riêng theo Nghị định số 145/2020/NĐ-CP [17]. Mức độ số hóa thấp khiến việc thu thập chỉ báo phải thực hiện trực tiếp tại hiện trường và qua kiểm tra hồ sơ.

3. Phương pháp nghiên cứu và khung chỉ số đề xuất

3.1. Quy trình chung

Nghiên cứu triển khai theo quy trình ở Hình 2, trong đó bước một và bước ba mang tính lý thuyết, bước hai và bước bốn mang tính thực nghiệm. Phạm vi bài báo dừng ở kết quả vòng thứ nhất.



Hình 2. Quy trình nghiên cứu bốn bước

3.2. Lựa chọn kỹ thuật Delphi mờ

Kỹ thuật Delphi mờ, phát triển từ lý thuyết tập mờ [18] và các biến thể tích hợp mờ của Delphi cổ điển [19, 20], được chọn cho giai đoạn sàng lọc vì hai lý do: đánh giá của chuyên gia mang bản chất ngôn ngữ với ranh giới giữa các bậc không rõ ràng, nên biểu diễn bằng số mờ giữ lại phần thông tin bị mất khi quy đổi thẳng sang số thực; và phương pháp bảo tồn ý kiến thiểu số qua cận dưới, cận trên của số mờ tam giác thay vì triệt tiêu bằng trung bình.

3.3. Công thức Delphi mờ

Đánh giá của chuyên gia i đối với chỉ báo j được biểu diễn bằng số mờ tam giác, chuẩn hóa về đoạn đơn vị, theo quy trình Delphi mờ [20, 21]:

$$\tilde{A}_{ij} = (l_{ij}, m_{ij}, u_{ij}); \quad 0 \le l_{ij} \le m_{ij} \le u_{ij} \le 1 \tag{1}$$

trong đó l là cận dưới; m là giá trị điển hình; u là cận trên của đánh giá; i là chỉ số chuyên gia; j là chỉ số chỉ báo. Đánh giá mờ của nhóm chuyên gia đối với chỉ báo j được tổng hợp theo trung bình từng thành phần, trong đó K là số chuyên gia trả lời hợp lệ đối với chỉ báo j:

$$\tilde{A}_j = \left(\frac{1}{K} \sum l_{ij}, \frac{1}{K} \sum m_{ij}, \frac{1}{K} \sum u_{ij} \right) \tag{2}$$

Khoảng cách d giữa đánh giá cá nhân và đánh giá nhóm được đo bằng khoảng cách đỉnh theo phương pháp Vertex [20]:

$$d(\tilde{A}_{ij}, \tilde{A}_j) = \sqrt{\frac{1}{3}\left[(l_{ij} - l_j)^2 + (m_{ij} - m_j)^2 + (u_{ij} - u_j)^2 \right]} \tag{3}$$

Tỷ lệ chuyên gia đạt ngưỡng khoảng cách, dùng làm thước đo mức đồng thuận đối với chỉ báo j, được xác định theo:

$$C_j = \frac{1}{K} \sum I[d(\tilde{A}_{ij}, \tilde{A}_j) \le \delta] \tag{4}$$

trong đó C_j là tỷ lệ chuyên gia đạt ngưỡng khoảng cách đối với chỉ báo j; I là hàm chỉ báo; δ là ngưỡng khoảng cách được khóa trước khi phân tích, lấy bằng 0,20 theo giá trị thông dụng trong các nghiên cứu Delphi mờ [20, 21]. Điểm giải mờ của chỉ báo j được tính theo phương pháp trọng tâm:

$$S_j = \frac{l_j + m_j + u_j}{3} \tag{5}$$

Chỉ báo được giữ lại khi thỏa mãn đồng thời $C_j \ge 0{,}75$ và $S_j \ge \alpha$, cặp điều kiện được khuyến nghị trong hướng dẫn thực hành Delphi mờ [21]. Giá trị α không ấn định theo một nghiên cứu đơn lẻ mà được khảo sát bằng phân tích độ nhạy trên dải 0,50 đến 0,70 (Mục 4.3, Hình 5).

3.4. Thang đo, ánh xạ sang số mờ tam giác và quy đổi thang đo

Mức quan trọng của từng chỉ báo được đánh giá trên thang Likert 7 bậc, từ 1 là rất thấp đến 7 là rất cao. Thang 7 bậc được chọn vì độ phân giải cao hơn giúp phân biệt tốt hơn giữa các chỉ báo trong một tập lớn khi phần lớn đánh giá tập trung ở nửa trên của thang, và vì so với thang 5 bậc, thang 7 bậc cho trung bình đã hiệu chỉnh cao hơn và độ nhọn thấp hơn, tức phân bố đáp án gần chuẩn hơn [22].

3.5. Thống kê mô tả bổ trợ và kiểm định đồng thuận

Tính tương đương của hai đợt thu thập được kiểm định trước khi gộp dữ liệu. Điểm trung bình toàn bảng của từng chuyên gia đạt 4,556 ở đợt quy đổi và 4,714 ở đợt thu trực tiếp, kiểm định Mann-Whitney cho $U = 41{,}5$ và $p = 0{,}459$; ở cấp từng chỉ báo, sau hiệu chỉnh Benjamini-Hochberg với tỷ lệ phát hiện sai 5% không còn khác biệt có ý nghĩa. Hai đợt tương đương về mức trung bình nhưng chưa tương đương về trật tự ưu tiên, với tương quan hạng $\rho = 0{,}176$ và $p = 0{,}156$; vì đợt thu trực tiếp chỉ gồm 7 phiếu, đây được đọc là hệ quả của cỡ mẫu hơn là bằng chứng về sai lệch do quy đổi, và là một lý do độc lập để chỉ phát biểu T2 ở mức sàng lọc sơ bộ.

Mỗi kiểm định trả lời một câu hỏi riêng, chọn theo bản chất dữ liệu là thứ bậc, cỡ mẫu nhỏ, không giả định phân phối chuẩn. Mann-Whitney U xét khả năng gộp hai đợt thu thập. Tương quan hạng Spearman đo mức ổn định của trật tự ưu tiên chỉ báo giữa các kịch bản phân tích (Mục 4.4). Hệ số hòa hợp Kendall đo mức hội tụ của 22 chuyên gia trên toàn bộ tập chỉ báo. Kruskal-Wallis dò sai lệch hệ thống giữa năm nhóm chủ thể. Bootstrap và jackknife đo độ ổn định của quyết định giữ hay loại từng chỉ báo.

Bootstrap lấy mẫu lại có hoàn lại từ 22 chuyên gia, giữ nguyên vectơ điểm của từng phiếu để bảo toàn tương quan nội tại giữa các chỉ báo, với $B = 2.000$ lần lặp và hạt giống ngẫu nhiên cố định; tiêu chí sàng lọc được áp lại từ đầu ở mỗi lần lặp, và tần suất một chỉ báo được giữ trên 2.000 lần lặp là đại lượng báo cáo. Jackknife bỏ-một thực hiện 22 lần trên 21 chuyên gia. Hai mức bootstrap 0,80 và 0,60 là các mức tiền định dùng để phân loại kết quả.

Mức độ hòa hợp giữa các chuyên gia trên toàn bộ tập chỉ báo được đánh giá bằng hệ số hòa hợp Kendall, xác định theo công thức (6), trong đó W là hệ số hòa hợp Kendall; m là số chuyên gia; n là số chỉ báo được xếp hạng; R_j là tổng hạng của chỉ báo j; $\bar{R}$ là trung bình các tổng hạng; T_i là số hiệu chỉnh cho các hạng bằng nhau của chuyên gia i:

$$W = \frac{12 \sum (R_j - \bar{R})^2}{m^2(n^3 - n) - m \sum T_i} \tag{6}$$

Ý nghĩa thống kê của W được kiểm định bằng đại lượng chi bình phương bằng $m(n-1)W$ với $n-1$ bậc tự do. Hệ số alpha Cronbach chỉ dùng để đánh giá tính nhất quán nội tại của thang đo; hệ số này không đo mức đồng thuận giữa các chuyên gia và không được dùng cho mục đích đó ở bất kỳ đâu trong bài báo. Chỉ báo được giữ lại theo tiêu chí thống kê khi thỏa mãn đồng thời trung bình từ 4,75 trở lên và hệ số biến thiên không vượt quá 0,30. Cả hai được khóa trước khi mở khảo sát. Ngưỡng 4,75 là ảnh của mức 3,50 trên thang Likert 5 bậc, mức tối thiểu thông dụng để giữ một mục đánh giá trong nghiên cứu Delphi, khi ánh xạ tuyến tính sang thang 7 bậc: $1 + (3{,}50 - 1) \times (7 - 1)/(5 - 1) = 4{,}75$, tương ứng 62,5% dải đo. Ngưỡng 0,30 đặt chặt hơn mức 0,50 thường dùng làm ranh giới đồng thuận, nhằm chỉ giữ các chỉ báo phân tán ý kiến thấp; độ nhạy của kết quả với cặp ngưỡng này được khảo sát ở Mục 4.3.

3.6. Chọn mẫu chuyên gia

Chuyên gia được chọn theo chủ đích với hai tiêu chí: kinh nghiệm nghiên cứu hoặc thực hành trong lĩnh vực xây dựng và an toàn lao động từ 5 năm trở lên, hoặc dưới 5 năm nhưng đang trực tiếp phụ trách công tác an toàn tại công trường và đã tham gia từ ba dự án trở lên; và giữ vai trò thuộc một trong năm nhóm chủ thể chính của dự án, gồm chủ đầu tư; tư vấn giám sát và quản lý dự án; nhà thầu chính; cơ quan quản lý nhà nước về xây dựng và lao động; cơ sở đào tạo và nghiên cứu. Cỡ mẫu mục tiêu là 20 đến 30 chuyên gia. Delphi mờ không suy luận từ mẫu ngẫu nhiên nên cỡ mẫu xác định theo cấu trúc hội đồng, không bằng tính toán năng lực thống kê: hội đồng phải phủ đủ năm nhóm nêu trên với ít nhất hai người mỗi nhóm, trong đó nhà thầu chính và tư vấn giám sát chiếm tỷ trọng lớn hơn, nên cận dưới ở khoảng 20 người; cận trên 30 người phản ánh giới hạn thực tế của một vòng Delphi có kiểm soát, khi hội đồng lớn hơn làm chậm vòng phản hồi và tăng bỏ cuộc mà ít cải thiện độ ổn định của kết quả. Độ ổn định của kết quả sàng lọc vì vậy được kiểm tra hậu định bằng bootstrap và jackknife (Mục 3.5 và Mục 4.4). Chuyên gia tham gia tự nguyện, được thông báo trước về mục đích nghiên cứu và phạm vi sử dụng dữ liệu; phiếu trả lời được xử lý ẩn danh và chỉ báo cáo ở dạng tổng hợp.

3.7. Nguyên tắc xây dựng và cấu trúc phân cấp của khung

Khung được xây dựng theo ba nguyên tắc: bám cấu trúc phân tầng chỉ báo kết quả và chỉ báo hoạt động của OECD; mỗi chỉ báo có định nghĩa vận hành đầy đủ gồm công thức tính, đơn vị, chiều tốt/xấu, tần suất đo, nguồn dữ liệu, bằng chứng kiểm chứng và một cặp vai trò nhập liệu - phê duyệt xác định (Mục 3.9); và khung có thể tùy biến theo loại hình dự án, giai đoạn phát triển văn hóa an toàn và chủ thể áp dụng. Khung phân vai cho năm chủ thể: chủ đầu tư hoặc ban quản lý dự án; tư vấn giám sát và tư vấn quản lý dự án; nhà thầu chính, với bộ phận an toàn tách khỏi bộ phận thi công; nhà thầu phụ; và tư vấn thiết kế đối với khía cạnh I. Cơ quan quản lý nhà nước chỉ tiếp nhận kết quả. Vai trò nhập liệu thuộc chủ thể trực tiếp tạo ra dữ liệu gốc, vai trò phê duyệt thuộc chủ thể có thẩm quyền kiểm chứng bằng chứng (Mục 3.9). Bảng 2 trình bày cấu trúc khung đề xuất (tập T0).

Bảng 2. Cấu trúc khung đề xuất (T0): 9 khía cạnh, 28 tiêu chí (TC), 95 chỉ báo (CB)

Mã	Khía cạnh	Số TC	Số CB	Căn cứ
A	Cam kết và sự tham gia của lãnh đạo	3	12	[10, 11]
B	Quản lý rủi ro và lập kế hoạch	4	14	[10, 11, 13]
C	Điều kiện làm việc và môi trường	4	14	[4]
D	Năng lực và hành vi lực lượng lao động	6	17	[4, 17]
E	Sự tham gia của người lao động	3	9	[8, 10, 14]
F	Quản lý chuỗi cung ứng	3	9	[5]
G	Giám sát và đo lường hiệu quả	2	12	[6, 10, 13]
H	Học hỏi và cải tiến liên tục	2	7	[10, 12]
I	An toàn trong thiết kế và giai đoạn tiền xây dựng	1	1	[4, 15]
	Tổng	28	95	

Định nghĩa vận hành của từng chỉ báo được công bố cùng dữ liệu và mã nguồn phân tích (Mục 4.2).

3.8. Rút gọn tập chỉ báo phục vụ khảo sát và định nghĩa các tập T0–T3

Đưa toàn bộ khung đề xuất T0 gồm 95 chỉ báo vào bảng hỏi sẽ vượt quá khối lượng câu hỏi hợp lý của một vòng khảo sát, nên khung T0 được thu hẹp thành tập khảo sát Vòng 1 (T1) gồm 67 chỉ báo (Hình 3). Mọi đại lượng thống kê ở Mục 4 được tính trên 67 chỉ báo của tập T1.

Việc rút gọn từ T0 xuống T1 dùng phương pháp thảo luận nhóm chuyên gia (focus group) với bốn chuyên gia. Nhóm rà soát toàn bộ 95 chỉ báo của T0 theo năm tiêu chí sàng lọc cơ sở nêu dưới đây; một chỉ báo chỉ bị loại khi cả bốn chuyên gia cùng thống nhất rằng nó không đạt ít nhất một tiêu chí, trường hợp còn ý kiến khác nhau được giữ lại để đưa vào khảo sát. Thảo luận loại 28 chỉ báo và giữ 67 chỉ báo cho T1. Bốn chuyên gia này đồng thời là thành viên hội đồng Delphi mờ ở Vòng 1, nên bước rút gọn và bước đánh giá không hoàn toàn độc lập. Trong 67 chỉ báo của T1 có hai cặp trùng nội dung do sai sót khi triển khai bảng hỏi: F06 được hỏi lại dưới mã I01 và F07 dưới mã I03. Dữ liệu được giữ nguyên; hai cặp lặp này được dùng hậu kiểm để ước lượng độ tin cậy trong cùng người đánh giá (Mục 4.4).

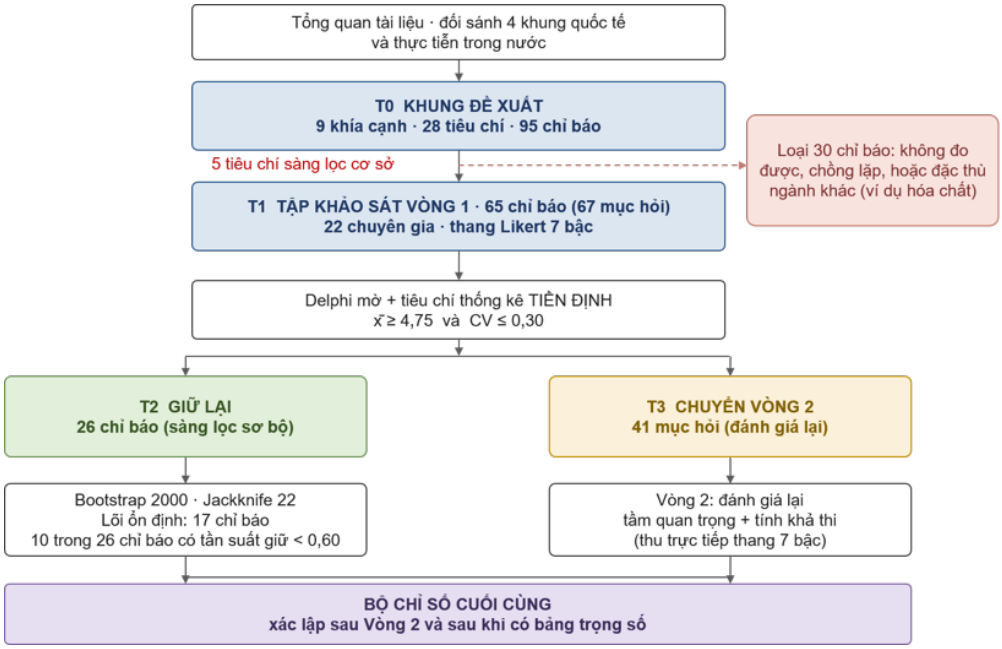
Kết quả khảo sát Vòng 1 với 22 chuyên gia phân T1 thành T2 gồm 26 chỉ báo vượt ngưỡng thống kê và T3 gồm 41 chỉ báo cần được đánh giá lại ở Vòng 2. T2 chưa phải bộ chỉ báo cuối cùng, còn T3 không bị loại mà được giữ lại để tiếp tục xem xét trước khi xác lập bộ chỉ báo và trọng số chính thức.

Năm tiêu chí sàng lọc cơ sở gồm: đo lường được bằng dữ liệu thu thập tại công trường, loại các chỉ báo mô tả trạng thái nhận thức không có thang đo vận hành; khả thi thu thập, không đòi hỏi thiết bị

hoặc phần mềm chuyên biệt chưa phổ biến; không chồng lấp; phù hợp bối cảnh thi công dân dụng và công nghiệp, loại các chỉ báo đặc thù dầu khí hoặc hóa chất; và có giá trị phân biệt giữa công trường quản lý tốt và quản lý kém, loại các chỉ báo đạt giá trị gần tuyệt đối ở mọi dự án.

3.9. Gắn chỉ báo với trách nhiệm dữ liệu

Điểm khác biệt của khung so với các danh mục tiêu chí đang dùng trong nước là mỗi chỉ báo được gắn đồng thời chủ thể nhập liệu và chủ thể phê duyệt, hai vai trò không trùng nhau đối với chỉ báo rủi ro cao; chẳng hạn chỉ báo tỷ lệ tuần tra an toàn của lãnh đạo do nhà thầu chính nhập liệu và chủ đầu tư phê duyệt.

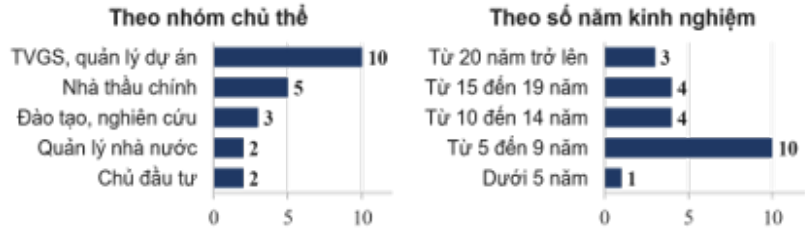


Hình 3. Sơ đồ quá trình sàng lọc chỉ báo qua các tập (T0 đến T3)

4. Kết quả khảo sát chuyên gia và thảo luận

4.1. Hồ sơ chuyên gia

Vòng khảo sát thứ nhất thu được 22 phiếu hợp lệ, không có chỉ báo nào khuyết dữ liệu; mẫu bao gồm cả năm nhóm chủ thể nhưng phân bố không cân bằng giữa các nhóm và các mức kinh nghiệm (Hình 4): nhiều nhất là nhóm tư vấn giám sát (TVGS) và quản lý dự án, phần lớn có từ 5 đến 9 năm kinh nghiệm, một người dưới 5 năm thuộc diện tiêu chí thay thế nêu ở Mục 3.6. Có 21 trong 22 người đã tham gia từ 3 dự án trở lên. Cơ cấu nghiêng về TVGS và nhà thầu với 68% tổng số phiếu là hạn chế được xem xét ở Mục 4.4 và 4.7.



Hình 4. Hồ sơ 22 chuyên gia tham gia khảo sát vòng thứ nhất

4.2. Kết quả ở cấp chỉ báo

Mức quan trọng trung bình của 67 chỉ báo, tính trên thang 7 bậc sau khi quy đổi 15 phiếu thang 5 bậc theo quy tắc tiền định ở Mục 3.4, dao động từ 3,64 đến 5,45 với trung bình chung 4,61; không có chỉ báo nào đạt mức 6,00 trở lên. Theo cặp ngưỡng tiền định, 26 chỉ báo được giữ và tạo thành tập T2 (xem Bảng 3). Bốn chỉ báo cao nhất là C05, A06, D01 và C06; bốn chỉ báo thấp nhất là E06, G06, G04 và G07.

Bảng 3. Danh mục 26 chỉ báo thuộc tập T2

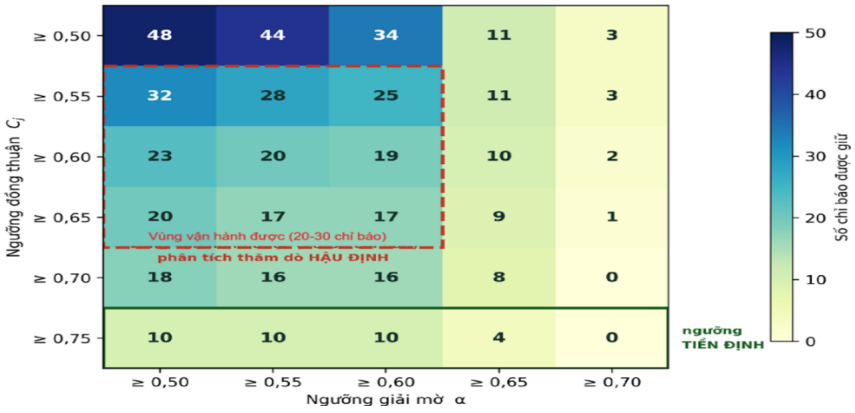
Mã	Khía cạnh	Nội dung chỉ báo	$\bar{x}$; CV
A04	A	Tỷ lệ giải ngân ngân sách an toàn so với kế hoạch	$\bar{x}$ = 4,77;CV = 0,258
A05	A	Tỷ lệ nhân sự chuyên trách trên tổng lao động	$\bar{x}$ = 5,05;CV = 0,270
A06	A	Tỷ lệ hoàn thành kế hoạch huấn luyện ATLĐ	$\bar{x}$ = 5,32;CV = 0,275
A07	A	Tỷ lệ quản lý cấp trung/cao có chỉ số đánh giá hiệu quả an toàn lao động	$\bar{x}$ = 4,91;CV = 0,259
A08	A	Tỷ lệ công việc rủi ro cao được lãnh đạo cấp cao xem xét	$\bar{x}$ = 4,95;CV = 0,282
B01	B	Tỷ lệ công việc rủi ro cao có lập Đánh giá rủi ro/an toàn công việc	$\bar{x}$ = 4,77;CV = 0,250
B05	B	Tỷ lệ tuân thủ giấy phép làm việc	$\bar{x}$ = 4,91;CV = 0,280
B06	B	Tỷ lệ biện pháp kiểm soát áp dụng đúng tại hiện trường	$\bar{x}$ = 4,82;CV = 0,269
B07	B	Tỷ lệ phân tích công việc an toàn/giấy phép làm việc phổ biến cho tổ đội trước khi làm việc	$\bar{x}$ = 4,86;CV = 0,264
B08	B	Tỷ lệ hoàn thành kế hoạch diễn tập ứng phó khẩn cấp	$\bar{x}$ = 4,86;CV = 0,293
B09	B	Tỷ lệ trang thiết bị ứng phó khẩn cấp sẵn sàng hoạt động	$\bar{x}$ = 4,82;CV = 0,261
B10	B	Tỷ lệ nhân sự chủ chốt được huấn luyện ứng phó khẩn cấp	$\bar{x}$ = 5,00;CV = 0,269
C02	C	Tỷ lệ tuân thủ về giao thông & lối đi bộ	$\bar{x}$ = 5,09;CV = 0,284
C03	C	Tỷ lệ tuân thủ hệ thống bảo vệ chống rơi ngã	$\bar{x}$ = 5,18;CV = 0,194
C05	C	Tỷ lệ khu vực nguy hiểm có biển báo/rào chắn	$\bar{x}$ = 5,45;CV = 0,194
C06	C	Tỷ lệ thiết bị cần kiểm định có hồ sơ hợp lệ	$\bar{x}$ = 5,23;CV = 0,228
C07	C	Tỷ lệ tuân thủ ATLĐ hệ thống điện tạm	$\bar{x}$ = 5,14;CV = 0,235
D01	D	Tỷ lệ tuân thủ huấn luyện bắt buộc theo quy định	$\bar{x}$ = 5,27;CV = 0,263
D03	D	Tỷ lệ hành vi an toàn quan sát được	$\bar{x}$ = 4,86;CV = 0,248
D04	D	Tỷ lệ tuân thủ thiết bị bảo hộ cá nhân	$\bar{x}$ = 4,77;CV = 0,266
F03	F	Tỷ lệ Nhà thầu phụ đạt đánh giá ATLĐ trước khi ký hợp đồng	$\bar{x}$ = 4,86;CV = 0,193
F04	F	Tỷ lệ hoàn thành kế hoạch thanh tra/kiểm tra Nhà thầu chính với Nhà thầu phụ	$\bar{x}$ = 4,77;CV = 0,214
F05	F	Tỷ lệ tham dự họp điều phối ATLĐ chung (Nhà thầu chính/phụ)	$\bar{x}$ = 4,77;CV = 0,296
F06	F	Tỷ lệ thiết kế chính được rà soát yêu cầu ATLĐ	$\bar{x}$ = 4,82;CV = 0,209
G01	G	Tỷ lệ hoàn thành kế hoạch thanh tra/giám sát chủ động	$\bar{x}$ = 5,05;CV = 0,225
H03	H	Số bài học kinh nghiệm được phổ biến	$\bar{x}$ = 4,77;CV = 0,241

4.3. Sàng lọc và phân tích độ nhạy của ngưỡng

Mục này phân định tường minh hai loại ngưỡng. Cặp ngưỡng tiền định dùng để sàng lọc là trung bình $\geq 4{,}75$ và hệ số biến thiên $\leq 0{,}30$, cùng với ngưỡng khoảng cách $\delta = 0{,}20$ và ngưỡng đồng thuận $C_j \geq 0{,}75$ của quy trình Delphi mờ [20, 21]; toàn bộ các giá trị này được xác định và khóa trước khi mở khảo sát. Ma trận độ nhạy theo cặp ngưỡng ở Hình 5, trừ hàng $C_j \geq 0{,}75$, là phân tích thăm dò hậu định, thực hiện sau khi đã có dữ liệu và không có hiệu lực sàng lọc.

Theo tiêu chí thống kê tiền định, 26 trong 67 chỉ báo của T1 được giữ, tạo thành T2. Theo tiêu chí Delphi mờ tiền định với $C_j \geq 0{,}75$ và $\alpha \geq 0{,}60$, chỉ 10 chỉ báo được giữ; ở mức $\alpha = 0{,}70$ không còn chỉ báo nào. Như vậy ngưỡng $C_j = 0{,}75$ loại 16 trên 26 chỉ báo khi áp cho hội đồng 22 chuyên gia thuộc năm nhóm chủ thể.

Ma trận cho thấy kết quả sàng lọc nhạy với cả hai ngưỡng, rõ nhất ở bước nhảy giữa $\alpha = 0{,}60$ và $\alpha = 0{,}65$, nơi số chỉ báo được giữ giảm 47% đến 68% tùy mức C_j (Hình 5). Vùng cho tập giữ ở quy mô vận hành được, khoảng 20 đến 30 chỉ báo, gồm các cặp C_j từ 0,60 đến 0,65 kết hợp α từ 0,50 đến 0,55; trong đó cặp $C_j \geq 0{,}60$ và $\alpha \geq 0{,}55$ trùng lặp nhiều nhất với tiêu chí thống kê, giữ 20 chỉ báo và 19 trong số đó cũng được tiêu chí thống kê giữ. Đây là quan sát hậu định trên cùng một tập dữ liệu và chủ yếu đến từ trục α: điểm giải mờ S_j gần như là biến đổi đơn điệu của điểm trung bình, tương quan hạng 0,998, nên chỉ trục C_j mang thông tin độc lập với tiêu chí thống kê.



Hình 5. Bản đồ nhiệt số chỉ báo được giữ theo cặp ngưỡng

4.4. Mức độ đồng thuận và độ nhạy của kết quả theo cỡ mẫu

Hệ số hòa hợp Kendall trên toàn bộ 67 chỉ báo đạt $W = 0{,}081$ ($\chi^2 = 117{,}32$; $df = 66$; $p < 0{,}001$); ở cấp nhóm đạt $W = 0{,}097$ ($\chi^2 = 17{,}10$; $df = 8$; $p = 0{,}029$). Cả hai có ý nghĩa thống kê nhưng độ lớn rất thấp, thuộc mức hòa hợp yếu; do 22 người xếp hạng trên nhiều bậc tự do khiến p luôn nhỏ, kết quả được diễn giải theo độ lớn W, không theo giá trị p.

Bootstrap 2.000 lần giữ 5 chỉ báo với tần suất $\geq 0{,}80$ và 16 chỉ báo với tần suất $\geq 0{,}60$; 10 chỉ báo thuộc T2 có tần suất dưới 0,60. Jackknife 22 lần cho 37 chỉ báo từng được giữ ít nhất một lần, nhưng chỉ 17 chỉ báo giữ ở cả 22 lần (20 chỉ báo đổi trạng thái). Lõi ổn định 17 chỉ báo này hẹp hơn hẳn tập T2 (26 chỉ báo): ranh giới của T2 nhạy với cỡ mẫu ở mức không thể bỏ qua.

Sai lệch giữa nhóm chủ thể và độ tin cậy trong cùng người đánh giá cũng được kiểm tra: Kruskal-Wallis cho $H = 4{,}443$ ($df = 4$; $p = 0{,}349$; $\varepsilon^2 = 0{,}026$), với năng lực phát hiện chênh lệch giữa các nhóm không vượt quá 11,2%. Hệ số tương quan nội lớp cho hai cặp lặp (F06-I01: 0,301; F07-I03: 0,432; trung bình 0,366) thuộc mức kém [23], sai số đo tiêu chuẩn 0,94–1,02 bậc.

Cặp F06-I01 minh họa trực tiếp hệ quả của sai số đo: F06 ($\bar{x} = 4{,}82$) vượt ngưỡng và thuộc T2, còn bản lặp I01 ($\bar{x} = 4{,}50$) bị loại, dù khoảng cách 0,32 bậc nhỏ hơn sai số đo tiêu chuẩn 0,94 bậc —

phản ánh một ngưỡng tuyệt đối cắt qua vùng nhiễu đo lường hơn là khác biệt thực chất. Ở cấp trung bình 22 chuyên gia, sai số chuẩn còn 0,21 bậc, nhưng 32/67 chỉ báo (48%) nằm trong dải 4,75 ± 0,21 quanh ngưỡng, đây là lý do lõi ổn định jackknife chỉ gồm 17 chỉ báo. Vòng 2 sẽ chủ đích đưa 2–3 cặp chỉ báo lặp vào bảng hỏi kèm định nghĩa vận hành ngắn gọn.

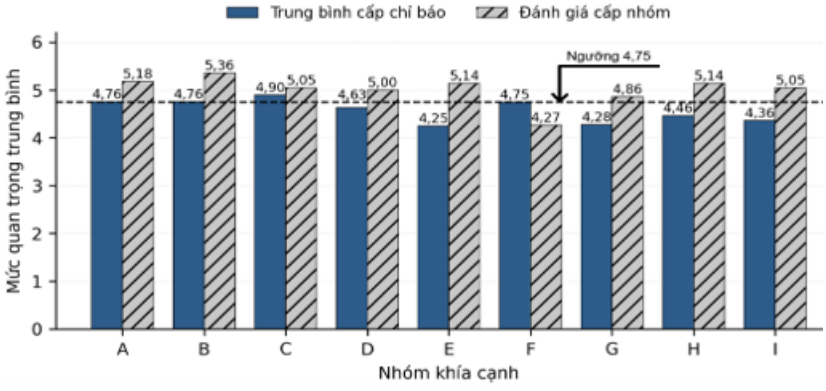
Kết quả sàng lọc còn được kiểm tra độ nhạy với quy tắc quy đổi thang đo (Mục 3.4) qua ba kịch bản: phân bổ tỷ lệ cố định (kết quả chính), quy đổi tuyến tính, và chỉ 7 phiếu thu trực tiếp trên thang 7 bậc. Trật tự ưu tiên gần như không đổi giữa hai kịch bản đầu (Spearman $\rho = 0{,}987$; $p < 0{,}001$, cùng 22 phiếu gộp — hệ số cao vì áp trên cùng tập phiếu; kiểm tra độc lập giữa hai đợt thu thập cho hệ số thấp hơn, Mục 3.5). Ngược lại, điểm tuyệt đối dịch chuyển có hệ thống: quy tắc tuyến tính cho trung bình thấp hơn 0,305 điểm, đủ để số chỉ báo vượt ngưỡng 4,75 giảm từ 26 xuống 6; kịch bản thứ ba (7 phiếu) chỉ là kiểm tra hướng. Phép quy đổi vì vậy bảo toàn trật tự ưu tiên nhưng không bảo toàn vị trí tuyệt đối, nên số 26 chỉ báo luôn cần kèm điều kiện ngưỡng và quy đổi.

4.5. Kết quả ở cấp nhóm

Bảng 4. Tổng hợp kết quả theo 9 nhóm khía cạnh

Mã	Khía cạnh	Số chỉ báo	TB cấp chỉ báo	Đánh giá cấp nhóm	Chênh lệch	Giữ theo thống kê
A	Cam kết và sự tham gia của lãnh đạo	8	4,76	5,18	+0,42	5
B	Quản lý rủi ro và lập kế hoạch	10	4,76	5,36	+0,60	7
C	Điều kiện làm việc và môi trường	9	4,90	5,05	+0,15	5
D	Năng lực và hành vi lực lượng lao động	9	4,63	5,00	+0,37	3
E	Sự tham gia của người lao động	7	4,25	5,14	+0,89	0
F	Quản lý chuỗi cung ứng	7	4,75	4,27	−0,48	4
G	Giám sát và đo lường hiệu quả	8	4,28	4,86	+0,58	1
H	Học hỏi và cải tiến liên tục	6	4,46	5,14	+0,68	1
I	An toàn trong thiết kế và giai đoạn tiền xây dựng	3	4,36	5,05	+0,69	0
	Tổng	67	4,61	5,01	+0,40	26

Bảng 4 cho thấy 8 trong 9 nhóm có đánh giá cấp nhóm cao hơn trung bình cấp chỉ báo, lớn nhất ở nhóm E; nhóm duy nhất có chênh lệch âm là F (Hình 6, chi tiết ở Mục 4.7).



Hình 6. So sánh mức quan trọng trung bình ở cấp nhóm và cấp chỉ báo

4.6. Ý kiến mở của chuyên gia

Phần câu hỏi mở tập trung vào ba chủ đề: nguồn lực tài chính, khi chi phí an toàn không tách riêng khỏi chi phí chung nên dễ bị điều chuyển; vị thế của bộ máy an toàn, khi cán bộ an toàn thường chỉ kiêm nhiệm trên hồ sơ; và tính trung thực của dữ liệu, khi nhà thầu có xu hướng che giấu sự cố. Hai chủ đề đầu liên quan trực tiếp đến chỉ báo A04 và A05; chủ đề thứ ba là căn cứ thực tiễn cho nguyên tắc tách vai trò ở Mục 3.9 (xem Mục 4.7(b) và 4.7(c)).

4.7. Thảo luận

a. Độ tin cậy dữ liệu là ràng buộc chi phối, biểu hiện ở hiện tượng báo cáo thiếu

Phân bố kết quả theo nhóm cho thấy một trật tự ưu tiên rõ: nhóm C về điều kiện làm việc cùng nhóm A và nhóm B dẫn đầu ở cấp chỉ báo, trong khi nhóm E và nhóm I không giữ được chỉ báo nào, nhóm G và nhóm H mỗi nhóm chỉ giữ một (Bảng 4). Ba nhóm sau đòi hỏi hệ thống dữ liệu và quy trình phân tích mà các công trường trong nước chưa vận hành; trật tự này phản ánh mức độ trưởng thành của thực tiễn quản lý, không phản ánh giá trị khoa học của các chỉ báo [4]. Ràng buộc chi phối phía sau trật tự đó là độ tin cậy của dữ liệu: ý kiến mở của chuyên gia (Mục 4.6) quy về cùng hiện tượng, nhà thầu có xu hướng che giấu sự cố nên số liệu sự cố và sự cố cận nguy bị báo cáo thiếu.

b. Tách vai trò nhập liệu khỏi vai trò phê duyệt

Phần lớn chỉ báo sớm trong khung được tính từ dữ liệu do chính đơn vị bị đánh giá tự ghi nhận, nên khi bộ chỉ báo gắn với thưởng phạt thì số liệu có nguy cơ bị làm nhẹ. Tách vai trò nhập liệu khỏi vai trò phê duyệt kiểm soát trực tiếp động cơ đó: bên tạo số liệu độc lập với bên xác nhận.

Chênh lệch có hệ thống giữa đánh giá cấp nhóm và cấp chỉ báo tương đồng với cơ chế này, lớn nhất ở nhóm E với 0,89 điểm: chuyên gia thừa nhận tầm quan trọng về nguyên tắc nhưng hoài nghi khả năng đo lường trung thực, vì chỉ báo nhóm này đều do chính đơn vị bị đánh giá tự khai báo. Nhóm F là ngoại lệ duy nhất, chênh lệch âm 0,48 điểm và điểm cấp nhóm thấp nhất nhưng vẫn giữ 4 trên 7 chỉ báo; khi phần lớn khối lượng thi công đi qua chuỗi nhà thầu phụ nhiều tầng, đây nên được hiểu là một khoảng trống nhận thức về an toàn trong chuỗi cung ứng.

c. Điểm thấp của khía cạnh G

Các chỉ báo thấp nhất của nhóm G đều là chỉ báo muộn hoặc chỉ báo về chất lượng hệ thống giám sát, và phân tán ý kiến lớn nhất toàn tập T1. Kết hợp với ý kiến mở về việc nhà thầu che giấu sự cố, một cách giải thích khả dĩ là chuyên gia đánh giá thấp không phải vì các chỉ báo này kém quan trọng mà vì không tin dữ liệu đầu vào phản ánh đúng thực tế.

d. Hàm ý triển khai và giới hạn của kết quả

Về hàm ý triển khai, tập T2 tập trung vào các yếu tố hữu hình tại hiện trường và thiếu hẳn hai khía cạnh E và I, khác với nguyên tắc phân tầng của OECD [10]. Nghiên cứu vì vậy đề xuất không áp dụng ngưỡng một cách cơ học: mỗi khía cạnh giữ tối thiểu một chỉ báo đại diện, chọn theo tiêu chí quy định trước gồm điểm trung bình cao nhất trong khía cạnh, hệ số biến thiên nhỏ hơn khi chênh lệch dưới 0,10, tính ổn định trong jackknife, và ưu tiên chỉ báo có bên phê duyệt độc lập với bên nhập liệu; chỉ báo bị loại vì lý do khả thi đo lường chuyển sang diện rà soát ở Vòng 2. Riêng khía cạnh E và I, Vòng 2 phải bổ sung chỉ báo để hai khía cạnh này có mặt trong bộ chỉ báo cuối cùng.

Nghiên cứu có năm hạn chế. Thứ nhất, cơ cấu mẫu nghiêng về nhóm tư vấn giám sát và nhà thầu, và cỡ mẫu hiện tại không đủ để phát hiện khác biệt giữa các nhóm chủ thể. Thứ hai, mức đồng thuận Kendall thấp và lõi ổn định bootstrap còn hẹp, nên tập chỉ báo được giữ chỉ là kết quả trung gian. Thứ ba, độ tin cậy trong cùng người đánh giá đo từ hai cặp chỉ báo lặp chỉ ở mức kém, sai số đo xấp xỉ một bậc, độc lập với vấn đề đồng thuận. Thứ tư, tiêu chí về giá trị phân biệt hiện mới chỉ được áp dụng

theo phán đoán chuyên gia; đợt thí điểm bốn tuần tại một công trường ở Hà Nội chưa đủ quy mô để kết luận. Thứ năm, một phần dữ liệu được quy đổi từ phiên bản khảo sát trước, và kết quả phân tích độ nhạy cho thấy hệ quả của việc quy đổi chỉ giới hạn ở các kết luận dựa trên ngưỡng tuyệt đối.

5. Kết luận

Nghiên cứu đã đề xuất một khung chỉ số ba cấp đánh giá kết quả thực hiện an toàn trên công trường xây dựng, gồm 9 khía cạnh, 28 tiêu chí và 95 chỉ báo (T0), với điểm khác biệt về cấu trúc là gắn mỗi chỉ báo với một cặp vai trò nhập liệu và phê duyệt tách biệt nhằm giảm rủi ro làm sai lệch số liệu báo cáo.

Khảo sát Vòng 1 trên tập rút gọn 67 chỉ báo (T1) với 22 chuyên gia thuộc năm nhóm chủ thể cho mức quan trọng trung bình dao động từ 3,64 đến 5,45; theo tiêu chí thống kê tiền định, 26 chỉ báo được giữ (T2). Phân tích độ nhạy theo cỡ mẫu cho thấy chỉ 16 trong 26 chỉ báo của T2 có tần suất bootstrap từ 0,60 trở lên, chỉ 17 chỉ báo được giữ ở mọi lần lặp jackknife, và việc bỏ một chuyên gia đủ làm 20 chỉ báo đổi trạng thái. T2 vì vậy được báo cáo là kết quả sàng lọc sơ bộ gắn với một cặp ngưỡng cụ thể, không phải bộ chỉ báo đã được xác nhận.

Chênh lệch hệ thống giữa đánh giá cấp nhóm và cấp chỉ báo có ở tám trong chín nhóm, lớn nhất ở khía cạnh sự tham gia của người lao động; một cách giải thích khả dĩ là giới hạn của hệ thống đo lường hiện tại.

Lời cảm ơn

Nghiên cứu này do Bộ Xây dựng tài trợ trong đề tài mã số RD 29-24.

Dữ liệu

Danh mục đầy đủ 67 chỉ báo kèm trung bình, độ lệch chuẩn, hệ số biến thiên, C_j và S_j của Phụ lục được cung cấp chi tiết tại: https://app.box.com/s/h37zvgdf28smbhv95risv3zh7ldb1bgv.

Tuyên bố về việc sử dụng trí tuệ nhân tạo tạo sinh và các công nghệ hỗ trợ bởi AI

Trong quá trình chuẩn bị bản thảo, các tác giả đã sử dụng Gemini để soát lỗi chính tả. Các tác giả đã kiểm tra và chỉnh sửa nội dung và chịu hoàn toàn trách nhiệm về tính chính xác, tính nguyên bản và tính toàn vẹn của bài báo.

Tài liệu tham khảo

[1] Bộ Lao động - Thương binh và Xã hội (2021–2025). *Thông báo tình hình tai nạn lao động các năm 2020–2024 (số 565/TB-LĐTBXH, 843/TB-LĐTBXH, 1229/TB-LĐTBXH, 1136/TB-BLĐTBXH, 630/TB-BLĐTBXH)*. Hà Nội.

[2] Bộ Nội vụ (2026). *Thông báo số 2477/TB-BNV ngày 19/3/2026 về tình hình tai nạn lao động năm 2025*. Hà Nội.

[3] Hallowell, M. R., Hinze, J. W., Baud, K. C., Wehle, A. (2013). Proactive construction safety control: Measuring, monitoring, and responding to safety leading indicators. *Journal of Construction Engineering and Management*, 139(10):04013010.

[4] Xu, J., Cheung, C., Manu, P., Ejohwomu, O. (2021). Safety leading indicators in construction: A systematic review. *Safety Science*, 139:105250.

[5] Manu, P., Mahamadu, A. M., Phung, V. M., Nguyen, T. T., Ath, C., Heng, A. Y. T., Kit, S. C. (2018). Health and safety management practices of contractors in South East Asia: A multi-country study of Cambodia, Vietnam, and Malaysia. *Safety Science*, 107:188–201.

[6] Oswald, D., Zhang, R. P., Lingard, H., Pirzadeh, P., Le, T. (2018). The use and abuse of safety indicators in construction. *Engineering, Construction and Architectural Management*, 25(9):1188–1209.

[7] Awolusi, I., Marks, E., Hainen, A., Alzarrad, A. (2022). Incident analysis and prediction of safety performance on construction sites. *CivilEng*, 3(3):669–686.

[8] Hollnagel, E. (2014). *Safety-I and Safety-II: The past and future of safety management*. Ashgate Publishing, Farnham, UK.
[9] Lingard, H., Hallowell, M., Salas, R., Pirzadeh, P. (2017). Leading or lagging? Temporal analysis of safety indicators on a large infrastructure construction project. *Safety Science*, 91:206–220.
[10] OECD (2008). *Guidance on developing safety performance indicators related to chemical accident prevention, preparedness and response - Guidance for industry*. Series on Chemical Accidents No. 19, OECD Publishing, Paris, France.
[11] HSE, CIA (2006). *Developing process safety indicators: A step-by-step guide for chemical and major hazard industries (HSG254)*. Health and Safety Executive, UK.
[12] Bellamy, L. J., Sol, V. M. (2012). *A literature review on safety performance indicators supporting the control of major hazards*. National Institute for Public Health and the Environment, Bilthoven, Netherlands.
[13] CCPS (2022). *Process safety metrics guide for leading and lagging indicators (Version 4.1)*. Center for Chemical Process Safety, AIChE, New York, USA.
[14] Phinias, R. N. (2023). Benefits and challenges relating to the implementation of health and safety leading indicators in the construction industry: A systematic review. *Safety Science*, 163:106131.
[15] Nguyễn, T. V., Hồ, N. V. (2021). Xây dựng khung đánh giá sự quản lý an toàn lao động trên các công trường xây dựng tại Thành phố Hồ Chí Minh. *Tạp chí Khoa học và Công nghệ - Trường Đại học Công nghiệp Thành phố Hồ Chí Minh*, 53A(05):16–28.
[16] McGibbon, S., Hare, B. (2026). Construction project leaders' perceptions of respirable crystalline silica exposure and control: A safety leadership perspective in the UK social housing sector. *Journal of Science and Technology in Civil Engineering (STCE)*, 20(2S):14–28.
[17] Chính phủ (2020). *Nghị định số 145/2020/NĐ-CP ngày 14/12/2020 quy định chi tiết và hướng dẫn thi hành một số điều của Bộ luật Lao động về điều kiện lao động và quan hệ lao động*. Hà Nội.
[18] Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3):338–353.
[19] Murray, T. J., Pipino, L. L., van Gigch, J. P. (1985). A pilot study of fuzzy set modification of Delphi. *Human Systems Management*, 5(1):76–80.
[20] Ishikawa, A., Amagasa, M., Shiga, T., Tomizawa, G., Tatsuta, R., Mieno, H. (1993). The max-min Delphi method and fuzzy Delphi method via fuzzy integration. *Fuzzy Sets and Systems*, 55(3):241–253.
[21] Yin, T. T., Hanif, H. (2024). Fuzzy Delphi method: A step-by-step guide to obtain expert consensus. *International Journal of Academic Research in Business and Social Sciences*, 14(4):1109–1116.
[22] Dawes, J. (2008). Do data characteristics change according to the number of scale points used? An experiment using 5-point, 7-point and 10-point scales. *International Journal of Market Research*, 50(1): 61–104.
[23] Koo, T. K., Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*, 15(2):155–163.