Acta Social Science & Humanities
An International Journal

OPEN ACCESS

ISSN 3117-860X
Original Research Article

Check for updates

Self-Reported Internship Supervision Activities: An Exploratory Within-Supervisor Comparison

Nova M. Cabias*, Christian Jay Pungtilan, and Gerlie Miole

College of Industrial Technology, Mariano Marcos State University, Philippines

Corresponding author: **Nova M. Cabias** | E-mail: **nmcabias@mmsu.edu.ph**

Citation: Nova M. Cabias, Christian Jay Pungtilan, and Gerlie Miole (2026). Self-Reported Internship Supervision Activities: An Exploratory Within-Supervisor Comparison. *Acta Social Science & Humanities: An International Journal.* **DOI: https://doi.org/10.51470/SSH.2026.5.2.35**

Received 17 June 2026 | Revised 19 July 2026 | Accepted 22 August 2026 | Available Online 29 September 2026

Abstract

The overall scores of supervision could mask variations between activities. This research looked at the size of differences, supervisor dispersion of differences, and the sensitivity of those differences depending on item construction. An exploratory, post-hoc, quantitative secondary analysis was done using the full five-point frequency ratings from 200 industry supervisors in a Philippine university internship data set. Four task-clarification questions and five questions on feedback, check-in, explanation, and learning needs were created as content-based sets. Analyses were done for the mean paired difference and confidence interval, signed ranks and the direction of individual differences, and nine single-item exclusions and five alternative groupings. Routine guidance scores had an average of 4.758 (SD = 0.452), while that of developmentally supportive items was 4.426 (SD = 0.464). The mean difference between the two was 0.332 (95% CI [0.248, 0.415]); Wilcoxon W = 2,602.0; p < .001; rank-biserial correlation coefficient 0.650. There were 132 cases where the routine guidance averages were greater, 40 cases where developmentally supportive averages were greater, and 28 cases where the averages were identical. A significant mean difference was maintained in all test specifications although its value was variable. The differences reveal an unbalanced pattern of supervision activities within positive overall scores. Supervision activity should analyze particular support activities alongside the overall scores. The exploratory item groups cannot be seen as subscales and therefore their results cannot determine any effectiveness of the supervision process or its outcomes on the students.

Keywords: *work-integrated learning; internship supervision; industry supervisors; secondary analysis; paired comparison.*

Introduction

University internships link study with working. The role of the supervisor includes the provision of instructions, but might also include an explanation of procedures, discussion of progress and feedback. These actions can have more than one educational meaning. From the workplace learning point of view, attention is drawn not only to learning opportunities in a situation, but also to learners' involvement with them [1]. The question that arises regarding university internships is the following: to what extent does an overall good evaluation of supervision indicate the activities performed by supervisors?

In work-integrated learning assessment and supervision literature, different intentions behind assessment and supervision can be observed. A systematic review of work-integrated learning (WIL) assessment investigated the way in which the assessment was done and the intentions of using those ways, including support of learning [2]. A systematic review of placement supervision found that learner-centered research paid particular attention to supportive relationships and emotions, whereas researcher-centered investigations focused on expectations, goals, and links between theory and practice [3].

A framework for evaluating WIL curricula includes learning support, supervisor access, and preparation alongside authenticity and alignment [4]. Research on workplace-performance assessment also examines the challenges of obtaining useful evidence across placement contexts [5]. TThe above contributions help enable monitoring at the activity level, while keeping the empirical issue of activity rating within a supervisor's framework an open one.

A survey involving 323 co-operative education students revealed a significant disparity between the value attributed to the role played by supervision in supporting the students and the actual provision of such support to the students as feedback [6]. Another study from the Philippines involved using evaluation by industry partners of intern performance in improving the curriculum [7]. Neither comparison is equivalent to examining supervisors' reports of their own activities. The present analysis addresses that narrower question using a Philippine on-the-job training (OJT) dataset.

Workplace-feedback research describes the supervisor's role within a relationship involving students and academic staff [8].

© ***Authors:*** *Published in* ***Acta Social Science and Humanities: An International Journal*** *under the CC BY-NC-ND 4.0 license (http://creativecommons.org/licenses/by-nc-nd/4.0/). No commercial use or modifications permitted.*

Broader feedback literature examines the content and focus of feedback [9] and learners' opportunities to understand standards, evaluate their work, and act on information [10]. These distinctions matter here because a frequency rating does not establish the content of feedback or its use by a student.

Is there any possibility that the positive evaluation masks disparities in the performance of tasks by the same supervisors?An overall mean combines statements about different practices. Individual item means show which statements received higher ratings, but give no information about the uncertainty of a paired contrast or how many supervisors show each direction of difference. Classification also matters when statements could reasonably belong to either activity group. Examining these features can clarify what a favourable summary score represents without assuming that a lower relative rating indicates inadequate supervision.

The exploratory comparison uses four items on requirements, task purposes, performance expectations, and detailed directions to describe routine guidance. Five items on performance feedback, scheduled check-ins, technical explanations, learning-needs discussions, and timely appreciative and developmental feedback describe developmental support. These are labels for selected item content. Functions may overlap since task clarification may lead to learning while developmental feedback may take place in everyday tasks. Because the statements differ in wording and the occasions on which they apply, the analysis also examines alternative groupings.

The records come from the OJT Supervision Enhancement Framework (COSEF) project described in the related manuscript Developing a Work-Integrated Learning Supervision Framework for Philippine Industrial Technology Education: A Mixed-Methods Study. That study brings together student accounts, role-specific questionnaires, and expert appraisal to develop a supervision framework. It reports overall industry-supervisor ratings, selected item patterns, and a provisional measurement structure. This current study utilizes the same 200 industry-supervisor responses in order to investigate a post hoc issue related to the differences between certain activities amongst the respondents. In this case, the results include an estimate of the paired difference, distribution of the paired difference among the supervisors, as well as sensitivity analyses using different items.

This secondary analysis hence addresses the issue of what remains unknown regarding the frequency of certain supervisory activities from the summary score itself. This paper hence presents the results of the paired difference and the extent to which the difference is influenced by the choice of items. The results for internship supervision hence offer reasons why certain mentoring activities should be investigated alongside the summary score.

Research Objectives

The study sought to (1) describe industry supervisors' self-reported frequency ratings for the selected routine-guidance and developmental-support items; (2) estimate the magnitude and distribution of the difference between the two item-set averages within supervisors; and (3) examine whether the direction and magnitude of that contrast changed when individual items were omitted, or alternative content groupings were used.

Conceptual Framework

Routine guidance refers here to communicating requirements, task purposes, performance expectations, and practical instructions. Developmental support refers to the selected feedback, check-in, explanatory, and learning-needs activities. Reviews of WIL supervision and assessment support examining this range of activities and perspectives [2][3]. The researchers made the present classifications after inspecting the ratings. These content sets have not been independently validated as dimensions of supervision.

The distinction allows for shared functions. A conversation about assigned work may include both immediate instructions and feedback that supports further learning. The two averages therefore describe selected statements rather than separate forms of supervisory competence. Differences between them may reflect item wording and opportunities to perform the activities, as well as differences in reported practice.

The workplace-affordances perspective relates learning to the opportunities and guidance available at work and to individuals' participation in those opportunities [1]. Work on informal workplace learning also draws attention to learning embedded in everyday activity [11]. Routine instruction and developmental discussions may therefore overlap in practice. The questionnaire records supervisors' accounts, without establishing how students accessed or used the guidance. These perspectives inform interpretation; the analysis does not test a causal model of workplace learning.

Methodology

Research Design and Data Source

A cross-sectional quantitative secondary analysis included de-identified individual records from the 200 industry supervisors from the parent COSEF study. The parent study was carried out at a College of Industrial Technology in a public university in the Philippines together with its OJT host settings. The 200 industry supervisors were recruited via the Memorandum of Agreement (MOA) partnership of the College with its host organisations. The questionnaire was administered using paper format. The industry supervisors qualified for participation if the host organisations had partnership with the College for at least one year, were directly supervising students in the study period, and had supervised OJT students before. The parent manuscript states that all 200 industry supervisors were recruited and that data were collected from 17 March to 24 April 2025. Host organisation identification was not available in order to assess any potential clustering by host organisation. There was no further recruitment, participant contact, and data collection in the secondary analysis. The scores were re-calculated from I1-I30 in the designated respondent-level CSV file.

Instrument and Measures

Industry-supervisor questionnaire was developed within the parent project based on stakeholder descriptions of support for OJT. Statements of the questionnaire referred to the activities of the supervisors themselves. There were 30 items in the questionnaire, each of which had five frequency options in order of magnitude (1 = Never, 2 = Rarely, 3 = Sometimes, 4 = Often, and 5 = Always). An item map indicated the number of each statement as either I1-I30. Nine items served as the reference point: routine guidance consisted of I1 (requirements and expectations at work), I2 (reasons for task completion), I4 (performance expectations), and I6 (detailed instructions); developmental support included I10 (performance feedback on a regular basis), I18 (supervision meetings or consultations on schedule), I20 (detailed explanation of theories, procedures or skills), I22 (discussion of learning needs in detail), and I30 (appreciative and developmental feedback in a timely manner). A score was calculated as an arithmetic average for the selected items.

The routine guidance category indicated the immediate information needs for performing the tasks. The developmental support category covered performance appraisal, structured discussion, explanations, and discussion of learning needs. The assignments took into consideration the predominant role played by each item, given the fact that there is a possibility of overlap between the activities. I20 was included in the reference developmental support category as it dealt with explanations of knowledge and skill; this overlap with procedural instruction resulted in its omission as well as reallocation during the sensitivity analysis.

These nine items were chosen on the basis of analysis of the item ratings, so that the choice was post hoc. Nine items constituted a partial, illustrative comparison of the two functions rather than an exhaustive representation of either one. There were eight other statements involving role clarification, monitoring, orientation, coordination of work activities, theory application, discussion of concerns, professional behavior, and task coaching, all of which were included in expanded specifications. The rest dealt with more general relationship, psychosocial, ethical, or continuity functions and were incorporated into the full descriptive index. See Appendix B for the selected and excluded material and alternative categories. This content justification does not imply independent verification or equivalence of measurement. The sets also do not coincide with those factors isolated within the parent study itself: The four guidance items constitute the provisional performance structure factor in that study, while the five developmental support items cover four other provisional factors.

Data Analysis

Objective 1 required means and sample standard deviations to describe the selected items and the full questionnaire. For Objective 2, the difference between the mean for the routine guidance and the mean for the developmental support defined each paired difference. The main statistic was the mean paired difference with a 95% confidence interval based on a t-distribution. Frequencies of positive, negative, and zero differences provided information about variability among supervisors. A secondary two-sided Wilcoxon signed-rank test used the average ranks for ties, excluded zero differences, and relied on a tie-adjusted normal approximation without continuity correction. Since the reference scores included the averages of four and five integer-valued items, differences could be expressed exactly in units of 1/20 before being ranked. W is the smaller signed-rank sum. The matched pairs rank biserial correlation was computed as (positive rank sum - negative rank sum) / (positive rank sum + negative rank sum) [12]. The confidence interval estimated an arithmetic mean difference, whereas the signed-rank procedure focused on the ranked paired differences distribution. For Objective 3, nine specifications dropped one item from each respective reference set. Five additional specifications were: (a) Reassignment of I20 from developmental support to routine guidance; (b) Addition of I8, I19, I26, and I27 to routine guidance; (c) Addition of I7, I9, I21, and I28 to developmental support; (d) Combination of (b) and (c); and (e) Use of the extended sets from (d), including I28 under routine guidance. Each specification used the same 200 subjects. Prior to ranking, the numerator was calculated as the multiplication of the routine-item sum and the number of developmental items minus the multiplication of the developmental-item sum and the number of routine items, resulting in an exact integer numerator. When dividing the numerator by the product of the number of the two types of items, we arrived at the paired score difference. The process did not disrupt any ties. The 14 follow-up two-sided p values were adjusted together using the sequential procedure [13] for multiple testing; the original comparison was still the reference analysis. Mean-difference confidence intervals were marginal and were not adjusted for multiplicity. There was no substitute for the reference analysis on account of its p value. All calculations were done using Python with SciPy 1.17.0 [14]. The signed rank test calculations were cross-referenced with the built-in function. The correspondence between the objectives and analyses is provided in Appendix A. In Appendix C, the ancillary Spearman correlation coefficients of the overall 30-item average with three experience/exposure variables are provided separately.

Ethical Considerations

The parent study obtained the necessary ethical approval by the institutional committee in 2024 prior to conducting the data gathering process. Subjects were given the necessary information regarding the study, participation as a volunteer, confidentiality of information, and the option to withdraw without penalties, and their informed consent was obtained prior to participation. Permission for re-use of the data was granted by the leader of the project. No new participants were recruited for the secondary analysis.

Results

Participant Characteristics

The 200 supervisors reported a mean of 5.95 years of work experience (SD = 4.35; median = 5; range = 1–23) and 3.00 years supervising OJT students (SD = 2.74; median = 2; range = 1–18). The recorded number of OJT students supervised averaged 3.00 (SD = 1.80; median = 2; range = 1–17). All three profile variables were complete. These characteristics describe the participating supervisors in the study setting.

Data Characteristics and Item-Level Ratings

There were 200 fully recorded cases with regard to the 30 supervision questions, and thus a total of 6,000 responses without any missing data. The overall mean for 30 items was 4.497 on a five-point frequency scale. There were four (2.0%) respondents who marked maximum categories for all 30 questions, and the total number of maximum responses was 4,091 out of 6,000 responses (68.2%). Regarding Objective 1, the means of selected nine items varied between 4.255 and 4.835 (see Table 1). These values represent the frequencies of supervisors' own actions. Both the overall mean and several selected item patterns have been already reported in the parent manuscript as well.

Table 1: Ratings of Selected Supervisory Activities (N = 200)

Item	Supervisory activity	M	SD
I1	Clear instructions on requirements and workplace expectations	4.830	0.512
I2	Clarification of the purpose of assigned tasks	4.835	0.499
I4	Explanation of expected performance	4.695	0.666
I6	Detailed directions for workplace tasks	4.670	0.688
I10	Clear and consistent performance feedback	4.505	0.833
I18	Scheduled supervision meetings or check-ins	4.255	0.857
I20	Detailed explanation of theories, procedures, or skills	4.270	0.861
I22	Discussion of L JAILF learning needs	4.330	0.809
I30	Timely appreciative and developmental feedback	4.770	0.599

Note. Items I1, I2, I4, and I6 formed the exploratory routine-guidance set; I10, I18, I20, I22, and I30 formed the exploratory developmental-support set. M = mean; SD = standard deviation. The parent manuscript reports selected industry-supervisor item means to two decimal places. This table provides context for the secondary comparison and does not independently replicate those descriptive patterns.

The highest selected item mean was for clarifying the purpose of assigned tasks (I2, M = 4.835, SD = 0.499), followed by communicating requirements and workplace expectations (I1, M = 4.830, SD = 0.512). The lowest means were for scheduled supervision meetings or check-ins (I18, M = 4.255, SD = 0.857) and detailed explanations of theories, procedures, or skills (I20, M = 4.270, SD = 0.861). Within the developmental-support set, timely appreciative and developmental feedback (I30, M = 4.770, SD = 0.599) received the highest rating. No item-by-item significance tests were conducted; this ordering is descriptive.

Within-Supervisor Comparison

For Objective 2, respondent-level routine-guidance scores averaged 4.758 (SD = 0.452), whereas developmental-support scores averaged 4.426 (SD = 0.464).

The routine-minus-developmental mean paired difference was 0.332 scale points (SD = 0.597; t-based 95% CI [0.248, 0.415]). Routine-guidance scores were higher among 132 respondents (66.0%); developmental-support scores were higher among 40 (20.0%); and 28 (14.0%) had equal averages. Excluding the 28 zero differences left 172 pairs for the signed-rank calculation. The two-sided Wilcoxon signed-rank test was statistically significant (W = 2,602.0, p < .001), with a matched-pairs rank-biserial coefficient of 0.650 in the routine-guidance direction (Table 2).

Table 2: Within-Supervisor Comparison of Exploratory Item-Group Scores

Measure	Result
Routine guidance, M (SD)	4.758 (0.452)
Developmental support, M (SD)	4.426 (0.464)
Paired difference, M (SD)	0.332 (0.597)
95% CI for mean paired difference	0.248 to 0.415
Direction of difference	132 routine > developmental; 40 reverse; 28 equal
Wilcoxon signed-rank test	W = 2,602.0; p < .001 (two-sided)
Matched-pairs rank-biserial coefficient	0.650

Note. CI = confidence interval. The CI is t-based and estimates the arithmetic mean paired difference; it is not an interval estimate for the Wilcoxon statistic. The signed-rank calculation excluded 28 zero differences (nonzero n = 172). The positive rank-biserial coefficient indicates higher routine-guidance ratings. The item groups are exploratory, content-based composites.

Sensitivity to Item Composition

For Objective 3, all nine single-item omissions retained a positive mean paired difference, ranging from 0.289 to 0.418 points. Omitting I30 produced the largest difference (0.418), whereas omitting I18 produced the smallest (0.289). The five alternative groupings also retained positive differences, ranging from 0.168 to 0.304. Reassigning I20 to routine guidance reduced the difference to 0.195; expanding both sets reduced it to 0.168. Every marginal 95% confidence interval excluded zero, and all 14 Holm-adjusted follow-up p values were < .001 (Table 3).

Table 3: Sensitivity of the Within-Supervisor Contrast to Item Composition

Specification	Mean paired difference	95% CI	Holm-adjusted p
Original comparison	0.332	[0.248, 0.415]	Reference
Omit I1	0.307	[0.219, 0.395]	< .001
Omit I2	0.306	[0.220, 0.392]	< .001
Omit I4	0.352	[0.270, 0.435]	< .001
Omit I6	0.361	[0.276, 0.445]	< .001
Omit I10	0.351	[0.259, 0.443]	< .001
Omit I18	0.289	[0.206, 0.372]	< .001
Omit I20	0.293	[0.213, 0.372]	< .001
Omit I22	0.308	[0.223, 0.392]	< .001
Omit I30	0.418	[0.326, 0.509]	< .001
Move I20 to routine	0.195	[0.129, 0.261]	< .001
Expand routine set	0.196	[0.138, 0.254]	< .001
Expand developmental set	0.304	[0.232, 0.375]	< .001
Expand both sets	0.168	[0.122, 0.214]	< .001
Broad sets with task coaching in routine	0.185	[0.135, 0.235]	< .001

Note. Positive differences indicate higher routine-guidance averages. All specifications use N = 200. Confidence intervals are t-based estimates of the mean paired difference and are not adjusted for multiplicity. Holm adjustment applies to the 14 follow-up signed-rank tests. Appendix B specifies the item memberships. The original comparison is repeated as the reference.

Ancillary Associations With Experience

None of the Spearman correlations between the overall 30-item mean and years of work experience, years supervising OJT students, or recorded student count was statistically significant after Holm adjustment. Appendix C reports the coefficients, marginal bootstrap confidence intervals, and p values. These associations concern the overall index, not the routine-minus-developmental difference.

Discussion

Contribution of the Within-Supervisor Comparison

High overall ratings coexisted with differences between the selected activities. Routine-guidance ratings were 0.332 points higher on average on the five-category scale. Routine-guidance averages exceeded developmental-support averages for 132 of 200 supervisors (66%), while 40 (20%) showed the reverse and 28 (14%) had equal averages. Therefore, while the pattern was common, it was not universal in this sample. The parent COSEF manuscript includes overall and selected item ratings in its framework development study using a mixed-methods approach. The paired estimate, its variance, the direction counts, and the sensitivity analyses described above address a more precise question regarding variability among those supervisors.

As seen from the sensitivity analyses, there is a consistent direction along with a dependence on item content. In each specification tested, the estimate of the mean difference remains positive. It rises to 0.418 when the highly rated feedback item I30 is excluded, falls to 0.195 when item I20 is relabeled as routine guidance, and falls to 0.168 when both sets are enlarged. No single selected item determines the direction in these sensitivity analyses. Yet, the category labels did not fix the magnitude of the estimate – it depends on the activities selected and how they are classified.

Workplace Affordances and Individual Engagement

Workplace affordances provide one approach to examining the learning possibilities that the reported activities could present [1]. The instructions could allow the student to start working on the assignment, whereas the explanations and the discussion about progress could help the student to comprehend the assignment or to prepare for future assignments. Workplace learning studies that do not focus on formal learning include informal learning through typical interaction [11]. This is what the activities could accomplish, but the current study did not examine this issue.

The ratings raise questions about students' opportunities to complete tasks, review progress, and deepen understanding. They do not show students' access to guidance or their response to it. The distinction between available opportunities and learner participation [1] locates this boundary: a supervisor's account of providing guidance captures only part of the learning process. Engagement of students, mentorship, and results need to be proven separately.

Timely Feedback Within the Developmental Group

Appreciative and developmental feedback was highly rated among the less well-rated items in the developmental category. I30 averaged 4.770, compared with 4.255 for scheduled check-ins (I18) and 4.270 for detailed explanations (I20). The exclusion of I30 resulted in an increase in the paired difference by 0.418 points. This is significant since the mean for the group would imply that all developmental tasks were underreported. The item ordering is descriptive, with no item-level significance tests, and the scores do not establish whether the activities met students' needs.

One possible explanation is that supervisors provided feedback during tasks more often than through separate planned discussions. The questionnaire did not distinguish those interaction formats. I30 was also a combination of appreciative feedback and developmental feedback; thus, it is impossible to determine from this rating the extent of the role played by praise and information for improvement. Feedback is classified by researchers based on its emphasis and instructional value [9][10]; the reported frequency and promptness do not provide any of these.

Alternative Explanations for the Contrast

The activities vary in terms of when they become relevant. While requirements and instructions may need to be explained each time an activity is undertaken, a discussion on learning needs may become more relevant during certain placement stages. Coordination is also necessary for scheduled meetings. The scale did not provide any reference period, nor did it provide any indication of the frequency of the activities. Placement stage, task complexity, and interns' prior knowledge could affect the need for guidance, but were not measured.

Some statements could reasonably be assigned to either set. Detailed explanations in I20 may assist immediate task completion and longer-term understanding, while clarifying a task's purpose in I2 may itself be developmental. Having an average of four or five items allows us to compare the groups using the same number scale while not making the content nor endorsement easy for both the same. By shifting I20 to another group and increasing the number of items allowed us to see the shift due to our selection criteria. Independent assessment of the content as well as interviewing respondents regarding their comprehension of the questions will allow us to evaluate our classifications.

Overload and partnering can be considered possible explanations as well. The literature on workplace feedback mentions the issue of time and the variety of supervisor tasks [8]. In this case, time constraints from production could limit planned discussions, whereas access to information and assistance outside of discussions could have rendered them unnecessary. Responsibility for learning plan development and providing feedback could also be distributed among academic coordinators. None of these explanations have been tested: overload, informal contact, readiness of the supervisor, and the distribution of duties have not been documented.

Self-reports can be influenced by recall, social desirability response, halo effect, and difference in item interpretation [15]. Face-to-face delivery might have prompted positive responses, though this has not been considered. The maximum score comprised 68.2% of the 30-item responses, making discrimination difficult on the high end. This could be due to frequent activity, reporting bias, or both. Comparing scores from the same respondents eliminates the difference between different respondent samples, but not the effects of individual items, lack of equal opportunity, ceiling effect, or dependence within host organizations.

Comparison With Empirical Placement Studies

A student survey identified a gap between the importance and reported receipt of constructive feedback [6]. This makes it useful to distinguish a supervisor's report of timely feedback from a student's experience of its usefulness. However, that study examined student perceptions, whereas the present analysis uses supervisors' frequency ratings in Philippine industrial technology placements. The measures are not directly comparable. Matched student–supervisor accounts would be needed to examine agreement within the same placements.

A study of 99 host-organisation supervisors across 24 placement programmes examined assessment and feedback experiences, reporting positive experiences alongside differences from analogous student responses concerning relevance and transparency [16]. The cross-program assessment approach of the paper differs from that of the current within-supervisor activity comparison. It allows posing questions about the institution's organization and the perception of students, but does not show any conflict between the supervisors in this sample.

Moreover, research carried out in health fields addresses the issues of feedback other than its frequency. In particular, a study related to clinical placements involved occupational therapy students, their supervisors, university personnel and assessment documents in order to explore the understanding of their progress by students [17]. Another research was devoted to the experience of virtual placements of Australian students of health information management and their supervisors and the issue of informal communication loss [18].

Ancillary Associations With Experience

The additional correlations relate to the full 30-item index rather than to the chosen activity subsets or their difference (Appendix C). The correlation coefficients were –.082 for work experience, –.040 for supervisory experience, and .011 for student record count. After Holm adjustment, none of the coefficients were statistically significant. Zero values were included within all marginal bootstrap confidence intervals, allowing the possibility of an association of any sign; the interval for the work experience correlation coefficient was [–.222, .062]. This provides ambiguity regarding the associations without providing evidence of irrelevance of experience. Establishing equivalence would require valid equivalence intervals and a suitable test [19].

Indicators of experience capture exposure to work or supervision but not necessarily supervisor training. An overall average could reflect activities with different associations with experience. Possible ceiling effects could limit observable variations, as well as nonlinearities and workplace context effects. These factors were not examined. Since the correlation coefficients refer to the entire index, these cannot account for the discrepancy between the selected activity subsets or assess mentorship effectiveness or engagement or learning.

Implications and Further Research

Monitoring the internship might include brief descriptions of the guidance provided and required in different phases of placements, which might be used by students, supervisors and coordinators of academic programmes to discuss relevant tasks, feedback and access to the guidance. This suggestion is in line with the literature on collaborative feedback in the workplace [8], on supervisor availability and preparation [4] and on assessment in various placements [5][16]. It should be pilot-tested and evaluated. The current findings justify further investigation of the supervision issue rather than setting any required frequency of meetings.

Future research might involve the observation of both the provision of the guidance and the way it is used by students. Matched reports, observations and interaction logs might capture the opportunities for guidance, its provision, student access and the actions following. The feedback literacy research implies the need to treat provision and student uptake of the guidance as separate issues [20]. Placement phase and length, task complexity, supervisor preparation, workload and the host organization should also be captured. Independent sample with prespecified membership of items should verify the pattern observed.

Limitations

Selection and grouping based on the ratings occurred following the inspection of the ratings. Selection is based on overlapping functions, there is no independent validation of the grouped functions, and 14 of the post follow-ups are based on just some reasonable options. All analyses are done based on the same answers with their confidence intervals and corrected p-values being conditional on this post hoc decision. Independent validation by content review and prespecified test among another sample would improve inference. Calculation of means considers the categories numerically equidistant when there was nothing to support that difference between consecutive categories represented the same level of educational difference.

This analysis was done on one particular collaboration of institutions making it difficult to generalize the findings to another subject discipline or organizational sector or placement structure. Identifiers for the host organizations are not available. All uncertainty estimates ignore correlation in supervisors' answers since all are treated as independent even though there are chances that the answers from a single institution might be related.

Self report as well as ceiling effect in ratings makes it difficult to draw inferences.

Conclusion

These 200 industry supervisors provided higher ratings in mean item frequency for the selected routine guidance items compared to the developmental support items, despite positive general evaluations. The direction was opposite among respondents and continued positive among all specifications tested, depending on the item construction. The positive evaluation of appreciative and developmental feedback on time is also an indication of the importance of evaluating the composite together with individual items. This secondary analysis provides in addition the paired comparison, direction counts, and the sensitivity analyses to the related COSEF analysis. The workplace affordances approach sets up questions about possible opportunities provided by these activities, while access, involvement, mentoring, and learning outcomes are not measured.

Declarations

Data Ethics and Reuse. Ethical permission for the parent study was granted by the institution in 2024, and informed consent was given prior to participation. Permission for secondary use of de-identified data was given by the principal investigator. Identifiers of the review committee as well as the ethics approval numbers are not presented in this manuscript due to institutional confidentiality requirements. No additional contact of participants nor collecting new data took place.

Relation with previous research. This analysis makes use of the de-identified dataset comprising 200 responses of supervisors working in the 200 industries that were included in the COSEF project. The sampling, initial 30-item ratings, and some overall statistics as well as item-level descriptive statistics are reused in the analysis. The related manuscript deals with such topics as measurement, methodology integration, framework development, and expert assessment. The present article deals with paired contrast of activities, its distribution by supervisors, as well as its sensitivity to item inclusion. Appendix C presents auxiliary correlations with experience and exposure variables. This is not an independent sample.

Funding. The original OJT research project received institutional research funding. This article uses data collected through that project.

Competing interests. The authors declare no competing interests.

Availability of Data. No respondent-level data is available publicly due to confidentiality and data governance policies applied by the parent project.

Generative AI use. ChatGPT and Claude were used to help in draft writing, language revision, literature review search, and citation style formatting. The authors take responsibility for validation of source data, statistics, references, interpretations, and manuscript.

References

1. Billett S. Learning through work: workplace affordances and individual engagement. *Journal of Workplace Learning*. 2001;13(5):209–214. https://doi.org/10.1108/EUM0000000005548
2. Luk LYY, Chan CKY. Assessing student learning in work-integrated learning (WIL): a systematic review on assessment methods and purposes. *Studies in Educational Evaluation*. 2024;81:101345. https://doi.org/10.1016/j.stueduc.2024.101345
3. Hains-Wesson R, Fannon AM, Brekke Hauglid B, Graham J, Ademuyiwa I. Student and supervisor perceptions of effective placement supervision in work-integrated learning programs: a systematic literature review. *Teaching & Learning Inquiry*. 2026;14:1–21. https://doi.org/10.20343/teachlearninqu.14.15
4. Smith C. Evaluating the quality of work-integrated learning curricula: a comprehensive framework. *Higher Education Research & Development*. 2012;31(2):247–262. https://doi.org/10.1080/07294360.2011.558072
5. Jackson D. Challenges and strategies for assessing student workplace performance during work-integrated learning. *Assessment & Evaluation in Higher Education*. 2018;43(4):555–570. https://doi.org/10.1080/02602938.2017.1378618
6. Drewery D, Truong M, Fannon AM. How should work-integrated learning supervisors support their students? A concurrent triangulated mixed-method study. *Higher Education Research & Development*. 2025;44(5):1094–1110. https://doi.org/10.1080/07294360.2025.2486178
7. Bentor CTS, Sabonsolin JCM, Niez RA. Industry partners' feedback on students' on-the-job training: basis for curriculum enhancement and intervention. *International Journal of Multidisciplinary: Applied Business and Education Research*. 2022;3(12):2491–2497. https://doi.org/10.11594/ijmaber.03.12.02
8. Peach D, Ruinard E, Webb F. Feedback on student performance in the workplace: the role of workplace supervisors. *Asia-Pacific Journal of Cooperative Education*. 2014;15(3):241–252. https://www.ijwil.org/files/APJCE_15_3_241_252.pdf
9. Hattie J, Timperley H. The power of feedback. *Review of Educational Research*. 2007;77(1):81–112. https://doi.org/10.3102/003465430298487

10. Nicol DJ, Macfarlane-Dick D. Formative assessment and self-regulated learning: a model and seven principles of good feedback practice. *Studies in Higher Education*. 2006;31(2):199–218. https://doi.org/10.1080/03075070600572090
11. Eraut M. Informal learning in the workplace. *Studies in Continuing Education*. 2004;26(2):247–273. https://doi.org/10.1080/1580370420002025245
12. Kerby DS. The simple difference formula: an approach to teaching nonparametric correlation. *Comprehensive Psychology*. 2014;3:11.IT.3.1. https://doi.org/10.2466/11.IT.3.1
13. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. 1979;6(2):65–70. [suspicious link removed]
14. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods*. 2020;17(3):261–272. https://doi.org/10.1038/s41592-019-0686-2
15. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff NP. Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology*. 2003;88(5):879–903. https://doi.org/10.1037/0021-9010.88.5.879
16. O'Mahony T, Murphy C, Cole J. Reframing the reluctant assessor: host-organisation supervisor experiences of assessment and feedback in placement-based work-integrated learning. *Journal of Education and Work*. 2026;39(3):286–304. https://doi.org/10.1080/13639080.2026.2719488
17. Nicola-Richmond K, Lyons N, Ward N, Logan S, Ajjawi R. Feedback practices in clinical placement: how students come to understand how they are progressing. *Assessment & Evaluation in Higher Education*. 2025;50(2):323–335. https://doi.org/10.1080/02602938.2024.2400349
18. Prasad N, Maloof M, Gjorgioski S, Riley M. Work-integrated learning for health information management students: lessons learned from literature, and experiences of supervisors and students on virtual or remote placements. *Health Information Management Journal*. 2025;54(2):121–130. https://doi.org/10.1177/18333583241227002
19. Lakens D. Equivalence tests: a practical primer for t tests, correlations, and meta-analyses. *Social Psychological and Personality Science*. 2017;8(4):355–362. https://doi.org/10.1177/1948550617697177
20. Carless D, Boud D. The development of student feedback literacy: enabling uptake of feedback. *Assessment & Evaluation in Higher Education*. 2018;43(8):1315–1325. https://doi.org/10.1080/02602938.2018.1463354

Appendix A Mapping Research Objectives to Analyses

Table A1 maps the three research objectives to their analyses and findings. Tables 1–3 contain the corresponding numerical evidence.

Table A1: Research Objectives, Analyses, and Findings

Research objective	Statistical analysis	Reported finding
1 Describe ratings of selected routine-guidance and developmental-support items	Item means and standard deviations; overall 30-item mean; complete-response counts	N = 200. Selected item means: 4.255K4.835. Highest: task-purpose clarification, 4.835. Lowest: scheduled check-ins, 4.255. Timely appreciative and developmental feedback: 4.770. Overall mean: 4.497.
2 Examine the within-supervisor difference between the two exploratory item groups	Paired composite means; routine-minus-developmental differences; two-sided Wilcoxon signed-rank test; rank-biserial coefficient; t-based CI for the mean difference	Means: 4.758 and 4.426. Paired difference: 0.332, 95% CI [0.248, 0.415]. W = 2,602.0, p < .001; rank-biserial coefficient = 0.650. Routine higher: 132; developmental higher: 40; equal: 28.
3 Examine sensitivity to item omission and alternative content groupings	Nine single-item omissions and five alternative groupings; exact paired-difference ranks; Holm-adjusted follow-up tests; mean-difference CIs	Omissions: mean differences 0.289K0.418. Alternatives: 0.168K0.304. All confidence intervals above zero; all 14 Holm-adjusted follow-up p values < .001. Magnitude depends on item composition.

Interpretive boundary. Item ordering is descriptive. The paired comparison and sensitivity analyses concern the selected content sets. Their educational importance requires further evidence linking supervisory activities to students' experiences and learning.

Appendix B Item Selection and Alternative Specifications

Table B1 documents the content rationale for the reference sets. The rationale is an exploratory interpretation of the available item wording and does not establish independent content validation or preregistration. The nine-item comparison is illustrative, not exhaustive.

Table B1: Selected Items and Content Rationale

Item	Reference set	Statement and rationale
I1	Routine	I provide clear instructions on the requirements, documents, and workplace expectations that OJT students need to observe in the company. Immediate requirements and expectations.
I2	Routine	I clarify the purpose of the tasks and assignments given to OJT students. Purpose of assigned tasks; also potentially educative.
I4	Routine	I explain the level of performance expected from OJT students. Expected standard for current task performance.
I6	Routine	I provide detailed directions and instructions that help students perform effectively in the workplace. Directions for executing assigned work; also potentially instructional.
I10	Developmental	I provide clear and consistent feedback on students' work performance and assigned tasks. Review of performance through consistent feedback; also useful for immediate correction.
I18	Developmental	I conduct regularly scheduled supervision meetings, check-ins, or performance discussions with OJT students. Recurring planned supervision or performance dialogue.
I20	Developmental	I provide detailed explanations of theories, procedures, or skills that students need to apply in the workplace. Explanatory teaching of theory, procedure, or skill; overlaps with immediate instruction.
I22	Developmental	I discuss students' learning needs in sufficient depth to guide their development. In-depth discussion of learning needs and development.
I30	Developmental	I provide timely appreciative and developmental feedback on students' performance and progress. Appreciative and developmental feedback on progress; may occur during routine work.

Eight items omitted from the reference sets enter the alternative specifications because their content is relevant to the comparison: I7 (theory-application tips), I8 (role clarification), I9 (encouraging discussions), I19 (monitoring), I21 (professional-conduct insights), I26 (initial orientation), I27 (task, workload, and deadline discussion), and I28 (task coaching). Their omission from the reference comparison limits its scope; it does not imply that they lack educational value. Table B2 summarises the remaining omitted content.

Table B2: Other Omitted Items and Content Rationale

Other omitted items	Dominant content and reason
I3, I17	Appropriate treatment and respect: relational conduct rather than a specified instructional or review activity.
I5, I12	Empathy and reassurance: psychosocial encouragement beyond the focused activity contrast.
I11, I14, I15, I25	Boundaries, overwhelm, recovery, and distress: well-being and personal-distress support.
I13, I16, I23	Approachability, general support, and consistency: access and continuity across supervisory functions.
I24, I29	Fairness and confidentiality: ethical and relational conditions of supervision.

Table B3 specifies the alternative group memberships. Each single-item omission in Table 3 removes only the named item from the reference set and retains all other assignments.

Table B3: Alternative Item Group Memberships

Specification	Routine items	Developmental items
Move I20 to routine	I1, I2, I4, I6, I20	I10, I18, I22, I30
Expand routine set	I1, I2, I4, I6, I8, I19, I26, I27	I10, I18, I20, I22, I30
Expand developmental set	I1, I2, I4, I6	I7, I9, I10, I18, I20, I21, I22, I28, I30
Expand both sets	I1, I2, I4, I6, I8, I19, I26, I27	I7, I9, I10, I18, I20, I21, I22, I28, I30
Broad sets with task coaching in routine	I1, I2, I4, I6, I8, I19, I26, I27, I28	I7, I9, I10, I18, I20, I21, I22, I30

Appendix C Ancillary Associations With Experience

These ancillary analyses describe associations with the overall 30-item index. They are separate from the three main objectives and do not explain the contrast between the routine-guidance and developmental-support sets.

Spearman correlations were calculated between each respondent's 30-item mean and years of work experience, years supervising OJT students, and recorded student count. Holm adjustment treated the three two-sided p values as one family [13]. Marginal 95% percentile bootstrap confidence intervals used 10,000 respondent-level resamples and random seed 20261015; they were not adjusted for multiplicity. The student-count variable was analysed as recorded and was not interpreted as concurrent supervisory workload.

Table C1 reports the coefficients, confidence intervals, and p values. Every interval included zero, and none of the adjusted p values was below .05. These analyses do not test associations with either exploratory item-set mean or their paired difference.

Table C1. Spearman Correlations Between Experience Indicators and Overall Ratings

Experience or exposure indicator	Spearman	95% CI	p	Holm-adjusted p
Years of work experience	N0.082	ONPR 0.062]	0.251	0.752
Years supervising OJT students	N0.040	ONPR 0.101]	0.572	1.000
Recorded number of OJT students supervised	0.011	ONPR 0.159]	0.881	1.000

Note. CI = confidence interval. Two-sided p values are reported. The summary score is each respondent's mean across all 30 items. CIs are marginal 95% percentile bootstrap intervals from 10,000 respondent-level resamples and are not adjusted for multiplicity. Holm adjustment applies to the three p values.