IIAI Open Conference Publication Series
IIAI Letters on Institutional Research
Vol. 005, LIR446
DOI: https://doi.org/10.52731/lir.v005.446

Designing Behavioral Indicators for Assessing Learning Outcomes Based on Diploma Policy

Sayaka Matsumoto [*†], Kunihiko Takamatsu [*†],
Shotaro Imai [*], Tsunenori Inakura [*], Masao Mori [*]

Abstract

In higher education, assessing learning outcomes based on diploma policy (DP) is essential for ensuring educational quality. However, in Japan, such assessments often rely on subjective self-evaluation, which is prone to bias, contextual effects, and gaps between perception and actual behavior. This study proposes a method to visualize and deconstruct the abstract language of Diploma Policy (DP) statements using the principles of ontology engineering, followed by the development of concrete behavioral indicators through expert discussion. The target DP, defined by the former Tokyo Institute of Technology, spans the bachelor's, master's, and doctoral programs. A total of 24 behavioral indicators were developed across six domains: specialist skills, liberal arts (1 and 2), communication skills, and applied skills (inquiry/problem setting and practice/problem solving). These indicators visualize student competencies based on the frequency of observed behaviors, as an alternative or complement to self-perception. The framework is applicable not only at graduation but also during a program, allowing timely feedback for educational improvement and student support. This study presented a novel approach for evaluating DP achievement in a more objective and multidimensional manner. It also holds potential as a complementary framework for designing behavioral indicators and surveys in educational systems that already use outcome-based assessments.

Keywords: Diploma Policy (DP), Assessment of Learning Outcomes, Behavioral Indicators, Institutional Research (IR)

1 Introduction

In higher education, assessment of student learning outcomes plays a central role in improving instruction and ensuring educational quality. One core approach is the diploma policy (DP), which outlines the competencies required for graduation and degree referral.

In Japan, learning outcome assessments based on the DP have gained importance through the Ministry of Education's Guidelines for Academic Management [1][2]. However, many universities still rely on subjective self-assessment, typically asking students whether they "are able to" perform certain tasks at the time of graduation. Although these self-assessments are easy to implement, their validity and reliability have been questioned. In particular, asking students who have already been approved for graduation to confirm whether they have achieved specific outcomes can introduce logical contradictions [3].

[*] Institute of Science Tokyo, Tokyo, Japan
[†] These authors contributed equally to this work.

International frameworks place greater emphasis on objective evidence-based assessments. For example, the OECD's AHELO [4] and iPAL [5] projects evaluated student competencies using task-based methods. Regional frameworks such as the EQF [6], MQF [7], and HKQF [8] promote comparability by defining qualifications based on learning outcomes. The ASEAN Qualifications Reference Framework (AQRF) [9] likewise supports cross-country alignment of outcome-based qualifications.

These initiatives share a common feature: learning outcomes are not measured through students' self-perception alone but through demonstrated behaviors and deliverables. They highlighted the importance of capturing learning through observable performance rather than relying solely on internal judgments.

Despite the growing interest in outcome-based evaluations in Japan, several structural issues remain. First, DPs are defined independently by each institution and tend to use abstract and ambiguous language [3], making it difficult to design clear behavior-oriented questions. Second, although the MEXT guidelines [1][2] encourage outcome-based assessment and continuous improvement cycles, they do not mandate standardized evaluation methods. Consequently, most universities have broad discretion, and owing to limited resources, they often rely on easily administered self-report questionnaires. However, this approach presents several challenges: it is vulnerable to bias and over- or under-estimation; it may diverge from students' actual performance; and it is susceptible to mood, context, and framing effects.

Although self-assessment is valuable for capturing students' perceived growth, it is insufficient for Institutional Research (IR) or formal accountability, which require objective and verifiable indicators.

This study aimed to establish a new framework for evaluating DP achievement based on observable behaviors rather than subjective perception. To accomplish this, we adopted principles from ontological engineering [13] as a conceptual tool to systematically deconstruct abstract DP statements and visualize their internal structure. By translating DP statements into concrete behavioral components, this study sought to clarify which actions demonstrate achievement and to provide a more reliable and actionable basis for learning outcome assessment.

2 Methods

This study focused on the diploma policy (DP) of the former Tokyo Institute of Technology [14], which defined five competency domains across bachelor's, master's, and doctoral programs: specialist skills, liberal arts skills, communication skills, applied skills (inquiry and problem setting), and applied skills (practice and problem solving).

Structural commonalities in the DP statements were identified and deconstructed into behavioral elements. The survey items were designed to apply across programs, assuming more frequent behavioral expressions at higher academic levels.

Unlike the other domains, the Liberal Arts domain included two descriptors per level. Thus, we treated them as two subdomains: Liberal Arts 1 and 2.

To translate abstract competencies into behavioral indicators, we followed four steps:

(1) Parallel comparison of DP texts across programs to identify common terms and developmental distinctions (e.g., "possess a foundation," "deepen expertise," "generate knowledge");

(2) A conceptual diagram to illustrate the relationships between actors, actions, and objects, inspired by ontological engineering [13] as a conceptual framework.

(3) Tabular rewriting of objects and actions for clarity across levels.

(4) Decomposition criteria focusing on targets, actions, levels, and conceptual hierarchy.

From these, behavioral indicators were derived through collaborative author discussions. For example, in the specialist skills domain, we identified research- or development-related actions that embody DP. The indicators were designed to reflect actual behavior, not intent or ability, and to be concise, generalizable across programs, and suitable for surveys.

Each item was rated on a five-point frequency scale ranging from "Never" to "Always" to assess how frequently students exhibit each behavior. Sample items for specialist skills included writing reports, presenting and publishing papers, and teaching. This represents progression from basic to advanced competencies.

This methodology enables DP assessment based on observable behaviors rather than on subjective self-reporting. It combines conceptual structuring inspired by ontological engineering with evidence-based evaluation to offer an actionable framework for tracking student development.

3 Results and Discussion

This section focuses on specialist skills as an example of elemental decomposition. We compared diploma policy (DP) statements for bachelor's, master's, and doctoral programmes, identifying common targets and level-specific actions. Figure 1 illustrates the decomposition process and the results.

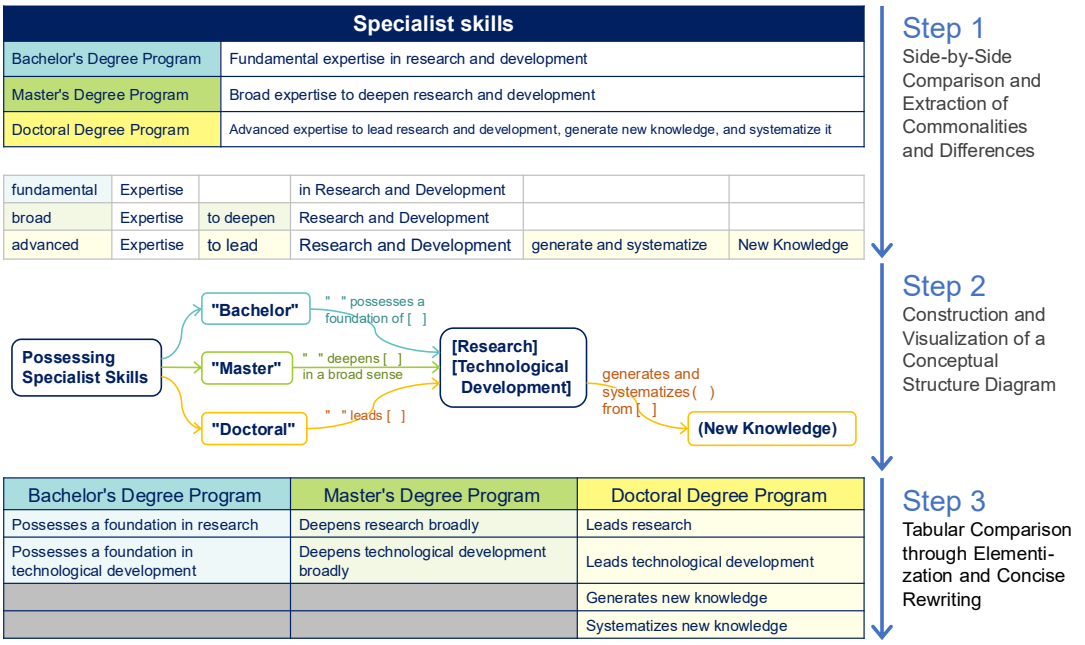

Figure 1: Conceptual Structure Diagram of the Element Decomposition for Specialist Skills

All levels share "research and development" as the core target. However, the associated actions evolve: "possess a foundation" (bachelor), "deepen expertise" (master), and "lead," "generate," and "systematize knowledge" (doctoral). These terms reflect the developmental progression of the DP.

To clarify this, we extracted the subject, object, and action from each level and visualized their relationships through a conceptual diagram inspired by ontological engineering. We also rewrote them into concise object-verb pairs and arranged them in a comparative table showing the progression in expected behaviors.

While this section presents specialist skills, the same procedure was applied to Liberal Arts 1, Liberal Arts 2, Communication skills, Applied skills 1 (Exploratory/Problem-Setting), and Applied skills 2 (Practical/Problem-Solving). A full list of behavioral indicators is presented in the next section.

To capture diploma policy (DP) competencies at the behavioral level, this study derived a set of behavioral indicators for each domain. The full list of indicators is shown in Table 1.

Table 1: List of Behavioral Indicators

Behavioral Indicators for Specialist Skills
1. Summarize knowledge in one's area of specialization in a report or paper
2. Present knowledge in one's area of specialization clearly to others
3. Write a research paper in one's area of specialization and publish it in a journal
4. Teach a lecture in one's area of specialization
Behavioral Indicators for Liberal Arts 1
1. Read books from a wide range of fields (e.g., history, literature) beyond one's own specialty
2. Express opinions or write essays on social issues using knowledge from one's specialty
3. Exchange ideas with people from culturally different backgrounds
4. Conduct academic activities (e.g., reading, presentations, writing) in a language other than one's native language
Behavioral Indicators for Liberal Arts 2
1. Engage in research or social activities with a strong sense of ethics
2. Try new things in daily academic or personal activities
3. Consistently investigate social issues and collect information to help solve them
4. Conduct joint research with groups that include people from other disciplines
Behavioral Indicators for Communication Skills
1. Explain one's claims, reasoning, and evidence in a logical and coherent manner
2. Adjust one's language and delivery based on the audience
3. Foster mutual understanding and respect among group or team members
4. Coordinate diverse opinions and lead the group toward better outcomes
Behavioral Indicators for Applied Skills 1 (Inquiry and Problem-Setting Skills)
1. Observe issues and objects from multiple perspectives
2. Explain phenomena logically and consistently according to established principles
3. Study how theories in one's discipline are constructed and applied
4. Identify the essence or universality of phenomena and formulate new research questions
Behavioral Indicators for Applied Skills 2 (Practice and Problem-Solving Skills)
1. Generate effective and innovative ideas without being bound by conventional thinking
2. Apply specialized knowledge to solve real-world problems
3. Combine knowledge and skills from multiple fields to address real-world problems
4. Present research findings at conferences or in papers and share them with society

All behavioral indicators developed in this study assessed competencies through observable behavior rather than through self-perception. The domains emphasize different aspects: specialist skills progress from foundational tasks to advanced outputs, Liberal Arts combine inner attitudes and external engagement, communication skills focus on interaction and leadership, Applied skills 1 stress theory and exploration, and Applied skills 2 focus on solving real-world problems. These indicators support behavior-level assessments and help visualize students' growth and strengths.

This study proposes a behavioral indicator-based assessment as an alternative to subjective self-assessment. While the latter captures students' perceptions and is easy to implement, it suffers from bias, inconsistency with actual behavior, and vulnerability to mood or context effects [10][11][12]. It also risks logical contradictions when used with students who are already deemed qualified to graduation [3].

In contrast, the proposed method evaluates competencies based on the frequency of observable behaviors. This enables objective, practice-based assessments, continuous monitoring during the program (not just at the end), and timely feedback on educational improvement and support. Institutional Research (IR) serves as a tool to track learning outcomes and enhancing quality assurance.

Both assessment methods, subjective self-assessment and behavioral indicator-based evaluation, have merits and limitations. Subjective self-assessment effectively captures students' perceived growth and is simple to implement. However, it is prone to bias, may not reflect actual behaviors, and can be influenced by context. In contrast, behavioral indicators offer objective, behavior-based evaluations and enable longitudinal tracking, but they require attention to differences in students' opportunities to engage in the assessed behaviors. Rather than treating these approaches as mutually exclusive, we advocate their complementary use. Combining self-perceived growth with evidence-based behavioral outcomes allows for a more robust and balanced assessment, supporting educational quality assurance, personalized learning, and career support.

While combining self-assessment and behavioral indicators may offer a more comprehensive view of student development, practical considerations must also be taken into account. Participation in student surveys related to learning outcomes is typically voluntary, and excessive length or complexity can discourage responses or reduce data quality. Therefore, in cases where only a limited number of items can be included, we argue that priority should be given to behavioral indicators. They provide more objective and verifiable evidence of learning, especially when compared to subjective perceptions. Nevertheless, we do not deny the value of self-assessment as a complementary measure, particularly in future studies aiming for multi-perspective evaluation.

Although developed in the Japanese context, the method also holds potential for application in countries with existing outcome-based systems as a framework for behavioral indicator design, item construction, and alignment with institutional goals.

4 Conclusion

To address the limitations of subjective self-assessment, this study proposes a new evaluation method based on observable behaviors. Using the diploma policy (DP) of the former Tokyo Institute of Technology, we adopted principles from ontological engineering as a conceptual tool to deconstruct DP competencies into target, action, and achievement levels. Through collaborative discussions, we developed 24 behavioral indicators across six domains and designed survey items to measure behavioral frequency. This resulted in a new framework for visualizing DP

achievement as a real-world behavior rather than as self-perception.

This method offers new perspectives for DP assessment. First, it enables an objective, evidence-based evaluation grounded in behavior. Second, it supports continuous monitoring during academic programmes, provides timely feedback, and enhances Institutional Research. Third, by combining behavioral data with self-assessment, institutions can gain a more holistic view of student development, supporting both quality assurance and learning. While both approaches have value, we believe behavioral indicators should be prioritized when item length must be limited, due to their objectivity and action-based nature.

However, this approach has certain limitations. Differences in students' opportunities to engage in target behaviors may affect outcomes, and adapting indicators to other institutions or disciplines may require modifications. In particular, future research should include pilot testing across diverse academic contexts to evaluate the reliability (e.g., internal consistency) and validity (e.g., content and criterion validity) of the indicators.

It is also important to recognize that while the ontological structuring used to visualize DP competencies can be widely applied, the process of generating specific behavioral indicators through expert discussion is inherently influenced by contextual factors. These include the characteristics of the institution, academic discipline, socio-cultural setting, and prevailing educational policies. As such, the indicators should not be treated as universally fixed; rather, they require periodic revision and triangulation with other data sources such as academic records or performance evaluations to ensure ongoing relevance and validity.

To address these challenges, the framework should remain flexible and allow for contextual customization. In addition to national use, we also explored international applications by aligning them with global practices and refining the structure as a tool adaptable to various educational systems.

Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 22H00077.

References

[1] Subcommittee on Universities, Central Council for Education, "Guidelines for academic management," [in Japanese], 2020. [Online]. Available: https://www.mext.go.jp/b_menu/shingi/chukyo/chukyo0/toushin/1411360_00001.html [Accessed: Apr. 28, 2025].

[2] Subcommittee on Universities, Central Council for Education, "Supplement to the guidelines for academic management," [in Japanese], 2023. [Online]. Available: https://www.mext.go.jp/content/20230228-mxt_daigakuc01-000004749_1.pdf [Accessed: Apr. 28, 2025].

[3] S. Matsumoto, K. Takamatsu, S. Imai, T. Inakura, and M. Mori, "Survey of current methods for assessing the achievement of diploma policy," [in Japanese], Proc. 13th Meeting on Japanese Institutional Research, pp. 133–139, 2024. https://doi.org/10.50956/mjir.13.0_133_1.

[4] OECD, Assessment of Higher Education Learning Outcomes (AHELO) Feasibility Study

Report, 2013.

[5] R. J. Shavelson, O. Zlatkin-Troitschanskaia, and J. W. Caudill, International Performance Assessment of Learning in Higher Education (iPAL). Springer, 2018.

[6] European Commission, The European Qualifications Framework for Lifelong Learning (EQF), 2008.

[7] Malaysian Qualifications Agency, Malaysian Qualifications Framework (2nd ed.), 2017.

[8] Education Bureau, HKSAR, Introduction to the Hong Kong Qualifications Framework (HKQF), 2008.

[9] ASEAN Secretariat, ASEAN Qualifications Reference Framework, 2015.

[10] G. T. L. Brown, H. L. Andrade, and F. Chen, "Accuracy in student self-assessment: Directions and cautions for research," Assessment in Education: Principles, Policy & Practice, vol. 22, no. 4, pp. 444–457, 2015. http://dx.doi.org/10.1080/0969594X.2014.996523.

[11] L. Jamain, A.-L. de Place, T. Bouffard, and P. Pansu, "Students' self-evaluation bias of school competence and its link to teacher judgment in elementary school," Social Psychology of Education, vol. 24, pp. 919–938, 2021. https://doi.org/10.1007/s11218-021-09638-7.

[12] J. Kruger and D. Dunning, "Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments," Journal of Personality and Social Psychology, vol. 77, no. 6, pp. 1121–1134, 1999. https://doi.org/10.1037/0022-3514.77.6.1121.

[13] T. Gruber, "A translation approach to portable ontology specifications," Knowledge Acquisition, vol. 5, no. 2, pp. 199–220, 1993. https://doi.org/10.1006/knac.1993.1008.

[14] Tokyo Institute of Technology, "Degree policy," [Online]. Available: https://www.titech.ac.jp/english/0/about/policies/education [Accessed: Apr. 30, 2025].