A Structured Approach for Aggregating Assessment Evidence for ABET Student Outcome 1

Dr. Jitesh H Panchal, Purdue University – West Lafayette (College of Engineering)

Dr. Jitesh H. Panchal is a Professor and Associate Head of Mechanical Engineering at Purdue University. He received his BTech (2000) from the Indian Institute of Technology (IIT) Guwahati, and MS (2003) and PhD (2005) in Mechanical Engineering from Georgia Institute of Technology. Dr. Panchal's research interests are in (1) design at the interface of social and physical phenomena, (2) computational methods and tools for digital engineering, and (3) secure design and manufacturing. He is a recipient of CAREER award from the National Science Foundation (NSF); Young Engineer Award, Guest Associate Editor Award, and three best paper awards from ASME; and was recognized by the B.F.S. Schaefer Outstanding Young Faculty Scholar Award, the Ruth and Joel Spira Award, and as one of the Most Impactful Faculty Inventors at Purdue University. He received the Distinguished Alumni award from IIT Guwahati. He is a co-author of two books and has co-edited one book on engineering systems design. He has served on the editorial board of international journals including ASME Journal of Mechanical Design, ASME Journal of Computing and Information Science in Engineering.

Dr. Carl Wassgren, Purdue University – West Lafayette (College of Engineering)

Dr. Carl Wassgren is the Charles E. and Louise W. Springer Professor for Excellence in Teaching and Associate Head of Mechanical Engineering at Indianapolis. He is a Fellow of Purdue's Teaching Academy and is in Purdue's Book of Great Teachers. He has received multiple teaching honors including the Charles B. Murphy Award, which is Purdue's highest undergraduate teaching honor. Professor Wassgren's research interests focus on modeling the dynamics of particulate and powder systems, with applications in mixing, segregation, coating, attrition, compaction, granulation, and characterization. This research has been applied to systems used to process pharmaceuticals, food and consumer products, agrochemicals, and biomass materials. He has received an NSF CAREER award and was a co-founder and Director of the Center for Particulate Products and Processes (CP3) at Purdue.

Dr. Eckhard A. Groll, Purdue University – West Lafayette (College of Engineering)

Dr. Eckhard A. Groll is the Reilly Distinguished Professor of Mechanical Engineering and the William E. and Florence E. Perry Head of Mechanical Engineering at Purdue University. He joined Purdue University as an Assistant Professor in 1994 and was promoted to Associate Professor in 2000, to Professor in 2005, to Reilly Professor in 2013, and to Reilly Distinguished Professor in 2024.

Dr. Sameer V. Naik, Purdue University – West Lafayette (College of Engineering)

A Structured Approach for Aggregating Assessment Evidence for ABET Student Outcome 1

Abstract

ABET requires engineering programs to provide direct evidence demonstrating that graduates have attained the knowledge and skills defined by the student outcomes. A persistent practical challenge for programs is determining how to sample and aggregate assessment evidence across courses and artifacts to reliably reflect students' ability to solve complex engineering problems. This paper develops an approach for aggregating assessment evidence aligned with ABET Student Outcome 1, which emphasizes the ability to identify, formulate, and solve complex engineering problems. The proposed approach introduces *outcome-weighted aggregation*, in which individual assessment artifacts such as exam problems, sub-problems, and homework assignments are weighted according to both their instructional importance and their relevance to characteristics of complex engineering problems as defined by ABET. By explicitly mapping artifacts to the cognitive actions embedded in Student Outcome 1 and to dimensions of problem complexity, the approach preserves transparency and traceability while enabling principled aggregation across assessments and courses. The application of the approach is demonstrated using a set of fundamental undergraduate mechanical engineering courses. It is argued that outcome-weighted aggregation across carefully selected assessments provides broader, more stable, and more representative evidence of outcome attainment than narrowly sampled artifacts alone, while remaining fully aligned with ABET's expectations for direct, documented assessment. The proposed approach offers engineering programs a defensible method for selecting and aggregating assessment evidence that balances specificity with reliability in support of continuous improvement.

1 Introduction

Continuous improvement in engineering education relies on systematic assessment of student outcomes. Within ABET-accredited programs, direct evidence of student performance is required to demonstrate that graduates attain the competencies specified in Criterion 3[1]. A widely cited guideline in ABET practice discourages the use of overall course grades as evidence of student outcome attainment[2]. The underlying concern is well founded: a course grade is typically designed to evaluate course-specific learning objectives and often aggregates performance across multiple outcomes, as well as factors unrelated to learning, such as attendance, participation, or extra credit. As a result, raw course grades are poorly aligned with any single ABET student outcome and lack the interpretability required for outcome-based assessment.

In response to this guidance, many programs rely almost exclusively on narrowly defined assessment artifacts including specific homework problems, exam questions, or rubric items as direct measures of student outcomes. These artifacts offer clear traceability between the assessment and the outcome being measured. However, this practice introduces its own challenges. Individual artifacts capture only a limited portion of the intended outcome, are sensitive to task design and grading variability, and often rely on small samples of student work. While such measures may be precise in their alignment, they are frequently limited in scope and statistical reliability. In emphasizing specificity, programs risk underrepresenting the breadth and integrative nature of certain outcomes.

This tension is particularly pronounced for ABET Student Outcome 1[3]: "an ability to identify, formulate, and solve complex engineering problems by applying principles of engineering, science, and mathematics"[4]. Unlike skill-specific outcomes, Outcome 1 represents a broad cognitive construct that develops across multiple contexts and over time. In mechanical engineering, this ability is cultivated through a sequence of core analytical courses that collectively define engineering fundamentals, e.g., statics, dynamics, thermodynamics, fluid mechanics, heat and mass transfer, and control systems. Mastery of Outcome 1 is therefore expressed not in isolated tasks, but through sustained performance across assessments that require integration of modeling, analysis, and problem-solving skills.

The purpose of this paper is to examine how assessment evidence from various courses can be used appropriately, and in an integrated manner, to assess Student Outcome 1 while remaining fully consistent with ABET's emphasis on direct, outcome-specific assessment. We argue that the central challenge lies not in aggregation itself, but in ensuring that aggregation is explicitly aligned with the outcome being measured. Rather than treating all assessment components equally or the weights used to assign course grades, we propose an outcome-aligned aggregation approach in which individual assessment artifacts, e.g., exam problems, sub-problems, and homework assignments, are weighted according to both their *instructional importance* and their *relevance to complex engineering problem solving* as defined by ABET. Under this formulation, artifacts that explicitly require students to identify, formulate, and solve complex engineering problems contribute more strongly to the overall outcome measure, while artifacts that primarily assess other learning objectives are down-weighted or excluded. This approach preserves the transparency and traceability expected in ABET assessment while enabling principled synthesis of evidence across assessments and courses.

2 Background

2.1 Assessment of ABET Outcomes

ABET accreditation is grounded in Criterion 4, which requires programs to demonstrate continuous improvement based on the assessment of student outcomes[1]. It emphasizes both direct and indirect measures of student achievement. Direct measures include rubric-based evaluation of assignments, projects, or exams, while indirect measures include surveys of students, alumni, and employers. Regardless of method, ABET emphasizes that evidence must be systematic, documented, and aligned with the intended outcome.

Over time, a common heuristic has emerged in accreditation practice: overall course grades should not be used to assess outcomes[2]. The rationale is that a final grade is an amalgam of multiple elements (exams, homework, projects, participation) that may map to more than one outcome or even to factors unrelated to outcomes (e.g., attendance or extra credit). Consequently, evaluators often recommend assignment-level evidence, arguing that it provides clearer alignment to the specific constructs being measured. This stance has been echoed in numerous ABET workshops and training sessions for program evaluators.

Literature on assessment highlights two important criteria: *validity*, or the degree to which evidence measures the intended construct, and *reliability*, or the stability of measurement across contexts[5]. While assignment-level data can provide precise alignment, assessment scholars have long emphasized that precision alone does not guarantee validity. According to Besterfield-Sacre et al.[6], validity in outcome assessment requires that evidence reflect the construct in its full scope. Narrow artifacts may satisfy alignment requirements but risk underrepresenting the breadth of knowledge and skills intended by an outcome. Reliability is also a concern: small samples, variability in instructor assignment design, and differences in grading practices can compromise the stability of assignment-level measures across time and cohorts.

In the engineering education literature, concerns about the overuse of narrow measures have been raised repeatedly. Shuman et al.[7] note that outcomes assessment must balance directness with authenticity, ensuring that evidence reflects sustained student performance in realistic contexts. Similarly, Rogers[2] highlights the need for multiple forms of evidence, cautioning against single artifacts that overemphasize compliance at the expense of meaningful evaluation. These perspectives suggest the need for aggregating diverse evidence from different courses.

2.2 Overview of ABET Student Outcome 1

As part of the transition from the earlier *(a)–(k)* outcomes to the current *(1)–(7)* framework, ABET refined both the wording and interpretation of Student Outcome 1 to place explicit emphasis on *complex engineering problems*. ABET further clarifies that complex engineering problems include “one or more of the following characteristics: involving wide-ranging or conflicting technical issues; having no obvious solution; addressing problems not encompassed by current standards and codes; including many component parts or sub-problems; involving multiple disciplines; or having significant consequences in a range of contexts”[4]. This clarification reflects a shift away from viewing engineering problem-solving as the execution of isolated analytical procedures, toward viewing it as an integrative cognitive activity that unfolds across multiple dimensions of complexity[8].

To illustrate this distinction, consider two representative examples from mechanical engineering. A typical textbook problem requiring students to compute the steady-state heat flux through a one-dimensional composite wall with prescribed boundary conditions, material properties, and geometry constitutes a *non-complex* engineering problem. While such a problem may be analytically challenging, it has a clearly defined solution pathway, relies on established models and assumptions, and involves a limited number of interacting components. In contrast, a problem requiring students to design a thermal management strategy for an electronic system—balancing conduction, convection, and radiation under uncertain operating conditions, accounting for material constraints, and evaluating performance trade-offs—exhibits multiple characteristics of a complex engineering problem. The latter problem does not admit a single obvious solution, requires integration across disciplines, and has consequences for system reliability and safety.

Each element of ABET's definition of complex engineering problems captures a distinct aspect of this broader cognitive demand. "Wide-ranging or conflicting technical issues" refers to situations in which multiple physical phenomena or performance objectives must be considered simultaneously, often requiring trade-offs rather than optimization of a single metric. "Having no obvious solution" emphasizes problems for which the governing approach or modeling assumptions are not immediately apparent, requiring students to exercise judgment in problem formulation. "Problems not encompassed by current standards or codes" involve scenarios where handbook solutions or prescriptive methods are insufficient, forcing engineers to reason from first principles or adapt existing models.

Similarly, "including many component parts or sub-problems" highlights problems in which partial solutions must be synthesized into a coherent whole, with errors or assumptions propagating across stages of analysis. "Involving multiple disciplines" recognizes that authentic engineering problems often span traditional curricular boundaries, e.g., combining mechanics, heat transfer, materials science, and control systems. Finally, "having significant consequences in a range of contexts" underscores that complex engineering problems are consequential, with implications for safety, performance, cost, sustainability, or societal impact. These characteristics align closely with research in cognitive psychology and engineering education on complex problem solving, which emphasizes uncertainty, interdependence, nonlinearity, and the need for iterative reasoning[9,10].

From this perspective, Student Outcome 1 is not a narrowly defined skill that can be captured by a single assignment or exam problem. Rather, it represents a latent cognitive ability that develops across multiple learning experiences and manifests through sustained performance on integrative tasks. This interpretation has important implications for assessment. Because complex engineering problem solving is inherently distributed across contexts and time, evidence of attainment must reflect not only isolated demonstrations of analytical proficiency, but also consistent performance across assessments that collectively embody the defining characteristics of complexity articulated by ABET.

3 Proposed Approach

The proposed approach operationalizes ABET Student Outcome 1 by explicitly decomposing both the outcome and the assessment artifacts used to evaluate it. Rather than treating complex engineering problem solving as a monolithic skill, we interpret Outcome 1 as comprising three interrelated cognitive actions—*identifying*, *formulating*, and *solving* engineering problems—each of which may be exhibited to varying degrees within a given assessment task. In parallel, we adopt ABET's definition of complex engineering problems as a set of six distinct characteristics, recognizing that no single problem or assignment is expected to exhibit all characteristics simultaneously.

For each assessment artifact considered in this study (exam problems, sub-parts, and selected homework assignments), we conduct a structured mapping along two dimensions. First, each problem or sub-problem was evaluated for whether it required students to identify relevant physical principles and constraints, formulate an appropriate mathematical or computational model, and/or execute a solution using engineering analysis. Second, the same artifacts were mapped to ABET's six characteristics of complex engineering problems (CEP):

1. wide-ranging or conflicting technical issues,
2. absence of an obvious solution path,
3. problems not fully encompassed by existing standards or codes,
4. multiple component parts or sub-problems,
5. involvement of multiple engineering disciplines, and
6. significant consequences in one or more contexts.

This mapping was performed at the problem and sub-problem level, to ensure transparency and traceability between evidence and outcome definitions. Importantly, the intent was not to label individual problems as either "complex" or "non-complex," but to characterize the *extent and manner* in which each artifact contributes evidence toward different aspects of complex engineering problem solving. Some artifacts emphasize formulation under uncertainty, others emphasize multi-step solution synthesis, and others highlight disciplinary integration or consequential decision-making.

Once outcome-aligned mappings have been established at the problem or sub-problem level, the next question is how to aggregate this evidence into a meaningful indicator of Student Outcome 1 attainment. We argue that aggregation should reflect two distinct considerations: *instructional weighting* and *outcome relevance*. The instructional weighting component is the weight originally assigned to an assessment artifact within the course. Exams, projects, homework assignments, and individual problems are typically weighted by the instructor to reflect their relative importance to the course learning objectives. These weights encode pedagogical intent and ensure that more consequential assessments exert proportionally greater influence on any aggregated measure. Retaining the original course-level weighting preserves alignment with the instructional context and avoids post hoc reinterpretation of student performance.

To quantify instructional weight, we compute the contribution of each problem (or sub-problem) to the overall curriculum by combining its weight within the course with the credit value of the course itself. Specifically, the percentage weight of a problem in the total course score is multiplied by the number of course credits. This yields a normalized measure of instructional importance that reflects both the instructor's grading intent and the curricular significance of the course. This formulation captures an important distinction in outcome assessment. A problem that constitutes a small fraction of a high-credit, analytically intensive course may represent more substantive evidence of outcome attainment than a problem with a similar percentage weight in a lower-credit course. By incorporating course credits, the instructional weight accounts for differences in contact hours, expected student effort, and curricular emphasis, enabling meaningful comparison and aggregation across courses with differing credit structures. This approach does not privilege any single course a priori; rather, it provides a transparent and quantitative mechanism for situating individual assessment artifacts within the broader curriculum.

The second component (outcome relevance) reflects the relevance of an assessment artifact to complex engineering problem solving as defined by ABET. As described earlier, each problem or sub-problem is mapped to the six characteristics of complex engineering problems. The degree of relevance can be operationalized by counting the number of characteristics exhibited by the artifact, or by assigning differentiated relevance scores based on the nature of the characteristics present. Artifacts that exhibit multiple characteristics—such as multi-step structure, disciplinary integration, and non-obvious solution paths—are treated as stronger evidence of Outcome 1 attainment than artifacts that assess primarily procedural or isolated analytical skills.

The combined contribution of assessment artifact i in course j to Student Outcome 1 is defined as

$$w_{ij}^{\text{SO1}} = w_{ij}^{\text{instr}} \times r_{ij}^{\text{CEP}}, \tag{1}$$

where $w_{ij}^{\text{instr}} = p_{ij} C_j$ represents the instructional weight based on the artifact's contribution to the course grade and course credit value, and r_{ij}^{CEP} denotes the relevance factor associated with the number (or weighted number) of complex engineering problem characteristics exhibited by the artifact.

Under this formulation, the contribution of a given artifact to Outcome 1 assessment is proportional to both its instructional weight and its relevance to complex engineering problem solving. Aggregation across problems, exams, and courses is therefore not a uniform averaging of scores, but a structured synthesis of weighted evidence that reflects both curricular intent and outcome alignment.

4 Illustrative Example

An overview of a Bachelor of Science in Mechanical Engineering (BSME) program at Purdue University is provided in Figure 1. The courses selected for this study—ME 31500 (Heat and Mass Transfer), ME 35400 (Machine Design), and ME 36500 (Systems, Measurements, and Control I)—were chosen because they collectively represent the core analytical foundation of the undergraduate mechanical engineering curriculum and provide repeated, authentic opportunities for students to demonstrate Student Outcome 1. Each course emphasizes rigorous problem solving grounded in engineering science and mathematics, while engaging different modes of complexity: ME 31500 focuses on multi-physics transport phenomena and non-linear modeling; ME 35400 emphasizes mechanical design under combined loading, failure criteria, and real-world constraints; and ME 36500 integrates dynamic systems modeling, control, and cross-domain abstraction. Together, these courses span multiple semesters and disciplinary perspectives, allowing complex engineering problem solving to be observed across diverse contexts rather than isolated tasks. As such, they provide a representative and complementary set of assessment environments for examining outcome-aligned aggregation of evidence for Student Outcome 1.

4.1 ME 31500: Heat and Mass Transfer

ME 31500 (Heat and Mass Transfer) is a 4-credit course focused on the fundamentals of heat transfer by conduction, convection, and radiation, as well as mass transfer by convection, with relevance to diverse engineering applications. Exam 3 is the final exam in the course. It emphasizes modeling and problem formulation in contexts where solution pathways are not immediately prescribed, requiring students to make informed assumptions, interpret physical significance, and integrate multiple modes of heat transfer. Problems span topics such as spectral emissivity and absorptivity, radiation exchange between surfaces, view factors, and numerical solution approaches, often situated in aerospace or industrial thermal environments. Collectively, the exam requires students to identify dominant physical mechanisms, formulate appropriate governing equations, and execute multi-step solutions with implications for system performance and safety, making it a strong source of direct evidence for complex engineering problem solving under Student Outcome 1. The exam consists of four problems. The characteristics of the problems, as they relate to Student Outcome 1 are listed in Table 1.

4.2 ME 35400: Machine Design

ME 35400 (Machine Design) is a 3-credit course focusing on the design, analysis, and selection of machine components, with a particular emphasis on fluctuating loading conditions. Building

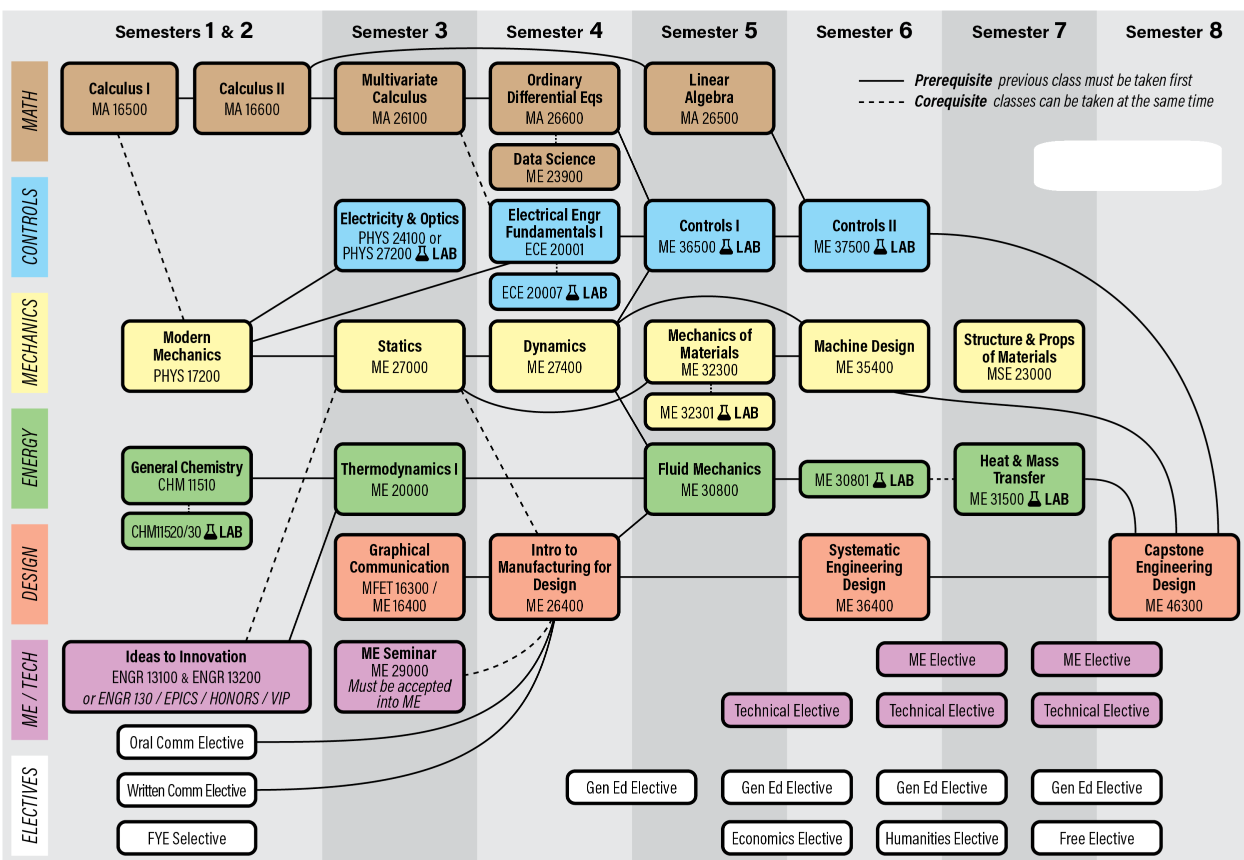

Figure 1: Courses and prerequisites in an BS in Mechanical Engineering Program

Table 1: ME 31500 - Heat and Mass Transfer (Exam 3)

Problem	Identify	Formulate	Solve	CEP-1 Wide / Conflicting Issues	CEP-2 No Obvious Solution	CEP-3 Beyond Standards	CEP-4 Many Sub-Problems	CEP-5 Multiple Disciplines	CEP-6 Significant Consequences
Problem 1	✓	✓	✓	—	Numerical solutions	—	Sensors, fins, coatings	Conduction, radiation, materials	Deep space application
Problem 2	✓	✓	✓	—	—	—	—	Fluid mechanics, heat transfer	—
Problem 3	✓	✓	✓	—	—	—	Multi-part, propagating calculations	—	Space applications, survivability
Problem 4	✓	✓	✓	—	—	Spectral radiation not catalog-driven	—	Combustion, radiation, materials	Manufacturing processes, industrial heating

upon foundational concepts from statics, dynamics, and mechanics of materials, the curriculum teaches students to apply engineering fundamentals to mechanical components and integrate them to form complete systems. Exam 2 in the course emphasizes multi-step problem solving involving combined loading, fatigue and static failure theories, contact mechanics, interference fits, and rolling-element bearings. Students are required to identify governing failure modes, formulate appropriate analytical models using established mechanics and materials principles, and synthesize solutions across multiple sub-problems whose results are interdependent. Many problems situate the analysis in design-relevant contexts where assumptions, safety factors, and parameter choices have direct implications for component performance and reliability. The exam provides direct evidence of students' ability to identify, formulate, and solve complex engineering problems that involve multiple interacting components and disciplinary considerations.

Table 2: ME 35400 - Machine Design (Exam 2)

Problem	Identify	Formulate	Solve	CEP-1 Wide / Conflicting Issues	CEP-2 No Obvious Solution	CEP-3 Beyond Standards	CEP-4 Many Sub-Problems	CEP-5 Multiple Disciplines	CEP-6 Significant Consequences
Problem 1	✓	✓	✓	—	—	—	—	Statics, mechanics, materials, fatigue	Failure prevention, safety
Problem 2	✓	✓	✓	—	—	—	Shafts, bearings, gears, keys	Mechanics, materials, manufacturing	Failure prevention, Safety
Problem 3	✓	✓	✓	—	—	—	Shafts and bearings	Materials, mechanics	Failure prevention, Safety
Problem 4	✓	✓	✓	Load and reliability	—	—	—	—	Failure prevention, Safety, Reliability
Problem 5	✓	✓	✓	Coefficient of friction, tolerances, temperature rise	—	—	—	Triboogy, Manufacturing tolerances, Fluid mechanics, Thermodynamics	Failure prevention, safety

4.3 ME 36500: Systems, Measurements, and Control I

ME 36500 is a three-credit junior-level core course designed to introduce students to the basic principles of systems modeling, analysis, and measurements, with coverage of digital data acquisition, statistical methods, and signal processing. The course emphasizes foundational topics, including system modeling, time and frequency domain analysis, and the statistical interpretation of experimental data. Homework 6 emphasizes system modeling and abstraction, requiring students to derive governing equations, reduce models to dominant dynamics, and interpret system behavior using frequency-domain and time-domain representations. Homework 7 extends this analysis to hydraulic, thermal, and signal-processing systems, including derivation of governing equations for lumped-parameter systems, application of Fourier series to represent periodic inputs and system responses, and analysis of system behavior through transfer functions.

Together, these assignments require students to identify relevant physical principles, formulate mathematical models under simplifying assumptions, and solve interconnected sub-problems whose results inform system-level behavior. They provide direct evidence of complex engineering problem solving involving multi-domain integration, modeling judgment, and abstraction. The details of individual problems are provided in Tables 3 and 4. Note that Problem 3 in Homework 7 is geared towards problem identification, formulation and solution, but does not relate to the six complex engineering systems characteristics.

Table 3: ME 36500 - Systems, Measurement, and Control I (Homework 6)

Problem	**Identify**	**Formulate**	**Solve**	**CEP-1 Wide / Conflicting Issues**	**CEP-2 No Obvious Solution**	**CEP-3 Beyond Standards**	**CEP-4 Many Sub-Problems**	**CEP-5 Multiple Disciplines**	**CEP-6 Significant Consequences**
Problem 1			✓	Accuracy vs. simplicity	Simplifying complex problems	—	—	—	—
Problem 2	✓	✓		Accuracy vs. simplicity	Multiple ways of simplification	—	—	Controls and Mechanics	—
Problem 3		✓		—	—	—	—	Electrical Engineering and Control	—
Problem 4	✓	✓		—	—	—	Multiple components in a system	Electro-mechanical Systems	—

Table 4: ME 36500 - Systems, Measurement, and Control I (Homework 7)

Problem	**Identify**	**Formulate**	**Solve**	**CEP-1 Wide / Conflicting Issues**	**CEP-2 No Obvious Solution**	**CEP-3 Beyond Standards**	**CEP-4 Many Sub-Problems**	**CEP-5 Multiple Disciplines**	**CEP-6 Significant Consequences**
Problem 1	✓	✓		—	—	—	—	Hydraulics and Control	—
Problem 2	✓	✓		—	—	No standards available for Pizza ovens	Multiple components with different materials	Thermo-dynamics, Heat transfer	Food preparation systems
Problem 3	✓	✓	✓	—	—	—	—	—	—

Table 5 illustrates how weights are computed for individual assessment artifacts in support of ABET Student Outcome 1. Each row corresponds to a specific problem from an exam or homework assignment drawn from ME 31500, ME 35400, and ME 36500. The *Points* and *Out of* columns indicate the number of points assigned to the problem and the total points for the corresponding assessment, respectively. The *Weight in Course* column represents the fraction of the overall course grade assigned to that exam/homework, while the *Course Credits* column reflects the credit value of the course within the curriculum. These quantities are combined to compute the *Instructional Weightage*, which captures the relative curricular importance of the problem by accounting for both instructor intent and course credit.

The *Identify*, *Formulate*, and *Solve* columns indicate the fraction of points within each problem dedicated to the corresponding cognitive actions embedded in Student Outcome 1. These fractions allow finer-grained attribution of outcome evidence to different aspects of complex engineering problem solving. The *CES Criteria Met* column records the number of ABET complex engineering problem characteristics exhibited by each problem. Finally, the *ABET Outcome 1 Weight* is computed as the product of the instructional weightage and the number of complex engineering criteria met, providing a quantitative measure of the strength of each artifact as evidence for Outcome 1. Although the weights shown in the table are unnormalized, normalization may be applied when comparing outcome evidence across semesters or cohorts.

Table 5: Evaluation of weights for ABET Student Outcome 1

Assessment Artifact	**Points**	**Out of**	**Weight in Course**	**Course Credits**	**Instructional Weightage**	**Identify**	**Formulate**	**Solve**	**CES Criteria Met**	**ABET Outcome 1 Weight**
ME 31500 Exam 3 Problem 1	25	100	25%	4	0.25	33%	33%	33%	4	1.000
ME 31500 Exam 3 Problem 2	15	100	25%	4	0.15	33%	33%	33%	1	0.150
ME 31500 Exam 3 Problem 3	30	100	25%	4	0.30	33%	33%	33%	2	0.600
ME 31500 Exam 3 Problem 4	30	100	25%	4	0.30	33%	33%	33%	3	0.900
ME 35400 Exam 2 Problem 1	18	100	24%	3	0.13	33%	33%	33%	2	0.259
ME 35400 Exam 2 Problem 2	28	100	24%	3	0.20	33%	33%	33%	3	0.605
ME 35400 Exam 2 Problem 3	18	100	24%	3	0.13	33%	33%	33%	3	0.389
ME 35400 Exam 2 Problem 4	18	100	24%	3	0.13	33%	33%	33%	2	0.259
ME 35400 Exam 2 Problem 5	18	100	24%	3	0.13	33%	33%	33%	3	0.389
ME 36500 Homework 6 Problem 1	25	100	1%	3	0.01	0%	0%	100%	2	0.015
ME 36500 Homework 6 Problem 2	25	100	1%	3	0.01	50%	50%	0%	3	0.023
ME 36500 Homework 6 Problem 3	25	100	1%	3	0.01	0%	100%	0%	1	0.008
ME 36500 Homework 6 Problem 4	25	100	1%	3	0.01	50%	50%	0%	2	0.015
ME 36500 Homework 7 Problem 1	33	100	1%	3	0.01	50%	50%	0%	1	0.010
ME 36500 Homework 7 Problem 2	33	100	1%	3	0.01	50%	50%	0%	4	0.040
ME 36500 Homework 7 Problem 3	34	100	1%	3	0.01	33%	33%	33%	0	0.000

The table also highlights how the proposed weighting scheme differentiates between assessment artifacts. In particular, problems from ME 31500 Exam 3 receive substantially higher ABET Outcome 1 weights than homework problems from ME 36500, primarily because the exam constitutes a larger fraction of the course grade and is situated within a higher-credit, analytically intensive course. By contrast, the homework assignments in ME 36500 carry relatively small instructional weight within the course and, consequently, contribute less to the aggregated outcome measure despite addressing relevant analytical concepts. This behavior is intentional and reflects the framework's goal of aligning outcome evidence with both instructional intent and curricular significance. Additionally, ME 365 Homework 7 Problem 3 receives an ABET Outcome 1 weight of zero because it does not exhibit any of the defined characteristics of complex engineering problems, illustrating how the framework explicitly excludes artifacts that do not contribute meaningful evidence toward Student Outcome 1. The framework effectively filters out low-impact assessments without requiring arbitrary exclusion decisions. As a result, programs can use the analysis as a decision-support tool to determine which artifacts meaningfully contribute to outcome assessment and whether certain low-weight artifacts should be excluded for practicality.

Having estimated the weights of each assessment artifact, we model the distribution of performance on each assessment artifact and use these distributions to infer the distribution of Student Outcome 1 attainment at the program level. With this aggregation approach, each artifact contributes evidence according to its outcome-aligned weight, allowing the aggregate outcome distribution to be constructed as a weighted combination of assessment-level distributions. This is in contrast to aggregating each student's performance and then analyzing the distribution of performance across the student body. Such aggregation is particularly well-suited to curricula in which students encounter outcome-relevant evidence through different courses and pathways.

Table 6 summarizes the estimated probabilities that a value drawn from the aggregated outcome distribution exceeds selected performance thresholds. As expected, the probability of exceeding a given threshold decreases monotonically as the threshold increases, reflecting the overall spread

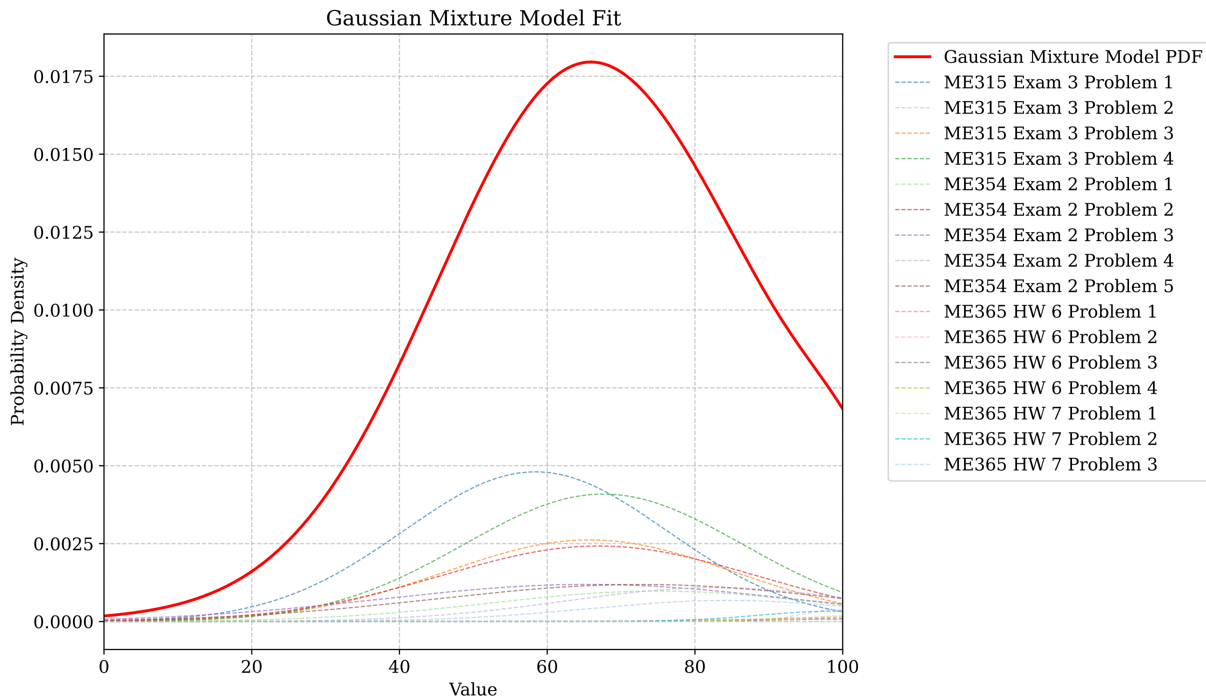

Figure 2: Weighted assessment components and the overall Gaussian Mixture model for Student Outcome 1

of outcome attainment across the program. For example, the model indicates a 70.62% probability that the aggregated outcome measure exceeds 50, while the probability of exceeding 80 drops to 20.82%. Such threshold-based interpretations provide an interpretable way to assess outcome attainment at the program level.

Table 6: Estimated Probability of Exceeding Performance Thresholds for Student Outcome 1

Threshold	P(Aggregate Performance $>$ Threshold)
50	70.62%
60	54.97%
70	37.15%
80	20.82%
90	8.38%

5 Discussion

5.1 Using Weighted Evidence to Inform Curricular Decisions

The probability distribution over the weighted performance measure provides a structured basis for decision-making at multiple levels of the curriculum. Rather than treating assessment results as binary indicators of attainment or non-attainment, the proposed approach enables programs to interpret performance trends in relation to where and how complex engineering problem-solving is emphasized within the curriculum.

At a basic level, weighted evidence can be aggregated to produce an overall indicator of Student Outcome 1 attainment for a cohort. Programs can track this indicator longitudinally to identify trends, evaluate the impact of curricular changes, and distinguish between random variation and systematic shifts in student performance. Beyond program-level summaries, the weighing structure supports diagnostic analysis. By decomposing the aggregate outcome measure back into its constituent components, programs can identify which courses, exams, or types of problems contribute most strongly to evidence of Outcome 1 attainment. For example, if a large fraction of the weighted evidence originates from a small number of courses, this may indicate over-reliance on particular curricular bottlenecks. Conversely, if evidence is broadly distributed but performance is consistently weaker in artifacts exhibiting certain characteristics of complex engineering problems (e.g., formulation under uncertainty or multi-disciplinary integration), this points to specific areas for instructional improvement.

The approach also enables comparison across different dimensions of complexity. Because relevance is quantified through alignment with individual characteristics of complex engineering problems, programs can examine performance conditioned on those characteristics. Such analyses can reveal whether students are proficient at solving multi-step problems yet struggle with problems lacking obvious solution paths, or whether disciplinary integration presents a recurring challenge. These insights can directly inform targeted curricular interventions, such as introducing earlier exposure to open-ended modeling tasks or strengthening integration across foundational courses.

Finally, the weighted aggregation approach provides a transparent rationale for curricular decisions documented in continuous improvement processes. Changes to course content, assessment design, or sequencing can be justified by demonstrating how they redistribute or strengthen weighted evidence of Student Outcome 1 attainment. In this way, the weights serve not merely as a scoring mechanism, but as a lens through which programs can evaluate alignment between curricular intent, assessment practice, and observed student performance.

5.2 From Weighted Aggregation to Probabilistic Inference

While the proposed weighting scheme is presented in a deterministic form, it naturally lends itself to more sophisticated probabilistic analysis as increasingly fine-grained assessment data are now readily available. Modern grading platforms, such as Gradescope, allow programs to collect large-scale, structured data at the level of individual sub-problems, rubric items, and performance dimensions across multiple courses and semesters. When combined with outcome-aligned weighting, this richness of data supports a shift from point estimates of outcome attainment toward probabilistic inference about students' underlying abilities.

In the deterministic formulation, weighted aggregation produces a composite indicator of Student Outcome 1 attainment by emphasizing assessment artifacts that are both instructionally important and strongly aligned with complex engineering problem solving. From a probabilistic perspective, this process can be interpreted as combining multiple noisy observations of a latent construct—students' ability to identify, formulate, and solve complex engineering problems. Each assessment artifact contributes evidence with a different degree of reliability and relevance, precisely the distinctions captured by the instructional and relevance weights. Thus, the weighting scheme can be viewed as an implicit model of observation quality, where more relevant and consequential artifacts exert greater influence on the inferred outcome.

As assessment data scale in both volume and granularity, probabilistic models become not only feasible but advantageous. With repeated observations across sub-problems, exams, and courses, programs can estimate uncertainty in outcome attainment, rather than relying solely on threshold-based judgments. For example, rather than concluding that an outcome is "met" or "not met," programs could estimate confidence intervals or posterior distributions reflecting the degree of certainty in that conclusion. Such representations are particularly valuable for complex outcomes like Student Outcome 1, where performance is inherently variable and context-dependent.

Fine-grained data enable disaggregation along meaningful dimensions of complexity. Probabilistic models can incorporate indicators of specific characteristics of complex engineering problems, allowing programs to estimate how students perform under different cognitive demands and settings (e.g., timed exams vs. take-home assignments). This opens the door to analyses such as identifying latent clusters of problem-solving ability, quantifying learning trajectories across the curriculum, or assessing the marginal contribution of individual courses to outcome attainment.

In this sense, the proposed approach serves as a bridge between traditional ABET-aligned assessment practices and more advanced machine-learning approaches. Programs may begin with transparent, weighted aggregation to support decision-making and continuous improvement, while progressively adopting probabilistic models as data maturity increases. The availability of large-scale, fine-grained assessment data lowers the barrier to such analyses, enabling engineering programs to move beyond static snapshots of outcome attainment toward dynamic, uncertainty-aware interpretations that better reflect the complexity of engineering learning itself.

6 Conclusion

This paper examined the challenge of assessing ABET Student Outcome 1 in a manner that is both pedagogically meaningful and methodologically defensible. We argued that the common prohibition against using aggregate course performance should be interpreted not as a rejection of aggregation itself, but as a caution against unaligned and opaque use of course-level measures. For outcomes such as complex engineering problem solving, which are inherently integrative and distributed across the curriculum, narrowly sampled artifacts alone may fail to capture the full breadth of student ability.

To address this challenge, we proposed an outcome-aligned aggregation approach that combines instructional weighting with explicit relevance to ABET's characteristics of complex engineering problems. By explicitly aligning assessment artifacts with both the cognitive actions embedded in Outcome 1 and the defining characteristics of complex engineering problems, this approach provides a structured, auditable, and ABET-aligned mechanism for selecting and justifying evidence sources. It supports aggregation across assessments where appropriate, while preserving the specificity and documentation required for continuous improvement. The approach is inherently hierarchical and can be applied at multiple levels of assessment granularity. At the most fine-grained level, individual sub-parts of a problem may be weighted according to the extent to which they assess specific components of Student Outcome 1, such as identifying governing principles, formulating appropriate models, or executing a solution. At higher levels, the same principle extends naturally to multiple problems within an exam and to multiple assessments across courses. In this formulation, aggregation is not a uniform averaging of scores, but a structured synthesis of weighted evidence that reflects both relevance and scope. By supporting weighting at multiple levels, the approach allows programs to balance specificity and robustness, ensuring that evidence of outcome attainment is both directly traceable and representative of sustained student performance on complex engineering problems.

The instructional weighting component should be interpreted as a transparent proxy for where outcome-relevant evidence is generated in the curriculum. While course credits and instructor-defined grading weights are used to reflect relative exposure and emphasis, they do not operate in isolation; outcome relevance (i.e., alignment with complex engineering problem characteristics) ultimately governs whether an artifact contributes meaningfully at all, and low-impact artifacts naturally receive negligible or zero weight, as illustrated in Table 5 . At the same time, the framework is intentionally flexible and can accommodate program-level refinements, such as normalizing weights across sections, using representative or common assessment artifacts, or replacing credits with alternative scaling factors (e.g., required vs. elective distinctions or faculty-defined weights). Although any weighting scheme could, in principle, be manipulated, the explicit and auditable nature of this formulation makes such changes visible and subject to program and accreditation review.

The proposed approach naturally positions programs to leverage increasingly available fine-grained assessment data from contemporary platforms (e.g., Gradescope) and to transition toward probabilistic interpretations of outcome attainment as data maturity grows. As a result, the proposed approach does not impose an undue administrative burden, but instead leverages existing infrastructure. Taken together, these contributions offer engineering programs a practical and theoretically grounded path toward more reliable, scalable, and meaningful assessment of complex engineering problem-solving outcomes.

References

[1] R. M. Felder and R. Brent, "Designing and teaching courses to satisfy the abet engineering criteria," *Journal of Engineering Education*, vol. 92, p. 7–25, Jan. 2003.

[2] G. Rogers, "Assessment 101: Assessment tips with gloria rogers, ph.d. using course or test grades for program assessment." Online, 2006. https://www.abet.org/wp-content/uploads/2015/04/using-grades-for-program-assessment-.pdf.

[3] B. G. Mobasseri, L. Klein, and E. S. C. Jr., "Challenges in designing complex engineering problems to meet abet outcome 1," in *2023 ASEE Annual Conference & Exposition*, no. 10.18260/1-2–43153, (Baltimore , Maryland), ASEE Conferences, June 2023. https://peer.asee.org/43153.

[4] ABET, "Criteria for accrediting engineering programs, 2025 – 2026." Online, 2025.

https://www.abet.org/accreditation/accreditation-criteria/criteria-for-accrediting-engineering-programs-2025-2026/.

[5] C. Owen, K. Scales, and M. Leonard, “Preparing for program accreditation review under abet engineering criteria 2000: Creating a database of outcomes and outcome indicators for a variety of engineering programs,” *Journal of Engineering Education*, vol. 88, p. 255–259, July 1999.

[6] M. Besterfield-Sacre, L. Shuman, H. Wolfe, C. Atman, J. McGourty, R. Miller, B. Olds, and G. Rogers, “Defining the outcomes: a framework for ec-2000,” *IEEE Transactions on Education*, vol. 43, p. 100–110, May 2000.

[7] L. J. Shuman, M. Besterfield-Sacre, and J. McGourty, “The abet “professional skills” - can they be taught? can they be assessed?,” *Journal of Engineering Education*, vol. 94, p. 41–55, Jan. 2005.

[8] C. P. Liew, S. H. Hamzah, M. Puteh, S. Mohammad, and W. H. W. Badaruzzaman, *A Systematic Approach to Implementing Complex Problem Solving in Engineering Curriculum*, p. 880–891. Springer International Publishing, 2020.

[9] J. E. Davidson and R. J. Sternberg, eds., *The Psychology of Problem Solving*. Cambridge University Press, June 2003.

[10] D. R. Woods, A. N. Hrymak, R. R. Marshall, P. E. Wood, C. M. Crowe, T. W. Hoffman, J. D. Wright, P. A. Taylor, K. A. Woodhouse, and C. K. Bouchard, “Developing problem solving skills: The mcmaster problem solving program,” *Journal of Engineering Education*, vol. 86, p. 75–91, Apr. 1997.