5-11-2023

CO-CREATING SUSTAINABLE AND DYNAMIC CURRICULUM: A DECISION SUPPORT SYSTEM FOR ASSURING GRADUATE CAPABILITY

Bhavani Sridharan
Australian Catholic University, bhsridharan@acu.edu.au

Jamie Mustard
Deakin University, jamie.mustard@deakin.edu.au

CO-CREATING SUSTAINABLE AND DYNAMIC CURRICULUM: A DECISION SUPPORT SYSTEM FOR ASSURING GRADUATE CAPABILITY

Research Paper

Bhavani Sridharan, Australian Catholic University, Australia, Bhsridharan@acu.edu.au

Jamie Mustard, Deakin University, Australia, jamie.mustard@deakin.edu.au

Abstract

An important challenge facing higher education institutions is authentically assuring graduate capabilities, including the assessment of 'softer skills'. Many institutions have developed procedures through curriculum reform and renewal processes, but too often these processes involve isolated approaches such as 'ticking the box', use of summary statistics and gap analysis which are both imprecise and piecemeal. Despite growing interest in assuring employable graduate capabilities, there has been little research into using a structured holistic approach engaging with multiple key stakeholders. This research proposes a decision support system adopting a multi-criteria decision approach (MCDA) to appraise whole of degree level outcomes and the robustness of assessment policy decisions. The objective is to authentically assure graduate capability by considering multiple key criteria to reduce arbitrariness of current approaches to assessment design. Multiple scenarios are presented to explain the application of MCDA in evaluating assessment benchmarks using simple to complex criteria.

Keywords: Assuring graduate capability; curriculum reform; multi criteria decision approach; assessment benchmark; quality of higher education.

1 INTRODUCTION

Soft skills represent "a dynamic combination of cognitive and meta-cognitive skills, interpersonal, intellectual and practical skills" (Succi and Canovi, 2020). The European LifeComp framework highlighted integration of three core skills namely learnability skills (learning to learn), personal and social competencies (Sala et al., 2020). They underscored the importance of lifelong learning skills as against episodic learning. Moreover, employers have been expressing their dissatisfaction with higher education institutions for not sufficiently preparing graduates equipped with these transferable skills (Succi and Canovi, 2020). The covid pandemic along with digital disruptions highlighted a deficit in digital skills among graduates including ethical aspects of using technologies (Erstad and Voogt, 2018).

The importance of integrating soft skills in curriculum is recognized by higher education institutions to equip graduates with the challenges of the word of work (Oliver and Jorre de St Jorre, 2018). For example, the European Commission has announced 2023[1] as the 'Year of Skills' to deal with the misalliance between employers' needs and the education sector's preparation of graduates. Other innovations in fulfilling the employability agenda include: curriculum reform (Oliver, 2013); embedding soft skills in stem education (Villán-Vallejo et al., 2022); constructive alignment (Biggs, 2014);

[1] See European Year of Skills 2023 (europa.eu)

curriculum mapping (Watson et al., 2020) (Lam and Tsui, 2016); evidencing and assuring learning outcomes (Larkin and Richardson, 2013); and closing the loop (French et al., 2014).

These initiatives have been instrumental in raising the prominence of graduate capabilities in higher education curricula. However, serious implementation concerns exist. Boud and Brew (2017) emphasise the need for proper training for new teachers in the profession of higher education, which is crucial for transferring competencies to graduates. Larkin and Richardson (2013) indicate a significant gap in evaluating constructive alignment initiatives from a student outcomes perspective. Boud and Falchikov (2007) point to a lack of evidence of its systematic implementation and even less evidence of its evaluation. Oliver (2013) stresses that even though curriculum mapping is a useful exercise for closing the loop, it can have undesirable impacts if academics do not understand it, or find it an administrative burden. French et al. (2014) and Hager (2006) argue curriculum mapping has potential to reduce higher education to a ‘tick-list’ of skills and competencies. Greene and Saridakis (2008) posit that Universities have underestimated the policy changes required to realistically embed graduate capabilities in curricula.

These inadequacies in curriculum transformation highlight the significant problem caused by the lack of a structured holistic approach considering multiple assessment parameters to achieve a scalable sustainable assessment reform process. For example, across the sector, students are generally assessed and evidenced on their discipline-specific knowledge and required to obtain a minimum pass mark in each of the contributing subjects. This is a valid system of assessment, where there is only one associated variable for each unit (discipline-specific knowledge) involved. To embed generic learning outcomes, this research proposes a systematic approach to making decisions taking into account assessment coverage, importance and standards. The objective is to develop and test a framework with the help of a decision support system using course level test data at multiple levels of complexity to authentically assure graduate capability. This will reduce arbitrariness of current approaches to assessment decisions. Accordingly, the research question posed is:

> *How does embedding simple to complex attributes and benchmarking standards at multiple levels facilitate authentically assuring graduate capability in higher education?*

Throughout the remainder of this paper ‘course’ refers to a degree program and ‘unit’ refers to a component of a course contributing towards the degree (subject). Courses are viewed as both a collection of units and a collection of graduate attributes. Graduate attributes includes both discipline specific technical knowledge and soft skills.

2 REVIEW OF LITERATURE

Graduate learning outcomes (GLOs), refer to the essential knowledge and skills for ongoing success in the current dynamic working environment (Succi and Canovi, 2020). This requires that graduates be equipped with not just discipline-specific knowledge but also other generic (‘soft’) skills to support students preparation for society and personal development throughout their life and livelihood (Makhachashvili, 2021, Sala et al., 2020). As an example, the importance of communication, teamwork, organizational skills are identified as the most critical for information technology professionals’ digital literacy in order to meet the challenges of digital disruptions (Stal and Paliwoda-Pękosz, 2019).

The prominence of such generic skills has come about through: the changing requirements in the employment market; prevalence of demand and supply imbalances in graduate capabilities; and inability of graduates to adapt to the needs of the employment market. Numerous studies have reflected the seriousness of these issues (Jackson, 2013, Makhachashvili, 2021). The need for alignment of curriculum to meet the gaps between employers’ expectations and the actual potential of recruits have been identified including specifically for Information Systems/Technology professionals (Aljohani et al., 2022, Dubey and Tiwari, 2020). Similarly, Erstad and Voogt (2018) underscore disjunction between research findings and actual practices creating tensions around preparing students for the 21st century living and unmet ‘soft’ skills needs of employers. IT executives identified a range of soft skills that are most difficult to find and most important for the organizational success including problem solving, collaboration and communication skills (Kappelman et al., 2019, Kappelman et al., 2016)

Reacting to this, national quality assurance agencies and international accreditation bodies have mandated that institutions explicitly instil these graduate capabilities and more effectively evidence student learning outcomes. Responding to this directive, institutions are developing comprehensive policies and procedures to enhance the quality of higher education by embedding; graduate attributes with explicit coverage of soft skills in the curriculum reform process (Oliver, 2013, Succi and Canovi, 2020) and internal and external quality assurance programs (French et al., 2014). This often involves contextualized constructive alignment of learning outcomes at multiple levels – university level objectives, course level, unit level, assessment level, and evidence level (Sridharan et al., 2015).

The underpinning theoretical model adopted in this research is constructive alignment theory (Biggs and Tang, 2015). The three main constructive alignment process are: development of intended learning outcomes; mapping of teaching activities; and aligning assessment methods to learning outcomes (Biggs, 2014). The effectiveness of such an approach has been demonstrated across the higher education including Management Information Systems curriculum (Zhang et al., 2022). They found superior learning effectiveness from a constructively aligned compared to non-aligned curriculum. Based on this constructive alignment process, the curriculum mapping initiatives are scaffolded through planning, executing, evidencing and ultimately closing the loop, as discussed below.

The planning process entails identifying the contributing units at a course level and selecting the GLO covered by each contributing unit (Fig. 1). Execution of this curriculum mapping process requires articulating and aligning Biggs' (2014) three aspects of the constructive alignment: Unit Learning Outcomes (ULOs); how to educate those ULOs (teaching and learning methods); and how to explicitly evaluate whether those intended ULOs are realized. A common method for measuring and collecting student achievements is through use of an analytic rubric, which provides a means to disaggregate and measure each learning outcome. However, measurement difficulties are recognized by scholars across the sector (French et al., 2014) (Succi and Wieandt, 2019). Kirkpatrick's four level evaluation model is a powerful and practical model to evaluate training in terms of criteria at four levels: reaction, learning, behaviour and results (Kirkpatrick and Kirkpatrick, 2016), which provides an interesting complement to the proposed model in this study.

Evidencing is the next step by which outcomes are measured (Lawrence et al., 2011) through Assurance of Learning (AoL), which evaluates whether the intended GLO is actually implemented in reality at course level. This requires documenting: a list of GLOs covered in each contributing unit; how teaching and learning activities are scaffolded/developed in each of the contributing units; and overall evaluation of whether the majority of students have achieved the minimum requirement.

Closing the loop for continuous improvement is an integral aspect of curriculum mapping. Lawrence et al. (2011) emphasise acting on assessment findings including analysing information collected from AoL and taking corrective action as needed.

Curriculum mapping through quality assurance and accreditation frameworks provides some guidance in enhancing learning outcomes. The approach undoubtedly provides a simple and systematic mechanism to ensure necessary graduate attributes are covered at a whole of course level (Jackson, 2013). However, prior literature indicates the insufficiency of this model (Akdur, 2021, Aljohani et al., 2022), highlighting issues including misalignment between key competencies taught and required for students' career success, ineffectiveness of the closing the loop process and lack of benchmark standards to ensure quality of programs.

Addressing this issue, Cordiner et al. (2007) propose a multi-dimensional approach to integrate constructively aligned curricula, involving the inter-relationship of objectives across various levels. However, for practical implementation, incorporating the critical dimensions of constructive alignment at each level are required. Filling this gap, Sridharan et al. (2015) proposed a holistic multi-level, multi-dimensional course alignment model to facilitate development of an aligned curriculum. However, these theoretical models have not yet translated into actual practice owing to lack of effective process tools.

Closing the loop involves taking corrective action through systematic evidence, data collection and analysis to make worthwhile improvements. Although seemingly simple, this stage of the process proves problematic and is often the least developed in many institutions (French et al., 2014). Various issues including weight allocation, complexity, standards, assessment of multiple skills (both discipline and soft skills) are acknowledged as critical (Boud and Falchikov, 2007). Nevertheless, very little has been done to date for continuous improvement with robust objective standards and review mechanisms.

Units/GLOs	Unit-1	Unit-2	Unit-3	Unit-4	Unit-5	Unit-6
1.Discipline specific	☑	☑	☑	☑	☑	☑
2.Communication	☑	☑	☑	☑	☑	☑
3.Digital Literacy	☑	☑	☐	☐	☐	☐
4.Critical Thinking	☐	☐	☐	☐	☑	☑
5.Problem Solving	☐	☐	☑	☑	☐	☑
6.Self-management	☐	☑	☐	☑	☐	☐
7.Team Work	☑	☑	☐	☐	☐	☐
8.Global Citizenship	☐	☐	☐	☐	☐	☑

Figure 1. Example of a whole of course mapping of learning outcomes.

3 BENCHMARKING SOLUTION

The crux of benchmarking is a continuous and systematic approach in evaluating strategies, products and processes for improvement (Spendolini, 1992). Example initiatives include: consensus moderation (Orr, 2007), internal and external benchmarking to achieve equivalence in assessment standards and achievement, intra-institutional benchmarking of student learning outcomes (Thomas et al., 2013), post-marking moderation (Sadler, 2009), and internal benchmarking through course review, cross-program review, assessment outcome review, and rubric benchmark exercises (Riebe and Jackson, 2014). Many of these approaches fail to consider holistic benchmarking approaches which encompass multiple aspects at multiple levels.

The importance of embedding both objective and subjective measures in benchmarking and closing the loop activities is acknowledged as instrumental in enhancing and assuring graduate capabilities. Along these lines, Campbell and Oblinger (2007) pointed out that "in higher education many institutional decisions are too important to be based only on intuition, anecdote, or presumption; critical decisions require facts and the testing of possible solutions. Reports based on data and statistical analysis represent an improvement over intuition". Scott and Ofori-Dankwa (2006) propose "quantifiable, numerical, and presumably 'objective' measures" in the closing of the loop process. Specifically, in relation to a course approach to curriculum, if we cannot measure the performance of learning outcomes at course level, how can we make correct policy decisions?

In summary, this research contends that course level quality assurance is a precursor to assuring success at student level. While many of the current curriculum enhancement approaches are necessary, they do not offer sufficient conditions for authentically assuring learning outcomes. These models have failed to consider a number of crucial variables at multiple levels simultaneously to realise the vision of assuring graduate capabilities. This paper aims to provide a more holistic approach addressing all three concerns, namely: alignment of learning outcomes, closing the loop and setting benchmark standards at multiple levels.

4 METHODLOLOGY

This research adopts the Multi-Criteria Decision Analysis (MCDA) to fulfil the objective of multi-level benchmarking for authentically assuring graduate capabilities. MCDA is an "umbrella term to describe a collection of formal approaches which seek to take explicit account of multiple criteria in helping individuals or groups explore decisions that matter" (Belton and Stewart, 2002). MCDA aids stakeholders to organize and synthesize information when they are confronted with complex scenarios with multiple competing alternatives to aid in making robust decisions (Zeleny, 1982).

The application of the MCDA approach has been used to support decisions in fields such as health, business, economics, logistics, and finance. However, to our knowledge, the full potential of this approach has not been exploited to resolve the challenges of modern higher education reform processes. The application of MCDA to a higher education assessment and curriculum reform context is appropriate, given the presence of conflicting multiple criteria where trade-off is critical; importance of considering both objective and subjective measures; guiding the process of quantifying and dealing with uncertainty; and involvement of multiple stakeholders in the decision-making process.

To systematically implement the MCDA approach (Dodgson et al., 2009), breaking down a complex process into multiple steps is fundamental. For the purposes of this research, we have incorporated an eight-step process (Fig. 2).

Kirkpatrick's training evaluation model lens is a very useful starting point by considering the four levels of evaluation (reaction, learning, behaviour and result). For example, Step-1 relates to 'Reaction Level' criteria of how well learners can benefit based on the learning outcomes at course and unit levels. Steps 2 and 3 relates to 'Learning Level' criteria considering the nuances of taxonomy levels incorporated at multiple levels. Steps 4 to 6 relates to 'Behavioural Level' criteria by deep diving into evaluation surrounding application knowledge in all three learning domain – cognitive, affective and psychomotor. Finally Steps 7 and 8 relates to holistic 'Results Level' to take corrective actions.

Step	Objective
1. Problem definition	Evaluate whether or not each GLOs are assessed adequately both at course level and unit level to ensure that the course has met the benchmark expectations to assure the graduate capabilities to students.
2. Identification of Evaluation Criteria and Sub-criteria	Coverage of GLOs (criteria) and taxonomy levels (sub-criteria) at both course and unit level.
3. Identification of Alternatives	Alternatives are arrived at by choosing the core units contributing towards each GLOs that are contextualized to ULOs and assessment criteria.
4. Data Collection of measurement of evaluation criteria	Collection of two types of data: (a) factual assessment criteria data (or performance data) for each unit across all GLOs (criteria); and (b) benchmark data for interpretation and evaluation of performance.
5. Aggregation Method and model construction	Use simple additive methods with varying complexities (linear addition, weighted, normalised methods) based on the scenario for aggregating and evaluating the performance.
6. Perform sensitivity analysis	Use sensitivity analysis to: identify and close the gaps by modifying the benchmark parameter levels to arrive at the desired outcomes; assess if each of the contributing units are meeting threshold benchmark expectations; and adjust to varying requirements of courses and market demands by modfying the benchmark parameters levels and exploring the results. (will be tested in a follow-on paper).
7. Analysis	Analyse the results to support decision making and provide corrective recommendations to relevant stakeholders.
8. Closing the loop	Close the loop for continuous improvement decisions based on analysis in step-7

Table 1. MCDA steps and objectives.

4.1 Objective

Overall objective is to assure graduate capabilities, to prepare students to meet the dynamic expectations of the labour market.

4.2 Criteria

Criteria is typically derived from either a university or course level agenda covering the GLOs (See Fig. 1). All learning outcomes would often be viewed as critical, but the practical implementation requires identification of those most important in a given domain to effectively distribute limited resources in teaching, assessing and evidencing. The responsibility of making policy decisions on coverage, importance and complexity/standard of each GLO relies heavily on key stakeholders.

4.3 Sub-criteria

Sub-criteria is obtained using the three main learning domains: cognitive, affective and psychomotor from Anderson et al. (2001). Each of the GLOs can be classified into these domains albeit with some overlap. For example, discipline-specific knowledge, critical thinking and problem-solving sit under the cognitive domain, while communication and digital literacy mainly fall under the psychomotor domain. Critical thinking, problem-solving, self-management, teamwork and global citizenships pertain to the affective domain (and can also be embedded within the cognitive domain). This classification is essential as the level descriptors and expectations are different for different learning domains. For illustrative purposes, Fig. 2 shows an example of six increasing levels of complexity in order to fit a broader cross section of learning domains.

4.4 Options

Ultimately, decisions need to be made on which units are to cover which GLOs, the extent of coverage and complexity at which each of the GLOs is pitched in terms of teaching, assessing and evidencing learning outcomes. These decisions are to be taken through negotiation and compromise between key stakeholders. Some basic principles of relevance, fairness, and equity in sharing the load are to be applied. As an illustration, six units are listed as options to cover the required GLOs in Fig. 2.

4.5 Indicators

Objective and subjective measures are considered when embedding attributes such as coverage, importance, and complexity/standards. The coverage of each GLO is measured simply as either included or not. Based on the importance of a particular GLO for a given course, weight for each GLO is measured by distributing the total of 100% across each of the GLOs. The sub-criteria are derived based on a hierarchy of complexity measuring surface learning to deep learning skills. This will entail collection of two types of data relating to these indictors: factual performance data for each unit across all GLOs (criteria); and benchmark data for evaluation of performance.

4.6 Aggregation methods

The aggregation methods used include a linear additive, simple average, normalised weighted and simple weighted average methods for both obtaining the threshold benchmark targets at multiple levels (course, unit and GLO), and for measuring actual performance (Goodwin and Wright, 2014).

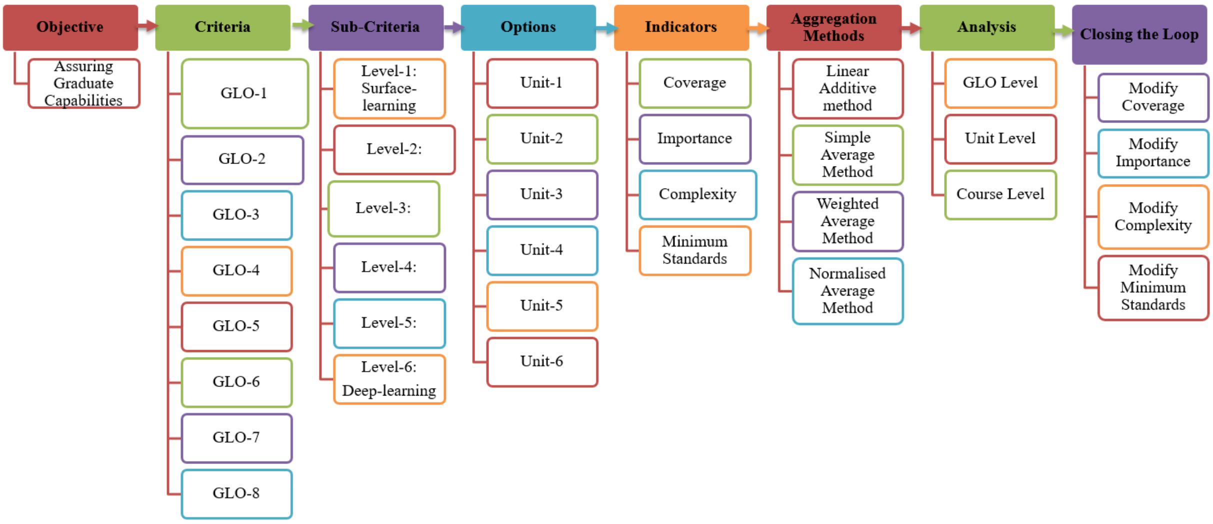


Figure 2. MCDA for assessment benchmarking.

4.7 Analysis

Results are analysed by comparing against benchmark measures to identify gaps at multiple levels considering multiple attributes (simple to complex). Following this, conversation with relevant stakeholders is essential to decide on strategies and recommendations for modification at various levels.

4.8 Closing the Loop Decisions

Closing the loop decisions can be achieved through multiple iterations of sensitivity analysis in identifying the gap and modifying benchmark parameters to arrive at desired outcomes. This requires a unit leader taking necessary adjustment decisions, considering multiple attributes, for all the contributing units that are not meeting the threshold benchmark expectations, to ensure the gap is closed.

5 RESEARCH PROPOSTITIONS

The analytical framework underpinning the conceptual framework is ADDIE (Analysis, Design, Development, Implement and Evaluate) model of Instructional Systems Design (ISD). Considering the complexities of evaluating programmatic outcomes before delivery of higher education, the conceptual framework is evaluated considering multiple criteria and at multiple levels.

For illustrative purposes, we will consider the following scenario in the implementation of MCDA namely: (1) Six core units contributing to the course; (2) Eight GLOs to be covered across a course; (3) Minimum benchmark weight allocation as follows: 60% for discipline-specific knowledge; 20% for generic skills; and 20% which is flexible and at the discretion of the unit leader; and (4) Subjective aspects of constructive alignment of graduate attributes at various levels (course, unit, assessment and rubric), considering multiple criteria (coverage, weight, complexity) are in place (Sridharan et al., 2015). This conceptual framework can be used alongside Kirkpatrick's (2016) four-level model to assess outcome upon delivering the training and education in higher education context. For example, all four levels of learning model can be applied to evaluate the quality assurance process considering multiple criteria as indicated above.

Four propositions are drawn to demonstrate how iterative incremental improvements can be accomplished by embedding simple to more complex assessment parameters to close the gap in the current curriculum practices. Fig. 3 presents the conceptual framework embedding these antecedent parameters and associated propositions towards enhancing quality assurance. Critical issues relating to simple propositions leads to incrementally embedding more complex parameters and associated propositions. The following section illustrates the role of each followed by a rationale for embedding more complex attributes.

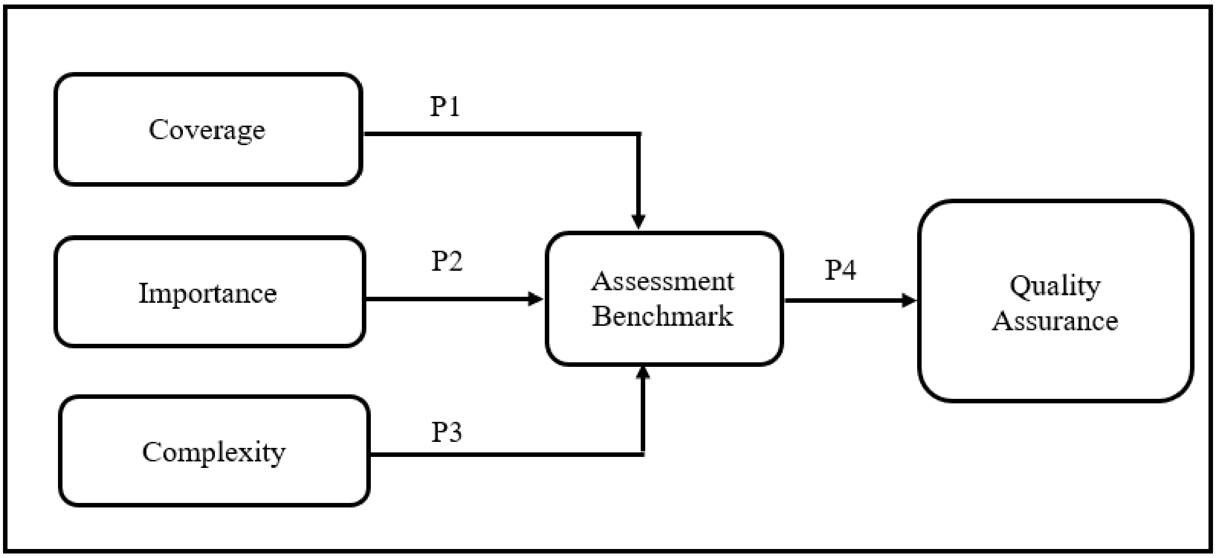


Figure 3. Conceptual framework embedding antecedent parameters and propositions.

5.1 Proposition 1: Embedding GLO coverage is the minimum requirement for assessment benchmarking model

This proposition begins with inclusion of a single attribute (GLO count) considered at course and unit level. Decision makers such as a course director set the minimum benchmark coverage based on the perceived importance of a given GLO. Typically discipline-specific knowledge is covered in all units in a course. Communication skills may also have high coverage due to its often-intrinsic requirement for completion of assessment tasks. Other learning outcomes' coverage will vary depending on the discipline, equity and fairness considerations to provide a wide-ranging experience to students.

As an example, Fig. 4(a) shows the initial parameters for which a course is designed along with benchmark coverage decisions. The minimum benchmark coverage at GLO and unit level are given in column-2 and the last row respectively. The shaded cells indicate the initial allocation of GLOs to units. By comparing actual count against the required benchmark count at GLO level (row total) and unit level (column total), any gaps can be quickly identified. Fig. 4(a) highlights all but one GLO (communication) meets the minimum coverage. At the unit level, two of the six units have failed to deliver the required number of GLOs. Decision makers then iteratively adjust the actual coverage until the benchmarks are satisfied. In this case, additional GLOs were added to the first two units (green shade) as shown in Fig. 4(b). Note that this scenario does not preclude units covering additional GLOs. In the example, unit-3 assesses GLO-7 and unit-6 covers GLO-5 even though they are not required to do so.

This approach, by itself, is too simplistic to authentically realise the vision of assuring graduate attributes. This does not provide an accurate measure of performance as it does not take into account the weight contributed by each unit to each GLO. For instance, a course can seem satisfactory even with an assessment allocation of as little as 1% for one or more GLOs. To overcome this issue, the next proposition below adds weight allocation for each GLO.

	(a) Pre-evaluation								(b) Post-evaluation						
Units/GLOs	**Benchmark**	**UNIT 1**	**UNIT-2**	**UNIT-3**	**UNIT-4**	**UNIT-5**	**UNIT-6**	**Actual**	**UNIT-1**	**UNIT-2**	**UNIT-3**	**UNIT-4**	**UNIT-5**	**UNIT-6**	**Actual**
GLO-1	6	1	1	1	1	1	1	6	1	1	1	1	1	1	6
GLO-2	6	0	0	1	1	1	1	4	1	1	1	1	1	1	6
GLO-3	2	1	1	0	0	0	0	2	1	1	0	0	0	0	2
GLO-4	2	0	0	0	0	1	1	2	0	0	0	0	1	1	2
GLO-5	2	0	0	1	1	0	1	3	0	0	1	1	0	1	3
GLO-6	2	0	1	0	1	0	0	2	0	1	0	1	0	0	2
GLO-7	2	1	1	1	0	0	0	3	1	1	1	0	0	0	3
GLO-8	1	0	0	0	0	0	1	1	0	0	0	0	0	1	1
Actual	23	3	4	4	4	3	5	23	4	5	4	4	3	5	25
Benchmark	23	4	5	3	4	3	4	23	4	5	3	4	3	4	23

Figure 4. Performance matrix embedding GLO coverage.

5.2 Proposition 2: Embedding GLO weights can further enhance assessment benchmarking model

In this proposition, a course level benchmark weight for each GLO is decided by domain experts based on their judgment (ideally in consultation with industry experts) of the importance contextualized to the discipline requirement. Fig. 5(a) below compares actual GLO percentage contribution (calculated using simple average method) against benchmark level contribution both at GLO and unit level. GLO level performance is assessed by comparing the average contribution to each GLO across all units (row average) against the minimum requirement for each GLO. The results in Fig. 5(a) show that the benchmark has not been achieved for GLO-2 and GLO-8.

Unit level performance for generic skills is assessed by comparing the actual total weight covering generic skills (GLO-2 to GLO-8: column total) against a benchmark weight set for generic skills for each unit. As shown in Fig. 5(a), unit-2 has insufficient weight allocation for the overall generic learning outcomes with 4.5% below required benchmark (24.5%).

The benchmark is determined by the weighting required for each GLO that a unit is required to cover. For example, all units are expected to cover GLO-2 and therefore an aggregate weight of 30% (6 units × 5% weight) is equally distributed across all six units. On the other hand, for GLO-5, the aggregate benchmark weight is 21% (6 units × 3.5% weight). Since only two units (unit 3 and 6) are allocated to cover GLO-5, both are expected to cover 10.5% each. It is anticipated that generic skills contributions are not necessarily equal with some units contributing more to compensate for other less contributing units. Unit leaders have 20% discretion over and above benchmark expectations. For example, unit-3 is not required to contribute to GLO-7, but doing so by using the 20% buffer.

Decision makers, after doing some sensitivity analysis, iteratively adjust the actual coverage until all benchmarks at overall GLO and unit level are eventually satisfied as shown in Fig. 5(b). In this instance, unit-2 was adjusted by adding additional weighting to GLO-2, which resolved the issue at overall level. Insufficient coverage of GLO-8 was resolved by increasing the weighting in unit-6.

Although an improvement on proposition 1, embedding GLO weight is not sufficient to authentically realise the vision of assuring graduate attributes as the level of complexity of assessment tasks contextualized to requirements is not considered. As an example, an introductory health subject may require significant rote learning of medical facts and thus lower order skills are considered important. A higher level of complexity could be the focus of a final year capstone unit involving complex problems, where integrating knowledge and skills from multidimensional and relational perspectives are critical. The third proposition below allows for a better measure of performance assessment by embedding both complexity and importance of task.

Units/GLOs	Benchmark	(a) Pre-evaluation							(b) Post-evaluation						
		UNIT-1	UNIT-2	UNIT-3	UNIT-4	UNIT-5	UNIT-6	Actual	UNIT-1	UNIT-2	UNIT-3	UNIT-4	UNIT-5	UNIT-6	Actual
GLO-1	60.0	78.0	80.0	66.5	76.0	81.0	77.0	76.4	78.0	72.0	66.5	76.0	81.0	73.0	74.4
GLO-2	5.0	2.0	2.0	5.0	4.0	4.0	6.0	3.8	2.0	10.0	5.0	4.0	4.0	6.0	5.2
GLO-3	2.0	8.0	5.0	0.0	0.0	0.0	0.0	2.2	8.0	5.0	0.0	0.0	0.0	0.0	2.2
GLO-4	4.0	0.0	0.0	0.0	0.0	15.0	12.0	4.5	0.0	0.0	0.0	0.0	15.0	12.0	4.5
GLO-5	3.5	0.0	0.0	22.5	10.0	0.0	3.0	5.9	0.0	0.0	22.5	10.0	0.0	3.0	5.9
GLO-6	1.5	0.0	8.0	0.0	10.0	0.0	0.0	3.0	0.0	8.0	0.0	10.0	0.0	0.0	3.0
GLO-7	3.0	12.0	5.0	6.0	0.0	0.0	0.0	3.8	12.0	5.0	6.0	0.0	0.0	0.0	3.8
GLO-8	1.0	0.0	0.0	0.0	0.0	0.0	2.0	0.3	0.0	0.0	0.0	0.0	0.0	6.0	1.0
Actual (Generic Skills)		22.0	20.0	33.5	24.0	19.0	23.0	23.6	22.0	28.0	33.5	24.0	19.0	27.0	25.6
Benchmark (Generic Skills)	20.0	20.0	24.5	15.5	20.0	17.0	23.0	20.0	20.0	24.5	15.5	20.0	17.0	23.0	20.0

Figure 5. Performance matrix embedding weight

5.3 Proposition 3: Embedding different levels of complexity can significantly enhance assessment benchmarking model

In this proposition, both complexity and importance at each level of complexity are incorporated into the model. The rationale for introducing complexity is, in many instances, the level pitched may not develop students' higher-order thinking skills. This may not necessarily be a problem at the unit level, however, if this happens consistently across many assessments and units, we cannot authentically say that graduating students have sufficient higher order skills.

In some instances, importance supersedes complexity (eg. hygiene education for medical students is lower order learning but very important within the discipline). To incorporate this, flexibility and discretion are permissible for both course directors and individual unit leaders to allocate appropriate weights based on course and unit requirements. To enable benchmark evaluation for this proposition, taxonomy adjusted benchmark values and actual values are to be computed.

For the purpose of this paper, a high level description, rather than the detailed calculations is provided: (1) Set unit level benchmarks by effectively breaking down the marks allocated for each Unit/GLO across the different levels of complexity. (2) A simple linear scale is used for assigning a value to each taxonomy level based on complexity. (3) Imputation of weights for each level using normalised relative average method. (4) Decide how to allocate the benchmark weight across the six levels of complexity depending on the importance to the discipline. (5) Using the benchmark weight and the original weight allocation for each GLO, the benchmark weight for each GLO and each level is derived. (6) The derived benchmark weight from step-4 along with normalised taxonomy weight from step-2 are used to derive taxonomy adjusted minimum benchmark weight across all levels. (7) Deriving taxonomy adjusted benchmark weight at unit level requires total number of units contributing to the course, course level coverage count for each GLO, course level derived taxonomy adjusted benchmark weight for each GLO and whether or not the chosen unit is expected to contribute to a specific GLO. (8) Deriving the taxonomy adjusted actual values. (8) Decide how to distribute the contribution level of each GLO across the different levels. Finally multiplying the actual mark allocation at each level (from step-7 above) by normalised taxonomy weights, we derive the taxonomy-adjusted mark for each GLO and each unit. The total for each GLO is derived by using simple additive method

To facilitate benchmark comparison against actual contribution, step-8 is repeated for each unit/GLO to derive taxonomy adjusted actual contribution. The results (Fig. 6) are evaluated at GLO level (row average) and unit level (column total). The results indicate that five out of eight GLOs have satisfied benchmark expectations and three GLOs have failed to meet the course level expectations. At the unit level, three units have surpassed the benchmark expectations and three units have failed to meet the benchmark expectations.

Units/GLOs	UNIT-1	UNIT-2	UNIT-3	UNIT-4	UNIT-5	UNIT-6	Actual	Benchmark
GLO-1	98.6	67.4	48.9	77.4	112.3	88.3	82.1	59.1
GLO-2	0.6	4.3	7.1	4.0	4.6	1.7	3.7	4.9
GLO-3	9.1	2.0	0.0	0.0	0.0	0.0	1.9	2.0
GLO-4	0.0	0.0	0.0	0.0	21.4	4.6	4.3	3.9
GLO-5	0.0	0.0	29.3	15.7	0.0	0.9	7.6	3.5
GLO-6	0.0	9.1	0.0	14.3	0.0	0.0	3.9	1.5
GLO-7	3.4	1.4	5.1	0.0	0.0	0.0	1.7	3.0
GLO-8	0.0	0.0	0.0	0.0	0.0	2.3	0.4	1.0
Actual	111.7	84.3	90.4	111.4	138.3	97.7	105.6	98.6
Benchmark	98.6	103.0	94.1	98.6	95.6	101.5	98.6	

Figure 6. Performance matrix embedding complexity level for all GLOs.

This high-level view provides a course director with a broad understanding of potential gaps in coverage of GLOs across different levels of complexity. However, this does not indicate how units are performing at generic skills owing to possible 'overshadowing' of GLO-1's contribution. To identify performance at generic skills level, evaluation of results excluding the contribution from GLO-1 for each unit is derived as shown in Fig. 7. The analysis reveals three units failed to meet the benchmark expectations for generic skills. Even though unit-1 met the expectations at an overall level, failed at generic skills level due to 'overachieving' in GLO-1.

Units/GLOs	UNIT-1	UNIT-2	UNIT-3	UNIT-4	UNIT-5	UNIT-6	Actual	Benchmark
GLO-2	0.6	4.3	7.1	4.0	4.6	1.7	3.7	4.9
GLO-3	9.1	2.0	0.0	0.0	0.0	0.0	1.9	2.0
GLO-4	0.0	0.0	0.0	0.0	21.4	4.6	4.3	3.9
GLO-5	0.0	0.0	29.3	15.7	0.0	0.9	7.6	3.5
GLO-6	0.0	9.1	0.0	14.3	0.0	0.0	3.9	1.5
GLO-7	3.4	1.4	5.1	0.0	0.0	0.0	1.7	3.0
GLO-8	0.0	0.0	0.0	0.0	0.0	2.3	0.4	1.0
Actual	13.1	16.9	41.6	34.0	26.0	9.4	23.5	19.7
Benchmark	19.7	24.2	15.3	19.7	16.8	22.7	19.7	

Figure 7. Performance matrix embedding complexity level for generic skills (excluding GLO-1).

Even though the analysis provides insight into gaps in generic skills performance, it is critical to identify unit level performance for each GLO to facilitate closing the loop process. Accordingly, the fourth proposition below facilitates micro-level benchmarking comparison to isolate problem areas to close the loop.

5.4 Proposition 4: Micro-level benchmarking against minimum standards for each unit and GLO to facilitate closing the loop process realistically

This proposition highlights the importance of micro-level analysis, given many of these issues cannot be isolated within the initial broad analysis. Evaluation process requires a simple comparison of benchmark values against the derived taxonomy adjusted actual values for each GLO and each unit. Shaded cells in Fig. 8 highlight areas where actual is below minimum benchmark standards. For example, the results indicate that all units, except unit 3, have failed to meet benchmark expectations for communication skills.

Units/GLOs	Benchmark	UNIT-1	UNIT-2	UNIT-3	UNIT-4	UNIT-5	UNIT-6
		Actual					
GLO-1	59.1	98.6	67.4	48.9	77.4	112.3	88.3
GLO-2	4.9	0.6	4.3	7.1	4.0	4.6	1.7
GLO-3	2.0	9.1	2.0	0.0	0.0	0.0	0.0
GLO-4	3.9	0.0	0.0	0.0	0.0	21.4	4.6
GLO-5	3.5	0.0	0.0	29.3	15.7	0.0	0.9
GLO-6	1.5	0.0	9.1	0.0	14.3	0.0	0.0
GLO-7	3.0	3.4	1.4	5.1	0.0	0.0	0.0
GLO-8	1.0	0.0	0.0	0.0	0.0	0.0	2.3

Figure 8. Detailed performance matrix for each unit and GLO.

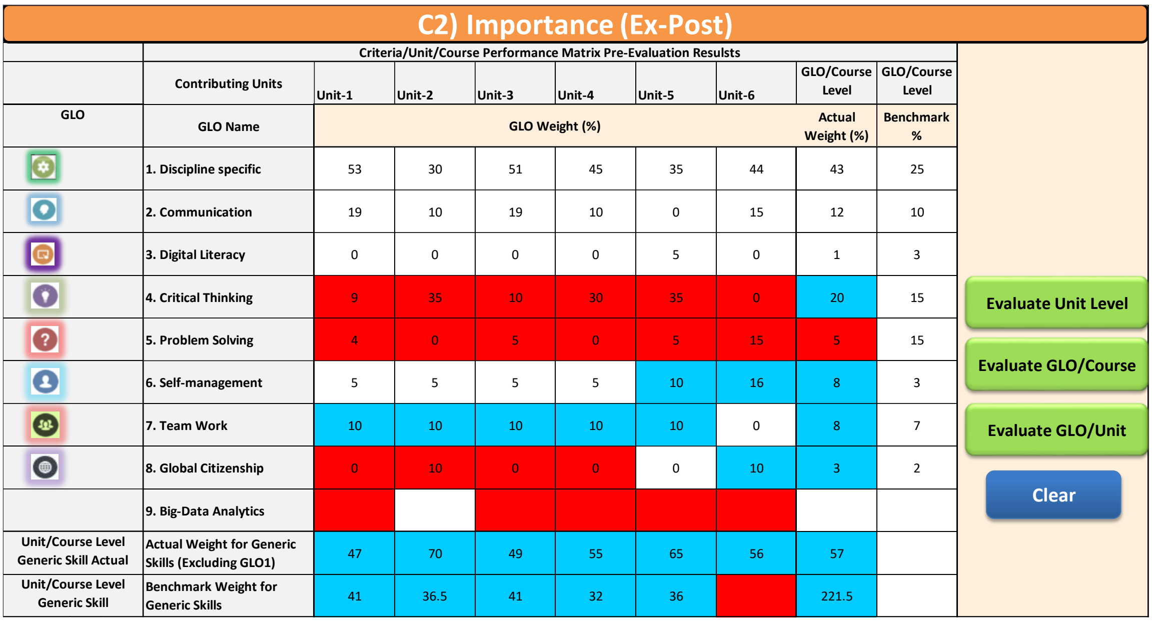
Course directors can use these results to analyse and identify poor performing GLOs and units and negotiate with individual unit coordinators to close any gaps. There could be instances where the results are not satisfactory prima facie owing to the choice of complexity level, but may well be appropriate for the unit, and require course directors' judgment. Returning to an earlier example of understanding hygiene procedures in medical education which falls under lower order skills owing to the rote learning nature, the importance is very high and as a result this model may give 'false' negative outcomes. In such cases, a course director would simply increase the weight for this task to compensate for the lower performance.

Given the above analysis, course directors can iteratively perform sensitivity analysis of allocations across units and complexity levels to close any gaps. We propose incorporating human judgement based on institutional, disciplinary, policy and systemic dimensions by adjusting the parameters relevant to the local institution setting.

6 DECISIONS SUPPORT SYSTEM

The proposed framework is tested through a Decision Support System (DSS) using course level test data with the following modules: Benchmark, Diagnose, Intervene and Evaluate. A sample screen shot is shown in Figure 11. The blue shaded cells indicate meeting the benchmarks and red shade indicate not meeting the benchmark at various levels.

The *benchmark* module considers course level and unit level benchmarks are decided through negotiation and agreement process with key stakeholders such as course coordinators and unit coordinators. The *diagnosis* module identifies gaps in fulfilling course level, unit level and graduate attribute level gaps. Here ex-ante represents predicted values based on the benchmark levels. The *intervene* model helps in making modifications to the contributions at various levels to see the impact of changes at multiple levels. Here the ex-post (as shown in Fig. 9) refers to actual values upon implementation of benchmark decisions. The evaluation module provide a holistic comparative view of predicted (ex-ante) and actual (ex-post) values identifying gaps in assurance of soft skill development at multiple levels. Upon evaluation, the closing the loop cycle can start all over for continuous improvement process.


C2) Importance (Ex-Post)

	Criteria/Unit/Course Performance Matrix Pre-Evaluation Resulsts								
	Contributing Units	Unit-1	Unit-2	Unit-3	Unit-4	Unit-5	Unit-6	GLO/Course Level	GLO/Course Level
GLO	GLO Name	GLO Weight (%)						Actual Weight (%)	Benchmark %
	1. Discipline specific	53	30	51	45	35	44	43	25
	2. Communication	19	10	19	10	0	15	12	10
	3. Digital Literacy	0	0	0	0	5	0	1	3
	4. Critical Thinking	9	35	10	30	35	0	20	15
	5. Problem Solving	4	0	5	0	5	15	5	15
	6. Self-management	5	5	5	5	10	16	8	3
	7. Team Work	10	10	10	10	10	0	8	7
	8. Global Citizenship	0	10	0	0	0	10	3	2
	9. Big-Data Analytics								
Unit/Course Level Generic Skill Actual	Actual Weight for Generic Skills (Excluding GLO1)	47	70	49	55	65	56	57	
Unit/Course Level Generic Skill	Benchmark Weight for Generic Skills	41	36.5	41	32	36		221.5	

Figure 9. Sample screenshot of Assuring Graduate Capability Framework DSS.

7 CONTRIBUTIONS, LIMITATIONS AND CONCLUSION

The major contribution of this research is the development of an overarching assessment benchmarking model integrating multiple critical dimensions towards realising the vision of authentically assuring graduate capabilities. This is first research to use a robust and sophisticated MCDA methodology to solve contemporary higher education problems. This research also provides a pragmatic sustainable assessment benchmarking approach that can be adapted to meet the requirements of program level outcomes and the dynamic needs of the employment market. Finally, this study aims to provide the missing link in authentically assuring graduate attributes by linking multiple key dimensions of curriculum enhancement process.

One limitation of this research is the assumption that it is possible to measure all soft skills in an objective way. Nevertheless, scholars have provided innovative ways of assessing soft skills through appropriate assessment designs. Also, effective use of this approach depends on the valuable judgment and collaboration amongst the decision-making team in establishing criteria, relative importance and complexity levels. Another challenge in implementation of this model may arise from availability of granular data. Future research will be carried out by implementing and evaluating a system for both diagnostic and evaluation process using comprehensive course level data.

This research proposes a novel assessment benchmarking approach supporting authentically assuring graduate capabilities at course level. By scaffolding multiple measurable and subjective criteria at multiple levels, this research is one step closer to realising the vision of authentically developing generic skills, where a significant gap has been found in the literature.

REFERENCES

Akdur, D. 2021. Skills gaps in the industry: Opinions of embedded software practitioners. *ACM Transactions on Embedded Computing Systems (TECS),* 20**,** 1-39.

Aljohani, N. R., Aslam, A., Khadidos, A. O. & Hassan, S.-U. 2022. Bridging the skill gap between the acquired university curriculum and the requirements of the job market: A data-driven analysis of scientific literature. *Journal of Innovation & Knowledge,* 7**,** 100190.

Anderson, L. W., Krathwohl, D. R. & Bloom, B. S. 2001. *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives,* New York, Longman.

Belton, V. & Stewart, T. 2002. *Multiple criteria decision analysis: an integrated approach,* Norwell, MA, Kluwer Academic Publishers.

Biggs, J. 2014. Constructive alignment in university teaching. *HERDSA Review of Higher Education,* 1**,** 5-22.

Biggs, J. & Tang, C. 2015. Constructive alignment: An outcomes-based approach to teaching anatomy. *Teaching anatomy: A practical guide***,** 31-38.

Boud, D. & Brew, A. 2017. Learning to teach as the development of practice. *In:* LEIBOWITZ, B., BOZALEK, V. & KAHN, P. (eds.) *Theorising Learning to Teach in Higher Education: Sociomaterial, Social practice and Realist approaches.* London: Routledge.

Boud, D. & Falchikov, N. 2007. *Rethinking Assessment in Higher Education Learning for the Longer Term,* London, Routledge.

Campbell, J. P. & Oblinger, D. G. 2007. Top-Ten Teaching and Learning Issues, 2007. *Educause Quarterly,* 30**,** 15-22.

Cordiner, M., Stenzel, D. J. & Hafner, L. M. 2007. Levelling expectations across different years in an undergraduate degree to support work-related learning: evidence for a multi-pronged approach. *ATNEAC2007 Book of Proceedings.* QUT, Brisbane: Kelvin Grove.

Dodgson, J., Spackman, M., Pearman, A. & Phillips, L. 2009. *Multi-criteria Analysis: A Manual,* Department for Communities and Local Government: London.

Dubey, R. S. & Tiwari, V. 2020. Operationalisation of soft skill attributes and determining the existing gap in novice ICT professionals. *International Journal of Information Management,* 50**,** 375-386.

Erstad, O. & Voogt, J. 2018. The twenty-first century curriculum: issues and challenges. *Springer International Handbooks of Education***,** 19-36.

French, E., Summers, J., Kinash, S., Lawson, R., Taylor, T., Herbert, J., Fallshaw, E. & Hall, C. 2014. The practice of quality in assuring learning in higher education. *Quality in Higher Education,* 20**,** 24-43.

Goodwin, P. & Wright, G. 2014. *Decision Analysis for Management Judgment,* New York, John Wiley and Sons.

Greene, F. J. & Saridakis, G. 2008. The role of higher education skills and support in graduate self-employment. *Studies in Higher Education,* 33**,** 653-672.

Hager, P. 2006. Nature and development of generic attributes. *In:* HAGER, P. & HOLLAND, S. (eds.) *Graduate attributes, learning and employability.* Dordrecht: Springer.

Jackson, D. 2013. Business graduate employability–where are we going wrong? *Higher Education Research & Development,* 32**,** 776-790.

Julian, S. D. & Ofori-Dankwa1, J. C. 2006. Is accreditation good for the strategic decision making of traditional business schools? *Academy of Management Learning & Education,* 5**,** 225-233.

Kappelman, L., Johnson, V., Torres, R., Maurer, C. & Mclean, E. 2019. A study of information systems issues, practices, and leadership in Europe. *European Journal of Information Systems,* 28**,** 26-42.

Kappelman, L., Mclean, E., Johnson, V. & Torres, R. 2016. The 2015 SIM IT Issues and Trends Study. *MIS Quarterly Executive,* 15.

Kirkpatrick, J. D. & Kirkpatrick, W. K. 2016. *Kirkpatrick's four levels of training evaluation,* Association for Talent Development.

Lam, B. H. & Tsui, K. T. 2016. Curriculum mapping as deliberation – examining the alignment of subject learning outcomes and course curricula. *Studies in Higher Education,* 41**,** 1371-1388.

Larkin, H. & Richardson, B. 2013. Creating high challenge/high support academic environments through constructive alignment: student outcomes. *Teaching in higher education,* 18**,** 192-204.

Lawrence, K. E., Reed, K. L. & Locander, W. 2011. Experiencing and measuring the unteachable: Achieving AACSB learning assurance requirements in business ethics. *Journal of Education for Business,* 86**,** 92-99.

Makhachashvili, R. 2021. Soft Skills and ICT Tools for Final Qualification Assessment in Universities of Ukraine and India in COVID-19 Framework. *Psychology and Education,* 58**,** 849-861.

Oliver, B. 2013. Graduate attributes as a focus for institution-wide curriculum renewal: innovations and challenges. *Higher Education Research & Development,* 32**,** 450-463.

Oliver, B. & Jorre De St Jorre, T. 2018. Graduate attributes for 2020 and beyond: Recommendations for Australian higher education providers. *Higher Education Research & Development,* 37**,** 821-836.

Orr, S. 2007. Assessment moderation: constructing the marks and constructing the students. *Assessment & Evaluation in Higher Education,* 32**,** 645-656.

Riebe, L. & Jackson, D. 2014. The use of rubrics in benchmarking and assessing employability skills. *Journal of Management Education,* 38**,** 319-344

Sadler, R. 2009. Grade integrity and the representation of academic achievement. *Studies in Higher Education,* 34**,** 807-826.

Sala, A., Punie, Y., Garkov, V. & Cabrera, M. 2020. LifeComp: The European framework for personal, social and learning to learn key competence. Joint Research Centre (Seville site).

Spendolini, M. J. 1992. *The Benchmarking Book,* New York, American Management Association.

Sridharan, B., Leitch, S. & Watty, K. 2015. Evidencing learning outcomes: a multi-level, multi-dimensional course alignment model. *Quality in Higher Education,* 21**,** 171-188.

Stal, J. & Paliwoda-Pękosz, G. 2019. Fostering development of soft skills in ICT curricula: a case of a transition economy. *Information Technology for Development,* 25**,** 250-274.

Succi, C. & Canovi, M. 2020. Soft skills to enhance graduate employability: comparing students and employers' perceptions. *Studies in higher education,* 45**,** 1834-1847.

Succi, C. & Wieandt, M. 2019. Walk the talk: soft skills' assessment of graduates. *European Journal of Management and Business Economics,* 28**,** 114-125.

Thomas, J. P., Pondish, C. & Secolsky, C. 2013. Issues in Institutional Benchmarking of Student Learning Outcomes Using Case Examples. *Research in Higher Education Journal,* 20.

Villán-Vallejo, A., Zitouni, A., García-Llamas, P., Fernández-Raga, M., Suárez-Corona, A. & Baelo, R. 2022. Soft Skills and STEM Education: Vision of the European University EURECA-PRO. *BHM Berg-und Hüttenmännische Monatshefte,* 167**,** 485-488.

Watson, E. G. S., Steketee, C., Mansfield, K., Moore, M., Dalziel, B., Damodaran, A., Walker, B., Duvivier, R. J. & Hu, W. 2020. Curriculum mapping for health professions education: A typology. *Focus on Health Professional Education: A Multi-Professional Journal,* 21**,** 91-113.

Zeleny, M. 1982. *Multiple Criteria Decision Making,* New York, McGraw-Hill.

Zhang, H., Su, S., Zeng, Y. & Lam, J. F. 2022. An experimental study on the effectiveness of students' learning in scientific courses through constructive alignment—a case study from an MIS course. *Education Sciences,* 12**,** 338.