Board 73: AI Skills-based Assessment Tool for Identifying Partial and Full-Mastery within Large Engineering Classrooms

Mr. Amirreza Mehrabi, Purdue University

I am Amirreza Mehrabi, a Ph.D. student in Engineering Education at Purdue University, West Lafayette. Now I am working in computer adaptive testing (CAT) enhancement with AI and analyzing big data with machine learning (ML) under Prof. J. W. Morphew at the ENE department. My master's was in engineering education at UNESCO chair on Engineering Education at the University of Tehran. I pursue Human adaptation to technology and modeling human behavior(with machine learning and cognitive research). My background is in Industrial Engineering (B.Sc. at the Sharif University of Technology and "Gold medal" of Industrial Engineering Olympiad (Iran-2021- the highest-level prize in Iran)). Now I am working as a researcher in the Erasmus project, which is funded by European Unions (1M $_European Union & 7 Iranian Universities) which focus on TEL and students as well as professors' adoption of technology(modern Education technology). Moreover, I cooperated with Dr. Taheri to write the "R application in Engineering statistics" (an attachment of his new book "Engineering probability and statistics.")

Dr. Jason Morphew, Purdue University

Jason W. Morphew is an Assistant Professor in the School of Engineering Education at Purdue University. He earned a B.S. in Science Education from the University of Nebraska and spent 11 years teaching math and science at the middle school, high school, and community college level. He earned a M.A. in Educational Psychology from Wichita State and a Ph.D. from the University of Illinois Urbana-Champaign.

Work-In-Progress: AI skill-based assessment tool for identifying partial and full mastery within Large Engineering Classrooms

Abstract

In the evolving landscape of engineering education, there's a pronounced shift from traditional assessment to mastery-based curricula that evaluates student skill mastery rather than a simple percentage correct. This shift in assessment is often aligned with curricular shift from passive to active learning pedagogies, which help to foster design thinking, and accentuates creative and iterative problem-solving. However, for mastery-based instruction to be effective, educators must have access to assessments that measure individual nuances and differences in conceptual understanding for each student to design instruction to individually promote skill mastery and provide individual and meaningful feedback. The purpose of this ongoing study is to introduce the Partial-Mastery Cognitive Diagnosis model as an Artificial Intelligent-driven tool to optimize and assess skill mastery within large engineering classroom assessments. The model classifies specific cognitive errors made by students and defines new ways of identifying students who have not fully mastered a skill but have an explainable cognition error. This study presents the results of a theoretical simulation study that aims to examine the potential for Partial-Mastery Cognitive Diagnosis model to classify students who have partially, but not fully, mastered skills in engineering courses. The results can enable educators identify students' current conceptual models in order to create targeted interventions to rectify misconceptions, to adapt curriculum based on student mastery, and to provide individualized feedback.

Keyword: cognitive diagnosis model; mastery learning; adaptive learning; adaptive assessment; artificial intelligence in education

Introduction

Research indicates that college and engineering students often lack essential skills required by employers, such as communication, decision-making, problem-solving, leadership, emotional intelligence, and social ethics [1], [2]. This gap between college preparation and career demands is particularly evident in the engineering field, where technical knowledge is prioritized over soft skills like creativity, innovation, leadership, management, and teamwork [3]. Moreover, the shift from traditional instruction to skill-based curricula has gained momentum in educational settings to center the learner in education. This approach encourages students to engage in hands-on activities, problem-solving, critical thinking, design thinking, domain-specific skills, and knowledge-domain skills. While most of these skills can be achieved in an active learning setting, assessment is an essential pillar of active learning.

To be effective, assessment should be closely tied to learning objectives, which detail the skills to be learned and their intended outcomes. Notably, each skill can be broken down into different steps or levels, adding granularity to the learning and assessment process. Therefore, a primary goal of such assessment is to verify the level of skill mastery attained [4]. This approach not only measures progress but also provides a clear view of the learner's journey towards skill proficiency. In doing so, assessments serve as a crucial tool, reflecting the learner's understanding and competence, and ultimately guiding them towards targeted improvement and development. However, in active learning the content is usually taught for a set amount of time, and a student's aptitude is based on how much they learned in that time. Conversely, mastery learning assumes that all students, given enough time and intervention, can eventually master the content [4], [5], [6]. Learning within mastery frameworks concerns itself with identifying learning trajectories and providing students with curriculum for gaining knowledge and skills, assessing mastery through formative assessments, and providing feedback to help students master one set of skills before moving on to the next set [7].

There have been many approaches for developing formative assessments to measure skill mastery, such as Cognitive Diagnostic Models (CDM). However, some of the assumptions of some of CDMs may not meet the requirements of mastery learning [8], [9]. One limitation of CDMs is the lack of precision for measuring mastery, which can limit the personalization of feedback. In this ongoing study, we initially explore perspectives on the evaluation of mastery learning. Subsequently, we discuss a new way to evaluate assessments in a way that more precisely classifies students based on their level of mastery rather than a simple dichotomous determination. This approach aims to comprehensively understand and evaluate how skills acquisition is conceptualized and assessed across different educational perspectives and assessment frameworks [10], [11].

Mastery Learning Theory

The concept of mastery learning is pivotal in education, offering a structured framework for ensuring that students achieve a high level of understanding before progressing. Bloom found that the most effective learning scenario involved one-on-one tutoring, where students first comprehend the material, undergo formative assessments, and receive proportional corrective activities if needed. Mastery learning aims for all students to master the course material, requiring instructors to clearly define learning outcomes, organize topics into intervals, and provide enrichment or remediation based on students' demonstrated mastery in formative assessments [12]. Prior to moving onto more intricate topics. Additionally, it underscores the importance of fostering abilities critical for analytical thinking and real-world application, such as analytical problem-solving and experimental methods, to guarantee learners are well-equipped to utilize their learning in practical scenarios [6]. Bloom's mastery learning model encapsulates the core tenets of this educational approach by emphasizing the definition, planning, teaching, and grading for mastery. It recognizes the assessment of mastery as a crucial pillar, addressing the research question of how student mastery is monitored and identifying the assessment models capable of measuring it [6].

Mastery Learning Steps

Bloom Mastery learning commences with the precise delineation of learning objectives, a step also referred to as standard setting [13]. Educators undertake the responsibility of identifying essential concepts, knowledge, or skills and determining the proficiency level at which students should engage with this new information. The formulation of these objectives adheres to the SMART criteria—specific, measurable, achievable, realistic, and timely model—emphasizing minimum acceptable performance within a specific context [6].

The second pillar of mastery learning involves formative assessments, serving to furnish feedback on progress toward the predefined learning outcomes [12]. These assessments are strategically designed to pinpoint and rectify any learning difficulties, thereby guiding future efforts. Notably, formative assessments should align with the complexity and format of the learning objectives, ensuring coherence between instructional content and assessment criteria. Through recurrent cycles of this process, the mastery learning model systematically steers learners toward their educational goals ([14], [15]; Figure 1).

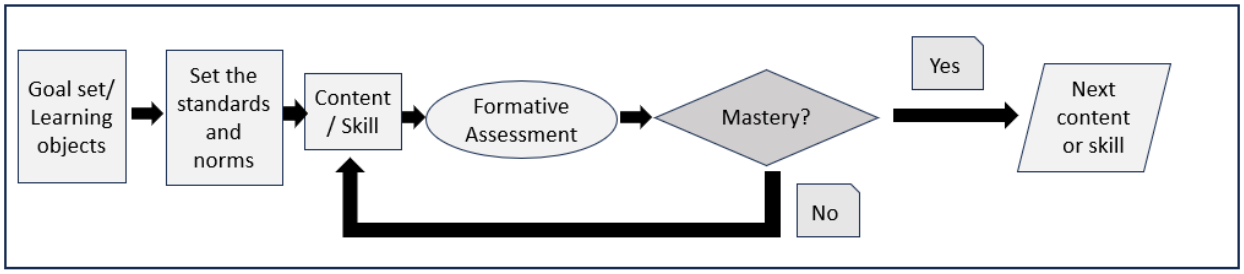

Figure 1. The Mastery Learning model of assessment (adapted from the designed model by [6])

Figure 1 overlooks a crucial aspect, with the pivotal layer being Bloom's mastery theory, particularly in the context of feedback [15]. To elucidate, in the feedback loop responding to instances of "not mastery," the return path to the learning object and material is acknowledged. However, within a comprehensive session of adaptive learning, which embodies the primary objective of mastery learning, the distinction between two students who provide incorrect answers is imperative. In Bloom's envisioned system, students are intended to receive tailored feedback, thereby aligning with the assessment model depicted in Figure 2 within the Mastery Learning framework [6].

Moreover, learning management systems enhanced with artificial intelligence, such as IBM Watson Media and Google Cloud Video Intelligence API, leverage advanced artificial intelligence and machine learning algorithms for content-based video search, enabling diverse applications from surveillance to media. As Staples [7] mentioned, the optimized feedback in Bloom Mastery Learning approach is an undeniable requirement to focus on the problematic sections. Consequently, Figure 2 aligns more with both the recent developments in artificial intelligence and relies on the mastery learning theory [16], [17].

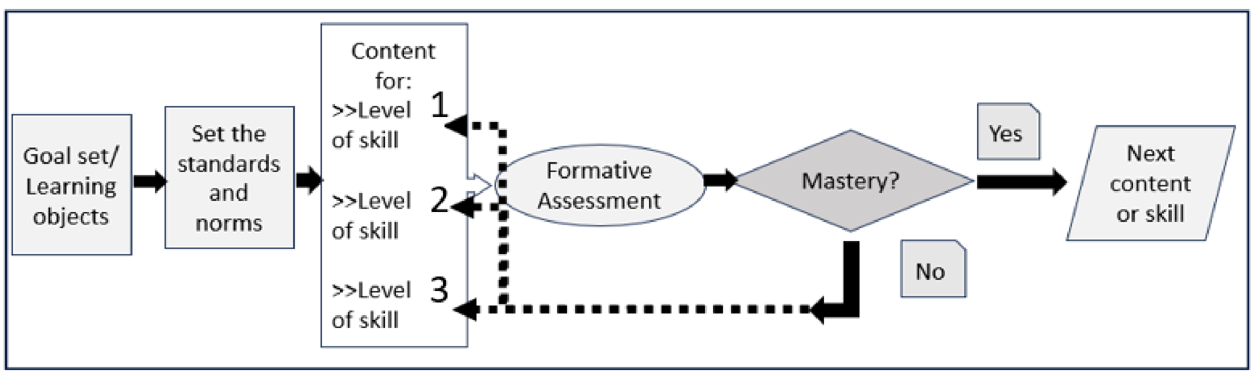

Figure 2. The Mastery Learning model of assessment for adaptive learning systems

Mastery Learning Assessment

In accordance with the fundamental pillars of Bloom's mastery learning, assessment should adopt an adaptive approach tailored to individual students, eschewing comparison in favor of replacing standard norms [12], [13], [18]. Thus, in various class scenarios, the assessment should be aligned with the specific purposes of the course and the material, as outlined in Table 1.

Within the domain of mastery learning measurement methods (Table 1), mastery testing encompasses two primary approaches. The initial approach entails setting a grade threshold to distinguish mastery, while the second involves assessing each individual's performance relative to the class-specific mean, identifying a specific subset of students recognized as upper-level masters in the respective topic. These approaches provide flexibility for application across diverse population sizes. There are some requirements for this approach like the standard questions and clear relation between the question with the learning object, content, or/and skills. While this method widely utilized, the conventional practice of assigning a standard percentage or weighted percentage grade fails to discern skill or concept level mastery and fails to provide a standardized means of comparing mastery levels across different classes. This becomes particularly pertinent in the context of nationwide programs where the assessment of skill or concept mastery achievements necessitates a standardized approach [15], [19].

Table 1. Different methods of assessing the Mastery Learning Model [12], [13], [15], [19]

Measurement Method	Description	Key Features
Criterion-Referenced Testing	Evaluates if students have achieved specific learning objectives to the desired level of mastery, independent of peer performance.	- Direct comparison against standards - No peer comparison
Formative Assessments	Ongoing assessments that provide feedback on student understanding and guide instructional adjustments.	- Continuous feedback loop
Rubrics and Scoring Guides	Provide explicit criteria and expectations for assignments or tasks, outlining what mastery looks like for each objective.	- Detailed evaluation criteria - Transparent grading process
Performance-Based Assessments	Require students to apply knowledge and skills in practical scenarios, demonstrating their mastery through real-world or simulated tasks.	- Application of knowledge - Real-world or simulated scenarios
Portfolio Assessments	A collection of a student's work over time, offering a comprehensive view of their learning journey and progression towards mastery.	- Holistic view of student's work - Evidence of learning progression
Mastery Tests	Tests designed to measure if students have achieved a high level of understanding, with a high threshold for passing.	- High standards for passing - Administered after instruction and correction
Self and Peer Assessments	Encourages reflective learning and meta-cognition by having students evaluate their own or peers' work against mastery criteria.	- Promotes self-reflection and meta-cognitive skills
Reassessment Opportunities	Allows students multiple attempts to demonstrate mastery, providing additional instruction and chances to be reassessed if initial mastery is not achieved.	- Supports mastery over time

Conversely, Item Response Models (IRMs), present a comprehensive analysis by delineating item (question) features such as difficulty, discrimination, and guessing parameters. Although IRMs are effective in assessing proficiency across various contexts, they encounter challenges when classifying mastery situations. This is mainly due to their inability to acknowledge that different skills can underlie a question. Another type of model used in mastery learning are CDMs [20] which classify students in two class of mastery (mastery, non-mastery). Indeed, IRMs are used to model the relationship between individuals' latent traits and their responses to test items, providing a nuanced understanding of ability. CDMs go beyond IRMs by diagnosing specific cognitive skills contributing to test performance, offering targeted insights for instructional purposes in educational assessment. CDMs have been utilized to detect distinct patterns of attribute mastery, providing valuable insights into students' cognitive abilities and learning outcomes [21]. These models employ discrete latent variables to categorize students into profiles indicating their mastery status of each attribute based on their question responses [22], [23]. The use of CDMs allows for the assessment of students' mastery of specific cognitive skills, providing fine-grained information about their learning strengths and weaknesses and assess skill mastery in students, providing a framework for evaluating changes in skill profile mastery over time [21]. In CDMs, diagnoses encompass identifying specific cognitive skills or attributes that a student has either mastered or not mastered within a given domain. For example, in a mathematics assessment of fifth grade, diagnoses might include whether a student has mastered concepts like multiplication, division, fractions, or algebraic equations, allowing for a nuanced understanding of the student's proficiency levels across various skills within the subject area.

A pivotal mathematical model within CDMs is the Deterministic Inputs, Noisy "and" Gate (DINA) model, which assesses mastery or non-mastery statuses across multiple cognitive skills based on raw question responses [21], [24]. The DINA model, a latent class model, classifies students into skill mastery profiles based on their responses to exam questions, with each question having a specific relation to one or more skills [21], [24]. The linkage between questions and their corresponding intended skills are captured in a Q-matrix, a matrix of ones and zeros indicating which questions require a particular skill in order to correctly answer a question, facilitating precise skill mastery assessment [25]. Dina model can accept the dichotomous responses matrix and Q-matrix [21], [24]. While CDMs in general, and the DINA model in particular, have proven valuable in evaluating skill mastery, certain limitations warrant consideration. The DINA model's reliance on dichotomous responses (i.e., correct or incorrect) assumes an all-or-nothing attribute paradigm, potentially oversimplifying skill mastery complexities, particularly when students exhibit partial mastery [5], [25]. Furthermore, the binary nature of mastery or non-mastery statuses in the DINA model and CDMs in general may oversimplify the patterns of partial mastery displayed by most students [5].

Critiques of CDMs extend to their static nature of capturing the dynamic evolution of skill mastery over time. In other words, CDM-based assessments provide static snapshots that may not fully capture students' abilities. The DINA model's assumption of equal skill mastery needed to correctly answer different questions may oversimplify the multidimensionality of skills and impact the accuracy for assessments to measure cognitive abilities [18]. In response to these limitations, researchers have introduced advanced models like Generalized Deterministic-Input, Noisy “And” gate (GDINA) [9], Higher-Order–Deterministic-Inputs, Noisy “And” gate (HO-DINA) [26], [27], aiming to address the simplifying assumptions underlying classical CDMs [18], [25]. Specifically, GDINA model are able to accept polytomous responses from students and consider the higher order skills by polytomous labels of skills while do not accept the partially mastery on skill profile. These models, while more intricate, attempt to overcome issues such as polytomous responses and varying skill levels under each question. However, their implementation necessitates large sample sizes for stable estimates, rendering them susceptible to model misspecification and misfit. The challenge of identifying the optimal model for data fitting, coupled with the difficulty of providing a class that represents partial mastery, underscores the complexity and potential oversimplification inherent in employing single parametric CDMs [25], [28].

The CDM and DINA models strive to maintain the mastery learning assumption by imposing stringent criteria, including the necessity for correct responses to all questions associated with a particular skill for mastery classification. Indeed, a main assumption of DINA model is that achieving a correct response to an item necessitates that an examinee possessing all requisite skills avoids errors attributable to slipping, while an examinee deficient in one or more of the necessary skills must correctly guess to provide a correct answer. However, these models encounter limitations in terms of computational complexity, inflexible mastery/non-mastery categorization, and the challenge of capturing higher levels of skills underlying a question. In particular, two questions can require one skill but at different levels of application [8], [29]. Indeed, CDM and DINA models are computationally very hard to implement, and sensitive to differences in student population and estimation models. Williams et al., [14] and Youn et al. [30] both found that machine learning algorithms, particularly Naive Bayes, can accurately predict cognitive impairment based on neuropsychological data as machine learning methods can react faster to any type of diagnosis. Bratić [31] and Ma et al., [32] further support applying machine learning to satisfy the need for early diagnostic practice or understanding any specific diagnosis while the exam is administered.

Attempts had been made to classify the students based on the assumption of CDM and specifically DINA models to use the benefits of model parameters like guessing and slip parameters. Indeed, within the DINA model, the slip parameter is associated with a "mastery" student responding incorrectly due to slip, and the guess parameter is associated with a "non-mastery" student answering correctly due to guessing. However, these attempts use simpler models that accept more variety of inputs like Support Vectors Machines models [33], but do not attempt to add additional classification groups. However, supervised machine learning models, exemplified by Support Vector Machines and Naive Bayes, necessitate a response variable for training. This variable may represent the mastery and non-mastery status of students. However, this labeling is challenged by the stringent assumption of the DINA model regarding non-mastery labels, rendering it unreliable and precluding the assurance of accurate non-mastery identification. Furthermore, this assumption does not account for the existence of a middle ground denoting partial mastery. In the realm of CDMs, no model accommodates more than two classes, specifically signifying mastery and non-mastery. It is imperative to underscore that these supervised models can solely be trained using the available information on guessing and slip parameters [34][35].

The K-means algorithm, an unsupervised learning method, emerges as a solution to these constraints by offering a means to classify students' learning behaviors without the need for predefined categories [34]. This algorithm analyzes patterns in students' response data to group them into clusters based on similarities in their learning behaviors or ability levels [34]. Such an approach allows for the identification of nuanced learning states beyond the binary mastery/non-mastery paradigm, facilitating a more detailed understanding of student performance and potential areas for intervention [34]. By employing K-means, educators and researchers can better tailor educational content and strategies to meet the diverse needs of students, as demonstrated by Lakshman et al. [35] in their application of K-means for classifying student abilities within an Item Response Theory framework. This method's flexibility and ability to uncover hidden patterns in data make it an invaluable tool for enhancing the personalization and effectiveness of educational interventions, ultimately contributing to improved student learning outcomes.

Partial-Mastery Cognitive Diagnosis Model Construction

K-mean

The K-means clustering methodology comprises several distinct steps aimed at partitioning a dataset into ' K ' clusters. In the initialization phase, K centroids are selected, denoted as $C^{(0)} = \left\{c_1^{(0)}, c_2^{(0)}, \ldots, c_k^{(0)}\right\}$, with the option for random or systematic centroid assignment. The assignment step involves associating each data point $\boldsymbol{x}_i$ with the nearest centroid c_j, determined by $j = \arg\min_l d\left(\boldsymbol{x}_i, c_l^{(t)}\right)$, where d represents the chosen distance function, commonly the Euclidean distance. The formula for Euclidean distance is $d(x_i, c_l) = \sqrt{\sum_{m=1}^{n}\left(x_{i,m} - c_{l,m}\right)^2}$, with n being the dimensionality of the data [36], [37].

Subsequently, the centroids are updated in the next step, denoted as $C^{(t+1)} = \left\{\frac{1}{|S_j|}\sum_{x_i \in S_j} x_i \,\middle|\, j = 1,2,\ldots,k\right\}$, calculated as the mean of the data points in each cluster j, represented by S_j. Iterative execution of assignment and update steps continues until convergence, marked by minimal changes between consecutive centroids or reaching a specified iteration limit [36].

The final step entails assigning data points to clusters based on converged centroids, concluding the clustering process. The choice of the distance function is integral, with metrics like Euclidean distance, Manhattan distance, or cosine similarity applied based on data characteristics. Determining the optimal number of clusters K is pivotal, often addressed through techniques like the Elbow Method or Silhouette Analysis, and methodology robustness is enhanced by multiple initializations and averaging results [34], [37].

Integration of Cognitive Diagnosis Into the K-mean Model

The methodology introduces a weighting scheme (W_j) or each assessment question, meticulously calculated as inversely proportional to its associated guessing (G_j) and slipping parameter (L_j) as Eq 1.

$$W_j = \frac{1}{1+G_j+L_j} \qquad \text{(Eq 1)}$$

Subsequently, the adjustment of student skill mastery levels ($\hat{S}_{ij}$) is executed through a process that harmoniously blends these weights with the q-matrix (Eq. 2 & Table 2). This ensures that only pertinent questions inform the adjusted mastery level for each skill, articulated as:

$$\hat{S}_{ij} = \frac{\sum_{m=1}^{M} W_m Q_{mj} S_{ij}}{\sum_{m=1}^{M} Q_{mj}} \qquad \text{(Eq 2)}$$

The mastery levels (S_{ij}) for student i in skill j, where m is the total number of questions, constitute a crucial preprocessing step. This process aligns the clustering with the educational context, creating clusters that more accurately reflect the educational constructs measured. By adjusting skill mastery levels to mitigate guessing and slipping effects, the methodology ensures that subsequent K-means clustering is grounded in a representation of student abilities that closely approximates true skill mastery. The resulting matrix of adjusted skill mastery levels serves as a robust foundation for clustering, enabling the identification of meaningful student groups based on skill profiles. These groups inform targeted instructional strategies, fostering a personalized learning environment that addresses the diverse educational needs of students.

Table 2: Sample of Q-matrix for simulation

question	Skill number one	Skill number two	Skill number three
1	0	0	1
2	2	1	1
3	2	1	0
4	0	1	1
⋮	⋮	⋮	⋮
15	2	0	1

Note: *Number 2 means that question requires a higher level of skill mastery to correctly solve.*

Customization of K-Means Clustering

The core modification to the K-means clustering algorithm involved the introduction of a custom distance function, designed to compute the dissimilarity between student skill profiles and cluster centroids. This function, $D(S_i, C_k)$, calculates the weighted Euclidean distance between a student's skill profile S_i and a cluster centroid C_k:

$$D(S_i, C_k) = \sqrt{\sum_{j=1}^{n} W_j \cdot \left(\hat{S}_{ij} - C_{kj}\right)^2} \qquad \text{(Eq 3)}$$

In Eq. 3, n represents the number of $skills$, and the weights adjust the influence of each skill in the distance calculation based on the associated guessing and slipping probabilities. The use of the Q-matrix (Table 2) in modulating the distance function allows the clustering process to reflect the specific contributions of different skills to the assessment questions, thereby aligning the clustering more closely with educational theory and practice.

By applying this custom distance function (Eq.3) in K-means, the algorithm (Eq. 4) tries to minimize the total within-cluster variance. Where C_k is the set of points in cluster *k*, and μ_k is the centroid of cluster *k*. The algorithm proceeds by assigning points to clusters and updating centroids based on this weighted distance (Eq. 3).

$$\text{Min} \sum_{k=1}^{K} \sum_{x \in C_k} d_w(\hat{S}_{ij}, \mu_k, W_j) \qquad \text{(Eq 4)}$$

Simulation Methodology

Our study embarked on a path to refine the traditional K-means clustering algorithm by weaving in intricate elements of Cognitive Diagnostic Models (CDMs), with a specific lens on the unique landscape of educational assessment data. Therefore, we first simulate the responses and Q-matrix (question-skill) by considering the guessing and slip parameters and computing the probability of the response of each student to the question [34], [37]. In our response simulation, as cited in [9], a simulated population of more than 200 individuals adequately covers a comprehensive range of response scenarios that account for both slip and guessing parameters. The research presented in [9] and [38] further supports the premise that simulating responses can yield models closely mirroring empirical conditions. This approach enhances the validity of simulation, especially within the realms of CDMs and IRMs, by providing a rigorous method for model verification.

To create a realistic educational dataset, we simulated the responses of 200 students across a suite of 15 assessment questions. This choice of question numbers was motivated by the desire to mirror realistic assessment conditions, minimizing the potential for student fatigue and providing a reasonable parameter of guessing and slip. The simulation was meticulously designed to encapsulate the guessing and slipping parameters emblematic of the GDINA model, aiming to forecast the probability of each student's correct response to the individual questions. also, three skills designed vary from simple to higher-order application of skills [34], [37].

Results

To evaluate the effectiveness of our clustering approach, we calculated several fit statistics, including the Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Index, for each cluster. These metrics provided quantitative insights into the cohesiveness and separation of the clusters, enabling a comprehensive assessment of the clustering quality [34], [37]. The Silhouette Score helps us understand how well our clustering worked by measuring how close each data point is to others in its own cluster compared to those in the nearest cluster. A higher score means the clusters are clear and distinct, with each cluster's members being close together and far from other clusters [34], [37]. The Calinski-Harabasz Index, also known as the Variance Ratio Criterion, looks at how separate clusters are by comparing the variance (or spread) of data points between clusters to the variance within each cluster [34], [37]. A higher index means the clusters are well-spaced out, indicating a good clustering job. The Davies-Bouldin Index checks how compact (tight) and separated the clusters are by comparing distances within and between clusters [34], [37]. Lower scores here mean the clusters are tight and well-separated, which is what we want. However, the RMSEA doesn't work well for checking how good our clustering is because it's meant for a different type of analysis involving models and observed data, not for grouping data points based on similarity [34], [37]. Herein, we discuss the clustering outcomes for each skill, elucidating the educational implications and insights derived from these results.

Clustering Quality and Segmentation

The results across all three skills exhibit a pattern of high Silhouette Scores, indicative of cohesive and distinct clusters (around 0.5). The Calinski-Harabasz Indices underscore the effectiveness of our methodology in achieving significant separation and density among clusters, a testament to the integration of CDM parameters into the clustering process. The Davies-Bouldin Indices, consistently below 0.5 for all skills, validate the distinctiveness of the clusters formed, with minimal overlap in the skill mastery profiles of different clusters [34], [35], [37].

These findings underscore the value of incorporating educational assessment characteristics, such as the Q-matrix, guessing, and slipping probabilities, into clustering algorithms. By doing so, we have demonstrated the potential to achieve a more granular and educationally meaningful segmentation of student skill mastery. This enhanced clustering approach not only facilitates a deeper understanding of student skill profiles but also offers actionable insights for personalized instruction and targeted educational interventions (Table 3).

Table 3. The model fit results

Skills	Fit Statistics	Test value
Skill number one	Silhouette Score	0.62
	Calinski-Harabasz Index	838.12

	Davies-Bouldin Index	0.472
Skill number two	Silhouette Score	0.61
	Calinski-Harabasz Index	826.30
	Davies-Bouldin Index	0.481
Skill number three	Silhouette Score	0.60
	Calinski-Harabasz Index	707.28
	Davies-Bouldin Index	0.499

Post-hoc classification

In our study, we utilized the K-means clustering algorithm to organize students into distinct groups based on their proficiency across various skills, as detailed in Table 4. This classification enabled us to discern clear patterns in the skill mastery levels of the student cohort, facilitating an insightful comparison between students' cluster assignments and their actual skill competencies as depicted in Table 5 and Figure 4. Table 5, specifically indicates the relation between the clustering result of the K-means and mastery situation which comes from the post-hoc analysis. Such an analysis provides a granular view of how student proficiencies are distributed among the identified clusters, highlighting the diversity in mastery levels within the studied population.

Table 4: Students cluster dependency of each skill

Students	Skill number one cluster	Skill number two cluster	Skill number three cluster
1	1	2	1
2	2	1	3
3	1	1	2
4	1	3	2
5	3	3	2
⋮	⋮	⋮	⋮
200	1	1	3

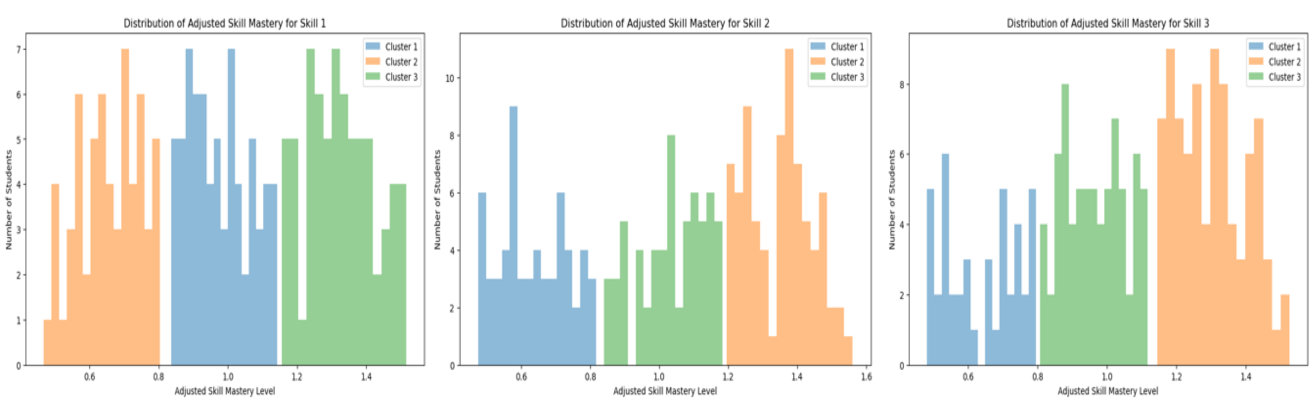

Figure 3. Distribution of clusters for each skill number one (left), skill number two(middle), skill number three (right)

The robustness of resultant clusters hinges on the distribution and scale of the dataset, as this algorithm operates under the assumption that clusters exhibit uniform dimensions. In the context of CDM, preprocessing techniques extend to the qualitative assessment of Q-matrices, in parallel with CDM, K-means clustering depends on the clustering of the initial placement of centroids. So, in the unclear clustering situation of mastery situation for skill number two (Figure 4) the CDM researchers recommend adjustments to the Q-matrix for educators, reflecting the pivotal role of this matrix in refining outcomes.

Table 6 provides an overview of diverse proficiency levels for all three skills among 15 questions. Notably, the middle group of mastery for the skill number three (Partial Mastery ≈50%) is larger than other two portions, which violates the model balancing for the k-means algorithm. In this situation, the center of the second group in skill number three (C_{23} in Eq. 3) will be at the middle and its performance for finding the center of the other two clusters (Master (C_{13} in Eq. 3) and Non-Master (C_{33} in Eq. 3)) decrease. Indeed, model select the centers of two other clusters in a non-accurate place which is very close to the center of the second cluster (middle group).

Table 5: Students Skill Mastery Profile

Students	Skill number one cluster	Skill number two cluster	Skill number three cluster
1	Mastery	Non-Mastery	Non-Mastery
2	Partial Mastery	Partial Mastery	Partial Mastery
3	Partial Mastery	Non-Mastery	Mastery
4	Non-Mastery	Non-Mastery	Mastery
5	Non-Mastery	Mastery	Mastery
⋮	⋮	⋮	⋮
200	Partial Mastery	Non-Mastery	Partial Mastery

Figure 4. Clustering by mastery level for each skill. Skill number one (left), Skill number two (middle), Skill number three (right)

Table 6. Portion of different levels of each skill in 15 questions

	Skill number one	Skill number two	Skill number three
Mastery	50.0%	25.0%	31.3%
Partial Mastery	31.3%	31.3%	50.0%

No Mastery	12.5%	37.5%	12.5%

Discussion

Our study shows how a Partial-Mastery Cognitive Diagnosis model that incorporates K-means clustering can effectively group students based on how well they've mastered various skills. By adding special parameters from CDMs into our analysis, we've been able to form meaningful groups of students that really show the range of skills they've mastered, including those who have only partially mastered skills. This method is a step forward in educational data analysis and adaptive learning. This model gives us a clear picture of student skills, helping educators understand and meet the different needs of their students.

This attention to detail significantly enhances the effectiveness of the K-means model, particularly in scenarios where the distinctions between different skill levels of partially mastery model are minimal [39]. Indeed, the key to better clustering is not just increasing the number of students; it's equally crucial to pay attention to skill level distribution among questions [40]. This attention to detail significantly enhances the effectiveness of the K-means model, particularly in scenarios where the distinctions between different skill levels are minimal.

Employing the K-means clustering method allows educators to tailor educational experiences, ensuring that each student's unique strengths are identified needs and their unique needs are addressed. This personalized approach not only can facilitate the creation of targeted lessons but also opens the door to recognizing the multifaceted nature of student learning. Beyond merely categorizing students by their mastery levels, this method allows for the identification of specific learning trajectories, preferences, and obstacles that individual students may face.

Using this K-means clustering approach, instructors can see exactly what skills each student has mastered, and the skills that are they are in the process of mastering. By having a skill profile, instructors can identify student needs, which helps them create interventions and lessons that are personalized for each student. For example, engineering students who have partially mastered integration can receive extra support with their calculus skills while still receiving appropriate instruction in a content topic such as Thermodynamics. In addition, students who have mastered integration might be given more challenging Thermodynamics problems to keep them engaged. In this way, Partial-Mastery Cognitive Diagnosis models allow a sophisticated way to use formative assessment data to guide their teaching. This data-driven approach shows promise for allowing instructors in large-enrollment courses to more easily individualize instruction to improve learning for all students by making sure everyone gets support that's tailored to their ability and skill mastery levels [39], while helping classrooms become more inclusive by recognizing and supporting diverse students' learners.

Looking ahead, future directions for this work could involve integrating K-means clustering with other machine-learning techniques to refine the understanding of student learning behaviors further. Deep learning could help predict student performance trends, enabling proactive adjustments to teaching strategies. Moreover, exploring the correlation between clustered learning styles and long-term academic outcomes could yield valuable insights into how personalized education impacts student success beyond the classroom.

Additionally, the integration of this approach within digital learning platforms could facilitate real-time and focused feedback loops, where instructional content is dynamically adjusted based on the evolving needs of the student body. Such an adaptive educational environment could revolutionize the traditional one-size-fits-all teaching model, fostering a more inclusive and effective learning atmosphere that celebrates and cultivates individual differences. This expansion provides a more comprehensive view of the potential impact of the K-means clustering approach on education, highlighting its capacity to foster a deeper understanding of student learning and outlining future avenues for research and application.

References

[1] A. Adeseye, “Migrant Remittance and Household Expenditure Pattern in Nigeria,” Open J Polit Sci, vol. 11, no. 01, pp. 73–98, 2021, doi: 10.4236/OJPS.2021.111006.

[2] M. Y. Kamarudin, N. M. R. N. Yusoff, H. Yamat, and K. Abdul Ghani, “Inculcation of Higher Order Thinking Skills (HOTS) in Arabic Language Teaching at Malaysian Primary Schools,” Creat Educ, vol. 07, no. 02, pp. 307–314, 2016, doi: 10.4236/CE.2016.72030.

[3] H. Baytiyeh and M. Naja, “Identifying the challenging factors in the transition from colleges of engineering to employment,” European Journal of Engineering Education, vol. 37, no. 1, pp. 3–14, Mar. 2012, doi: 10.1080/03043797.2011.644761.

[4] T. R. Guskey, “Closing Achievement Gaps: Revisiting Benjamin S. Bloom’s ‘Learning for Mastery,’” http://dx.doi.org/10.4219/jaa-2007-704, vol. 19, no. 1, pp. 8–31, Nov. 2007, doi: 10.4219/JAA-2007-704.

[5] M. Hansen, L. Cai, S. Monroe, and Z. Li, “Limited-information goodness-of-fit testing of diagnostic classification item response models,” British Journal of Mathematical and Statistical Psychology, vol. 69, no. 3, pp. 225–252, Nov. 2016, doi: 10.1111/BMSP.12074.

[6] M. Winget and A. M. Persky, “A Practical Review of Mastery Learning,” Am J Pharm Educ, vol. 86, no. 10, p. ajpe8906, Dec. 2022, doi: 10.5688/AJPE8906.

[7] M. Staples, “Personalized Feedback: Testing a Tutoring System That Was Informed by Learning Analytics,” Instructional Psychology and Technology Graduate Student Projects, Jun. 2019, Accessed: Mar. 09, 2024. [Online]. Available: https://scholarsarchive.byu.edu/ipt_projects/19

[8] J. Liu, W. Tang, X. He, B. Yang, and S. Wang, “Research on DINA Model in Online Education,” Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol. 340, pp. 279–291, 2020, doi: 10.1007/978-3-030-63955-6_24.

[9] J. de la Torre, “The Generalized DINA Model Framework,” Psychometrika, vol. 76, no. 2, pp. 179–199, 2011, doi: 10.1007/S11336-011-9207-7.

[10] N. M. N. Mathivanan, N. A. Nor, and R. M. Janor, “Improving Classification Accuracy Using Clustering Technique,” Bulletin of Electrical Engineering and Informatics, vol. 7, no. 3, pp. 465–470, Sep. 2018, doi: 10.11591/EEI.V7I3.1272.

[11] S. Rakic, N. Tasic, U. Marjanovic, S. Softic, E. Lüftenegger, and I. Turcin, “Student Performance on an E-Learning Platform: Mixed Method Approach,” International Journal of Emerging Technologies in Learning (iJET), vol. 15, no. 02, pp. 187–203, 2020, doi: 10.3991/IJET.V15I02.11646.

[12] J. R. Frank, L. Snell, R. Englander, and E. S. Holmboe, “Implementing competency-based medical education: Moving forward,” Med Teach, vol. 39, no. 6, pp. 568–573, Jun. 2017, doi: 10.1080/0142159X.2017.1315069.

[13] B. J. Zimmerman and M. K. Dibenedetto, “Mastery learning and assessment: Implications for students and teachers in an era of high-stakes testing,” Psychol Sch, vol. 45, no. 3, pp. 206–216, Mar. 2008, doi: 10.1002/PITS.20291.

[14] J. A. Williams, A. Weakley, D. J. Cook, and M. Schmitter-Edgecombe, “Machine Learning Techniques for Diagnostic Differentiation of Mild Cognitive Impairment and Dementia,” Expanding the Boundaries of Health Informatics Using Artificial Intelligence: Papers from the AAAI 2013 Workshop, 2013, Accessed: Feb. 05, 2024. [Online]. Available: www.aaai.org
[15] T. R. Guskey, “Mastery Learning,” Encyclopedia of the Sciences of Learning, pp. 2097–2100, 2012, doi: 10.1007/978-1-4419-1428-6_1553.
[16] A. McDonald, H. McGowan, M. Dollinger, R. Naylor, and H. Khosravi, “Repositioning students as co-creators of curriculum for online learning resources,” Australasian Journal of Educational Technology, vol. 37, no. 6, pp. 102–118, 2021, doi: 10.14742/AJET.6735.
[17] L. Chen, P. Chen, and Z. Lin, “Artificial Intelligence in Education: A Review,” IEEE Access, vol. 8, pp. 75264–75278, 2020, doi: 10.1109/ACCESS.2020.2988510.
[18] M. L. Graber, N. Franklin, and R. Gordon, “Diagnostic Error in Internal Medicine,” Arch Intern Med, vol. 165, no. 13, pp. 1493–1499, Jul. 2005, doi: 10.1001/ARCHINTE.165.13.1493.
[19] C. L. O’brien, M. Adler, W. C. Mcgaghie, C. Laird O’brien, M. Adler, and W. C. Mcgaghie, “Assessment in Mastery Learning,” pp. 89–107, 2020, doi: 10.1007/978-3-030-34811-3_5.
[20] C. F. Collares, “Cognitive diagnostic modelling in healthcare professions education: an eye-opener,” Advances in Health Sciences Education, vol. 27, no. 2, pp. 427–440, May 2022, doi: 10.1007/S10459-022-10093-Y.
[21] X. Qiao and H. Jiao, “Explanatory Cognitive Diagnostic Modeling Incorporating Response Times,” J Educ Meas, vol. 58, no. 4, pp. 564–585, Dec. 2021, doi: 10.1111/JEDM.12306.
[22] A. Berestova, G. Burdina, L. Lobuteva, and A. Lobuteva, “Academic Motivation of University Students and the Factors that Influence it in an E-Learning Environment,” The Electronic Journal of e-Learning, vol. 20, no. 2, pp. 201–210, Accessed: Feb. 11, 2023. [Online]. Available: www.ejel.org
[23] J. W. Morphew, M. Silva, G. Herman, and M. West, “Frequent mastery testing with second-chance exams leads to enhanced student learning in undergraduate engineering,” Appl Cogn Psychol, vol. 34, no. 1, pp. 168–181, Jan. 2020, doi: 10.1002/ACP.3605.
[24] L. Bradshaw, A. Izsák, J. Templin, and E. Jacobson, “Diagnosing Teachers’ Understandings of Rational Numbers: Building a Multidimensional Test Within the Diagnostic Classification Framework,” Educational Measurement: Issues and Practice, vol. 33, no. 1, pp. 2–14, 2014, doi: 10.1111/EMIP.12020.
[25] W. Ma, “Evaluating the Fit of Sequential G-DINA Model Using Limited-Information Measures,” Appl Psychol Meas, vol. 44, no. 3, pp. 167–181, May 2020, doi: 10.1177/0146621619843829.
[26] R. Liu, “Misspecification of Attribute Structure in Diagnostic Measurement,” Educ Psychol Meas, vol. 78, no. 4, pp. 605–634, Aug. 2018, doi: 10.1177/0013164417702458.
[27] P. Zhan, W. Ma, H. Jiao, and S. Ding, “A Sequential Higher Order Latent Structural Model for Hierarchical Attributes in Cognitive Diagnostic Assessments,” Appl Psychol Meas, vol. 44, no. 1, p. 65, Jan. 2020, doi: 10.1177/0146621619832935.
[28] W. Ma and W. Guo, “Cognitive diagnosis models for multiple strategies,” British Journal of Mathematical and Statistical Psychology, vol. 72, no. 2, pp. 370–392, May 2019, doi: 10.1111/BMSP.12155.

[29] T. O. Başokçu, T. Öğretmen, and H. Kelecioğlu, "Model Data Fit Comparison between DINA and G-DINA in Cognitive Diagnostic Models," Education Journal, vol. 2, no. 6, p. 256, 2013, doi: 10.11648/J.EDU.20130206.18.
[30] Y. C. Youn et al., "Detection of cognitive impairment using a machine-learning algorithm," Neuropsychiatr Dis Treat, vol. 14, p. 2939, 2018, doi: 10.2147/NDT.S171950.
[31] B. Bratić, V. Kurbalija, M. Ivanović, I. Oder, and Z. Bosnić, "Machine Learning for Predicting Cognitive Diseases: Methods, Data Sources and Risk Factors," J Med Syst, vol. 42, no. 12, Dec. 2018, doi: 10.1007/S10916-018-1071-X.
[32] H. Ma, Z. Huang, H. Zhu, W. S. Tang, H. Zhang, and K. Li, "Predicting examinee performance based on a fuzzy cloud cognitive diagnosis framework in e-learning environment," Soft comput, vol. 27, no. 24, pp. 18949–18969, Dec. 2023, doi: 10.1007/S00500-023-08100-4.
[33] L. Yang and M. Deng, "Based on k-Means and Fuzzy k-Means Algorithm Classification of Precipitation," 2010 International Symposium on Computational Intelligence and Design, vol. 1, pp. 218–221, 2010, doi: 10.1109/ISCID.2010.72.
[34] A. Likas, N. Vlassis, and J. J. Verbeek, "The global k-means clustering algorithm," Pattern Recognit, vol. 36, no. 2, pp. 451–461, Feb. 2003, doi: 10.1016/S0031-3203(02)00060-2.
[35] B. V. Lakshman, J. V. Wijekulasooriya, and M. Sandirigama, "Modified item response theory (IRT) model and k means clustering for agent based e learning system," Proceedings of the 2018 International Conference On Communication, Computing and Internet of Things, IC3IoT 2018, pp. 169–173, Jul. 2018, doi: 10.1109/IC3IOT.2018.8668154.
[36] B. Georgescu, I. Shimshoni, and P. Meer, "Mean shift based clustering in high dimensions: a texture classification example," Proceedings Ninth IEEE International Conference on Computer Vision, vol. 1, pp. 456–463, 2003, doi: 10.1109/ICCV.2003.1238382.
[37] C. Hua, H. Wu, Q. Chen, and T. Wada, "Object Tracking with Target and Background Samples," IEICE Trans Inf Syst, vol. E90-D, no. 4, pp. 766–774, 2007, doi: 10.1093/IETISY/E90-D.4.766.
[38] J. de la Torre, "DINA Model and Parameter Estimation: A Didactic," Journal of Educational and Behavioral Statistics, vol. 34, no. 1, pp. 115–130, Mar. 2009, doi: 10.3102/1076998607309474.
[39] A. M. Baswade, K. D. Joshi, and P. S. Nalwade, "A Comparative Study Of K-Means And Weighted K-Means For Clustering," International Journal of Engineering Research & Technology (IJERT), vol. 1, no. 10, 2012, Accessed: Feb. 04, 2024. [Online]. Available: www.ijert.org
[40] T. Li and Y. Chen, "An improved k-means algorithm for clustering using entropy weighting measures," Proceedings of the World Congress on Intelligent Control and Automation (WCICA), pp. 149–153, 2008, doi: 10.1109/WCICA.2008.4592915.
[41] C. Liu and Y. Cheng, "An Application of the Support Vector Machine for Attribute-By-Attribute Classification in Cognitive Diagnosis," Appl Psychol Meas, vol. 42, no. 1, pp. 58–72, Jan. 2018, doi: 10.1177/0146621617712246