

Next Skill Evaluation & Targeting via Map-Based Analytics

Jeff Hoover, Amy Clark
University of Kansas: ATLAS

Abstract

This study integrates assessment data with learning maps, using a neural network to recommend individualized next skills for instruction and assessment. Results indicate high-certainty, expert-validated recommendations, with empirical support for foundational assumptions. This demonstrates the potential for AI to support more personalized learning-maps-based instruction and assessment.

1 Introduction

Efforts to individualize instruction and assessment have accelerated as adaptive learning systems demonstrate the value of tailoring supports to learners' current knowledge and likely next steps (e.g., Li et al., 2023; Zhang et al., 2025). Advances in artificial intelligence (AI) have further extended these capabilities (e.g., Deckker & Sumanasekara, 2025; Wang et al., 2024). Learning progressions and tools for sequencing learning also demonstrate potential for individualizing instruction for students (e.g., Alonzo, 2018).

Although educators recognize that academic skills are interrelated and that individualized instruction supports learner growth, implementation has traditionally been constrained by educators' time and capacity. Research-based learning maps formalize learning progressions for skill acquisition. Recent advances in AI and machine learning enable these maps to be paired with assessment data to evaluate the learning progressions, detect complex patterns in skill acquisition, and predict likely next skills. Building on the demonstrated benefits of adaptive systems, these advances create new opportunities to extend adaptive systems and further individualize learning.

2 Context

The Dynamic Learning Maps® (DLM®) system is built on research-based learning maps that depict student learning targets from early pre-academic skills through college and career readiness expectations (DLM Consortium, 2022). The DLM assessments measure clusters of highly related skills in these maps, called linkage levels. Linkage levels are available at different levels for academic content standards, called Essential Elements (EEs), with levels increasing in depth, breadth, and complexity leading to the grade-level academic learning target. In mathematics, there are five linkage levels for each EE. A subset of DLM states delivery assessments using an instructionally embedded model, whereby short testlets are delivered throughout the year as learning occurs. Teachers select which EEs and linkage levels they want to assess based on the student's learning goals and instruction provided to date.

The DLM assessments are scored with diagnostic classification models (DCMs; Bradshaw, 2016). DCMs rely on a Q-matrix to map assessment items to measured attributes (Tatsuoka, 1983). The reliance on theoretically specified attributes along with *a priori* rationale for mapping items to attributes allows for reporting results at a fine grain-size. More specifically, DCMs are a form of latent class analysis (e.g., Macready & Dayton, 1977) that estimate the conditional probability of a learner providing a correct response given their latent class and the base rate of attribute mastery. Results reported from a DCM-based assessment center around the estimated probability that a learner has mastered each of the assessed attributes given the Q-matrix and learner's item responses (Bradshaw, 2016). Scoring rules are then applied to dichotomize attribute mastery

such that each learner is classified as either a master or nonmaster of each attribute. For example, .50 is often used as a scoring rule (e.g., Templin & Bradshaw, 2013), where probabilities of attribute mastery of .50 or greater are interpreted as mastery and probabilities less than .50 are interpreted as nonmastery. The attribute mastery statuses are compiled into an attribute mastery profile, reflecting what learners know and can do.

3 Neural Networks

Neural networks are highly flexible machine learning models that are designed after biological neural networks (Alzubi et al., 2018; Ding et al., 2013; Dreiseitl & Ohno-Machado, 2002). Neural networks consist of input, hidden, and output layers of nodes (Abraham, 2005; Bi et al., 2019). The nodes in adjacent layers are connected via weighted connections (Kubat, 2017), and these weights are estimated as model parameters (Kotsiantis, 2007). An activation function at each node transforms the weighted input that is summed from each preceding node (Bi et al., 2019). Feedforward neural networks are common, as these use learning algorithms such as backpropagation to refine model parameters (Abraham, 2005; Kotsiantis, 2007; Neocleous & Schizas, 2002). The model parameters are refined over a pre-specified number of epochs (i.e., the number of times the data is passed through the neural network to further refine the parameter estimates; Brownlee, 2018).

The architecture of a neural network is specified based on the intended uses of the model. For example, a non-binary model output necessitates one node in the output layer for each possible classification, and the activation function should be consistent with non-binary classifications (Aggarwal, 2018). Decision-making around the number of hidden layers and the number of nodes per hidden layer is also critical (Jain et al., 1996). Because the weighted connections between nodes in adjacent layers are model parameters, adding hidden layers and nodes per hidden layer increases the number of model parameters, meaning unnecessary increases in these factors increases the likelihood of overfitting the model (Abraham, 2005; Ding et al., 2013; Kotsiantis et al., 2006). However, the hidden layers allow for modeling non-linearity, so performance suffers if an insufficient number of hidden layers is used to model nonlinear trends in the data (Dreiseitl & Ohno-Machado, 2002), meaning choosing the

number of hidden layers and the number of nodes per hidden layer is a delicate process with significant implications for model performance (e.g., Jain et al., 1996).

We developed and evaluated a neural network approach that integrates operational assessment data with learning map structures to recommend the next skill a student is most likely to master. The goal is to demonstrate how AI can leverage cognitive models and longitudinal assessment data to support personalized instruction.

4 Objectives

This project had three aims:

1. Evaluate assumptions supporting skill recommendation
2. Examine the feasibility of a neural network modeling approach
3. Assess the reasonableness of model recommendations

5 Methods

The pilot evaluation used data from 38,712 students participating in the DLM mathematics assessments in Grades 4, 7, and 10 between 2016 and 2025. Testlet-level mastery was estimated using operational scoring models. Each testlet was scored to produce linkage level mastery estimates. For students with multiple attempts at a given linkage level, analyses were limited to the most recent attempt.

5.1 Foundational Assumptions

Three assumptions were evaluated: 1) mastery of linkage levels is related, 2) there is a relationship between the highest linkage level mastered across EEs, and 3) ordering effects influence mastery likelihood. These assumptions were assessed using the correlations between linkage level mastery statuses, correlations between the highest linkage level mastered across EEs, and absolute differences in conditional mastery probabilities for both possible testing orders of each linkage level pair.

Pearson correlations were calculated to evaluate the first two assumptions, and the distributions of the correlation coefficients were analyzed to determine whether there was a nonnegligible relationship between the mastery indicators. The distributions were interpreted as the strength of the relationship between the mastery indicators (Schober et al., 2018).

Conditional mastery probability was defined as the probability of mastering the second assessed linkage level given mastery of the first. The absolute difference of the conditional probabilities reflects the degree to which mastery is order-dependent. For example, if the probability of mastering one EE at Initial Precursor after mastering another EE at Initial Precursor is 0.30, and the reverse ordering yields a probability of 0.50, the absolute difference is 0.20. Analyses of ordering effects were restricted to linkage level pairs with correlations greater than 0.50 to focus on linkage level pairs where there was a substantive relationship. We used the Wald test (Agresti, 2019) to determine whether the absolute difference of the conditional probabilities was statistically different from 0.

5.2 Feasibility

A feedforward neural network was estimated using operational assessment data and learning map information to predict the probability that a student would master each possible EE and linkage level combination within a grade (Kotsiantis, 2007; Kotsiantis et al., 2006). Predictors included data from previously tested EE and linkage level combinations. More specifically, the predictors included the dichotomous linkage level mastery statuses, percent correct, and base rate of mastery. Additional predictors captured structural relationships within the learning map, including correlations between the highest linkage level mastered across EEs, correlations between EE and linkage level combinations, cosine similarity ratings among linkage level descriptors, and conditional mastery rates (i.e., the proportion of students mastering an EE and linkage level who mastered the previously tested EE and linkage level).

The neural network architecture consisted of a single, fully-connected hidden layer with 100 nodes, selected to provide model flexibility while limited risk of overfitting (Heaton, 2005). Models were trained for 100 epochs, allowing repeated passes through the training data to update model weights (Brownlee, 2018). The training duration was selected to ensure adequate learning given the large sample sizes available at each grade. The connection weights in the model were randomly drawn so that they span the range of plausible values, aiming to increase the likelihood of model convergence. Each node in the neural network uses a sigmoid activation function, which maps the mean of the summation of linear predictors to

the mean of a binary response variable (DeMaris, 1995; Peng et al., 2002). The sigmoid function is ideal, since linkage level mastery is reported as a binary outcome (i.e., mastery or nonmastery of the linkage level). The model was trained using the backpropagation algorithm, where a loss function is applied to identify sources of misclassifications and the connection weights are refined to minimize loss (Bi et al., 2019). We extracted the raw outputs from the neural network, which provide the log-odds of the outcome associated with each node in the output layer, and we used the log-odds estimates to calculate the probability of mastering each linkage level given the model inputs.

Feasibility was assessed with model convergence and recommendation certainty. Model convergence was measured with a dichotomous indicator reflecting whether model parameter estimates stabilize, indicating the observed epoch-to-epoch changes in the parameter estimates are less than the pre-specified convergence threshold. Recommendation certainty was quantified using the highest estimated probability of mastery for a linkage level for each learner. Model convergence was considered a prerequisite for feasibility, as nonconverging models may not produce reliable predictions. High-certainty recommendations indicate meaningful differentiation between potential next skills, supporting instructional usefulness. Further, high-certainty recommendations are critical, as low-certainty recommendations approximate random selection which may provide limited instructional value for educators.

5.3 Reasonableness

Two subject matter experts who were not part of the research team reviewed a subset of learning profiles, which indicated the skills that were mastered as well as skills that were assessed but not mastered, model-based recommendations for next skills to prioritize, and AI-generated rationales for the recommendations informed by the learning map structure and conceptual relationships among skills. The experts evaluated whether the recommendations were instructionally appropriate and provided an explanation of their evaluation to support future refinements of the model.

6 Results

Results are shared for each objective.

6.1 Foundational Assumptions

Figure 1 displays the distribution of correlations between linkage level mastery statuses. Across grades, the distributions were centered near 0.20, with a substantial number of linkage level pairs exhibiting higher correlations. These patterns support the assumption that mastery relationships exist among linkage levels.

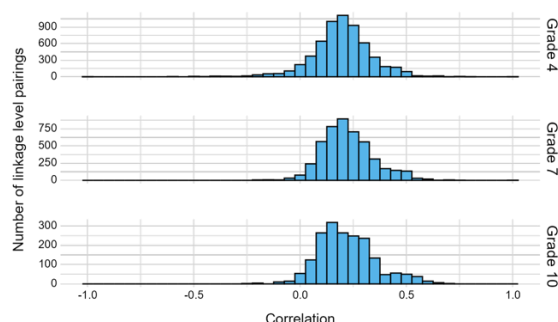


Figure 2: Distribution of the Correlations Between Linkage Level Mastery Statuses by Grade

Figure 2 presents the distribution of correlations between the highest linkage level mastered across EEs. The distributions for Grades 4 and 7 were similar, with centers near 0.40, while Grade 10 exhibited a distribution with a slightly lower center near 0.30. Despite this difference, all grades demonstrated meaningful relationships between the highest mastered linkage levels across EEs. These patterns support the assumption that there is a relationship between the highest linkage level mastered across EEs

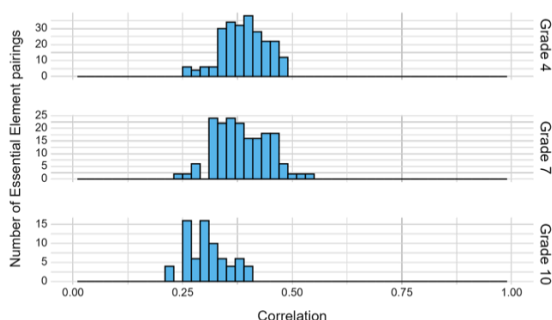


Figure 1: Distribution of the Correlations Between the Highest Linkage Level Mastered Across Essential Elements by Grade

A total of 124 linkage level pairs exhibited correlations greater than 0.50. Among these linkage level pairs with a substantive relationship, absolute differences in conditional mastery probabilities ranged from 0 to 1.00, with a median of 0.17 and a mean of 0.20. Of these 124 pairs, 34 (27%) demonstrated statistically significant

absolute differences in conditional probabilities. Most statistically significant pairs ($n = 29$; 85%) were supported by sample sizes of at least 1,000 students. Collectively, these findings provide empirical support for an ordering effect among linkage levels, wherein mastery of one linkage level differentially increases the likelihood of mastering another.

6.2 Feasibility

All models converged successfully, indicating confidence in interpreting model outputs. The models produced high-certainty recommendations for almost all learners (Table 1). Across grades, the vast majority of learners received a recommendation with a predicted mastery probability of at least 0.80: 97% in Grade 4, 98% in Grade 7, and 98% in Grade 10. The mean probability of each learner's most certain recommendation was 0.97, 0.98, and 0.98 for Grades 4, 7, and 10, respectively. Taken together, these results indicate that using a neural network to recommend next likely mastery targets based on operational and learning map data is technically feasible.

Grade	< .50	.50 $\geq x <$.80	$\geq .80$
4	.03	<.01	.97
7	.02	<.01	.98
10	.02	<.01	.98

Table 1: Proportion of Learners' Most Certain Predictions by Grade

6.3 Reasonableness

The subject matter experts consistently reported the recommendations were sensible, instructionally grounded, and aligned with observed learners' mastery profiles. Recommendations targeting one linkage level above demonstrated mastery were viewed as mostly accurate and instructionally appropriate. Recommendations for untested EEs were considered appropriate when there was clear similarity to skills the student had already mastered. Reviewers emphasized that recommendations for retesting previously assessed but unmastered linkage levels were appropriate when the underlying skill was considered instructionally critical for future learning. Recommendations targeting one linkage level down from a previously assessed but

unmastered linkage level were viewed as context dependent. The experts indicated that such recommendations were appropriate when the lower linkage level aligned with patterns observed across other mastered skills or signaled a potential knowledge gap requiring reinforcement.

Overall support for the recommendations was high, suggesting broad acceptance of the model's functionality. There was feedback suggesting that the model recommendations could be improved by increasing the coherence of recommendations so that the recommendations were a conceptually related set of skills. More specifically, the experts noted potential value in framing recommendations around the integration of key concepts or relationships across linkage levels to better support student learning. One comment explicitly suggested prompting the AI agent to generate recommendations that highlight conceptual connections among skills, rather than presenting recommendations as isolated targets.

In summary, subject matter expert feedback strongly supported the overall logic, usefulness, and instructional relevance of NEXTMAP recommendations. Suggested refinements primarily focused on improving coherence, consistency, and conceptual framing of the recommendations rather than on fundamental concerns about recommendation accuracy or validity.

7 Discussion

Using data and existing learning maps, we were able to identify patterns in student mastery and use those patterns to recommend next instructional steps. Quantitative and qualitative data demonstrated instructional recommendations were reasonable next learning targets for students based on their mastery patterns to date. Findings from our study show promise for combining rich, theoretical cognitive models with neural networks to provide more personalized learning and assessment opportunities. As learning progressions and learning maps continue to gain traction, this study demonstrates one way their use can be extended to personalize instruction without increasing burden on educators to reflect on learning to date and select next targets from those mastery patterns. Particularly in elementary grades and for special educators, who may not receive extensive subject matter training during their teacher preparation programs, approaches like ours can support conceptually grounded instructional approaches that support identifying

and making connections among academic content.

AI-enabled models can help bridge the gap between theoretical cognitive models and real-time instructional decision making, offering a scalable path toward more coherent, personalized, and cognitively grounded support for diverse learners across educational systems.

References

- Abraham, A. (2005). Artificial neural networks. In P. H. Sydenham & R. Thorn (Eds.), *Handbook of Measuring System Design* (pp. 901-908). John Wiley & Sons. <https://doi.org/10.1002/0471497398.mm421>
- Aggarwal, C. C. (2018). *Neural networks and deep learning*. Springer. <https://doi.org/10.1007/978-3-319-94463-0>
- Agresti, A. (2019). *An introduction to categorical data analysis* (3rd ed.). John Wiley & Sons.
- Alonzo, A. C. (2018). An argument for formative assessment with science learning progressions. *Applied Measurement in Education*, 31(2), 104–112. <https://doi.org/10.1080/08957347.2017.1408630>
- Alzubi, J., Nayyar, A., & Kumar, A. (2018). Machine learning from theory to algorithms: An overview. *Journal of Physics: Conference Series*, 1142(1), 012012. <https://doi.org/10.1088/1742-6596/1142/1/012012>
- Bi, Q., Goodman, K. E., Kaminsky, J., & Lessler, J. (2019). What is machine learning? A primer for the epidemiologist. *American Journal of Epidemiology*, 188(12), 2222–2239. <https://doi.org/10.1093/aje/kwz189>
- Bradshaw, L. (2016). Diagnostic classification models. In A. A. Rupp & J. Leighton (Eds.), *The handbook of cognition and assessment: Frameworks, methodologies, and applications* (1st ed., pp. 297–327). John Wiley & Sons. <https://doi.org/10.1002/9781118956588.ch13>
- Brownlee, J. (2018). What is the Difference Between a Batch and an Epoch in a Neural Network. *Machine learning mastery*, 20(1), 1-15.
- Deckker, D., & Sumanasekara, S. (2025). ChatGPT and the evolution of AI-powered tutoring systems. *EPRA International Journal of Environmental Economics, Commerce and Educational Management*, 71–87. <https://doi.org/10.36713/cpra21332>

- DeMaris, A. (1995). A tutorial in logistic regression. *Journal of Marriage and the Family*, 57(4), 956–968. <https://doi.org/10.2307/353415>
- Ding, S., Li, H., Su, C., Yu, J., & Jin, F. (2013). Evolutionary artificial neural networks: A review. *Artificial Intelligence Review*, 39, 251–260. <https://doi.org/10.1007/s10462-011-9270-6>
- Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: A methodology review. *Journal of Biomedical Informatics*, 35(5-6), 352–359. [https://doi.org/10.1016/S1532-0464\(03\)00034-0](https://doi.org/10.1016/S1532-0464(03)00034-0)
- Dynamic Learning Maps Consortium. (2022). 2021–2022 technical manual—Instructionally embedded model. University of Kansas, Accessible Teaching, Learning.
- Heaton, J. (2005). *Introduction to neural networks with Java* (2nd ed.). Heaton Research, Inc.
- Jain, A. K., Mao, J., & Mohiuddin, K. M. (1996). Artificial neural networks: A tutorial. *Computer*, 29(3), 31–44. <https://doi.org/10.1109/2.485891>
- Kotsiantis, S. B. (2007). Supervised machine learning: A review of classification techniques. In I. G. Maglogiannis (Ed.), *Emerging artificial intelligence applications in computer engineering: Real word AI systems with applications in ehealth, HCI, information retrieval and pervasive technologies* (pp. 3-24). IOS Press.
- Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: A review of classification and combining techniques. *Artificial Intelligence Review*, 26, 159–190. <https://doi.org/10.1007/s10462-007-9052-3>
- Kubat, M. (2017). *An introduction to machine learning* (2nd ed.). Springer International Publishing AG. <https://doi.org/10.1007/978-3-319-63913-0>
- Li, X., Xu, H., Zhang, J., & Chang, H. (2023). Deep Reinforcement Learning for Adaptive Learning Systems. *Journal of Educational and Behavioral Statistics*, 48(2), 220–243. <https://doi.org/10.3102/10769986221129847>
- Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, 2(2), 99–120. <https://doi.org/10.3102/10769986002002099>
- Neocleous, C., & Schizas, C. (2002) Artificial Neural Network Learning: A Comparative Review. In: I. P. Vlahavas & Spyropoulos C.D. (eds) *Methods and Applications of Artificial Intelligence* (pp. 300-313). Springer, Berlin, Heidelberg. https://doi.org/10.1007/3-540-46014-4_27
- Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1), 3–14. <https://doi.org/10.1080/00220670209598786>
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & analgesia*, 126(5), 1763-1768.
- Tatsuoka, K. K. (1983). Rule space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20(4), 345–354. <https://www.jstor.org/stable/1434951>
- Templin, J., & Bradshaw, L. (2013). Measuring the reliability of diagnostic classification model examinee estimates. *Journal of Classification*, 30(2), 251–275. <https://doi.org/10.1007/s00357-013-9129-4>
- Wang, X., Huang, R. “Tammy,” Sommer, M., Pei, B., Shidfar, P., Rehman, M. S., Ritzhaupt, A. D., & Martin, F. (2024). The Efficacy of Artificial Intelligence-Enabled Adaptive Learning Systems From 2010 to 2022 on Learner Outcomes: A Meta-Analysis. *Journal of Educational Computing Research*, 07356331241240459. <https://doi.org/10.1177/07356331241240459>
- Zhang, Y., Zhang, L., Zhang, H., & Wu, X. (2025). Development of a Computerized Adaptive Assessment and Learning System for Mathematical Ability Based on Cognitive Diagnosis. *Journal of Intelligence*, 13(9), 114. <https://doi.org/10.3390/jintelligence13090114>