Evaluating learning progression hierarchies with diagnostic models

NCME 2025

Auburn Jimenez, W. Jake Thompson, and Brooke Nash

University of Kansas, ATLAS

PIE

Pathways for Instructionally Embedded Assessment

KU Achievement & Assessment Institute

The University of Kansas

Accessible Teaching, Learning, & Assessment Systems

Learning progression-based assessments

- In recent years, there has been a demand for assessments that provide more actionable and fine-grained information about students' knowledge, skills and understandings that can be used to inform teaching and learning activities.

- Learning progressions are one type of cognitive learning model that can be used to develop assessment systems, and these models describe one or more paths of skill acquisition within a domain (Simon, 1995; Alonzo & Steedle, 2009).

- Learning progressions-based assessments can benefit teachers and students by:
    1. Allowing teachers to see where their students are within the progression, and
    2. Offering recommendations for precursor skills to focus on if learning targets have not yet been mastered and successor skills when learning targets have been mastered.

Research Purpose

- In the present study, we will demonstrate an empirical method to learning progression validation using diagnostic classification models (DCMs; also called cognitive diagnosis models).

- The method is applied in the context of the Pathways for Instructionally Embedded (PIE) assessment system, a Competitive Grants for State Assessments funded project aimed at developing a proof-of-concept innovative assessment model based on learning progressions (known as learning pathways).

PIE Learning Pathway View

PIE.5.NF.A.3

Mathematics

Number Sense and Operations in Fractions (NF)

Grade 5

This document provides (a) the target grade-level content standard; (b) three levels of a learning pathway aligned with the learning target; (c) the knowledge, skills, and understandings associated with each level; and (d) a map view of the full learning pathway.

Learning Target

5.NF.A. Understand the relationship between fractions and decimals (denominators that are factors of 100).

3. Compare and order fractions and/or decimals to the thousandths place using the symbols >, = or <, and justify the solution.

Learning Pathway in Three Levels

The learning pathway presents three vertical levels that consist of knowledge, skills, and understandings that build toward and meet the learning target. **Level 1** represents emerging concepts and skills related to the learning target. **Level 2** represents concepts and skills approaching the learning target. **Level 3** represents the learning target and aligns with the grade-level content standard.

Level 1
Compare unit fractions.

↓

Level 2
Compare fractions with like denominators and like decimals.

↓

Level 3
Compare and order fractions and/or decimals up to the thousandths place using symbols (<, >, =) and justify the solution.

Item Field Test

- Held from March 4 – March 22, 2024 in 5th grade math classrooms in Missouri.
- A total of 1,708 5th grade students participated in the item field test.
- Students were tested on all 3 pathway levels of 2 learning pathways as well as the third pathway level of a selected third learning pathway.
- Sample size for each standard ranged from 76 to 203 students.

Study Context and Research Question

- We are interested in answering the following research question:

Is the hierarchical structure proposed for the PIE learning pathways empirically supported by evidence obtained using a diagnostic modeling framework?

Diagnostic classification models

- DCMs describe a family of multidimensional psychometric models that define mastery on multiple latent variables or skills of interest, presented in terms of a mastery profile (Rupp et al., 2010).

- DCMs provide many benefits when used for calibration and scoring purposes, including:
  1. Supporting instruction by reporting fine-grained results at the skill level.
  2. Testing different learning progression structures.

- Different parameterizations can be defined to reflect different assumptions about how the items and skills are related.

LCDM and HDCM parameterizations

- In the present study, we consider two parameterizations:
  1. Log-Linear cognitive diagnosis model (LCDM; Henson, Templin, & Willse, 2009)
  2. Hierarchical diagnostic classification model (HDCM; Templin & Bradshaw, 2014)

- The LCDM is a general DCM that contains all possible mastery profiles.

- The HDCM is a constrained DCM that restricts the skill space to only contain mastery profiles that are compatible with a proposed skill hierarchy.

Methods: Empirical evaluation of learning progressions

- Thompson and Nash (2022) outline a DCM-based diagnostic framework for empirically evaluating learning progressions by estimating and comparing a full model and a reduced model.
  - Full model: an unconstrained LCDM
  - Reduced model: an HDCM that depicts the proposed hierarchical structure of the learning progression

- Because the HDCM is nested within an LCDM, we can directly compare whether the exclusion of unallowed mastery profiles in the HDCM substantially impacts the model fit.

Methods: Model fit

- Model fit can be evaluated in terms of absolute fit and relative fit.

- To assess absolute fit, we computed posterior predictive model checks at the model-level for the raw score distribution, and we report posterior predictive p-values (*ppp*-values; e.g. Thompson, 2024).

- To assess relative fit, we used to information criterion:
    1. Leave-one-out cross validation (LOO; Vehtari, Gelman, & Gabry, 2017)
    2. Widely applicable information criterion (WAIC; Watanabe, 2010)

Results: Absolute fit

- Adjusted *ppp*-values were computed using the Holm correction (Holm, 1979) to control for family-wise error rates associated with testing of multiple models across the 25 learning pathways.

- Adequate model fit was demonstrated for 21 (84%) of the 25 learning pathways when fit using an LCDM.

- Adequate model fit was demonstrated for 20 (80%) of the 25 learning pathways when fit using an HDCM.

Results: Relative Fit

- Models with poor fit were not included in the model comparison approach.

- We concluded a significant difference in fit if the absolute difference between competing models was greater than 2.5 times the standard error of the difference (Bengio & Grandvalet, 2004).

- The HDCM model was the overwhelming favorite and was preferred in 19 of the 21 evaluated learning pathways.

Takeaways and Next Steps

- More complex skill acquisition pathway designs
  - We assumes a strict hierarchy of pathway level mastery in the PIE assessment system.
  - Future applications may consider more complex skill acquisition pathways (i.e. alternate pathways)

- Adaptive testing environments
  - Customized assessment for each student based on their current needs while simultaneously minimizing testing time.

THANK YOU!

Email questions to Auburn Jimenez at auburn.jimenez@ku.edu

PIE

Pathways for Instructionally Embedded Assessment

References

Bengio, Y., & Grandvalet, Y. (2004). No unbiased estimator of the variance of k-fold cross-validation. Journal of Machine Learning, 5, 1089–1105. Retrieved from http://www.jmlr.org/papers/v5/grandvalet04a.html

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*. doi:10.18637/jss.v076.i01

Dynamic Learning Maps Consortium. (2016). 2014–2015 Technical manual–Integrated model. University of Kansas, Center for Educational Testing and Evaluation. https://dynamiclearningmaps.org/sites/default/files/documents/publication/Technical_ Manual_IM_2014-15.pdf

Dynamic Learning Maps Consortium. (2022, December). *2021–2022 Technical Manual—Year-End Model*. University of Kansas, Accessible Teaching, Learning, and Assessment Systems.

Henson, R. A., Templin, J. L., and Willse, J. T. (2009). Defining a family of cognitive diagnosis models using log-linear models with latent variables. *Psychometrika 74*(2), 191–210. https://doi.org/10.1007/s11336-008-9089-5

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70. http://www.jstor.org/stable/4615733

Missouri Department of Elementary and Secondary Education. (2021). Priority standards for leveraging learning in mathematics: Grades K–12. https://dese.mo.gov/college-careerreadiness/curriculum/academic-standards/priority-standards

Rupp, A. A., & Templin, J. (2008). Unique characteristics of cognitive diagnosis models: A comprehensive review of the current state-of-the-art. *Measurement: Interdisciplinary Research and Perspectives*, *6*, 219–262.

Simon, M. (1995). Reconstructing mathematics pedagogy from a constructivist perspective. *Journal for Research in Mathematics Education, 26*(2), 114–145. https://doi.org/10.2307/749205

Stan Development Team (2024). RStan: the R interface to Stan. R package version 2.32.6. https://mc-stan.org/.

Swinburne Romine, R., Andersen, L., Schuster, J., & Karvonen, M. (2018). Developing and evaluating learning map models in science: Evidence from the I-SMART project. Accessible Teaching, Learning, and Assessment Systems (ATLAS), the University of Kansas. https://ismart.works/sites/default/files/documents/Publications/ISMART_Goal_1_Technical_Report_FINAL_0.pdf

Swinburne Romine, R., Schuster, J., Karvonen, M., Thompson, W. J., Erickson, K., Simmering, V., & Bechard, S. (2025). Learning Maps as Cognitive Models for Instruction and Assessment. *Education Sciences*, *15*(3), 365. https://doi.org/10.3390/educsci15030365

Templin, J., and Bradshaw, L. (2014). Hierarchical diagnostic classification models: A family of models for estimating and testing attribute hierarchies. *Psychometrika 79*(2), 317–339. https://doi.org/10.1007/s11336-013-9362-0

Thompson, W. J. & Nash, B. (2022). A diagnostic framework for the empirical evaluation of learning maps. *Frontiers in Education, 6*, Article 714736. https://doi.org/10.3389/feduc.2021.714736

Thompson, W. J. (2024). An Evaluation of Methods for Assessing Model Fit for Bayesian Diagnostic Classification Models. https://doi.org/10.35542/osf.io/ytjq9

Vehtari, A., Gelman, A., and Gabry, J. (2017). Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Statistics and Computing, 27*(5), 1413–1432. https://doi.org/10.1007/s11222-016-9696-4

Watanabe, S. (2010). Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *Journal of Machine Learning Research, 11*(116), 3571–3594. http://jmlr.org/papers/v11/watanabe10a.html