

Abstract submitted to, accepted by National Council on Measurements in Education (NCME) conference “Classroom assessment as a learning Experience”, September 2019. Paper subsequently withdrawn.

Classroom Learning as Assessment Experience: Continuous Measurement and Diagnosis in Mathematics Curricula.

IBM-Watson Education piloted an artificial-intelligence (AI)-assisted continuous cognitive-diagnostic capability at ten U.S. K-12 systems nationwide during academic years 2017-2019. Demonstrating the feasibility of competency-based personalized learning at-scale was the objective. These pilots employed **criterion-referenced assessment results — including those from classroom-learning activities** — for diagnostic purposes.

Maximizing each student’s opportunity to learn [Elliott2016] requires differentiation — sometimes personalization — of instruction according to individual needs. Optimally achieving this in turn requires a detailed, continuously-evolving picture of each student’s state of proficiency. Instruction is adapted based on individual strengths and weaknesses, among other considerations.

The span and dimensionality of an individual learner’s proficiency-measurand space is considerable [Bhola2015]. Exhaustive measurement at a level of confidence commensurate with state-accountability exams is expensive [Wainer2015], and can only be accomplished infrequently. Such assessments moreover inevitably under-sample students’ proficiency space.

Measurements with confidence levels that are lower than those from accountability exams can also be considered. Trading off individual-measurement confidence can lead to greater frequency and coverage density. As evidentiary measurements accrue, a more-complete, higher-confidence overall proficiency picture emerges. Accountability-exam-level confidence measurements can be combined with others of lower confidence to assemble a comprehensive, continuously-evolving picture.

Fusing together three well-known educational frameworks leads to just such a continuous-diagnostic methodology. Figure 1 illustrates. Curriculum Alignment (CA) provides the overarching integrative framework for Curriculum Instruction and Assessments (CIA) [Porter2001]. It contains important parallels to Evidence-Centered Design (ECD), Figure 2, which enjoys near-paradigmatic status as a framework for Diagnostic Cognitive Modeling (DCM) [Mislevy2003, Rupp2010, Haertel2016]. **CIA thereby becomes a system for continuous diagnostic measurement.**

The CA *Intended Curriculum* (ECD Proficiency Model) specifies proficiency measurands. These prescribe knowledge the learner obtains from the curriculum. Educational learning standards typically provide proficiency specifications within U.S. K-12 systems.

Often a policy statement, the *Intended Curriculum* also describes the high-level curricular organization. This typically appears in the form of academic courses and their constituent instructional units. The structural view can be thought of as analogous to an ECD Assembly Model.

Understanding by Design (UbD) frames the CA *Enacted Curriculum*, analogous to the ECD Task Model. UbD considers "...evidence gathered through a variety of formal and informal assessments during a unit of study or a course ... «including» traditional quizzes and tests, performance tasks and projects, observations and dialogues, as well as students' self-assessments gathered over time" [Wiggins2005]. **Recognizing that UbD casts classroom learning in part as evidentiary accrual represents the key point.** Marzano further describes sharpening of instructional targeting to improve evidentiary confidence from learning activities [Marzano2008].

The CA *Assessed Curriculum* — analogous to the ECD Evidence Model — describes the aggregation of the entire body of evidence attesting to each learner's overall state of proficiency. The ECD Evidence Model also specifies how to translate this body of evidence into the *Learned Curriculum*.

The *Learned Curriculum* represents a view of a student's actual state of proficiency, as well as associated degrees of confidence thereabout. It contains estimates of all measurands — proficiency factors — within the curricular span of interest. DCMs infer this from the body of evidence in-aggregate.

Proficiency-state estimates from a suitably-designed DCM are not limited to directly-measured learning standards. Learning-standard progressions describe how mastering one standard depends upon proficiency in prerequisites [Stevens2002]. Zimba built a directed-graph view of the progression of Common-Core State Standards (CCSS) for Math [Zimba2012]. This allows for a DCM based on a Bayesian Network [Rupp2010, Almond2015].

Aggregating a mixed-confidence body of evidence to a DCM requires particular attention. The Theory of Evidence (ToE) gives a formal probabilistic framework for this [Shafer1976, Salicone2007]. Almond described a graphical-modeling approach based on ToE [Almond1995]. IBM's continuous-diagnosis pilot employs a Bayesian-Network adaptation that functionally approximates Almond's approach.

Interpretation of DCM proficiency-state estimates for distinct learning standards varies depending on their graphical location with respect to directly-measured standards. Location in terms of directed-graph topological order is key. Estimates for not-assessed

learning standards topologically preceding directly-measured ones are retrospectively diagnostic. They indicate likely strengths or weaknesses. Estimates for which all measurements are topological predecessors in contrast indicate readiness to engage the associated content.

Retrospective diagnosis allows teachers to triangulate areas of weakness for individual learners. Localized weaknesses are addressed through incorporating scaffolding into instruction, or by providing supplementary remedial instruction. The forward-looking estimates represent a “readiness barometer” regarding what comes next.

Given its use of mixed-confidence evidence, IBM’s continuous-diagnosis framework may not achieve accountability-exam confidence levels potentially provided by a Through-Course Summative Assessment (TCSA) regime [Mislevy2002]. Its confidence levels do however support its intended purpose of instructional tailoring.

References

- [Almond1995] R. A. Almond, *Graphical Belief Modeling*, London, UK: Chapman-Hall, 1995.
- [Almond2015] R. A. Almond, et al, *Bayesian Networks in Educational Assessments*, New York: Springer, 2015, <https://www.springer.com/us/book/9781493921249>.
- [Bhola2003] D. Bhola, et al, “Aligning tests with states' content standards: Methods and issues,” *Education Measurement: Issues and Practices*, National Council on Measurements in Education, Fall 2003, <http://ibm.biz/NCME-ExamAlgmt>.
- [Elliott2016] S. N. Elliott and B. J. Bartlett, “Opportunity to Learn”, *Oxford Handbooks Online*, May 2016, <http://ibm.biz/Oxford-Hbk-OTL>.
- [Haertel2016] G. Haertel, et al, “General introduction to evidence-centered design”, *Meeting the Challenges to Measurement in an Era of Accountability*, New York: Routledge, for the National Council on Measurement in Education, 2016, <https://amzn.to/2OhMMx7>.
- [Hack2009] Hack, S., *Evidence and Inquiry*, expanded edition, New York: Prometheus, 2009.
- [Marzano2008] R. J. Marzano and M. W. Haystead, *Making Standards Useful in the Education*, Alexandria, VA: ASCD, 2008, http://ibm.biz/ASCD_Marzano_Stds_Classroom.
- [Mislevy2003] R. J. Mislevy, et al, “A brief introduction to evidence-centered design,” Research Report RR-03-16, Educational Testing Service (ETS), July 2003, http://ibm.biz/ETS_ECD_RschRep.
- [Mislevy2002] R. J. Mislevy and R. Zwick, “Scaling, linking, and reporting in a periodic assessment,” *Journal of Education Measurement*, National Council on Measurements in Education, Summer 2002, <http://ibm.biz/NCME-TCSA>.
- [Kahneman2011] Kahneman, D., *Thinking Fast and Slow*, New York: Macmillan, 2011, <https://amzn.to/2tSGHy8>.

- [Porter2001] A. C. Porter and J. L. Smithson, "Defining, developing, and using curriculum indicators," Report RR-048, Consortium for Policy Research in Education, University of Pennsylvania, December 2001, <http://ibm.biz/PorterCurriculumAlignment>.
- [Porter2002] A. C. Porter, "Measuring the content of instruction: Uses in research and practice," *Educational Researcher*, October 2002 31(7)3-14, http://ibm.biz/EdRschr_Meas_Cont_Instruction.
- [Rupp2010] A. Rupp, et al, *Diagnostic Measurement: Theory, Methods, and Applications*, New York: Guilford, 2010, <https://amzn.to/2AnMHVv>.
- [Salicone2007] S. Salicone, *Measurement Uncertainty: An Approach via the Mathematical Theory of Evidence*, New York: Springer Science+Business Media, LLC, 2007.
- [Shafer1976] G. Shafer, *A Mathematical Theory of Evidence*, Princeton, NJ: Princeton University Press, 1976.
- [Stevens2002] Stevens, S., et al, (2002), "Using learning progressions to inform curriculum, instruction and assessment design," Center for Highly Interactive Educations, Curricula & Computing in Education, School of Education/College of Engineering, University of Michigan, Retrieved on July 24, 2018 from <http://ibm.biz/CHICCCE-Progressions>.
- [Tomlinson2014] C. A. Tomlinson, *The Differentiated Classroom*, second edition, Alexandria, VA: ASCD, 2014, http://ibm.biz/ASCD_Tomlinson_DiffCR.
- [Wainer2015] H. Wainer and R. Feinberg, "For want of a nail: Why unnecessarily long tests may be impeding the progress of Western civilization," *Significance*, The Royal Statistical Society, February 2015, <http://ibm.biz/RSS-Significance-TestEfficiency>.
- [Wiggins2005] G. Wiggins, J. McTighe, *Understanding by Design*, second edition, Alexandria, VA: ASCD, 2005, http://ibm.biz/ASCD_UBD_Book.
- [Zimba2012] J. Zimba, "A graph of the content standards", Student Achievement Partners, June 7, 2012, <http://ibm.biz/Zimba-Wiring-Diagram>.

Figure 1 — IBM's continuous-diagnostic framework employs Curriculum Alignment [Porter2001] as its overarching integrative structure.

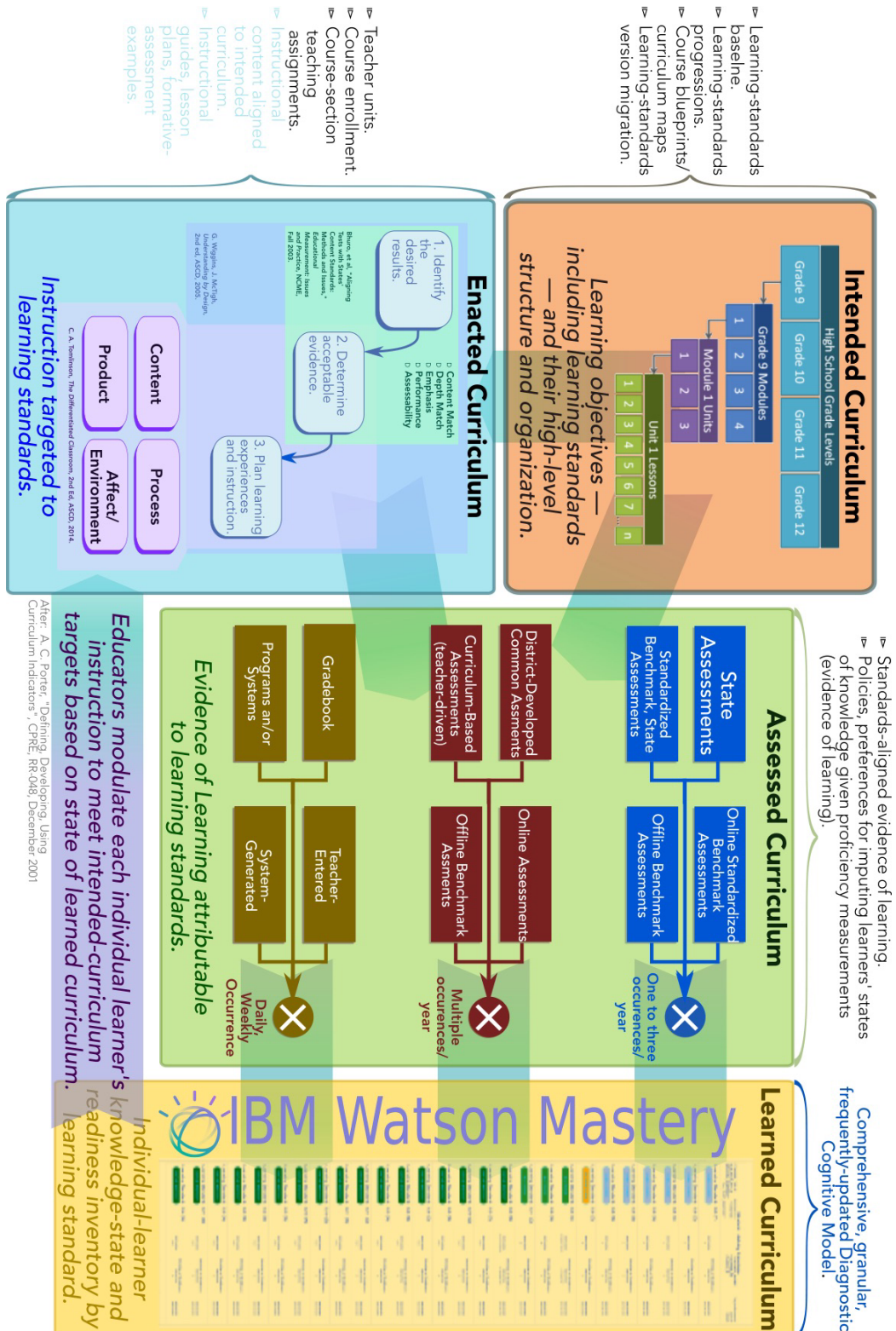
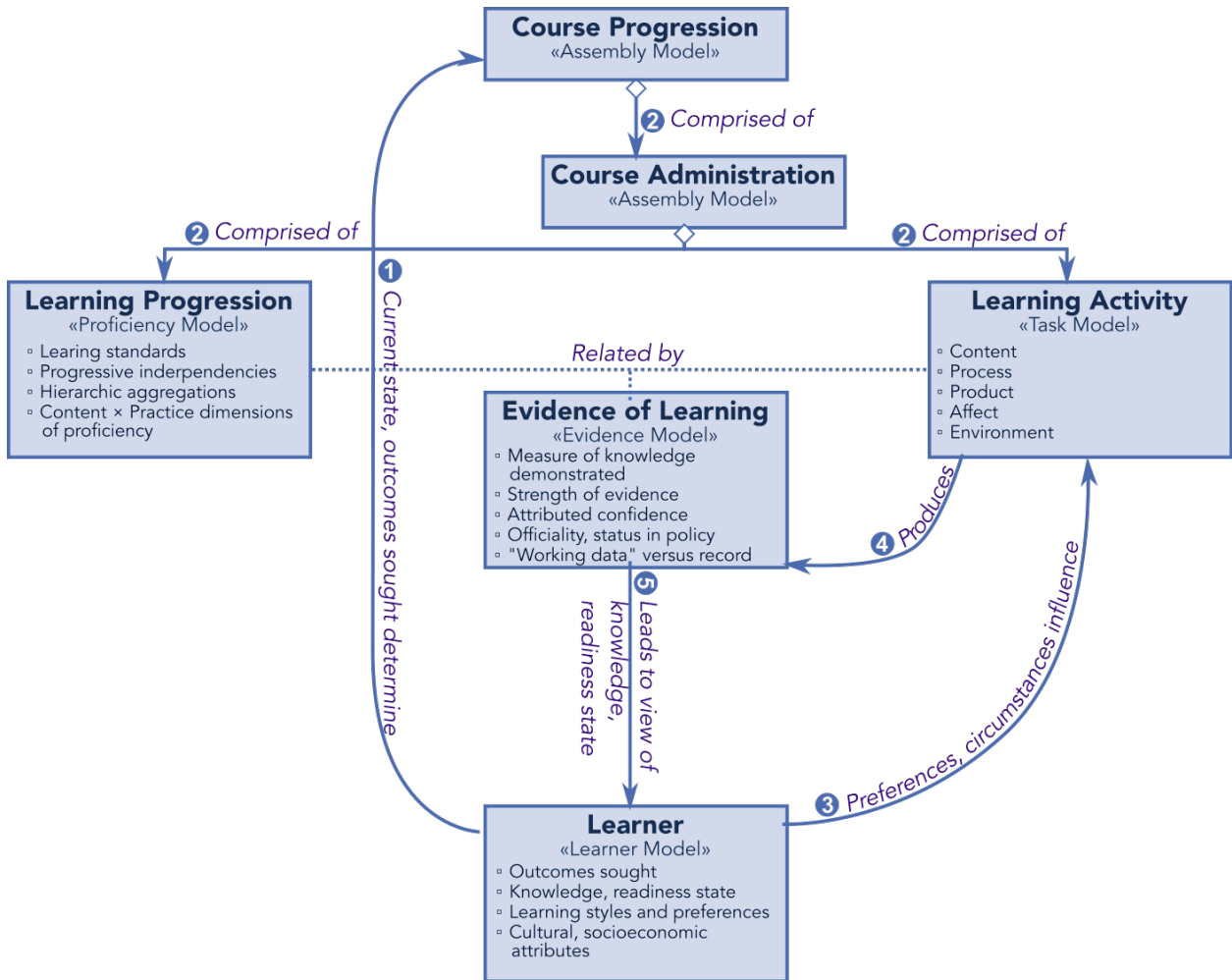


Figure 2 — IBM's continuous-diagnostic framework extends Evidence-Centered Design [Mislevy2003, Rupp2010, Haertel2016] into a framework for adaptive personalization of instruction according to individual-student states of proficiency.



10:20 AM Fri Mar 8

Georgetown University M x International Journal of A x What questions do you ha x NCME Special Conferenc x

colorado.edu

97%

University of Colorado Boulder

Center for Assessment, Design, Research and Evaluation (CADRE)

SCHOOL OF EDUCATION

[Home](#) [About CADRE](#) [Projects](#) [Resources](#) [Working with CADRE](#) [NCME Classroom Assessment Conference 2019](#)

[Home](#) > [NCME Special Conference on Classroom Assessment Proposal Submission Form](#) > NCME Special Conference on Classroom Assessment Proposal Submission Form

NCME Special Conference on Classroom Assessment Proposal Submission Form

Thank you for your submission! Decisions on proposals will be sent via email by no later than May 30, 2019.

[Go back to the form](#)

CADRE
University of Colorado Boulder
School of Education
249 UCB
Boulder, CO 80502-0249

Be Boulder.

University of Colorado Boulder
© Regents of the University of Colorado