Developing and Refining a Model for Measuring Implementation Fidelity for an Instructionally Embedded Assessment System

Jennifer L. Kobrin, Meagan Karvonen, Amy K. Clark, W. Jake Thompson, and Jennifer Burnes

Accessible Teaching, Learning, and Assessment Systems (ATLAS), The University of Kansas

Presented at:

Abstract

We developed and refined a model for measuring implementation fidelity for an instructionally embedded assessment system for students with significant cognitive disabilities. A logic model describing intended use of the assessments was used to identify indicators based on data currently available in our assessment system. We established criteria for sufficient implementation based what we consider minimum requirements for intended use of the system, and practices that we believe support strong implementation. We conducted exploratory analyses using the indicators to iteratively refine the logic model and criteria. This research demonstrates one way to evaluate components of an assessment's theory of action by contributing evidence on the extent to which it is implemented as intended to provide all students opportunity to learn.

Background

By design, instructionally embedded assessments do not merely serve as an indicator of student achievement; they are intendedto lead directly to action on the part of the teacher and student. In cases where assessment systems are intended to serve as agentsfor action, it is incumbent upon the test developer to develop a theory of action documenting what needs to be in place for the desired effects to occur, as well as the ways in which improper implementation may lead to unintended negative consequences (NCME, 2018).

The concept of implementation fidelity, common in evaluation research, can be used to guide the evaluation of action mechanisms in an assessment's theory of action. Century et al. (2010) developed a conceptual framework for implementation fidelity that includes structural and instructional components. The structural components represent what a teacher needs to do (procedural) and know (educative) to administer a program or intervention with fidelity, and the instructional components represent the actions, behaviors, and interactions teachers (pedagogical) and students are expected to engage in to implement aprogram or intervention with fidelity.

While measuring implementation fidelity is common in educational and health evaluation, it is not prevalent in educational assessment. This study illustrates a process for defining and evaluating a model of implementation fidelity appropriate for use inan instructionally embedded assessment system. Once it is fully developed and validated, we intend to use the model to evaluateclaims in the assessment's theory of action as well as identify areas where teachers may need additional support to use the instructionally embedded assessments with maximum fidelity.

Context: Dynamic Learning Maps Instructionally Embedded Assessment System

The purpose of the Dynamic Learning Maps® (DLM®) alternate assessment is to measure alternate academic achievement standardsin English language arts (ELA) and mathematics for students with significant cognitive disabilities who cannot meaningfully access general education assessments, even with accommodations.

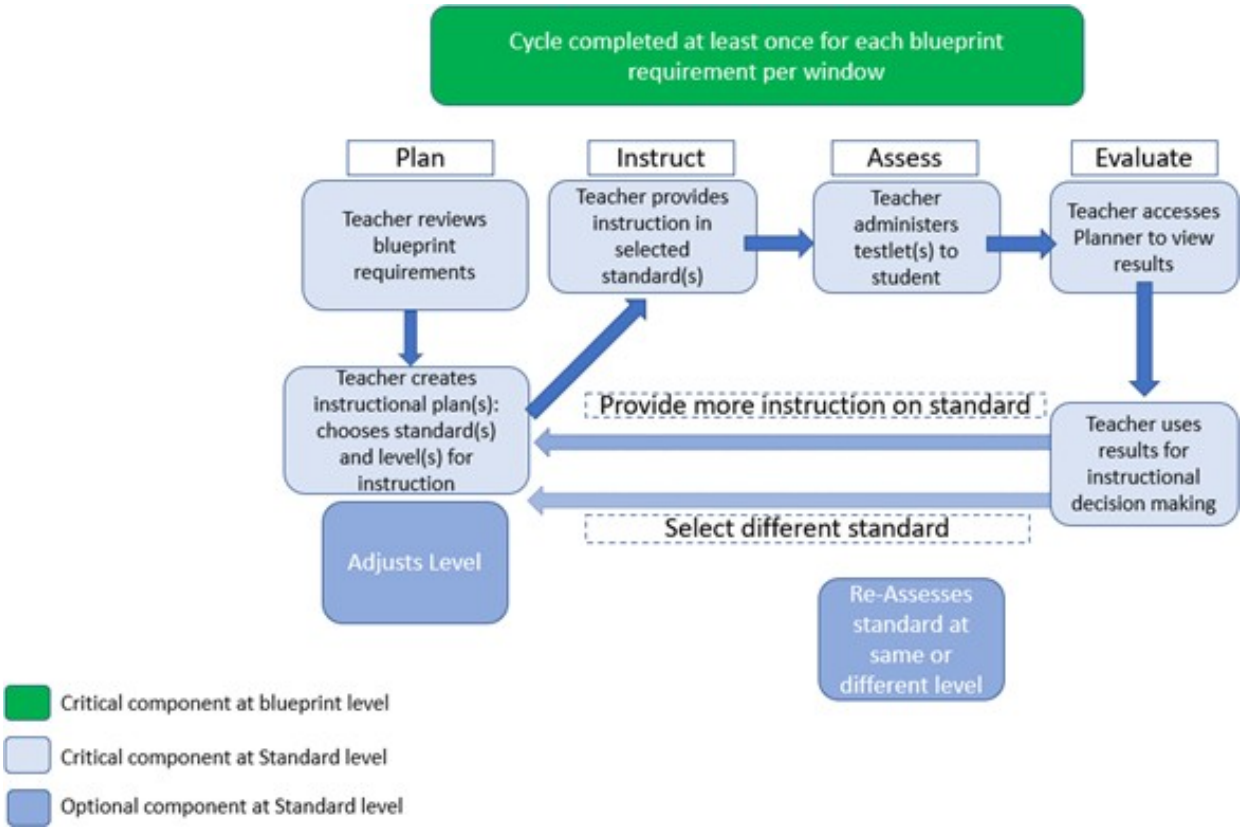
- States participating in the DLM Consortium can choose between two assessment models, and five states use an instructionally embedded model.
- States adopting the instructionally embedded model administer assessments on content standards of the teacher's choosing within blueprint constraints.
    - The blueprints are organized by groups of related standards in each subject and grade. Teachers choose which standards to assess within the constraints.

- The assessments are teacher-driven, allowing flexible selection of standards, complexity levels for assessment, and administration timing, with the expectation that the teacher covers blueprint requirements.
- Teachers receive annual training and have access to numerous resources to support their administration of assessments as intended.

Logic Model Development

The DLM theory of action represents a causal model for how DLM assessments are intended to achieve desired goals and outcomes and explains both how and why the desired change is expected to occur. One of the claims in the DLM theory of actionis that educators administer assessments with fidelity. We developed a logic model identifying the critical and optional components of implementation, based on the DLM assessment manual (Dynamic Learning Maps Consortium, 2019), other assessment documentation, and discussions with DLM staff.

For each selected standard in the assessment blueprint, teachers go through a cycle of instruction and assessment that includes five steps: 1) Plan, 2) Instruct, 3) Assess, 4) Evaluate, and 5) Provide more instruction if needed or select a different standard forinstruction. These steps are intended to be completed at least once for each standard and may be repeated at the teacher's discretion.



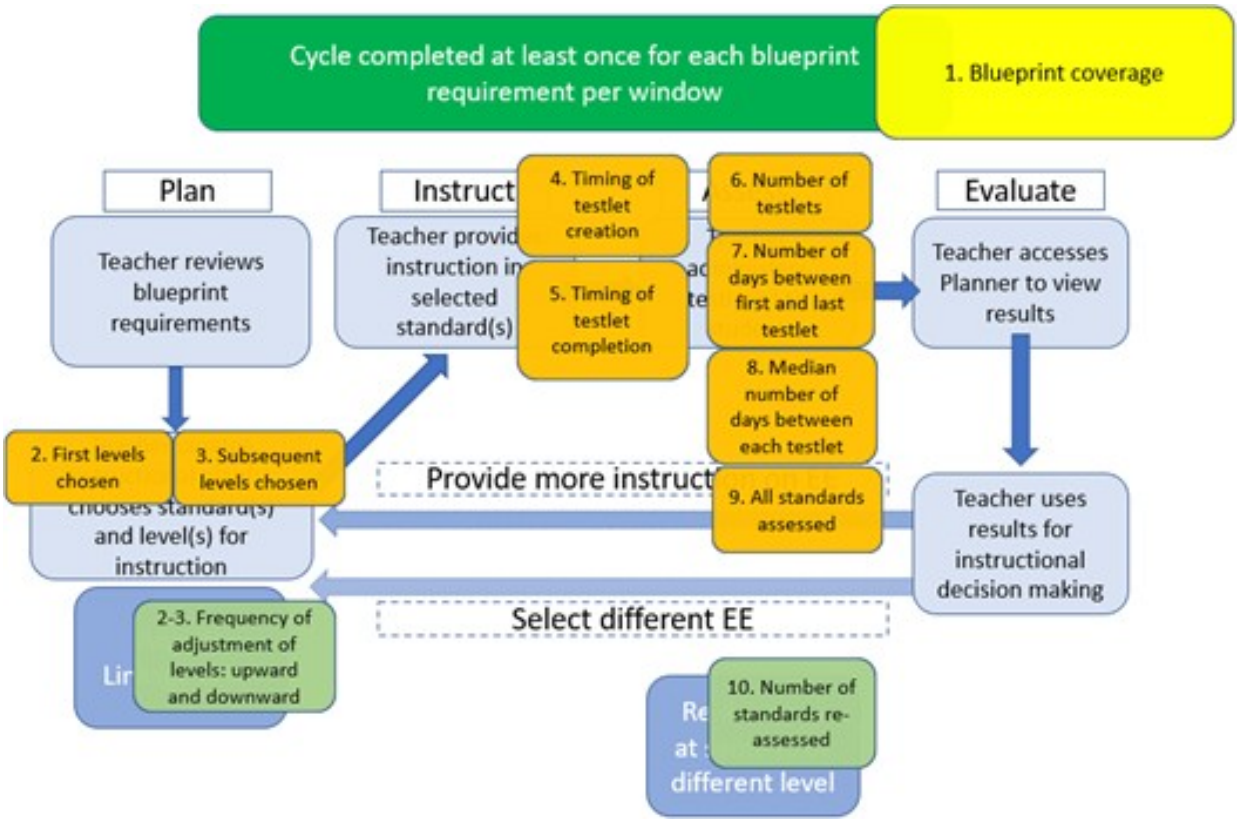
The following table shows the alignment of each component of the instructionally embedded logic model to Century et al.'s (2010) critical components and indicates whether the component is required or optional.

Step	Century et al. (2010) Critical Component(s)	Required or Optional	Description
Plan	Structural – Procedural	Required	Completing blueprint requirements and creating instructional plans
	Instructional – Pedagogical	Optional	Adjusting levels for assessment
Instruct	Instructional – Pedagogical	Required	Providing instruction on selected standard(s)
Assess	Structural – Procedural	Optional	Administering assessment(s) following published procedures
Evaluate	Instructional – Pedagogical	Optional	Viewing reports and using results to make instructional decisions
Re-Assess	Structural – Procedural	Optional	Administering assessment(s) following published procedures
	Instructional – Pedagogical	Optional	Choosing to re-assess students at the same level or a different level to assess mastery of progress
Outside-System*	Structural – Educative	Required	Completing required training to administer assessments
	Instructional – Student Engagement		Students interact with the system to show their knowledge, skills, and understanding

***Note**: These critical components are separate claims in the DLM's theory of action.*

Identification of Indicators, Criteria, and Implementation Levels

After identifying the critical and optional components of implementation, we identified indicators based on data currently available in our assessment system. This enabled us to see where there are gaps in data collection to evaluate the logic model(e.g., the Evaluate stage).



After identifying the indicators, we determined combinations of indicators that suggest implementation fidelity (or a lack thereof) based on what we consider minimum requirements for intended use of the system, and practices that support strong implementation based on our theory of action. We used these criteria to define three preliminary implementation levels (Level 1, Level 2, Level 3).

Level	Criteria	Rationale
1	Blueprint coverage not met	The blueprint describes the minimum requirements for assessment.
	All testlets assigned and completed within a one-week period	Completing all assessments within a one-week period might suggest that the teacher did not provide adequate amount of instruction on the standards that were assessed.
	Assessment of all possible standards	Assessing all standards may represent a misunderstanding of requirements and/or not linking assessment and instruction.
3	Met or exceeded blueprint coverage	The blueprint describes the minimum requirements for assessment. Teachers can choose to exceed.
	Time between first and last testlet is at least 60 days	The instructionally embedded window was 102 days in fall 2019; assessment over 60 days represents about 60% of the window, suggesting full use of the window for instruction and assessment on standards.
	Median days between testlets suggests adequate time for instruction	After each standard is selected in the DLM system, we expect teachers to provide instruction on that standard so that students have maximum opportunity to demonstrate their knowledge, skills, and understandings on the assessment. If a student assesses on standards in close succession, this could suggest that an adequate amount of instruction for each standard is not taking place.
	At least one standard is assessed more than once	Re-assessment may indicate that teachers are reteaching material and providing students with additional opportunity to learn the content of the standard.

***Note**: Cases meeting any of the criteria for Level 1 were placed in that level. Cases must have met all criteria for Level 3 to be placed in that level. All cases not meeting the Level I or Level 3 criteria were placed in Level 2.*

Exploratory Analyses

We conducted analyses to examine differences in the logic model indicators by implementation level. Data were obtained from the DLM instructionally embedded assessment system for the fall 2019 administration and represented 14,021 students in grades 3-11. Based on the current criteria, 8,602 students (31.1%) were in Level 1, 18,945 (68.4%) were in Level 2, and 152 (0.5%) were in Level 3. We computed effect sizes and odds ratios to examine differences in the indicators among implementation levels.

Implementation Indicators		Level 1		Level 2		Level 3	
		ELA[a] (*n*=3,570)	Math (*n*=5,032)	ELA (*n*=10,329)	Math (*n*=8,616)	ELA (*n*=96)	Math (*n*=56)
1. Blueprint coverage	% Not met[c]	41.8	62.9	N/A	N/A	N/A	N/A
	% Met[b]	50.7	27.1	79.9	72.0	60.4	50.0
	% Exceeded	7.5	10.0	20.1	28.0	39.6	50.0
2. First levels chosen	Average % accepted recommended level	43.2	51.3	46.4	57.7	49.8	56.6
	Average % adjusted upward	29.8	23.5	24.7	17.9	22.7	9.6
	Average % adjusted downward	27.0	25.2	28.9	24.4	27.5	33.8
3. Subsequent levels chosen for same standard	Average % accepted recommended level	31.7	31.0	31.8	30.4	35.4	35.6
	Average % adjusted upward	41.6	44.6	37.6	43.1	34.2	27.2
	Average % adjusted downward	26.7	24.4	30.7	26.5	30.4	37.2
4. Timing of testlet creation	% in first 20% of the window	4.0	5.0	9.0	7.0	4.0	0
	% in last 20% of the window	26.0	24.0	4.0	5.0	0	0
5. Timing of testlet completion	% in first 20% of the window	4.0	3.0	1.0	1.0	0	0
	% in last 20% of the window	28.0	30.0	13.0	16.0	0	0
	% All testlets completed within one week[c]	66.6	49.6	N/A	N/A	N/A	N/A
	% All testlets completed within two weeks	70.8	57.4	14.1	13.3	0	0
	% of students whose median days between testlets suggests adequate time for instruction[b]	6.9	5.5	20.7	13.1	100	100
6. Number of testlets (M, SD)		6.8 (2.9)	7.2 (4.3)	7.4 (3.3)	7.9 (4.4)	9.6 (2.1)	9.7 (1.6)
7. Number of days between first and last testlet[b] (M, SD)		14.0 (22.7)	13.7 (20.8)	28.4 (22.4)	29.2 (23.3)	72.2 (7.7)	74.9 (8.3)
8. Average % standards re-assessed[b] (M, SD)		4.2 (15.6)	3.4 (14.2)	3.9 (14.3)	3.3 (13.9)	28.5 (19.5)	27.1 (19.9)

Note. [a]English language arts; [b]Indicator used to define Level 3; [c]Indicator used to define Level 1.

The results show that many of the variables differentiate the three levels according to our hypotheses. Most implementationindicators distinguish between Levels 1 and 3.

A key finding is that cases may not clearly fall in one implementation level; rather, teachers seem to exhibit a combination of practices, some that demonstrate higher fidelity to intended practice and some that do not. For example, 6.9% of the Level 1 cases had median days between testlets that suggests that the teacher spent adequate time for instruction on each standard, and4.2% were re-assessed on at least one standard. This finding warrants further investigation and may influence subsequent development of our model of implementation fidelity.

Discussion

This work explicitly connects the literature on theories of action for assessment systems with the implementation fidelity literature originating from the program evaluation field. Incorporating implementation fidelity frameworks into a theory of action facilitates measuring action mechanisms and making and testing if/then hypotheses about how critical implementation components are related to intended outcomes of an assessment.

Century et al.'s (2010) implementation fidelity framework guided the identification of indicators that are currently available from our assessment system and helped us evaluate where there are gaps. The indicators evaluated in this research study align most directly with Century et al.'s structural/procedural components, that is, they reflect assessment fidelity including the basic steps teachers follow to set up instructional plans and administer the assessments. Some of the indicators address instructional/pedagogical components reflecting teacher actions and behaviors related to the instruction and assessment cycle which address the assessment system's theory of action. These instructional/pedagogical components are critical in embedded through-course and formative assessment systems as they represent teachers' use of assessment results for instructional decision-making.

Because these components are not directly measured in our assessment system, we are using indirect indicators to make inferences that need to be validated. For example, we used indicators on the amount of time between testlets to infer the amount of instruction on standards. We recognize that there are potential alternate hypotheses explaining teachers' decisions during their use of the assessment system.

In the next stage of research, we will more thoroughly explore alternative hypotheses and options for filling gaps in the indicators. We will collect qualitative data to further examine teachers' assumptions and motivations for making various choices in the assessment system to evaluate the extent to which our inferences about use of the system align with practice. We will continue to evaluate and refine our indicators and criteria as we learn more from future research.

Author Information

Jennifer L. Kobrin (corresponding author), Meagan Karvonen, Amy K. Clark, W. Jake Thompson, and Jennifer Burnes

Accessible Teaching, Learning, and Assessment Systems (ATLAS)
The University of Kansas, Lawrence, KS, United States

References

Century, J., Rudnick, M., & Freeman, C. (2010). A framework for measuring fidelity of implementation: A foundation for shared language and accumulation of knowledge. American Journal of Evaluation, 31(2), 199-218.

Dynamic Learning Maps Consortium (2019). Dynamic Learning Maps Test Administration Manual 2019-2020. Lawrence, KS: University of Kansas, ATLAS.

National Council on Measurement in Education (NCME, 2018). Position Statement on Theories of Action for Testing Programs. Retrieved from https://higherlogicdownload.s3.amazonaws.com/NCME/c53581e4-9882-4137-987b-4475f6cb502a/UploadedImages/Documents/NCME_Position_Paper_on_Theories_of_Action_-_Final_July 2018.pdf.