

Evaluating the Federal Innovative Assessment Demonstration Authority: Early Implementation and Progress of State Efforts to Develop New Statewide Academic Assessments

2023-004a
U.S. DEPARTMENT OF EDUCATION

A Publication of the National Center for Education Evaluation and Regional Assistance



U.S. Department of Education

Miguel Cardona
Secretary

Institute of Education Sciences

Mark Schneider
Director

National Center for Education Evaluation and Regional Assistance

Matthew Soldner
Commissioner

Marsha Silverberg
Associate Commissioner

Elizabeth Ty Wilde
Project Officer

Thomas Wei
K-12 Branch Chief

The Institute of Education Sciences (IES) is the independent, nonpartisan statistics, research, and evaluation arm of the U.S. Department of Education. The IES mission is to provide scientific evidence on which to ground education practice and policy and to share this information in formats that are useful and accessible to educators, parents, policymakers, researchers, and the public.

We strive to make our products available in a variety of formats and in language that is appropriate for a variety of audiences. You, as our customer, are the best judge of our success in communicating information effectively. If you have any comments or suggestions about this or any other IES product or report, we would like to hear from you. Please direct your comments to ncee.feedback@ed.gov.

This report was prepared for IES under Contract No. 91990019C0059 by Westat and HumRRO. The content of the publication does not necessarily reflect the views or policies of IES or the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

April 2023

This report is in the public domain. Although permission to reprint this publication is not necessary, it should be cited as:

Troppe, P., Osowski, M., Wolfson, M., Ristow, L., Lomax, E., Thacker, A., & Schultz, S. (2023). Evaluating the Federal *Innovative Assessment Demonstration Authority: Early Implementation and Progress of State Efforts to Develop New Statewide Academic Assessments* (NCEE 2023-004a). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance. <http://ies.ed.gov/ncee>

This report also is available on the IES website at <http://ies.ed.gov/ncee>.

Evaluating the Federal Innovative Assessment Demonstration Authority: Early Implementation and Progress of State Efforts to Develop New Statewide Academic Assessments

April 2023

Patricia Troppe
Michelle Osowski
Mary Wolfson
Liam Ristow
Erin Lomax
Westat

Arthur Thacker
Sheila Schultz
HumRRO

Contents

INTRODUCTION	1
Appendix A. IADA Program Description and IADA System Profiles.....	2
A.1. IADA Program Description and Requirements	2
A.2. Profiles of the Early IADA Systems	3
A.2.1. New Hampshire Performance Assessment of Competency Education (NH PACE)	7
A.2.2. Louisiana Educational Assessment Program (LEAP 2025) Humanities	11
A.2.3. North Carolina Personalized Assessment Tool (NC PAT)	16
A.2.4. Georgia MAP Assessment Partnership (GMAP) Through-Year Assessment	20
A.2.5. Georgia Navy	24
Appendix B. Study Methods.....	28
B.1. Data Sources and Collection Approach	28
B.1.1. State IADA Application Materials	28
B.1.2. State IADA Annual Performance Reports (APRs).....	29
B.1.3. Interviews with IADA System Officials.....	30
B.2. Analytic Approaches	31
B.2.1. Analyzing System Readiness at the Start of the IADA Demonstration Period	31
B.2.2. Analyzing System Status in 2020-21 and Progress Made.....	32
B.2.3. Analyzing System Challenges	43
Appendix C. Supplemental Information on Findings in the Report.....	44
Endnotes	58
References.....	66

List of Exhibits

Exhibit A.1.	Features of IADA Assessment systems as of 2020-21.....	4
Exhibit A.2.	Number of NH PACE districts and all districts in New Hampshire: 2018-19 and 2019-20.....	10
Exhibit A.3.	Planned development and administration of LEAP 2025 Humanities, by year (as of the Louisiana IADA application)	13
Exhibit A.4.	Actual development and administration of LEAP 2025 Humanities, by year (as of 2020-21)	13
Exhibit A.5.	Number of LEAP 2025 Humanities districts and all districts in Louisiana: 2018-19 and 2019-20	15
Exhibit A.6.	Planned development and administration of NC PAT, by year (as of the North Carolina IADA application)	18
Exhibit A.7.	Actual development and administration of NC PAT, by year (as of 2020-21).....	18
Exhibit A.8.	Number of NC PAT districts and all districts in North Carolina: 2019-20 and 2020-21	19
Exhibit A.9.	Planned development and administration of GMAP through-year assessments, by year (as of the Georgia application).....	22
Exhibit A.10.	Actual development and administration of GMAP through-year assessments, by year (as of 2020-21).....	22
Exhibit A.11.	Number of GMAP districts and all districts in Georgia: 2019-20 and 2020-21.....	23
Exhibit A.12.	Planned development and administration of Navy assessments, by year (as of the Georgia application)	26
Exhibit A.13.	Actual development and administration of Navy assessments, by year (as of 2020-21)	26
Exhibit A.14.	Number of Navy districts and all districts in Georgia: 2019-20 and 2020-21.....	27
Exhibit B.1.	IADA status indicators and thresholds, by major assessment implementation activity and sub-activity	33
Exhibit B.2.	Review protocol for the bias review indicator	40
Exhibit C.1.	2020-21 Status of IADA assessment systems, by indicators of initial planning.....	45

Exhibit C.2.	2020-21 Status of IADA assessment systems, by indicators of preparation for the first operational assessment	46
Exhibit C.3.	2020-21 Status of IADA assessment systems, by indicators of operational assessment administration	47
Exhibit C.4.	2020-21 Status of IADA assessment systems, by indicators of post-administration activities	48
Exhibit C.5.	2020-21 Status of IADA assessment systems, by indicators of continuous improvement activities.....	49
Exhibit C.6.	2020-21 Status of IADA assessment systems, by indicators of system scale-up.....	50
Exhibit C.7.	2019-20 Status of IADA assessment systems, by indicators of initial planning	51
Exhibit C.8.	2019-20 Status of IADA assessment systems, by indicators of preparation for the first operational assessment	52
Exhibit C.9.	2019-20 Status of IADA assessment systems, by indicators of operational assessment administration	53
Exhibit C.10.	2019-20 Status of IADA assessment systems, by indicators of post-administration activities	54
Exhibit C.11.	2019-20 Status of IADA assessment systems, by indicators of continuous improvement activities.....	56
Exhibit C.12.	2019-20 Status of IADA assessment systems, by indicators of system scale-up.....	57

INTRODUCTION

This appendix volume supplements the report NCEE 2023-004 analyzing the early progress of the first five assessment systems participating in the Innovative Assessment Demonstration Authority (IADA) created by the 2015 reauthorization of the Elementary and Secondary Education Act. These systems include: the New Hampshire Performance Assessment of Competency Education (NH PACE), the Louisiana Educational Assessment Program 2025 Humanities (LEAP 2025 Humanities), the North Carolina Personalized Assessment Tool (NC PAT), the Georgia MAP Assessment Partnership (GMAP) through-year assessment, and the Georgia Navy assessment system.¹ The volume documents the features of the IADA program and the five IADA systems, the study team’s research activities, and includes more detailed findings than were summarized in the report.

Appendix A. IADA Program Description and IADA System Profiles

Appendix A describes the IADA program and its requirements. This appendix also details each of the five IADA assessment systems included in the report.

A.1. IADA Program Description and Requirements²

According to legislation and other federal documents, the purpose of the IADA program is to encourage states to “establish, operate, and evaluate an innovative assessment system, including for use in statewide accountability systems, with the goal of using the innovative assessment system after the demonstration authority ends to meet the academic and statewide accountability system requirements.”³ IADA offers certain flexibilities, currently for up to seven states or consortia of states⁴ to make it easier for them to further “develop and administer high-quality, valid, and reliable assessments that measure student mastery of challenging State academic standards, improve the design and delivery of large-scale assessments, and better inform classroom instruction, ultimately leading to improved academic outcomes for all students.”⁵ The IADA program is authorized by the Elementary and Secondary Education Act (ESEA), as reauthorized by the Every Student Succeeds Act.⁶

IADA intends to be flexible in ways that allow states to meet the broad goals of the program. For example, it allows states to explore the use of any assessment design⁷—such as those that are competency-based, instructionally embedded, interim during the year, end-of-year cumulative, or performance-based.⁸ The leeway to tailor the IADA assessment systems by varying the timing and number of tests, for instance, may help educators better identify individual student learning needs and whether students are ready to demonstrate proficiency.⁹

However, assessment systems under IADA must also meet the requirements for federal accountability reporting under the ESEA and a specified timeline for moving toward using the assessment system statewide. States are expected to use their IADA-approved assessment for accountability in participating schools and local education agencies (LEAs) during the first year of the demonstration authority and to scale up their IADA systems to statewide use within 5 years.¹⁰ The Secretary may extend the demonstration authority period by an additional 2 years if the state provides evidence that the IADA assessment system is meeting program requirements.¹¹ After the IADA system is scaled statewide, the state must submit evidence to the U.S. Department of Education to show the suitability of the IADA system as a replacement to the existing statewide assessment system. In carrying out their IADA activities, states are required to consult and communicate with stakeholders,¹² to develop and use IADA assessments,¹³ and to support and monitor the assessment systems and report on results.¹⁴

While states do not have to participate in the IADA program to develop, administer, and use innovative assessments, the program incentivizes participation by removing some potential barriers to implementing new accountability assessments. For example, states without IADA authority that are testing an innovative assessment with a subset of students must also have those students take the regular statewide assessment. States with IADA authority, however, would only be required to use one assessment for accountability, eliminating “double testing” of some students.

States or state consortia receive no new federal funding under IADA.¹⁵ However, they may use funds from other federal grant programs or apply for new funds to support the development and piloting of a new assessment system. For example, states may draw on funds reserved for state administration in Title I-A (Improving Basic Programs Operated by State and Local Educational Agencies) or in Title I-B (State Assessment Grants) of ESEA.¹⁶

A.2. Profiles of the Early IADA Systems

State application materials and annual performance reports (APRs) through 2020-21 provide a description of the five IADA systems included in this report, though there is variation in the comprehensiveness and level of detail in these materials (see Appendix B). Exhibit A.1 highlights similarities and differences in the key features of the assessment systems.¹⁷

Exhibit A.1. Features of IADA assessment systems as of 2020-21

Planned assessment features ^a	IADA Assessment System				
	New Hampshire Performance Assessment of Competency Education (NH PACE)	Louisiana Educational Assessment Program (LEAP) 2025 Humanities	North Carolina Personalized Assessment Tool (NC PAT)	Georgia MAP Assessment Partnership (GMAP) Through-Year Assessment	Georgia Navvy
Basic Assessment Structure <ul style="list-style-type: none"> Assessment type (and number) Item type Scoring method 	Curriculum-embedded, performance-based	Interim (3) and a shorter than the traditional summative (1)	Interim (3) and summative (1)	Interim (3)	Interim (Multiple)
	Tasks to demonstrate proficiency in a competency (e.g., presentation, project)	Multiple-choice/multiple-select ^b or technology-enhanced, ^c constructed-response, ^d and short essay/writing prompt for interims Multiple-choice/multiple-select or technology-enhanced and extended essay for the summative	Multiple-choice/multiple-select, technology-enhanced, and constructed-response	Multiple-choice/multiple-select, technology-enhanced, and constructed-response ^e	Multiple-choice/multiple-select
	Hand-scored using standardized protocols	Machine-scored and hand-scored	Machine-scored	Machine-scored for multiple-choice/multiple-select, hand-scored for constructed-response	Machine-scored
Assessment System Scope <ul style="list-style-type: none"> Content Grades 	English language arts (ELA), math, and science	Combines ELA and social studies assessment items into a single assessment	Math and ELA	ELA, math, and science	ELA, math, and science
	<ul style="list-style-type: none"> Grades 4 through 7 (ELA) Grades 3, 5 through 7 (math) Grade 8 (science) High school (course specific) <ul style="list-style-type: none"> ELA, math, and science 	<ul style="list-style-type: none"> Grades 3 through 8 High school (English I [Humanities I] and English II [Humanities II]) 	Grades 3 through 8	<ul style="list-style-type: none"> Grades 3 through 8 (ELA, math) Grades 5 and 8 (science) 	<ul style="list-style-type: none"> Grades 3 through 8 (ELA, math) Grades 5 and 8 (science) High school <ul style="list-style-type: none"> ELA and math, 2 courses for each content (e.g., English I and II) Science - 1 course

Exhibit A.1. Features of IADA assessment systems as of 2020-21 (continued)

Planned assessment features ^a	IADA Assessment System				
	New Hampshire Performance Assessment of Competency Education (NH PACE)	Louisiana Educational Assessment Program (LEAP) 2025 Humanities	North Carolina Personalized Assessment Tool (NC PAT)	Georgia MAP Assessment Partnership (GMAP) Through-Year Assessment	Georgia Navvy
Test Administration <ul style="list-style-type: none"> • Timing • Mode 	Flexible, teacher-determined as part of curriculum	Preset assessment windows	Preset assessment windows ^f	Preset assessment windows	Flexible, district-determined ^g
	Computer administration, paper and pencil, or other modes as determined by teacher (not adaptive ^h)	Computer administration (not adaptive)	Computer administration (adaptive, student results from interims determine form for the summative)	Computer administration (adaptive within a test)	Computer administration (not adaptive)
Test Scores that Contribute to Accountability	All interim proficiency determinations ⁱ	Interim and summative scores (approximately 67%-80% of students' summative scores are expected to be based on interims and remaining based on the summative)	Single summative assessment	All interim scores	All interim proficiency determinations

^a This table reports on features of the IADA assessments as planned by the system as of 2020-21 except for NH PACE. The features for NH PACE reflect the assessment as administered. No other system administered an operational assessment by 2020-21.

^b Students select one or more than one correct response from a set of responses for multiple-choice/multiple-select items.

^c Technology-enhanced items are computer-administered and more complex than multiple-choice/multiple-select items. For example, students may drag and drop answers from a list, order items, or highlight text to identify a central idea.

^d A constructed-response item requires a student to independently create a response, rather than select a response from a set of choice options. Examples of constructed-response items include fill-in-the-blank and short essay.

^e The GMAP through-year assessment will likely include constructed-response items on the fall and winter interims for formative purposes. If included, the formative constructed-response items will be scored by teachers for their own use. The 2020-21 annual performance report also suggests that GMAP districts are considering technology-enhanced items.

^f The North Carolina Department of Public Instruction plans two assessment windows for NC PAT: one large window for the three interims, giving teachers flexibility on administration timing, and a fixed window for the summative.

^g Districts administering Navvy may allow schools or teachers to determine the calendar days during which an assessment may be administered, or “the assessment windows.”

^h An adaptive test provides a tailored assessment based on a student’s performance on prior assessment questions (or prior assessment scores).

ⁱ A “proficiency determination” is an indication of whether the student has met a proficiency threshold (e.g., Proficient, Not Proficient).

SOURCE: Application materials and annual performance reports for New Hampshire, Louisiana, North Carolina, and Georgia through 2020-21.

Profiles of each assessment system provide more information that policymakers or state leaders may find useful. The profiles focus on the following areas:¹⁸

- **Rationale (subsection a).** While all IADA states applied for demonstration authority to improve their assessment systems and have broadly similar goals (e.g., generating more frequent and more useful data), their specific rationale for applying varied in substantive ways. These differences are important to document as they provide context that may help illuminate why states' implementation paths played out the way they did.
- **Related assessment work prior to IADA (subsection b).** Each of the IADA systems began its demonstration period with some elements already in place or under development to facilitate their approval for and work under IADA. The nature of that prior experience could influence how much progress they could or did make during the demonstration.
- **Description of the IADA assessments and timeline through 2020-21 (subsection c).** The IADA program expected state systems to choose and use innovative features that would improve the usefulness of assessments and to implement these new assessments statewide after 5 years. Within these broad parameters, states were free to decide what kind of assessments to implement and on what specific timelines. It is therefore of interest to describe what IADA states elected to focus on, as well as their timeline for development and scale-up, to better understand if and how these key aspects of participation varied across the five systems.
- **Assessment partners and their roles (subsection d).** All IADA systems received advice and expertise from partners as required. However, documenting who in particular was involved and in what role is of interest because assessment partners help shape the IADA systems, either by acting as contractors to help design and field a state's assessment system or by helping to build capacity among educators to design, build, and maintain an assessment system.
- **Expectations of staff (subsection e).** Assessment innovations require the support and participation of educators. Better understanding what states expected of them, and whether that differed across states, is important because implementation of IADA assessments can introduce trade-offs by, for example, increasing educators' burden with new test administration windows, and by requiring their participation in meetings or training/professional development. New assessments could also alter educator interactions with students and parents, or involve educators in test item development and scoring.
- **District involvement (subsection f).** The number of districts involved¹⁹ with the system each year can be compared with the total number in the state to better understand how far the initial IADA implementation reaches compared to the key program goal of statewide implementation.

A.2.1. New Hampshire Performance Assessment of Competency Education (NH PACE)

NH PACE AT A GLANCE	
First Year of Demonstration Authority	2018-19
Basic Assessment Structure	Curriculum-embedded, performance-based
<ul style="list-style-type: none"> • Assessment type • Item type • Scoring method 	Tasks to demonstrate proficiency in a competency (e.g., presentation, project) Hand-scored using standardized protocols
Assessment System Scope	English language arts (ELA), math, and science
<ul style="list-style-type: none"> • Content • Grades 	<ul style="list-style-type: none"> • Grades 4 through 7 (ELA) • Grades 3, 5 through 7 (math) • Grade 8 (science) • High school (course specific) <ul style="list-style-type: none"> – ELA, math and science
Test Administration	Flexible, teacher-determined
<ul style="list-style-type: none"> • Timing • Mode 	Computer administration or paper-and-pencil components, determined by teacher (not adaptive ^a) Multiple modalities (e.g., projects, presentations, skits)
Test Scores that Contribute to Accountability	All interim proficiency determinations ^b

^a An adaptive test provides a tailored assessment based on a student’s performance on prior assessment questions (or prior assessment scores).

^b A “proficiency determination” is an indication of whether the student has met a proficiency threshold (e.g., Proficient, Not Proficient).

SOURCE: Application materials and annual performance reports for New Hampshire through 2020-21.

A.2.1.a. Rationale

The New Hampshire Performance Assessment of Competency Education (NH PACE) is the assessment system aligned to the state’s larger educational reform effort that began before 2012. Before 2012, the state grew concerned that students graduating from New Hampshire schools were not properly prepared for jobs, careers, or college. In 2012, the state transitioned public education to a competency-based learning system.²⁰ Through NH PACE, New Hampshire aims to measure a student’s mastery of competencies,²¹ which include higher-order thinking²² and other skills, such as problem solving, communication, collaboration, and creativity²³ in addition to proficiency on the state’s content standards. The assessments intend to link curriculum and instruction by embedding performance tasks²⁴ into instruction so that the results from the assessments can provide instructionally useful information to teachers and students.

A.2.1.b. Related Assessment Work Prior to IADA

The New Hampshire Department of Education (NH DOE) already had experience designing, developing, and implementing an ESEA-approved assessment system using NH PACE prior to IADA. Beginning in 2012, a cohort of nine schools implemented competency-based education. In 2013, New Hampshire piloted the first versions of the NH PACE assessments. Beginning in 2014-15 through the next 3 school years (2015-16, 2016-17, and 2017-18), the U.S. Department of Education allowed a small-scale pilot of the NH PACE assessment system for instructional and accountability purposes in these nine schools. In 2017, the New Hampshire legislature gave permission and policy support to scale the NH PACE statewide.

A.2.1.c. Description of the IADA Assessments and Timeline Through 2020-21

Since 2012, the thrust of the NH PACE assessment system has been essentially unchanged. The NH PACE assessments use two types of performance-based tasks to assess student mastery in a competency-based learning system: (1) teacher-developed tasks for use in their classrooms (i.e., locally developed performance assessments) and (2) tasks developed collaboratively across participating districts for use in all classes within a given grade/subject (i.e., common tasks).

The core feature of the NH PACE assessment system is that local districts leverage the expertise of their teachers to design common performance assessment tasks intended to measure identified competencies and then to embed those assessment tasks into the curriculum. All districts that voluntarily agreed to use the NH PACE must use at least one of the commonly developed performance tasks to help in calibrating scores across districts and for scorer training. However, districts supplement those common tasks with locally developed tasks to assess the full range of district-required competencies for each grade and content area.²⁵ Teachers can use the assessment tasks when they teach the related content, the intent being that it will lead to performance assessment that is directly tied to the content and timing of classroom instruction. To score their locally developed performance assessments in each grade and content area, many NH PACE districts use the same scoring procedures required for scoring the NH PACE common tasks (i.e., use of a scoring rubric). Scores from each performance assessment contribute to the overall determination of a student's proficiency.

According to its IADA application, NH DOE intended to administer operational assessments²⁶ each year in all grades and subjects targeted for NH PACE, expand the use of the NH PACE assessments to additional districts to meet scale-up expectations, and continue to investigate the comparability of results to the traditional statewide assessment (NH SAS). NH PACE administered operational assessments in 2018-19, but only some assessments during 2019-20 due to pandemic disruptions. No district used NH PACE assessments in 2020-21. New Hampshire received approval from the U.S. Department of Education to pause its IADA demonstration authority that school year.

A.2.1.d. Assessment System Partners and Their Roles

The NH DOE had three external partners that provided technical and professional learning expertise: (1) the National Center for the Improvement of Educational Assessment (the Center for Assessment), (2) the New Hampshire Learning Initiative (NHLI), and (3) Demonstrated Success. The Center for Assessment provided lead technical and policy development support (e.g., working to ensure the quality and rigor of the NH PACE common tasks), designed methods to aid in evaluating comparability of student results across districts, and produced the technical documentation for NH PACE. NHLI led efforts to raise external funds to support the assessment pilot work and to coordinate task development (e.g., identifying the specific content, format, administration, and scoring protocol for each task). Demonstrated Success was responsible for data analysis; however, the company went out of business during the COVID-19 pandemic.

A.2.1.e. Expectations of Staff

Schools and districts participate in the NH PACE assessment system to varying degrees (i.e., use of results can range from formative/instructional monitoring to accountability).²⁷ In districts that use the results from the NH PACE assessment system for accountability purposes for at least one grade and content, district and school staff are expected to: (1) develop and adopt a coherent set of course and grade competencies that map to the state's academic standards, (2) participate in peer and expert reviews of performance tasks, (3) administer at least one NH PACE common task in classrooms and submit student work from that task for scoring calibration, and (4) participate in assessment-related professional development.

A.2.1.f. District Involvement

District participation is entirely voluntary. In 2018-19, 15 school districts participated in the NH PACE assessment system (including 14 regular public school districts and one charter district), out of 192 eligible districts (Exhibit A.2). These districts used NH PACE assessments during the school year, but not necessarily for accountability. In 2019-20, 13 out of 193 districts were involved (including 12 regular public school districts and one charter school district).

Exhibit A.2. Number of NH PACE districts and all districts in New Hampshire: 2018-19 and 2019-20

District type	NH PACE districts		All districts in state ^a	
	2018-19	2019-20	2018-19	2019-20
Total districts	15	13	192	193
Regular local school district ^b	14	12	165	165
Independent charter school district ^c	1	1	27	28

^a Excludes districts that do not offer grades 38 or high school, the grades eligible for NH PACE.

^b “Regular local school district” is an agency responsible for providing free public education for school-age children residing within its jurisdiction. This category excludes local supervisory unions that provide management services for a group of associated school districts; regional education service agencies that typically provide school districts with research, testing, and data processing services; state and federally operated school districts; and other agencies that do not fall into these groupings.

^c “Independent charter school district” is a school district that includes only charter schools, typically a single school that was authorized under the charter.

NOTES: The NH PACE districts used the NH PACE system during the school year. These districts may or may not have used NH PACE for accountability purposes. The NH PACE system was not active in the 2020-21 school year, and New Hampshire did not report any involved districts.

SOURCE: New Hampshire’s annual performance reports through 2020-21; National Center for Education Statistics, Common Core of Data.

A.2.2. Louisiana Educational Assessment Program (LEAP 2025) Humanities

LEAP 2025 Humanities AT A GLANCE	
First Year of Demonstration Authority	2018-19
Basic Assessment Structure	Interim (3) and a shorter than the traditional summative (1)
<ul style="list-style-type: none"> Assessment type (and number) Types of items Scoring method 	<ul style="list-style-type: none"> Multiple-choice/multiple-select^a or technology-enhanced,^b constructed-response,^c and short essay/writing prompt for the interims Multiple-choice/multiple-select or technology-enhanced, and extended essay for the summative
	Machine-scored and hand-scored
Assessment System Scope	Combines English language arts (ELA) and social studies assessment items into a single assessment
<ul style="list-style-type: none"> Content Grade(s) 	Grades 3 through 8 High school ^d
Test Administration	Preset assessment windows
<ul style="list-style-type: none"> Assessment timing Administration mode 	Computer administration (not adaptive ^e)
Test Scores that Contribute to Accountability	<ul style="list-style-type: none"> Interim and summative scores Approximately 67%-80% of students' summative scores are expected to be based on interims and remaining based on the summative

^a Students select one or more than one correct response from a set of responses for multiple-choice/multiple-select items.

^b Technology-enhanced items are computer-administered and more complex than multiple-choice/multiple-select items.

For example, students may drag and drop answers from a list, order items, or highlight text to identify a central idea.

^c A constructed-response item requires a student to independently create a response, rather than select a response from a set of choice options. Examples of constructed-response items include fill-in-the-blank and short essay.

^d LEAP 2025 Humanities assessments are proposed for high school students enrolled in English I (Humanities I) and English II (Humanities II) courses.

^e An adaptive test provides a tailored assessment based on a student's performance on prior assessment questions (or prior assessment scores).

SOURCE: Application materials and annual performance reports for Louisiana through 2020-21.

A.2.2.a. Rationale

LEAP 2025 Humanities, the innovative assessment system being developed in Louisiana under IADA, will be an alternative to the traditional statewide LEAP 2025 ELA assessment system. According to Louisiana's IADA application, the LEAP 2025 Humanities system is being designed to mitigate potential bias and promote fairness in Louisiana's state assessments. Unlike the traditional LEAP 2025 ELA assessments, LEAP 2025 Humanities uses texts introduced during classroom instruction, theoretically making a student's background knowledge and prior experience less likely to affect the measurement of academic proficiency. Using known texts also allows the assessment to measure higher-order thinking skills (e.g., synthesis and analysis) rather

than lower-level skills (e.g., recognition and recall).²⁸ This is intended to promote fairer, less biased assessments of student achievement.

LEAP 2025 Humanities is also intended to address stakeholder concerns about the balance between testing and instructional time. If the Louisiana Department of Education (LDOE) is able to align the LEAP 2025 Humanities assessments with the state’s ELA and social studies standards, then two assessments would be merged into one, thus reducing testing time.²⁹

A.2.2.b. Related Assessment Work Prior to IADA

LEAP 2025 Humanities builds on earlier work of the LDOE. In 2015, Louisiana began revising its academic standards (now called LEAP 2025 standards) and aligned the traditional LEAP ELA assessments to the revised standards, creating LEAP 2025 ELA assessments. In 2016-17, the LDOE also built a standards-aligned, text-based ELA curriculum aligned to the revised standards—the ELA Guidebooks 2.0. Like the LEAP 2025 ELA assessments, the LEAP 2025 Humanities will align to the new standards and curriculum. However, the LEAP 2025 Humanities assessment will also draw on the format and content of the ELA Guidebooks 2.0 curriculum, using the same texts specified in the guidebooks for the assessment. The LDOE is also exploring the feasibility of aligning the LEAP Humanities assessments to curriculum and standards in social studies as part of their IADA system, in order to reduce testing burden.

A.2.2.c. Description of the IADA Assessments and Timeline Through 2020-21

Louisiana is piloting the LEAP 2025 Humanities assessments as a potential alternative to separate ELA and social studies assessments in grades 3-8 and high school.³⁰ Each LEAP 2025 Humanities assessment will include three interim assessments and one summative assessment designed to be shorter than the state’s traditional end-of-year summative assessment. Both the summative and interim LEAP Humanities assessments are intended to have stronger alignment than Louisiana’s traditional assessments to texts used in classroom teaching.

The computer-administered LEAP 2025 Humanities assessment also intends to measure knowledge and skills not readily assessed through multiple-choice or multiple-select test items through extended essay/writing prompts, constructed-response, and other technology-enhanced items.³¹ Because some LEAP 2025 Humanities items will be hand-scored by a vendor, results will not be immediate; however, the results are meant to be timelier than those from the traditional ESEA summative assessments.

Annual determinations of academic proficiency will be calculated by combining the interim and summative test scores. LDOE expects approximately 67 to 80 percent of students’ summative scores to be based on interim assessments with the remainder from the summative assessment.

According to its IADA application, the LDOE planned a staggered schedule for LEAP 2025 Humanities (Exhibit A.3) with administration across all planned grades expected by year 4 of the IADA demonstration period.

Exhibit A.3. Planned development and administration of LEAP 2025 Humanities, by year (as of the Louisiana IADA application)

Year	Planned Assessment Development	Planned Administration of an Operational Assessment
2018-19	Humanities I and II (High School)	None
2019-20	Humanities Grades 6-8	Humanities I and II (High School)
2020-21	Humanities Grades 3-5	Humanities I and II (High School) Humanities Grades 6-8
2021-22 and 2022-23	None	Humanities I and II (High School) Humanities Grades 6-8 Humanities Grades 3-5

SOURCE: Louisiana’s application materials.

However, the LDOE’s timeline for development and administration shifted in the first 3 years of their demonstration authority (Exhibit A.3 compared with Exhibit A.4). In its application, LDOE initially planned to begin with the development, piloting, and administration of Humanities I and II assessments for aligned high school courses. However, according to the 2018-19 IADA annual performance report, LDOE instead started assessment development with Humanities Grade 7. Subsequently, in 2019-20 the LDOE began developing Humanities assessments for grades 6 and 8, and it continued this development work in 2020-21. At that time, the LDOE also adjusted their planned approach to the grade 5 assessment under IADA to align with their Wit and Wisdom curriculum.

Exhibit A.4. Actual development and administration of LEAP 2025 Humanities, by year (as of 2020-21)

Year	Assessment Development	Actual Administration of an Operational Assessment
2018-19	Humanities Grade 7	None
2019-20	Humanities Grades 6 and 8	Humanities Grade 7 (administration not complete due to COVID-19 disruptions)
2020-21	Humanities Grades 6 and 8	None

SOURCE: Louisiana’s annual performance reports through 2020-21.

A.2.2.d. Assessment System Partners and Their Roles

The LDOE has partners involved in both the day-to-day work of the system as well as contributing to its overall vision. Northwest Evaluation Association (NWEA) is responsible for developing assessment frameworks, assessment guides, assessment items aligned with the Louisiana content standards, and operational test forms for each grade/course in the LEAP 2025 Humanities pilot. They also are expected to ensure security³² during test administration and scoring; to design and produce district, school, teacher, and student reports; and to provide ongoing psychometric support for the demonstration. NWEA also manages demonstration activities across partners. NWEA will also administer LEAP 2025 Humanities assessments both during and after the IADA demonstration period.

The LDOE also partnered with the Johns Hopkins Institute for Educational Policy, the Center for Assessment, and Strategic Measurement & Evaluation to form the LDOE's Technical Advisory Committee (TAC). The TAC provides technical expertise and supports the evaluation of the LEAP 2025 Humanities IADA system, including ensuring the validity, reliability, and comparability of summative tests for students in the IADA and regular statewide assessment systems. The TAC also helps to ensure the quality and efficacy of resources provided to districts participating in the pilot. MZ Development is another partner, primarily hosting the student test, administrator, and scoring sites for LEAP 2025 Humanities assessments.

A.2.2.e. Expectations of Staff

The LDOE expects staff in LEAP 2025 Humanities schools and districts to participate in professional learning related to the system, to administer the assessments as specified by the state, and to provide feedback regarding assessment items. The pilot schools and districts are also expected to provide feedback on their experience administering the tests, and to participate in webinars on administering the assessment as well as reporting and using the results.

A.2.2.f. District Involvement

District involvement in the LEAP pilot is voluntary. In 2018-19, five school districts were involved in LEAP (including four public school districts and one charter school district), out of 202 eligible districts (Exhibit A.5). In 2019-20, 17 out of 193 eligible districts were involved (including 16 regular school districts and one charter district).

Exhibit A.5. Number of LEAP 2025 Humanities districts and all districts in Louisiana: 2018-19 and 2019-20

District type	LEAP 2025			
	Humanities districts		All districts in state ^a	
	2018-19	2019-20	2018-19	2019-20
Total districts	5	17	202	193
Regular local school district ^b	4	16	71	72
Independent charter school district ^c	1	1	121	111
Other school districts ^d	0	0	10	10

^a Excludes districts that do not offer grades 3-8 or high school, which are the grades eligible for LEAP 2025 Humanities.

^b “Regular local school district” is an agency responsible for providing free public education for school-age children residing within its jurisdiction. This category excludes local supervisory unions that provide management services for a group of associated school districts; regional education service agencies that typically provide school districts with research, testing, and data processing services; state and federally operated school districts; and other agencies that do not fall into these groupings.

^c “Independent charter school district” is a school district that includes only charter schools, typically a single school that was authorized under the charter.

^d “Other school districts” include specialized public school districts, state agency providing elementary and/or secondary level instruction, and other education agencies.

NOTES: For 2018-19, the number of districts is those that field-tested the grade 7 LEAP 2025 Humanities assessment in spring 2019. In 2019-20, the number of districts includes those that planned to field-test the grades 6 and 8 assessments, or planned to administer the full grade 7 LEAP 2025 Humanities assessment for accountability. The LEAP 2025 Humanities assessment system was not active in the 2020-21 school year.

SOURCE: Louisiana’s annual performance reports through 2020-21; National Center for Education Statistics, Common Core of Data.

A.2.3. North Carolina Personalized Assessment Tool (NC PAT)

NC PAT AT A GLANCE	
First Year of Demonstration Authority	2019-20
Basic Assessment Structure	Interim (3) and summative (1)
<ul style="list-style-type: none"> • Assessment type (and number) • Item type • Scoring method 	Multiple-choice/multiple-select, ^a constructed-response, ^b technology-enhanced items ^c
	Machine-scored
Assessment System Scope	Math and English language arts (ELA)
<ul style="list-style-type: none"> • Content • Grades 	Grades 3 through 8
Test Administration	Preset assessment windows
<ul style="list-style-type: none"> • Timing • Mode 	Computer administration (adaptive, ^d student results from interims determine student’s form for the summative)
Test Scores that Contribute to Accountability	Single summative assessment

^a Students select one or more than one correct response from a set of responses for multiple-choice/multiple-select items.

^b A constructed-response item requires a student to independently create a response, rather than select a response from a set of choice options. Examples of constructed-response items include fill-in-the-blank and short essay.

^c Technology-enhanced items are computer-administered and more complex than multiple-choice/multiple-select items. For example, students may drag and drop answers from a list, order items, or highlight text to identify a central idea.

^d An adaptive test provides a tailored assessment based on a student’s performance on prior assessment questions (or prior assessment scores).

SOURCE: Application materials and annual performance reports for North Carolina through 2020-21.

A.2.3.a. Rationale

In 2014, the North Carolina Department of Public Instruction (NCDPI) established a task force to help respond to stakeholder concerns about the amount of time spent each year on state assessments, which do not inform instruction. The task force ultimately recommended a new assessment model with three or four assessments administered throughout the school year that could also be used to inform instruction and produce a summative score. While increasing the number of state assessments administered throughout the school year may seem counterintuitive to reducing testing time, the intended increase in the relevance of those assessments for instruction appears to have been an important consideration for the task force’s recommendation.

A.2.3.b. Related Assessment Work Prior to IADA

In response to the 2014 task force recommendation, a small set of volunteer districts began piloting items in 2015-16 for interim assessments in both math (grades 3-8) and ELA (grades 4-8). These interim assessments, called NC Check-Ins, were designed to allow teachers to use assessment results to better inform their classroom instruction. In 2018-19, after 3 years of piloting, the state made NC Check-Ins interim assessments available to districts statewide. The state's IADA assessment—NC PAT—builds on the NC Check-Ins model and aims to leverage lessons learned from the development of NC Check-Ins to create an assessment system for instructional and accountability purposes.

A.2.3.c. Description of the IADA Assessments

The NC PAT is being developed as a computer-administered assessment system, including three interim assessments and an adaptive summative assessment. The interim assessments will be administered by classroom teachers as their instruction progresses throughout the school year. NCDPI plans for one large assessment window (e.g., October 1-May 31) for the three interim assessments, giving teachers flexibility on the timing of administration. NCDPI also will give schools flexibility to determine whether students will take the interim assessments in person or remotely, and whether to administer each interim assessment in a single day or over multiple days. Teachers will receive interim assessment results immediately to help inform instruction.

The NC PAT summative assessment is intended to be adaptive and will use scores from the interim assessments to determine an individual student's summative assessment form. The summative assessment will include both common items (given to all students) and targeted items (informed by the student's interim assessment results). For example, a student would get an easier or harder summative test form based on their interim assessment results. The intent is that a better "fitting" score will provide more accurate results of the student's performance than the traditional summative assessment without an adaptive feature. Students will need to complete at least two of the three interims to have their prior performance inform the summative form. The final score from the summative assessment will be used to make annual determinations of academic proficiency. Unlike the interim assessments, the assessment window for the summative assessment will be fixed to the last 10 days of the school year.

At the time of its application, the NCDPI expected all assessments to be operational by the 2023-24 school year, which would require fully developing assessments for math and ELA in several grades (Exhibit A.6). However, the NCDPI's timeline for development and administration of the NC PAT assessments shifted in its first 2 years of implementation. Due to disruptions in testing caused by the pandemic, NCDPI was unable to administer operational grade 4 math and grade 7 ELA assessments in 2020-21 (Exhibit A.7). However, the NCDPI's 2020-21 annual performance report indicates that it still anticipates statewide implementation of all assessments by 2023-24.

Exhibit A.6. Planned development and administration of NC PAT, by year (as of the North Carolina IADA application)

Year	Planned Assessment Development	Planned Administration of an Operational Assessment
2019-20	Grade 4 Math Grade 7 ELA	None
2020-21	Grades 6 & 7 Math Grades 4, 6, & 7 ELA	Grade 4 Math Grade 7 ELA
2021-22	Grades 3,5, & 8 Math Grades 3,5, & 8 ELA	Grades 4, 6, & 7 Math Grades 4, 6, & 7 ELA
2022-23	Grades 3,5, & 8 Math Grades 3,5, & 8 ELA	Grades 4, 6, & 7 Math Grades 4, 6, & 7 ELA
2023-24	None	Grades 3-8 Math and ELA

SOURCE: North Carolina’s application materials.

Exhibit A.7. Actual development and administration of NC PAT, by year (as of 2020-21)

Year	Assessment Development	Actual Administration of an Operational Assessment
2019-20	Grade 4 Math Grade 7 ELA	None
2020-21	Grade 4 Math Grade 7 ELA	None

SOURCE: North Carolina’s annual performance reports through 2020-21.

A.2.3.d. Assessment System Partners and Their Roles

The NCDPI partnered with the North Carolina State University (NCSU) to develop and maintain the online platform for NC PAT assessments. The NCDPI and the NCSU had previously partnered on the development of NC Check-Ins. The NCDPI has also engaged with other universities through their technical advisory board, which includes experts from: University of North Carolina at Chapel Hill, University of North Carolina at Charlotte, University of North Carolina at Greensboro, Howard University School of Education, and the Center for Assessment. The group plans to meet biennially to review the test development design and psychometric analyses and to ensure the innovative assessments meet technical standards.

A.2.3.e. Expectations of Staff

Once the NC PAT system is operational, staff will provide training to other staff as well as administer and use the assessments in classrooms. The NCDPI intends to use a “train the trainer” model, with expert consultants training regional coordinators, who train LEA coordinators, who then train school-level testing coordinators to train teachers. Teachers must administer the assessments and use the scores in their instructional planning. Trainings will be delivered through a combination of webinars and in-person workshops, and will be designed to address assessment

literacy (e.g., NC PAT’s purpose and design), data literacy (including how to read and use items in the NC PAT data reports), and test administration, such as accommodations for students with disabilities and English language learners.

A.2.3.f. District Involvement

In 2019-20, 33 school districts were involved in NC PAT (including 18 regular public school districts, 13 charter school districts, and 2 other school districts) out of 311 eligible districts (Exhibit A.8). In 2020-21, 24 out of 315 were involved (including 14 regular public school districts, 8 charter school districts, and 2 other districts). While district participation in the NC PAT pilot is voluntary, there are state laws in place that could require districts to participate, if necessary. However, the NCDPI’s preference is for participants to opt into the pilot. Thus far, it has maintained an entirely volunteer sample.

Exhibit A.8. Number of NC PAT districts and all districts in North Carolina: 2019-20 and 2020-21

District type	NC PAT districts		All districts in state ^a	
	2019-20	2020-21	2019-20	2020-21
Total districts	33	24	311	315
Regular local school district ^b	18	14	120	120
Independent charter school district ^c	13	8	187	190
Other school districts ^d	2	2	4	5

^a Excludes districts that do not offer grades 3-8, which are the grades eligible for NC PAT.

^b “Regular local school district” is an agency responsible for providing free public education for school-age children residing within its jurisdiction. This category excludes local supervisory unions that provide management services for a group of associated school districts; regional education service agencies that typically provide school districts with research, testing, and data processing services; state and federally operated school districts; and other agencies that do not fall into these groupings.

^c “Independent charter school district” is a school district that includes only charter schools, typically a single school that was authorized under the charter.

^d “Other school districts” include state or federal agencies providing elementary- and/or secondary-level instruction.

NOTES: The number of districts is the number that were involved in NC PAT activities, including workshops, interviews, cognitive labs, or communications.

SOURCE: North Carolina’s annual performance reports through 2020-21; National Center for Education Statistics, Common Core of Data.

A.2.4. Georgia MAP Assessment Partnership (GMAP) Through-Year Assessment

GMAP Through-Year Assessment AT A GLANCE	
First Year of Demonstration Authority	2019-20
Basic Assessment Structure	Interim assessments (3)
<ul style="list-style-type: none"> Assessment type (and number) Item type Scoring method 	Multiple-choice/multiple-select, ^a technology-enhanced, ^b and constructed-response ^c
	Machine-scored for multiple-choice/multiple-select; hand-scored for constructed-response
Assessment System Scope	ELA, math, and science
<ul style="list-style-type: none"> Content Grades 	<ul style="list-style-type: none"> Grades 3 through 8 (ELA, math) Grades 5 and 8 (science)
Test Administration	Fall, winter, spring
<ul style="list-style-type: none"> Timing Mode 	Computer administration (adaptive ^d within a test)
Tests Scores that Contribute to Accountability	All interim scores

^a Students select one or more than one correct response from a set of responses for multiple-choice/multiple-select items.

^b The 2020–21 APR suggests that the GMAP through-year assessment developers are considering technology-enhanced items. Technology-enhanced items are computer-administered and more complex than multiple-choice/multiple-select items. For example, students may drag and drop answers from a list, order items, or highlight text to identify a central idea.

^c The GMAP through-year assessment will likely include constructed-response items on the fall and winter interims for formative purposes. A constructed-response item requires a student to independently create a response, rather than select a response from a set of choice options. Examples of constructed-response items include fill-in-the-blank and short essay. If included, the formative constructed-response items will be scored by teachers for their own use.

^d An adaptive test provides a tailored assessment based on a student’s performance on prior assessment questions (or prior assessment scores).

SOURCE: Application materials and annual performance reports for Georgia through 2020–21.

A.2.4.a. Rationale

Georgia pursued IADA as part of state policymakers’ desire to maximize flexibility for state and local assessments under federal law and allow local districts to take the lead in developing innovative approaches to assessments.³³ The set of districts associated with the Georgia MAP Assessment Partnership (GMAP) sought to develop an assessment system that challenges students to develop higher order thinking skills, leads to reduced testing time, provides information to guide instruction throughout the school year, and allows students to monitor their own learning. The resulting GMAP through-year assessment system is one of two IADA assessment systems in Georgia.³⁴

A.2.4.b. Related Assessment Work Prior to IADA

Before IADA, the GMAP consortium districts independently sought out and contracted with NWEA to use the testing organization's MAP Growth interim assessment system. However, these districts did not have experience developing or designing an assessment system that would meet ESEA requirements. Through IADA, these districts wanted to build upon the design, functionality, and reporting features of the MAP Growth assessments to develop a new assessment also suitable for accountability purposes.³⁵

A.2.4.c. Description of the IADA Assessments and Timeline Through 2020-21

The GMAP through-year assessments will consist of three interim assessments, administered in the fall, winter, and spring. The assessments will include items from the existing MAP Growth interim assessment items and new items being written by Georgia classroom teachers and curriculum personnel to ensure full coverage of Georgia's academic standards. The interim assessment items will be machine scorable (e.g., multiple-choice/multiple-select) and adaptive based on individual student performance.³⁶ They will be administered via a computer-based testing platform. Teachers and students will be able to access their results within a short time after the assessment window closes. The system plans to include constructed-response items in the fall and winter interims that are intended to assess higher-order thinking skills.

The scores from the interim assessments are intended to help educators understand how well students are learning grade-level content throughout each year and across years and help students monitor their progress toward proficiency. The scores will be aggregated into a single summative scale score, thus eliminating the need for a separate summative assessment. The summative scale score will classify students into one of four proficiency categories.

According to its IADA application, the GMAP through-year assessment system is expected to be used as an operational assessment in 2023-24, replacing the regular statewide assessment in grades 3 through 8 (ELA and math) and grades 5 and 8 (science) in participating districts (Exhibit A.9). As of 2020-21, the state reported that no districts administered the GMAP through-year assessment, which aligns with the application plans (Exhibit A.10).

Exhibit A.9. Planned development and administration of GMAP through-year assessments, by year (as of the Georgia application)

Year	Planned Development	Planned Administration of an Operational Assessment
2019-20	Develop GMAP through-year assessments	None
2020-21	Develop GMAP through-year assessments	None
2021-22	Develop GMAP through-year assessments	None
2022-23	Develop GMAP through-year assessments	None
2023-24	Continue development of GMAP through-year assessments for science ^a	ELA & math ^b

^a Science is expected to be administered as an operational assessment in school year 2024-25.

^b Georgia has an added requirement for its IADA systems to demonstrate comparability with the statewide assessment before the state allows the GMAP through-year assessment to replace the traditional statewide assessment for accountability purposes. As a result, the GMAP districts will have to administer the through-year assessments for at least a year to generate evidence of comparability before the state allows it to be an operational assessment under IADA in lieu of the regular statewide assessment.

SOURCE: Georgia’s application materials.

Exhibit A.10. Actual development and administration of GMAP through-year assessments, by year (as of 2020-21)

Year	Actual Development	Actual Administration of an Operational Assessment
2019-20	Develop GMAP through-year assessment	None
2020-21	Develop GMAP through-year assessment	None

SOURCE: Georgia’s annual performance reports through 2020-21.

A.2.4.d. Assessment System Partners and Their Roles

The GMAP through-year assessment is a joint effort between a consortium of districts (GMAP districts) already using NWEA’s MAP Growth assessment and NWEA. Together, NWEA and the Georgia Center for Assessment are expected to provide support and resources to design, develop, implement, and train for the administration of GMAP through-year assessment system. For instance, NWEA is overseeing all aspects of the implementation of the GMAP through-year assessment system, and the Georgia Center for Assessment is providing professional development in assessment literacy and learning progressions.

A.2.4.e. Expectations of Staff

Staff from GMAP districts and schools will be expected to serve on an advisory group that reviews assessment items for alignment with state standards, participate in annual professional development, and administer the assessments.

A.2.4.f. District Involvement

During the IADA demonstration period, GMAP districts voluntarily participate by administering the GMAP through-year assessments and will use resulting summative score from the interims for

accountability once Georgia approves the assessments for this purpose. In 2019-20, 14 school districts participated (including 13 regular public school districts and one charter school district), out of 214 eligible districts (Exhibit A.11). In 2020-21, 20 districts participated (including 19 regular public school districts and one charter district) out of 219 eligible districts.

Exhibit A.11. Number of GMAP districts and all districts in Georgia: 2019-20 and 2020-21

District type	GMAP districts		All districts in state ^a	
	2019-20	2020-21	2019-20	2020-21
Total districts	14	20	214	219
Regular local school district ^b	13	19	180	180
Independent charter school district ^c	1	1	30	36
Other school districts ^d	0	0	4	3

^a Excludes districts that do not offer grades 3-8, which are the grades eligible for GMAP.

^b “Regular local school district” is an agency responsible for providing free public education for school-age children residing within its jurisdiction. This category excludes local supervisory unions that provide management services for a group of associated school districts; regional education service agencies that typically provide school districts with research, testing, and data processing services; state and federally operated school districts; and other agencies that do not fall into these groupings.

^c “Independent charter school district” is a school district that includes only charter schools, typically a single school that was authorized under the charter.

^d “Other school districts” include state agencies providing elementary- and/or secondary-level instruction.

NOTES: The number of districts is the number that were considered members of the GMAP consortium.

SOURCE: Georgia’s annual performance reports through 2020-21; National Center for Education Statistics, Common Core of Data.

A.2.5. Georgia Navy

Georgia Navy AT A GLANCE	
First Year of Demonstration Authority	2019-20
Basic Assessment Structure	Interim (Multiple)
<ul style="list-style-type: none"> • Assessment type (and number) • Item type • Scoring method 	Multiple-choice/multiple-select ^a
	Machine-scored
Assessment System Scope	ELA, math, and science
<ul style="list-style-type: none"> • Content • Grades 	<ul style="list-style-type: none"> • Grades 3 through 8 (ELA, math) • Grades 5 and 8 (science) • High school <ul style="list-style-type: none"> – ELA and math, 2 courses for each content (e.g., English I and II) – Science - 1 course
Test Administration	Flexible, district-determined ^b
<ul style="list-style-type: none"> • Timing • Mode 	Computer administration (not adaptive ^c)
Tests Scores that Contribute to Accountability	All interim proficiency determinations ^d

^a Students select one or more than one correct response from a set of responses for multiple-choice/multiple-select items.

^b Districts administering Navy may allow schools or teachers to determine the calendar days during which an assessment may be administered, or “the assessment windows.”

^c An adaptive test provides a tailored assessment based on a student’s performance on prior assessment questions (or prior assessment scores).

^d A “proficiency determination” is an indication of whether the student has met a proficiency threshold (e.g., Proficient, Not Proficient).

SOURCE: Application materials and annual performance reports for Georgia through 2020-21.

A.2.5.a. Rationale

Georgia pursued IADA as part of state policymakers’ desire to maximize flexibility for state and local assessments under federal law and allow local districts to take the lead in developing innovative approaches to assessments.³⁷ The set of districts that formed the Putnam Consortium and use the Navy assessments (Navy) sought to create an assessment system that can be aligned to each district’s adopted curriculum or instructional pacing guides; challenges students to develop higher order thinking skills; leads to reduced testing time; and provides information in real time that can be used to guide instruction and allow students to monitor their learning throughout the school year. The Navy assessment system is one of two IADA assessment systems in Georgia.³⁸

A.2.5.b. Related Assessment Work Prior to IADA

Earlier versions of the Navy assessment were developed and used to monitor student progress and provide more timely data to teachers in a small set of districts during the 2017-18 school year.

Navvy subsequently built on the earlier versions by developing more items and expanding to more grades and courses.

A.2.5.c. Description of the IADA Assessments and Timeline Through 2020-21

The Navvy assessment system is a collection of short assessments that can be uniquely combined to create interim assessments that are administered over the course of the school year. Specifically, each short assessment addresses a single content standard, and individual districts choose the standards to assess during each assessment window and according to their instructional pacing plans. Districts can assess as few as one standard on an assessment for any particular administration; however, all Navvy districts are expected to assess students on the same standards by the end of the year. The Navvy assessment includes an “on-demand” feature which allows schools and teachers to open a window to allow for individual or small groups of students to make up the assessment (e.g., absence on day of administration) or retest a standard after remediation or individualized instruction.

Navvy assessments consist of multiple-choice/multiple-select items written by regional education service agency curriculum directors and current and former Georgia master classroom teachers. Assessments are administered via a computer-based testing platform. Score reports for individual students and classes are immediately available after the administration, and teachers and students can learn whether the student demonstrated proficiency for each standard assessed.

For each required standard, students can attempt to demonstrate proficiency up to three times in an academic year. Each student’s status (demonstrates proficiency/did not demonstrate proficiency) for each required standard is recorded and can be used to inform instruction, but only the status on the student’s final attempt is used as part of the student’s aggregated result for the annual determination of student academic achievement.

According to its IADA application, the Navvy districts planned to use the already piloted assessments during the demonstration period, but additional development work would be required before the assessments could be approved by the state and operational for accountability (Exhibit A.12). The Navvy system did not plan to administer an operational assessment until 2023-24.

There is a Navvy assessment prepared for every required grade (e.g., 3-8, and once in high school) for ELA and math; and one per grade band for science (e.g., grades 3-5, 6-8, once in high school). However, determining how to compute the annual summative determination and finalizing cut scores based on achievement level descriptors is a key focus of efforts under IADA.³⁹ As of 2020-21, the state did not report administering any operational assessments, as expected (Exhibit A.13).

Exhibit A.12. Planned development and administration of Navvy assessments, by year (as of the Georgia application)

Year	Planned Development	Planned Administration of an Operational Assessment
2019-20	Develop Navvy assessments	None
2020-21	Develop Navvy assessments	None
2021-22	Develop Navvy assessments	None
2022-23	Develop Navvy assessments	None
2023-24	Develop Navvy assessments	ELA and math, grades 3-8, and two courses in high school; Science, grades 5 and 8 ^a

^a Georgia has an added requirement for its IADA systems. It requires the Navvy assessment system to demonstrate comparability with the statewide assessment before the state allows the Navvy assessment to be used in lieu of the traditional statewide assessment for accountability purposes. As a result, the Navvy districts will have to administer the interim assessments for at least a year to generate evidence of comparability before the state allows it to be an operational assessment under IADA in lieu of the regular statewide assessment.

SOURCE: Georgia’s application materials.

Exhibit A.13. Actual development and administration of Navvy assessments, by year (as of 2020-21)

Year	Actual Development	Actual Administration of an Operational Assessment
2019-20	Develop Navvy assessments	None
2020-21	Develop Navvy assessments	None

SOURCE: Georgia’s annual performance reports through 2020-21.

A.2.5.d. Assessment System Partners and Their Roles

The consortium of districts administering the Navvy assessments (the Putnam Consortium) has three external partners advising the research and development or supporting the implementation of the Navvy assessment system: Navvy Education LLC, Institute for Performance Improvement, and the Center for Assessment. Navvy Education currently plays the primary role of developing the standards-based assessment system and developing and facilitating professional development on the implementation of the assessments. The Center for Assessment provides technical support for IADA compliance (e.g., comparability, reliability, validity). The 2020-21 APR did not report on the role of the Institute for Performance Improvement.

A.2.5.e. Expectations of Staff

Navvy district leaders are expected to participate in the Putnam Consortium Innovative Assessment Leadership Team, and district staff are expected to participate in professional learning and administer the Navvy assessments.

A.2.5.f. District Involvement

District involvement in Navvy is voluntary. In 2019-20, 15 school districts were involved (including 14 regular public school districts and one charter district), out of 217 eligible districts (Exhibit A.14). In 2020-21, 13 out of 223 were involved (including 11 regular public school districts and 2 charter districts).

Exhibit A.14. Number of Navy districts and all districts in Georgia: 2019-20 and 2020-21

District type	Navy districts		All districts in state ^a	
	2019-20	2020-21	2019-20	2020-21
Total districts	15	13	217	223
Regular local school district ^b	14	11	180	180
Independent charter school district ^c	1	2	33	39
Other school districts ^d	0	0	4	4

^a Excludes districts that do not offer grades 3-8 or high school, which are the grades eligible for Navy.

^b “Regular local school district” is an agency responsible for providing free public education for school-age children residing within its jurisdiction. This category excludes local supervisory unions that provide management services for a group of associated school districts; regional education service agencies that typically provide school districts with research, testing, and data processing services; state and federally operated school districts; and other agencies that do not fall into these groupings.

^c “Independent charter school district” is a school district that includes only charter schools, typically a single school that was authorized under the charter.

^d “Other school districts” include state agencies providing elementary- and/or secondary-level instruction.

NOTES: For Navy, the number of districts is the number using the Navy assessment during the 2019-20 and 2020-21 school years. However, these districts did not use the Navy assessments for accountability purposes.

SOURCE: Georgia’s annual performance reports through 2020-21; National Center for Education Statistics, Common Core of Data.

Appendix B. Study Methods

This appendix describes how data on the IADA systems were collected and analyzed for the report.

B.1. Data Sources and Collection Approach

The study team collected data from state application materials, APRs, and interviews with system officials.

B.1.1. State IADA Application Materials

The study team retrieved IADA application materials for each of the five IADA systems that are the focus of this report (Louisiana, New Hampshire, Georgia GMAP through-year assessment, Georgia Navy, and North Carolina) from the U.S. Department of Education's IADA program [website](#), including state applications and addenda and responses to U.S. Department of Education application reviewer comments. State applications must include:

- evidence of expert and stakeholder input into the design;
- how the state will meet the technical requirements of IADA assessments;⁴⁰
- the rationale for the IADA system;
- plans to develop strategies for scoring the innovative assessments and training evaluators in such strategies, if applicable;
- scale-up strategies, criteria for selection of initial participating schools and districts, and plans for high-quality and consistent implementation (if the system will be administered initially in a subset of schools or districts);
- prior experience developing and implementing the components of the IADA assessments;
- state and local capacity to implement the system, and extent and depth of state and local support for the application;
- intended timeline and budget;
- plans for supports for educators, students, and parents;
- plans for annual evaluation and continuous improvement activities;

- descriptions of each participating district and school, and assurances from the district that it will comply with all requirements (if the system will be administered initially in a subset of schools or districts); and
- a discussion of the governance structure and levels of affiliate participation for applications from a consortium of states.

In the application, the state also must assure that it will continue to use the traditional statewide assessments in all nonparticipating schools and in participating schools in required grades and subjects where the IADA assessment is not administered; that all students and each subgroup of students are held to the same standards; and that it will report annually to the U.S. Department of Education on their IADA assessment system (see section below on annual performance reports for the contents of these reports).⁴¹

B.1.2. State IADA Annual Performance Reports (APRs)

The U.S. Department of Education office responsible for the IADA program (the IADA program office) shared state-submitted IADA APRs for the 2018-19, 2019-20, and 2020-21 school years with the study team. The 2021-22 APRs were not available in time to be included in this report. New Hampshire and Louisiana were the only IADA participants in 2018-19 and are the only systems with APRs for that year.⁴² All five IADA systems submitted APRs for 2019-20 and 2020-21.⁴³

In the APRs, states report their performance or progress on a number of their IADA system activities. The APR form is composed of separate sections, each of which allows states to provide narrative descriptions of their progress or submit supporting evidence on the following required topics:

- their progress against their implementation timeline;
- the performance of students in participating schools on the IADA assessments;
- feedback from educators, parents, and other stakeholders on their satisfaction with the IADA system;
- if the system is not yet implemented statewide, school demographic information for participating schools and progress in scaling up to additional districts and schools;
- processes, procedures, or steps followed to develop a valid, reliable, and comparable innovative assessment system;
- updates on meeting technical requirements related to the assessments;
- trainings for educators and others to implement the system;

- activities to familiarize students, educators, and others with the new assessment system;
- educators' use of the IADA assessment data;
- efforts to notify parents of students in participating schools about the IADA assessments;
- changes in consortium governance membership, if applicable;
- assurances from participating districts that they will comply with requirements;
- changes in budget; and
- starting with the 2019-20 school year, how the COVID-19 pandemic may have affected the development and implementation of their IADA systems.

States updated their first APR submission for subsequent annual submissions. If they made no additional progress on a particular activity, that part of the submission tended to remain unchanged across submissions. For activities that were not yet started, the sections were left blank, included text stating only plans, or were reported as not applicable. States submitted an APR irrespective of whether they administered an assessment or made progress from the prior year.

Since the APR form largely consisted of sections for open-ended narrative descriptions, the level of detail states provided varied. To ensure that the information was as complete and consistent as possible, the study team asked follow-up questions of states, when needed, to give them an opportunity to supplement initial APR responses. States provided more information, when possible, including additional documentation such as technical and training manuals covering test administration, accommodations, test security, and scoring; alignment and validity study reports; assessment schedules; examples of district reports/report cards; and training and data use information.⁴⁴

Types of challenges addressed in interviews

- Building capacity needed to participate in the IADA program (including challenges encountered prior to joining the program)
 - Gaining support from school districts
 - Developing assessments that meet IADA program requirements (e.g., validity, reliability, comparability)
 - Developing training materials for districts, schools, and teachers
 - Developing or adapting student accommodations and accessibility features
-

B.1.3. Interviews with IADA System Officials

The study team conducted semi-structured group telephone interviews in November and December 2020 with officials identified by each state or district consortium as most familiar with

the development and early implementation of the IADA system. Participants included state or regional consortia assessment directors, psychometric experts, state academic officers, and representatives from partner organizations. Four of the five interviews included multiple participants, ranging from three to five participants per interview, with an average of four.

Two trained interviewers with experience in state or district assessment policy led the interviews, which lasted 60-90 minutes and were recorded and transcribed. Interview questions covered activities from the IADA application period through the 2019-20 school year. Interviewers followed a systematic protocol focusing on system development and implementation challenges and solutions (see call-out box).⁴⁵ Interviewers also asked participants to discuss pandemic-related challenges not reported in the 2019-20 APR. For the multi-participant interviews, all participants were invited (but not required) to respond to each question, including offering clarifications or additions to other participants' comments. Therefore, the number of participants responding varied from question to question (for example, because only some participants had the experience to speak to a particular issue). There were no instances where a system's participants gave conflicting answers to the same question, making it possible to have a single response to each question from each IADA system.

B.2. Analytic Approaches

B.2.1. Analyzing System Readiness at the Start of the IADA Demonstration Period

The readiness or initial status of each IADA system at the start of its demonstration period was coded based on application materials. Reviewers examined the materials to determine if states reported having already conducted six generally agreed-upon major assessment implementation activities.⁴⁶ These six high-level activities include initial planning, preparation for the first operational assessment, operational assessment administered, post-administration activities, continuous improvement activities, and scale-up. Systems were coded for whether they: (1) definitively reported conducting the activity in their IADA application materials (a solid circle in Exhibit 2 in the report); (2) reported conducting at least part of the activity (a hashed circle); or (3) did not report conducting the activity or that activity was not yet applicable (an open circle).⁴⁷ Examples of activities that were not complete because they were not yet applicable could include post-administration and continuous improvement activities if the system did not yet administer an operational version of their IADA assessment.

Two study team members independently reviewed the application materials to determine the appropriate status code for each activity. When there were disagreements regarding whether an activity occurred, the principal investigator helped make a final determination. Because application materials varied in their comprehensiveness and level of detail, and the six major activities do not directly correspond to information the states were required to provide in the application, these determinations should be interpreted with caution.

B.2.2. Analyzing System Status in 2020-21 and Progress Made

To assess the early progress made during the demonstration period, the study team first coded the status of each system as of the end of the 2020-21 school year. This status was based on all APRs submitted through 2020-21. Because the five systems began participating in the IADA at different times, their 2020-21 status reflects 2 to 3 years of demonstration experience and APRs. System status in 2020-21 was determined for the same six major activities in the readiness analysis described above, to facilitate comparison over time. However, to code system status in 2020-21, the study team defined a formal set of “indicators” for each of the activities and then coded each indicator in turn.⁴⁸ While it was not possible to similarly code system readiness at a finer level of detail using these indicators—given the limitations of the application materials—nonetheless, comparing the measures of where each system started and their status at the end of 2020-21 (that is, comparing Exhibit 3 to Exhibit 2) offers a preliminary estimate of “progress” under the IADA demonstration.

Indicators of System Status

The study team developed 50 indicators to determine system status on assessment implementation activities.⁴⁹ The indicators collectively describe the criteria for measuring whether the IADA system had performed an implementation activity or met an assessment requirement (yes versus no) (see Exhibit B.1). Each indicator specifies the kinds of evidence or documentation that states could provide related to the activity and one or more minimum thresholds, where each threshold pertains to a different aspect or component of the indicator. To be coded as a “yes” on an indicator, all minimum thresholds had to have been met.

Most indicators had only a single threshold, but some had as many as six. The number of thresholds is a function of the complexity of the indicator. For example, “item alignment [to the content standards]” is one indicator that has only a single threshold. To earn a “yes” on the indicator, the state had to provide documentation that the IADA item development procedures included instructions to align items to content standards. A more complex indicator is “bias review,” which is about safeguards against bias in item development. This indicator has three thresholds that needed to be met to earn a “yes” for the indicator. To meet the first threshold, the state had to provide documentation of the criteria for review for bias. To meet the second threshold, the state had to provide documentation of steps in the item development process where potential bias is considered and where items may be flagged, revised, or eliminated due to potential bias. To meet the third threshold, the state had to provide documentation that the IADA test items were reviewed for bias by an external review committee (i.e., individuals who did not develop the items). See Exhibit B.2 for the indicator coding sheet for the bias review indicator, which is offered as an illustrative example of how the indicators were coded.

Exhibit B.1. IADA status indicators and thresholds, by major assessment implementation activity and sub-activity

Major activity, ^a sub-activity, and indicator	Indicator threshold(s) that must be met
Initial Planning of Assessment	
Test and item requirements established	
Test specifications ^b	State has described the purpose and intended uses of the IADA assessment; the content, format, and length of the assessment; psychometric characteristics of the items (e.g., item difficulty); delivery mode; administration procedures; and software and hardware requirements, as appropriate
Item specifications ^b	State has described the content standards or targets to be assessed, item types and scoring requirements, intended level of item complexity needed to assess the content standards, intended difficulty of the items, accessibility tools and features, and format for responding to each item (e.g., multiple-choice, essay)
Other initial planning completed	
IADA system timeline established ^b	State has described key IADA activities, along with the year the activity was completed or is planned to occur
Familiarization ^b	State has confirmed that the state, district, or school provided information to parents and students and training to staff on the IADA assessment system’s purpose and major features
Prepare for the First Operational Assessment	
Initial items developed	
Item and pool selection ^b	State has (1) shown that items were selected to support the item specifications, (2) shown the item pool was selected to support the test blueprint, and (3) described the algorithm used to select items, if computer-adaptive, and shown that it covers the test blueprint
Item writer and reviewer qualifications and experience	State has confirmed that item writers and reviewers are qualified and have experience with the content area(s) and grade(s) for which the test items need to be written or reviewed, including experience with different student populations (e.g., students with disabilities, English learners)
Item writer and reviewer instructions	State has (1) described the steps to develop the items, including tools and features to ensure their accessibility to all students; (2) described procedures to help ensure that the items developed adhere to item specifications; and (3) described how item development instructions assure content accuracy (e.g., use of current references, review of item content accuracy by internal or external panels)
Universal design ^b	State has described how the totality of the items developed strives to give all students equal opportunity to access test items and show what they know

Exhibit B.1. IADA status indicators and thresholds, by major assessment implementation activity and sub-activity (continued)

Major activity, ^a sub-activity, and indicator	Indicator threshold(s) that must be met
Prepare for the First Operational Assessment—continued	
Initial items developed—continued	
Bias review	State has (1) bias review criteria, (2) shown that potential bias is considered when developing items and where items may be flagged for revision or deletion, and (3) confirmed that items are reviewed for bias by an external committee or individuals
Sensitivity ^b	State has (1) sensitivity review criteria, (2) rules for reviewing items for sensitivity and potential offensiveness, and (3) a list of sensitivity reviewers and their areas of expertise
Item alignment ^b	State has shown that item development procedures include instructions to align items to content standards
Blueprint alignment ^b	State has shown that all major content domains or strands assessed by the items are aligned to the test blueprint
Construct relevance	State has described procedures to ensure the intended content is assessed and that the need for additional knowledge to correctly answer items is minimized
Items piloted	
Intended response ^b	State has demonstrated that the test items produce the type of responses they were intended to produce (e.g., based on results from cognitive labs, interviews, or think-aloud sessions)
Appropriate testing accommodations ensured	
Accommodations and supports for students with disabilities ^b	State has (1) shown that staff are trained to administer the IADA assessment supports and accommodations to students with disabilities, and (2) described how the IADA assessment accommodations and supports (e.g., assistive technology devices) measure the academic achievement of students with disabilities
Linguistic accommodations and tools for English learners ^b	State has (1) shown that staff are trained to administer the IADA assessment linguistic accommodations and tools to English learners, and (2) described how the IADA assessment linguistic accommodations and tools measure the academic achievement of English learners
Assessment administration procedures developed	
Test administration ^b	State has described the policies and procedures to ensure the assessment is administered in a standardized manner
Test security ^b	State has described the policies and procedures for handling irregularities during test administration (e.g., fire drill, loss of electricity, cheating)

Exhibit B.1. IADA status indicators and thresholds, by major assessment implementation activity and sub-activity (continued)

Major activity, ^a sub-activity, and indicator	Indicator threshold(s) that must be met
Prepare for the First Operational Assessment—continued	
Assessment administration procedures developed—continued	
Process for generating individual student reports ^b	State has described the (1) timeline for releasing individual student assessment reports, (2) expectations for delivering student assessment reports to parents, (3) procedures for protecting the security of assessment-related personally identifiable information, and (4) expectations for releasing IADA assessment results within the same timeline as the “regular” statewide assessment results
Professional development for educators conducted	
Teacher capacity	State has shown that at least 75% of eligible teachers received training (1) that familiarized them with the innovative assessment system, (2) in test security (e.g., handling and distributing assessment materials) for the innovative assessment system, (3) to provide accommodations to students with disabilities, (4) to provide accommodations to English learners, and (5) to use innovative assessment data to inform instruction
Principal and other school leader capacity	State has shown that at least 75% of eligible principals and other school leaders received training (1) that familiarized them with the innovative assessment system, (2) in test security (e.g., monitoring assessment administration), (3) to provide accommodations to students with disabilities, (4) to provide accommodations to English learners, (5) to use innovative assessment data for accountability of all students, and (6) to use innovative assessment data for accountability across student subgroups
Operational Assessment Administered	
Full assessment(s) administered	State has shown that it administered all IADA testing components (e.g., interim and summative assessments) within a given grade and subject in participating schools
Annual determinations calculated	State has shown that it calculated annual determinations of academic proficiency using the IADA assessments for students in participating schools

Exhibit B.1. IADA status indicators and thresholds, by major assessment implementation activity and sub-activity (continued)

Major activity, ^a sub-activity, and indicator	Indicator threshold(s) that must be met
Post-administration Activities	
Technical quality of scores ensured	
Scoring following standardized and reliable procedures ^b	State has (1) described the protocols and procedures to ensure standardized human scoring of items, and (2) demonstrated that item machine-scoring is reliable and objective
Technical adequacy of items ^b	State has shown (1) the items are technically adequate (e.g., based on field test results), and (2) there is a review and follow up of any operational items that function differently (e.g., items that function differently for student subgroups)
Item scaling and equating ^b	State has described the (1) procedures to place scores from different tests or test forms on a common score scale, and (2) process to establish equivalent scores on two test forms
Form equivalence ^b	State has shown that scores from two different test forms that assess the same content and are administered under the same conditions can be used interchangeably
Overall test function ^b	State has shown that the assessment (1) can differentiate students at different levels of ability, and (2) is precise when differentiating among these ability levels
Overall reliability ^b	State has demonstrated that the assessment measures the content domain well and does so for all students over time, across items, and/or across different scorers
Subgroup reliability ^b	State has demonstrated that the assessment measures the content domain well and does so for different subgroups of students over time, across items, and/or across different scorers
Decision consistency and accuracy ^b	State has (1) demonstrated that the assessment results in accurate classifications of student achievement or proficiency (e.g., Basic, Proficient, Advanced), and (2) described the extent to which students were correctly classified
Across-years scaling and equating ^b	State has described the (1) procedures to place test scores across different years on a common score scale, and (2) process to establish equivalent scores on two test forms across different years
Assessment system comparability ^b	State has shown that IADA results are comparable to those from the non-IADA system
Same innovative assessment used for all students ^b	State has confirmed that the same innovative assessment system was used to measure the achievement of all students who participated in the IADA pilot

Exhibit B.1. IADA status indicators and thresholds, by major assessment implementation activity and sub-activity (continued)

Major activity, ^a sub-activity, and indicator	Indicator threshold(s) that must be met
Post-administration Activities—continued	
Reports produced	
Comprehensible ^b	State has shown that (1) simple language and uniform format are used to report the IADA results, and (2) translations are available and accessible to those who need them
Academic content ^b	State has shown that (1) state, district, and school reports are generated annually, and (2) these reports include student performance in terms of the content and knowledge assessed and academic descriptions of what students can and cannot do, along with supporting information to interpret the results
Student report ^b	State has shown that the annual assessment reports include an annual summative achievement determination, indicators of annual student progress, and indicators for identifying students not making progress
School report ^b	State has shown that the annual reports include annual summative achievement results for schools disaggregated by student subgroups
District and state reports ^b	State has shown that (1) a summative achievement of annual progress is generated for each district for all IADA pilot students, (2) a summative achievement of annual progress is generated for each district for important IADA pilot student subgroups, (3) a state summative achievement of annual progress is generated for all IADA pilot students, and (4) a state summative achievement of annual progress is generated for all important IADA pilot student subgroups
Similar participation rates for IADA and non-IADA statewide assessments	State has confirmed that the participation rate of eligible students in the IADA assessment is equal to or greater than the participation rate of eligible students for the “regular” (i.e., non-IADA) statewide assessment, overall and for key subgroups
Statewide representation ^b	State has (1) confirmed the IADA participation of new LEAs and districts, (2) described how additional schools or LEAs contributed to IADA implementation, (3) shown that academic achievement and participation data are reported for all IADA pilot students and important student subgroups at the state, LEA, and school levels, and (4) shown that school demographic and enrollment data are reported for IADA pilot student subgroups in participating schools and LEAs

Exhibit B.1. IADA status indicators and thresholds, by major assessment implementation activity and sub-activity (continued)

Major activity, ^a sub-activity, and indicator	Indicator threshold(s) that must be met
Post-administration Activities—continued	
Data from operational assessment used	
Teacher commitment	State has shown that the majority of participating teachers at participating schools use innovative assessment data to inform instruction
Principal and other school leader commitment	State has shown that (1) the majority of principals and other school leaders in participating schools use innovative assessment data for accountability of all students, and (2) the majority of principals and other school leaders in participating schools use innovative assessment data for accountability across student subgroups
Continuous Improvement Activities	
Feedback on the system obtained	
Feedback from any participating teacher(s)	State has shown that feedback was solicited from any teachers in any participating local education agency (LEA) regarding their satisfaction with the innovative assessment system
Feedback from any participating principal(s)/school leader(s)	State has shown that feedback was solicited from any principals or other school leaders in any participating LEA regarding their satisfaction with the innovative assessment system
Feedback from any participating parent(s)/caregiver(s)	State has shown that feedback was solicited from any parents or caregivers in any participating LEA regarding their satisfaction with the innovative assessment system
Feedback from any participating teacher(s) in each LEA	State has shown that feedback was solicited from at least one teacher in each participating LEA regarding their satisfaction with the innovative assessment system
Feedback from any participating principal(s)/school leader(s) in each LEA	State has shown that feedback was solicited from at least one principal or other school leader in each participating LEA regarding their satisfaction with the innovative assessment system
Feedback from any participating parent(s)/caregiver(s) in each LEA	State has shown that feedback was solicited from at least one parent or caregiver in each participating LEA regarding their satisfaction with the innovative assessment system
Evaluation	State has (1) shown there is an annual evaluation of the IADA assessment system, and (2) described how data, feedback, and evaluation results are used to improve the IADA assessment system

Exhibit B.1. IADA status indicators and thresholds, by major assessment implementation activity and sub-activity (continued)

Major activity, ^a sub-activity, and indicator	Indicator threshold(s) that must be met
Scale-Up Activities	
Progress toward scaling up the system made	State has shown that the number of participating districts (i.e., those that used the IADA assessment for accountability in participating schools) has increased from one year to the next

^a Indicators were grouped into six generally agreed-upon major assessment implementation activities by assessment experts. These activities were related to initial planning, preparation for the first operational assessment, operational assessment administered, post-administration activities, continuous improvement activities, and scale-up. Major activities were divided into sub-activities where appropriate. For example, the major activity “initial planning of assessment” had two sub-activities: “test and item requirements established” and “other initial planning completed.”

^b This indicator is considered a technical indicator. Technical indicators are those related to practices and procedures to ensure high-quality assessment scores.

Exhibit B.2. Review protocol for the bias review indicator

Context and Key Definition(s)

According to the *Standards for Educational and Psychological Testing*,^a “bias in tests and testing refers to construct-irrelevant [i.e., invalid] components that result in systematically lower or higher scores for identifiable groups of examinees” (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 2014, p. 76). Simply put, bias is the presence of some characteristic of an item that introduces construct-irrelevance components and results in differential performance for individuals of the same ability that come from different ethnic, gender, cultural, or religious groups.

Bias comes in many forms. An item may be biased if it contains content, contexts, or language that is differentially familiar to subgroups of examinees, or if the item structure or format is differentially difficult for subgroups of examinees with the same level of achievement in the content. An item may be language-biased if it uses terms that are not commonly used statewide or if it uses terms that have different connotations in different parts of the state or it uses language unnecessarily sophisticated for the measurement target.

Bias Review for Default States, Louisiana, New Hampshire, Georgia GMAP, Georgia Navy, and North Carolina

Three thresholds are required for the state to meet this indicator:^b

- State provides documentation of the criteria for review for bias.
- State provides documentation of steps in the item development process where potential bias is considered and where items may be flagged, revised, or eliminated due to potential bias.
- State provides documentation that the IADA test items were reviewed for bias by an external review committee (i.e., individuals who did not develop the items).

Likely Source(s) of Bias Review Data

- Louisiana APR, Technical Manual
- New Hampshire APR, Technical Manual
- Georgia GMAP Through-Year Assessment APR, Technical Manual
- Georgia Navy APR, Technical Manual
- North Carolina APR, Technical Manual

Bias Review Clarification/Information Needed (if applicable)

Exhibit B.2. Review protocol for the bias review indicator (continued)

Bias Review Rating (check one)

- State met indicator
- State did not meet indicator
- Indicator not applicable
- Clarification needed
- Information not available (temporary rating that becomes “state did not meet indicator” once state confirms the information is not available)

^a American Educational Research Association, American Psychological Association & National Council on Measurement in Education (2014). *The Standards for Educational and Psychological Testing*. Washington, DC: American Educational Research Association.

^b The threshold asks the state to only provide documentation because there is flexibility in the field, based on literature, research, and conventions, regarding appropriate processes that are acceptable evidence based on the assessment purpose, design, and item types regarding the specific steps in the item development process where potential bias must be considered or when an external committee must review items for bias. The U.S. Department of Education’s Assessment Peer Review Process (2018) indicates documentation of review by an external committee is an example of bias review evidence.

To determine what thresholds to include in an indicator, as well as what documentation/evidence to require from states, and how many thresholds must be met to earn a “yes” for an indicator, the study team referred to the minimum requirements in the legislation and regulations. However, in many cases, these sources did not describe explicit thresholds, and so the study team relied on best practices for assessment development^{50,51} and guidance from the U.S. Department of Education’s assessment peer review process.⁵²

Each of the indicators were grouped into the six generally agreed-upon major assessment implementation activities by assessment experts (see Exhibit B.1 above and Exhibit 3 in the report). Major activities were divided into sub-activities where appropriate. For example, the major activity “initial planning of assessment” had two sub-activities: “test and item requirements established” and “other initial planning completed.” The extent to which a system had completed each activity/sub-activity was determined by whether they had a “yes” for all, some, or none of the indicators associated with the activity/sub-activity.

The Review Protocol

The study team systematically reviewed the APR materials using a protocol that described each indicator, its associated threshold(s), and the evidence required to evaluate whether the IADA system met each threshold (see Exhibit B.2 for an illustrative example). An indicator was met if all thresholds were met. When a system could not provide evidence for any of an indicator’s threshold(s), that indicator was marked not applicable or NA. An NA rating was recorded, most often, because a system experienced a delay in activities due to the COVID-19 pandemic or because, for another reason, the system had not yet implemented the threshold activities (e.g., the

system had not yet administered an operational assessment). Ultimately, only thresholds with a rating of “yes” counted toward an indicator.

Reviewer Training

Reviewers received 10 hours of training, including:

- An independent “prework” session prior to the group training where each reviewer learned about the IADA program, IADA systems, the evaluation, and federal requirements.
- A training on the IADA program requirements, IADA systems, the review protocol, and the APR form. Reviewers received the review protocol, the APRs, a summary of the APR form and what the states are supposed to report on, and any additional materials submitted by states after requests for clarification or additional information. Reviewers learned how to consistently interpret threshold requirements on technical indicators.
- A practice activity, where reviewers independently coded state APR responses with anticipated review challenges.
- A final training event where trainers probed reviewers’ understanding of the indicators, thresholds, and the APR.

Conducting the Review

Two teams of trained researchers reviewed the APR materials. One team focused on indicators related to professional development, data use, obtaining feedback, and scale-up activities, and the other team (with expertise in student assessments) focused on the remaining technical indicators.

Researchers double-coded all nontechnical indicators with 100 percent agreement, and the ratings were then reviewed by a senior researcher with student assessment expertise.

All technical indicators were rated by at least one researcher. More complex indicators (those with more variation across states) were rated by two researchers. For example, indicators related to reliability (complex) were rated by two researchers because systems measured reliability differently. Only one researcher completed the review of the indicator for “bias review” because it only required evidence of the completion of a bias review. For quality control, the principal investigator randomly shadow-scored technical indicators for each IADA system. Ratings matched for approximately 75 percent of the double-coded indicators.

The principal investigator adjudicated discrepancies in the technical indicator ratings, facilitated by reviewers’ notes and references to the supporting documents. When reviewers were unable to find the needed information in the APRs, the study team requested additional information from states. In some instances, relevant information from the 2020-21 APRs was not available, in which case reviewers referenced earlier APRs. For New Hampshire, the 2020-21 ratings are derived from

the 2018-19 and 2019-20 APRs because New Hampshire received approval to pause their IADA timeline for 2020-21 and did not administer the NH PACE assessment.⁵³

Researchers flagged outliers (e.g., if four of the five systems received a “met” rating but one received a “NA”) and verified that the threshold requirements had been interpreted and applied consistently. This resulted in rating adjustments for three indicators. Researchers then finalized and summarized the ratings.

Reporting Indicator Findings

The study team summarized the indicator ratings for each major activity and sub-activity (Exhibit 3). A filled circle shows that the system met all indicators; a hashed circle shows the system met at least one indicator; and an open circle indicates the system did not meet any indicators. Exhibits C.1 through C.12 show ratings by IADA system. Recall that “progress” is the difference between where each IADA system started and their status in 2020-21, defined by comparing analogous activities in Exhibits 2 and 3. For example, an open circle in Exhibit 2 and a filled circle in Exhibit 3 for the same activity would suggest progress. Note that no expectations were set for system progress under IADA other than statewide implementation within 5 years. There are no interim benchmarks, and the IADA systems vary considerably in scope. Interpretations of progress should be made with caution.

B.2.3. Analyzing System Challenges

Information on challenges related to assessment development and early implementation was obtained from the 2019-20 and 2020-21 APRs as well as the fall 2020 interviews with system officials. Two study team members reviewed the section of the APRs where states were expected to report system implementation challenges related to the COVID-19 pandemic.

Following the interviews, the study team used qualitative analysis software to organize and code transcribed responses to six interview questions that focused on challenges encountered. The study team used an inductive approach to develop codes based on a review of the transcribed responses. Specifically, an analyst and the interview task lead each independently reviewed a set of transcribed responses to identify challenges raised during the interview and developed codes to describe the challenges (e.g., funding limitations, time limitations, obtaining buy-in from stakeholders). Then, the analyst and task lead grouped similar codes to develop themes. For example, challenges related to engaging with school districts or addressing district concerns about readiness fit within the theme of “district recruitment and onboarding.” The interview task lead developed a matrix summarizing the codes (with examples of associated interview responses) and themes and shared and discussed the matrix with the larger project team. With input from the project team, the task lead refined the summary matrix by consolidating the themes into categories of challenges (see Exhibit 4) and counts of IADA systems that reported challenges related to each of the broad categories.

Appendix C. Supplemental Information on Findings in the Report

This appendix provides more detailed findings than were summarized in the evaluation report. In the evaluation report, indicators were grouped into six generally agreed-upon major assessment implementation activities by assessment experts. These activities were related to initial planning, preparation for the first operational assessment, operational assessment administered, post-administration activities, continuous improvement activities, and scale-up. Major activities were further divided into sub-activities where appropriate, and then indicator results were aggregated and reported by major activity and/or sub-activity (see Exhibit 3). This appendix provides results that are disaggregated to the individual indicator level. Each of the 12 exhibits shows the indicator-level results for one major activity (and associated sub-activities) in a particular year. Exhibits C.1 to C.6 provide system implementation status in 2020-21, and Exhibits C.7 to C.12 provide the analogous status for 2019-20.

Exhibit C.1. 2020-21 Status of IADA assessment systems, by indicators of initial planning

Sub-activity and indicator	NH PACE	LEAP 2025 Humanities	NC PAT	GMAP Through- Year Assessment	Navy
Test and item requirements established					
Test specifications	Met	Met	Met	Not Met	Not Met
Item specifications	Met	Met	Met	Not Met	Not Met
Other initial planning					
IADA system timeline established	Met	Met	Met	Met	Met
Familiarization	Met	Met	Not Met	Met	Met

NOTE: See Appendix B, Exhibit B.1 for descriptions of what threshold(s) need to be met for each indicator. A Met rating indicates the state provided documentation to fully address the threshold(s) associated with the indicator. A Not Met rating indicates the state did not provide the documentation needed or provided documentation that only partially addressed the threshold(s) associated with the indicator. A Not Applicable (NA) rating indicates the state did not conduct the activity and thus could not yet provide the necessary documentation to address the threshold(s) associated with the indicator or the activity was not yet planned to be completed or it was delayed due to COVID-19.

SOURCE: Derived from the 2020-21 IADA annual performance reports (APRs) and additional materials for Louisiana, North Carolina, and Georgia. For New Hampshire, the ratings are derived from the 2018-19 and 2019-20 APRs because New Hampshire received approval to pause their IADA timeline for 2020-21 and did not administer the NH PACE assessment.

Exhibit C.2. 2020-21 Status of IADA assessment systems, by indicators of preparation for the first operational assessment

Sub-activity and indicator	NH PACE	LEAP 2025 Humanities	NC PAT	GMAP Through-Year Assessment	Navvy
Initial items developed					
Item pool and selection	Met	Met	Not Met	Not Met	Not Met
Item writer and reviewer qualifications and experience	Met	Met	Met	Met	Met
Item writer and reviewer instructions	Met	Met	Met	Met	Met
Universal design	Met	Met	Met	Met	Met
Bias review	Not Met	Met	Met	Met	Met
Sensitivity	Not Met	Met	Met	Met	Met
Item alignment	Not Met	Met	Met	Met	Met
Blueprint alignment	Met	Met	NA	Not Met	Not Met
Construct relevance	Met	Met	Met	Met	Met
Items piloted					
Intended response	Met	Not met	Met	Met	Met
Appropriate testing accommodations ensured					
Accommodations and supports for students with disabilities	Met	Met	Not Met	Not Met	Met
Linguistic accommodations and tools for English learners	Met	Met	Not Met	Not Met	Met
Assessment administration procedures developed					
Test administration	Met	Met	Not Met	Not Met	Met
Test security	NA	Met	Not Met	Not Met	Met
Process for generating individual student reports	Met	Met	Not Met	Not Met	Not Met
Professional development for educators conducted					
Teacher capacity	Not Met	Not Met	Not Met	Not Met	Not Met
Principal and other school leader capacity	Not Met	Not Met	Not Met	Not Met	Not Met

NOTE: See Appendix B, Exhibit B.1 for descriptions of what threshold(s) need to be met for each indicator. A Met rating indicates the state provided documentation to fully address the threshold(s) associated with the indicator. A Not Met rating indicates the state did not provide the documentation needed or provided documentation that only partially addressed the threshold(s) associated with the indicator. A Not Applicable (NA) rating indicates the state did not conduct the activity and thus could not yet provide the necessary documentation to address the threshold(s) associated with the indicator or the activity was not yet planned to be completed or it was delayed due to COVID-19.

SOURCE: Derived from the 2020-21 IADA annual performance reports (APRs) and additional materials for Louisiana, North Carolina, and Georgia. For New Hampshire, the ratings are derived from the 2018-19 and 2019-20 APRs because New Hampshire received approval to pause their IADA timeline for 2020-21 and did not administer the NH PACE assessment.

Exhibit C.3. 2020-21 Status of IADA assessment systems, by indicators of operational assessment administration

Indicator	NH PACE	LEAP 2025 Humanities	NC PAT	GMAP Through- Year Assessment	Navvy
Full assessment(s) administered	Met	Not Met	Not Met	Not Met	Not Met
Annual determinations calculated	Met	NA	NA	NA	NA

NOTE: See Appendix B, Exhibit B.1 for descriptions of what threshold(s) need to be met for each indicator. A Met rating indicates the state provided documentation to fully address the threshold(s) associated with the indicator. A Not Met rating indicates the state did not provide the documentation needed or provided documentation that only partially addressed the threshold(s) associated with the indicator. A Not Applicable (NA) rating indicates the state did not conduct the activity and thus could not yet provide the necessary documentation to address the threshold(s) associated with the indicator or the activity was not yet planned to be completed or it was delayed due to COVID-19.

SOURCE: Derived from the 2020-21 IADA annual performance reports (APRs) and additional materials for Louisiana, North Carolina, and Georgia. For New Hampshire, the ratings are derived from the 2018-19 and 2019-20 APRs because New Hampshire received approval to pause their IADA timeline for 2020-21 and did not administer the NH PACE assessment.

Exhibit C.4. 2020-21 Status of IADA assessment systems, by indicators of post-administration activities

Sub-activity and indicator	NH PACE	LEAP 2025 Humanities	NC PAT	GMAP Through- Year Assessment	Navy
Technical quality of scores ensured					
Scoring following standardized and reliable procedures ^a	Met	NA	NA	NA	NA
Technical adequacy of items ^a	Met	NA	NA	NA	NA
Item scaling and equating ^b	NA	NA	NA	Met	NA
Form equivalence ^a	NA	NA	NA	NA	NA
Overall test functions ^a	NA	NA	NA	NA	NA
Overall reliability	Not Met	NA	NA	NA	NA
Subgroup reliability	Not Met	NA	NA	NA	NA
Decision consistency and accuracy reliability	Met	NA	NA	NA	NA
Across-years scaling and equating ^c	Not Met	NA	NA	NA	NA
Assessment system comparability	Met	NA	NA	NA	NA
Same innovative assessment used for all students ^a	Met	NA	NA	NA	NA
Reports produced					
Comprehensible	Met	NA	NA	Met	NA
Academic content ^c	Met	NA	NA	NA	NA
Student report	Met	NA	NA	NA	NA
School report	Not Met	NA	NA	NA	NA
District and state reports	Not Met	NA	NA	NA	NA
Similar participation rates for IADA and non-IADA statewide assessments	NA	NA	NA	NA	NA
Statewide representation	Not Met	NA	NA	NA	NA
Data from operational assessment used					
Teacher commitment	NA	NA	NA	NA	NA
Principal and other school leader commitment	NA	NA	NA	NA	NA

^a This indicator is not applicable for LEAP 2025 Humanities, NC PAT, GMAP through-year assessments, and Navy because these systems did not implement an operational assessment in 2020-21.

^b Indicator not applicable for NH PACE because the NH PACE IADA assessment system administers one form of the relevant grade and subject Common Performance Tasks to all students in participating schools/districts. The GMAP through-year assessment system provided this information about NWEA’s Measures of Academic Progress (MAP) assessments.

^c Not applicable = The IADA summative assessments for GMAP through-year assessment system, Navy, LEAP 2025 Humanities, and North Carolina have not been administered over multiple years. NH PACE described their method for examining test results across years based on administrations prior to IADA—because their assessments and system were largely unchanged.

NOTE: See Appendix B, Exhibit B.1 for descriptions of what threshold(s) need to be met for each indicator. A Met rating indicates the state provided documentation to fully address the threshold(s) associated with the indicator. A Not Met rating indicates the state did not provide the documentation needed or provided documentation that only partially addressed the threshold(s) associated with the indicator. A Not Applicable (NA) rating indicates the state did not conduct the activity and thus could not yet provide the necessary documentation to address the threshold(s) associated with the indicator or the activity was not yet planned to be completed or it was delayed due to COVID-19.

SOURCE: Derived from the 2020-21 IADA annual performance reports (APRs) and additional materials for Louisiana, North Carolina, and Georgia. For New Hampshire, the ratings are derived from the 2018-19 and 2019-20 APRs because New Hampshire received approval to pause their IADA timeline for 2020-21 and did not administer the NH PACE assessment.

Exhibit C.5. 2020-21 Status of IADA assessment systems, by indicators of continuous improvement activities

Sub-activity and indicator	NH PACE	LEAP 2025 Humanities	NC PAT	GMAP Through- Year Assessment	Navy
Feedback on the system obtained					
Feedback from any participating teacher(s) ^a	Met	NA	NA	NA	NA
Feedback from any participating principal(s)/school leader(s) ^a	Met	NA	NA	NA	NA
Feedback from any participating parent(s)/caregiver(s) ^a	Met	NA	NA	NA	NA
Feedback from any participating teacher(s) in each local education agency (LEA) ^a	Met	NA	NA	NA	NA
Feedback from any participating principal(s)/school leader(s) in each LEA ^a	Met	NA	NA	NA	NA
Feedback from any participating parent(s)/caregiver(s) in each LEA ^a	Met	NA	NA	NA	NA
Feedback on the system obtained					
Evaluation	Met	NA	NA	NA	NA

^aThe soliciting feedback indicators are not applicable for LEAP 2025 Humanities, GMAP through-year assessment system, Navy, and North Carolina in the 2020-21 program year because these systems did not have an operational assessment in 2020-21.

NOTE: See Appendix B, Exhibit B.1 for descriptions of what threshold(s) need to be met for each indicator. A Met rating indicates the state provided documentation to fully address the threshold(s) associated with the indicator. A Not Met rating indicates the state did not provide the documentation needed or provided documentation that only partially addressed the threshold(s) associated with the indicator. A Not Applicable (NA) rating indicates the state did not conduct the activity and thus could not yet provide the necessary documentation to address the threshold(s) associated with the indicator or the activity was not yet planned to be completed or it was delayed due to COVID-19.

SOURCE: Derived from the 2020-21 IADA annual performance reports (APRs) and additional materials for Louisiana, North Carolina, and Georgia. For New Hampshire, the ratings are derived from the 2018-19 and 2019-20 APRs because New Hampshire received approval to pause their IADA timeline for 2020-21 and did not administer the NH PACE assessment.

Exhibit C.6. 2020-21 Status of IADA assessment systems, by indicators of system scale-up

Indicator	NH PACE	LEAP 2025 Humanities	NC PAT	GMAP Through- Year Assessment	Navy
Progress toward scaling up the system made	Not Met	NA	NA	NA	NA

NOTE: See Appendix B, Exhibit B.1 for descriptions of what threshold(s) need to be met for each indicator. This indicator is only applicable for NH PACE because it was the only system with participating districts that planned to administer an operational assessment in 2 years (2018-19 and 2019-20). A Met rating indicates the state provided documentation to fully address the threshold(s) associated with the indicator. A Not Met rating indicates the state did not provide the documentation needed or provided documentation that only partially addressed the threshold(s) associated with the indicator. A Not Applicable (NA) rating indicates the state did not conduct the activity and thus could not yet provide the necessary documentation to address the threshold(s) associated with the indicator or the activity was not yet planned to be completed or it was delayed due to COVID-19.

SOURCE: Derived from the 2020-21 IADA annual performance reports (APRs) and additional materials for Louisiana, North Carolina, and Georgia. For New Hampshire, the ratings are derived from the 2018-19 and 2019-20 APRs because New Hampshire received approval to pause their IADA timeline for 2020-21 and did not administer the NH PACE assessment.

Exhibit C.7. 2019-20 Status of IADA assessment systems, by indicators of initial planning

Sub-activity and indicator	NH PACE	LEAP 2025 Humanities	NC PAT	GMAP Through- Year Assessment	Navy
Test and item requirements established					
Test specifications	Met	Met	Met	Not Met	Not Met
Item specifications	Met	Not Met	Met	Not Met	Not Met
Other initial planning					
IADA system timeline established	Met	Met	Met	Met	Met
Familiarization	Met	Met	Not Met	Met	Met

NOTE: See Appendix B, Exhibit B.1 for descriptions of what threshold(s) need to be met for each indicator. A Met rating indicates the state provided documentation to fully address the threshold(s) associated with the indicator. A Not Met rating indicates the state did not provide the documentation needed or provided documentation that only partially addressed the threshold(s) associated with the indicator. A Not Applicable (NA) rating indicates the state did not conduct the activity and thus could not yet provide the necessary documentation to address the threshold(s) associated with the indicator or the activity was not yet planned to be completed or it was delayed due to COVID-19.

SOURCE: Derived from 2018-19 IADA annual performance reports (APRs) and additional materials for New Hampshire and Louisiana; and the 2019-20 IADA APRs and additional materials for New Hampshire, Louisiana, North Carolina, and Georgia.

Exhibit C.8. 2019-20 Status of IADA assessment systems, by indicators of preparation for the first operational assessment

Sub-activity and indicator	NH PACE	LEAP 2025 Humanities	NC PAT	GMAP Through- Year Assessment	Navy
Initial items developed					
Item pool and selection	Met	Not Met	Not Met	Not Met	Not Met
Item writer and reviewer qualifications and experience	Met	Not Met	Met	Met	Met
Item writer and reviewer instructions	Met	Not Met	Met	Met	Met
Universal design	Met	Met	Met	Met	Met
Bias review ^a	Not Met	Met	NA	Met	Met
Sensitivity ^a	Not Met	Met	NA	Met	Not Met
Item alignment	Not Met	Met	Met	Met	Met
Blueprint alignment ^a	Met	Not Met	NA	Not Met	Not Met
Construct relevance	Met	Not Met	Met	Met	Met
Items piloted					
Intended response ^a	Met	NA	NA	Met	Met
Appropriate testing accommodations ensured					
Accommodations and supports for students with disabilities ^b	Met	Met	Not Met	Not Met	Met
Linguistic accommodations and tools for English learners ^b	Met	Met	Not Met	Not Met	Met
Assessment administration procedures developed					
Test administration ^b	Met	Met	Not Met	Not Met	Met
Test security ^c	NA	Met	Not Met	Not Met	Met
Process for generating individual student reports	Met	Met	Not Met	Not Met	Not Met
Professional development for educators conducted					
Teacher capacity	Not Met	Not Met	Not Met	Not Met	Not Met
Principal and other school leader capacity	Not Met	Not Met	Not Met	Not Met	Not Met

^a Not applicable = Item development for all statewide assessments in North Carolina occurred through 2021.

^b Not applicable = State summative assessments were not administered in 2019-20 due to COVID-19, including the IADA summative assessments, so information about this indicator has been delayed or is not available. NH PACE and LEAP 2025 Humanities ratings are based on the 2018-19 operational and pilot tests, respectively.

^c Not applicable = This indicator is not applicable for the state’s IADA assessment, or the state’s plans did not yet indicate IADA assessment administration, or security policies were not yet sufficiently established for review.

NOTE: See Appendix B, Exhibit B.1 for descriptions of what threshold(s) need to be met for each indicator. A Met rating indicates the state provided documentation to fully address the threshold(s) associated with the indicator. A Not Met rating indicates the state did not provide the documentation needed or provided documentation that only partially addressed the threshold(s) associated with the indicator. A Not Applicable (NA) rating indicates the state did not conduct the activity and thus could not yet provide the necessary documentation to address the threshold(s) associated with the indicator or the activity was not yet planned to be completed or it was delayed due to COVID-19.

SOURCE: Derived from 2018-19 IADA annual performance reports (APRs) and additional materials for New Hampshire and Louisiana; and the 2019-20 IADA APRs and additional materials for New Hampshire, Louisiana, North Carolina, and Georgia.

Exhibit C.9. 2019-20 Status of IADA assessment systems, by indicators of operational assessment administration

Indicator	NH PACE	LEAP 2025 Humanities	NC PAT	GMAP Through- Year Assessment	Navvy
Full assessment(s) administered	Met	Not Met	Not Met	Not Met	Not Met
Annual determinations calculated	Met	NA	NA	NA	NA

NOTE: See Appendix B, Exhibit B.1 for descriptions of what threshold(s) need to be met for each indicator. A Met rating indicates the state provided documentation to fully address the threshold(s) associated with the indicator. A Not Met rating indicates the state did not provide the documentation needed or provided documentation that only partially addressed the threshold(s) associated with the indicator. A Not Applicable (NA) rating indicates the state did not conduct the activity and thus could not yet provide the necessary documentation to address the threshold(s) associated with the indicator or the activity was not yet planned to be completed or it was delayed due to COVID-19.

SOURCE: Derived from 2018-19 IADA annual performance reports (APRs) and additional materials for New Hampshire and Louisiana; and the 2019-20 IADA APRs and additional materials for New Hampshire, Louisiana, North Carolina, and Georgia.

Exhibit C.10. 2019-20 Status of IADA assessment systems, by indicators of post-administration activities

Sub-activity and indicator	NH PACE	LEAP 2025 Humanities	NC PAT	GMAP Through- Year Assessment	Navy
Technical quality of scores ensured					
Scoring following standardized and reliable procedures ^a	Met	NA	NA	NA	NA
Technical adequacy of items ^a	Met	NA	NA	NA	NA
Item scaling and equating ^b	NA	NA	NA	Met	NA
Form equivalence ^a	NA	NA	NA	NA	NA
Overall test function ^a	NA	NA	NA	NA	NA
Overall reliability ^c	Not Met	NA	NA	NA	NA
Subgroup reliability ^c	Not Met	NA	NA	NA	NA
Decision consistency and accuracy ^c	Met	NA	NA	NA	NA
Across-years scaling and equating ^d	Not Met	NA	NA	NA	NA
Assessment system comparability ^e	Met	NA	NA	NA	NA
Same innovative assessment used for all students ^a	Met	NA	NA	NA	NA
Reports produced					
Comprehensible ^f	Met	NA	NA	Met	NA
Academic content ^c	Met	NA	NA	NA	NA
Student report ^c	Met	NA	NA	NA	NA
School report ^c	Not Met	NA	NA	NA	NA
District and state reports ^c	Not Met	NA	NA	NA	NA
Similar participation rates for IADA and non-IADA statewide assessments ^g	NA	NA	NA	NA	NA
Statewide representation ^h	Not Met	NA	NA	NA	NA
Data from operational assessment used					
Teacher commitment ⁱ	NA	NA	NA	NA	NA
Principal and other school leader commitment ⁱ	NA	NA	NA	NA	NA

^a Not applicable = State summative assessments were not administered in 2019-20 due to COVID-19, including the IADA summative assessments, so information about this indicator has been delayed or is not available.

^b Not applicable = State summative assessments were not administered in 2019-20 due to COVID-19, including the IADA summative assessments. Indicator not applicable for NH PACE due to the nature of the NH PACE assessment system. The GMAP through-year assessment system provided this information about NWEA’s Measures of Academic Progress (MAP) assessments.

^c Not applicable = State summative assessments were not administered in 2019-20 due to COVID-19, including the IADA summative assessments. The NH PACE indicator rating is based on the 2018-19 annual performance report (APR).

^d Not applicable = The IADA summative assessments for the GMAP through-year assessment system, Navy, LEAP 2025 Humanities, and North Carolina have not been administered over multiple years. NH PACE described their method for examining test results across years based on administrations prior to IADA—because their assessments and system were largely unchanged.

^e NH PACE was the only system with operational data to compare to the non-IADA system.

^f Not applicable = This indicator is not applicable for the state’s IADA assessment, or the state’s plans did not yet indicate IADA assessment administration, or score reports were not yet sufficiently established for review.

^g This indicator is not applicable for the 2019-20 program year for all IADA pilot systems since the spring 2020 statewide assessments were canceled due to COVID-19. Thus, there are no participation rates from the regular statewide assessments to compare to the IADA participation rates for spring 2020.

^h Not applicable = The state’s plans for its IADA scale-up were delayed due to COVID-19.

Exhibit C.10. 2019-20 Status of IADA assessment systems, by indicators of post-administration activities (continued)

ⁱThe commitment indicators are not applicable for all IADA systems in 2019-20 because the full program year did not occur. There could be unevenness in the data collection timelines for monitoring use of the data across the states. That is, the indicator asks about use of the data for a variety of purposes over the course of the program year. Some states may have planned to collect information on data use at the end of the school year. Also, if we look only at use of the data over the partial year (e.g., after one interim assessment), the data would be substantially limited and not reflect the full use of the data as the state intended.

NOTE: See Appendix B, Exhibit B.1 for descriptions of what threshold(s) need to be met for each indicator. A Met rating indicates the state provided documentation to fully address the threshold(s) associated with the indicator. A Not Met rating indicates the state did not provide the documentation needed or provided documentation that only partially addressed the threshold(s) associated with the indicator. A Not Applicable (NA) rating indicates the state did not conduct the activity and thus could not yet provide the necessary documentation to address the threshold(s) associated with the indicator or the activity was not yet planned to be completed or it was delayed due to COVID-19.

SOURCE: Derived from 2018-19 IADA annual performance reports (APRs) and additional materials for New Hampshire and Louisiana; and the 2019-20 IADA APRs and additional materials for New Hampshire, Louisiana, North Carolina, and Georgia.

Exhibit C.11. 2019-20 Status of IADA assessment systems, by indicators of continuous improvement activities

Sub-activity and indicator	NH PACE	LEAP 2025 Humanities	NC PAT	GMAP Through- Year Assessment	Navy
Feedback on the system obtained					
Feedback from any participating teacher(s) ^a	Met	NA	NA	NA	NA
Feedback from any participating principal(s)/school leader(s) ^a	Met	NA	NA	NA	NA
Feedback from any participating parent(s)/caregiver(s) ^a	Met	NA	NA	NA	NA
Feedback from any participating teacher(s) in each local education agency (LEA) ^a	Met	NA	NA	NA	NA
Feedback from any participating principal(s)/school leader(s) in each LEA ^a	Met	NA	NA	NA	NA
Feedback from any participating parent(s)/caregiver(s) in each LEA ^a	Met	NA	NA	NA	NA
Evaluation ^b	Met	NA	NA	NA	NA

^aThe soliciting feedback indicators are not applicable for the GMAP through-year assessment system, Navy, and North Carolina in 2019-20, and Louisiana in 2018-19 or 2019-20, because the systems did not have an operational assessment in those years. For New Hampshire, we assume that the state’s process for soliciting feedback on the satisfaction of the system from any teacher, parent/school leader, or parent/caregiver in 2018-19 held for 2019-20.

^b State summative assessments were not administered in 2019-20 due to COVID-19, including the IADA summative assessments, so information about this indicator has been delayed or is not available. Indicator rated if evidence was provided, but not if state delayed collecting evidence because of COVID-19.

NOTE: See Appendix B, Exhibit B.1 for descriptions of what threshold(s) need to be met for each indicator. A Met rating indicates the state provided documentation to fully address the threshold(s) associated with the indicator. A Not Met rating indicates the state did not provide the documentation needed or provided documentation that only partially addressed the threshold(s) associated with the indicator. A Not Applicable (NA) rating indicates the state did not conduct the activity and thus could not yet provide the necessary documentation to address the threshold(s) associated with the indicator or the activity was not yet planned to be completed or it was delayed due to COVID-19.

SOURCE: Derived from 2018-19 IADA annual performance reports (APRs) and additional materials for New Hampshire and Louisiana; and the 2019-20 IADA APRs and additional materials for New Hampshire, Louisiana, North Carolina, and Georgia.

Exhibit C.12. 2019-20 Status of IADA assessment systems, by indicators of system scale-up

Indicator	NH PACE	LEAP 2025 Humanities	NC PAT	GMAP Through- Year Assessment	Navy
Progress toward scaling up the system made	Not Met	NA	NA	NA	NA

NOTE: See Appendix B, Exhibit B.1 for descriptions of what threshold(s) need to be met for each indicator. This indicator is only applicable for NH PACE because it was the only system with participating districts that planned to administer an operational assessment in 2 years (2018-19 and 2019-20). A Met rating indicates the state provided documentation to fully address the threshold(s) associated with the indicator. A Not Met rating indicates the state did not provide the documentation needed or provided documentation that only partially addressed the threshold(s) associated with the indicator. A Not Applicable (NA) rating indicates the state did not conduct the activity and thus could not yet provide the necessary documentation to address the threshold(s) associated with the indicator or the activity was not yet planned to be completed or it was delayed due to COVID-19.

SOURCE: Derived from 2018-19 IADA annual performance reports (APRs) and additional materials for New Hampshire and Louisiana, and the 2019-20 IADA APRs and additional materials for New Hampshire, Louisiana, North Carolina, and Georgia.

ENDNOTES

¹ New Hampshire formally withdrew from the IADA program in spring 2022, and Georgia withdrew in February 2023. Because this report is based on data and analyses through the 2020-21 school year, the withdrawals do not affect the findings. Readers should nevertheless keep in mind that as background, the report includes summaries of states' plans for these IADA assessment systems in 2022 and beyond, based on APRs and program applications from 2020-21 and earlier. These plans could, of course, have changed for any of the IADA systems but have definitely changed for New Hampshire and Georgia.

² The requirements discussed in Appendix A are comprehensive and are derived from legislation and regulations: (1) the Elementary and Secondary Education Act as reauthorized by the Every Student Succeeds Act (ESEA) (2015) and (2) Innovative Assessment Demonstration Authority (IADA) (2017). Multiple study team members conducted a close reading of the education legislation and final regulations of the IADA program to ensure that this report included a comprehensive list of requirements. Specifically, study team members examined legislative requirements in ESEA (2015) Sec. 1111 and Sec. 1204 and regulatory requirements in IADA (2017) §200.104-200.108.

³ See: Office of Elementary and Secondary Education (OESE) (2020), p. 6.

⁴ No more than seven states may be approved for the demonstration authority including those states participating in a consortium. Consortia may not include more than four states.

⁵ See: IADA (2017).

⁶ See: ESEA (2015).

⁷ States may pursue any assessment design that meets the requirements of the IADA discussed in Appendix A. ESEA (2015), Sec. 1204(a)(1) specifically identifies “competency-based assessments, instructionally embedded assessments, interim assessments, cumulative year-end assessments, or performance-based assessments that combine into an annual summative determination for a student, which may be administered through computer adaptive assessments...” The program regulations (IADA 2017, §200[b][3][ii]) add that states may also use “another innovative assessment design that meets the requirements under §200.105(b).”

⁸ Competency-based assessments allow students to demonstrate that they have learned the expected knowledge and skills (i.e., a competency) needed to progress to the next academic content, grade, or level. Instructionally embedded assessments are woven into instruction. Interim assessments are administered multiple times throughout the year. Cumulative year-end assessments evaluate whether students have learned what they were expected to learn at the end of the school year. Performance-based assessments allow students to demonstrate mastery of academic content standards (i.e., what they should know and be able to do) by performing a specific task or series of tasks.

⁹ See: Konstantopoulos et al. (2016) and Perie et al. (2007).

¹⁰ In a letter to Chief State School Officers, the U.S. Department of Education noted that “only those States that wish to use the innovative assessment in place of the statewide assessment, including for the purposes of accountability and reporting under Title I, Part A, in at least one school, require innovative assessment demonstration authority” (OESE, 2016a, p. 4). Moreover, the [Notice Inviting Applications](#) (NIA) defines a participating school as one in which the IADA assessment is administered “instead of or in addition to” the traditional statewide assessment and “where the results of the school’s students on the innovative assessment system are used by its State and LEA for the purposes of accountability and

reporting” (Applications for New Authorities, 2018, p. 46927). Though this suggests that the program office expected applicants to administer a fully operational assessment within their first year, the program office approved applications where the system timeline indicated that it would use its first year for assessment planning, design, or development work rather than administering an operational assessment for accountability purposes.

¹¹ See: ESEA (2015), Sec. 1204(a)(2) and ESEA (2015), Sec. 1204(g). If the IADA assessment system is not approved for statewide use by the end of the 2-year extension period, a state may apply to the Secretary for a 1-year waiver under the general waiver authority granted by ESEA (2015), Sec. 8401. The 1-year waiver gives states time to undergo the peer review process as described in IADA (2017), §200.107(b).

¹² Requirements related to consultation and communication with stakeholders are found in ESEA (2015), Sec. 1204(e)(2)(A)(v); IADA (2017), §200.105(a); and IADA (2017), §200.106(b). States must provide evidence that they have developed an IADA assessment system in collaboration with multiple stakeholders and must obtain feedback from stakeholders about the IADA system. States must consult with experts in planning, development, and implementation of assessment systems. States must also consult with stakeholders with specific interests prior to IADA approval including: (1) representatives for economically disadvantaged students, students from major racial and ethnic groups, students with disabilities, and English learners; (2) teachers, principals, and other school leaders; (3) representatives of Indian tribes in the state; (4) students and parents; and (5) civil rights organizations. State applications must include a plan for collaborating with external partners, including the success of prior efforts to develop innovative assessment items and the strategies used to mitigate risks to successful implementation. They must also include evidence of state and local support, demonstrated by signatures from superintendents, presidents of local school boards, local teacher organizations, and other affected stakeholders.

¹³ Requirements related to the development and use of IADA assessments are found in ESEA (2015), Sec. 111(b)(2)(B); ESEA (2015), Sec. 1204(a) and (e); and IADA (2017), §200.104-106. In general, states must propose IADA assessments that adhere to many of the same requirements as regular statewide assessments in the ESEA. States must demonstrate that IADA assessment systems meet these requirements:

- Be aligned with the depth and breadth of challenging state academic standards, report student attainment of those standards, and identify whether students are performing on grade level;
- Be used for intended purposes, objectively measure academic achievement, not evaluate or assess personal or family beliefs and attitudes, and not disclose personally identifiable information;
- Be of adequate technical quality for each intended purposes, which includes generating results that are valid, reliable, and comparable for all students and each subgroup of students;
- Involve multiple measures of student academic achievement, including measures that assess higher-order thinking and may include measures of student growth;
- Provide for (a) participation of all students, (b) appropriate accommodations for students with disabilities, and (c) inclusion of English learners;
- Be administered through single summative assessment or through multiple statewide interim assessments throughout the academic year that results in a single, summative score;
- Assess comparability between IADA assessments and the regular statewide assessment;

- Provide for assessments of reading or language arts for any student attending school in the U.S. for 3 or more consecutive school years;
- Produce individual student interpretative, descriptive, and diagnostic reports in understandable language and uniform format that are consistent with the academic standards;
- Identify students not making progress;
- Enable results to be disaggregated within each state, LEA, and school by (a) each major racial and ethnic group, (b) economically disadvantaged students compared to students not economically disadvantaged, (c) students with disabilities compared to students without disabilities, (d) English proficiency status, (e) gender, and (f) migrant status;
- Enable itemized score analyses to be produced and reported;
- Be developed using principles that ensure maximum accessibility for all students; Describe the distinct purpose of each innovative assessment and how it will promote high-quality instruction;
- Develop and use standardized methods for scoring, ensuring consistency with nationally recognized professional standards and providing evidence of the validity, reliability, and comparability of annual summative determinations of achievement;
- Describe prior experience in developing and implementing innovative assessment items, including supports and appropriate accommodations for English learners and students with disabilities; and
- Describe the state and local capacity to implement an innovative assessment system, considering factors such as technological infrastructure, state and local laws, dedicated staff, expertise, and resources.

The IADA legislation, however, does provide states with some flexibility regarding the development and administration of IADA assessments. Specifically, the legislation permits two exceptions to the technical requirements for assessments of ESEA. First, a state approved for the IADA program is not required to use the same assessment for all students (IADA [2017], §200.77[b][1][i]). That is, during the demonstration period, states with IADA authority may use a combination of IADA assessments and regular statewide assessments, which removes the need to “double test” students. Second, a state approved for IADA is not required to develop and administer IADA assessments in all grades and subjects required by the ESEA (e.g., ELA and math assessments in grades 3 through 8 and once in high school; science assessments at least once during grades 3 through 5, once during grades 6 through 8, and once in high school) (ESEA [2015], Sec. 1111[b][2][v]). Under IADA, states may develop IADA assessments in a subset of grades and subjects (e.g., math assessments for grades 6, 7, and 8).

¹⁴ Requirements related to supporting and monitoring IADA assessments are found in ESEA (2015), Sec. 1204(e)(2)(B) and (C); IADA (2017), §200.106(c)-(e); and IADA (2017), §200.107. Specifically, states are required to do the following:

- Provide support and training to participating LEA staff, including the provision of training on standardized scoring methods consistent with nationally recognized professional standards;
- Engage and support teachers in developing and scoring IADA assessments;
- Inform parents participating in LEAs about the innovative assessment system at the beginning of each school year and acclimate students to the innovative assessment system;

- Ensure students with the most significant cognitive disabilities are assessed with alternate assessments;
- Describe strategies to ensure all students receive appropriate support and accommodations and to ensure professional development for school staff on providing accommodations for English learners and students with disabilities;
- Describe strategies and safeguards to support teachers and other staff in designing, developing, and implementing locally developed or locally scored items, ensuring objectivity in scoring (if applicable);
- Develop and implement strategies to scale-up the innovative assessment system to statewide implementation, including a plan for (1) selecting new and additional participating schools and districts, (2) ensuring that high-quality and consistent implementation will take place across demographically diverse LEAs and schools, and (3) setting annual benchmarks for scale-up activities;
- Assess results each year of the IADA program;
- Report information related to the IADA system annually to the Secretary, including
 - Demographics of participating LEAs;
 - Performance of all participating students and each group of students;
 - Feedback about satisfaction with the innovative assessment system;
 - Activities that occur each year of the demonstration authority period;
 - The parties responsible for each activity;
 - How the consortium’s member states will implement activities at different paces and will implement interdependent activities (if applicable); and
 - Budget information, including the degree to which funding is contingent upon future appropriations at the state or local level or additional commitments from non-public sources.
- Evaluate the IADA program, including a description of whether the evaluation will be conducted by an independent third party and the likelihood that the evaluation will sufficiently determine the validity, reliability, and comparability to the statewide assessment system;
- Continuously improve the IADA program based on annual evaluation data and feedback from other sources;
- Report progress in scaling up the innovative assessment system, including
 - Ensuring demographic representativeness of participating schools; and
 - Securing an assurance from each participating district that it will comply with IADA requirements; and
- Monitor and ensure the quality of the IADA system, including
 - Ensuring all students in each subgroup of students receive instructional support;
 - Ensuring each LEA has technological infrastructure to implement the innovative assessment system;

- Holding all schools in participating LEAs accountable;
- Describing how the state will hold all students and each subgroup of students to the same high standard as other students in the state; and
- Describing how the inclusion of additional LEAs contributes to progress toward achieving high-quality and consistent implementation.

¹⁵ While states or state consortia may apply for the demonstration authority, as of December 2022, no consortia had applied for or received the authority. As such, the remainder of this section uses “states” in lieu of “states or state consortia.”

¹⁶ In 2020, the U.S. Department of Education made supporting IADA pilot implementation, preparation of an IADA application, or the development of innovative assessment item types and design approaches a specific funding priority that states could apply for under the Competitive Grants for State Assessments program. The Louisiana, Massachusetts, Nebraska, and Texas Departments of Education each received grants of approximately \$3 million, and the Hawaii Department of Education received a \$500,000 grant (Applications for New Awards, 2020; OESE, 2022).

¹⁷ Comparisons should only be made with caution since the nature, scope, and goals of each system are unique.

¹⁸ Readers interested in learning more about each IADA system can review the state application materials on this U.S. Department of Education page, Innovative Assessment Demonstration Authority: <https://oese.ed.gov/offices/office-of-formula-grants/school-support-and-accountability/iada/>.

¹⁹ Counts of involved districts are drawn from IADA system APRs. The systems reported districts involved in test administration as well as other activities such as providing input on assessment system design or training. Subsection f of the system profiles refers to district “involvement” rather than “participation” because under IADA statute, “participating” districts are those using the innovative assessments for accountability. In their APRs, states reported the number of districts involved in a broader range of activities that may or may not have used the assessment for accountability purposes.

²⁰ Competency-based learning or competency-based educational (CBE) systems expand requirements of what students must know and be able to do to be considered proficient in a content or course. CBE systems are intended to empower teachers and students to reconceptualize how and when students can demonstrate their knowledge and mastery of the competencies. Transitioning to a CBE system requires a thorough review of the existing curriculum, instructional pacing guides, materials, and resources; and planning of assessment activities beyond traditional ways.

²¹ Competencies are the expected knowledge and skills that are needed for students to progress to the next academic content, grade, or level. In New Hampshire, each district adopts course-specific competencies that detail specific behaviors (e.g., critical thinking, communication, reasoning), skills (e.g., using one or many skills in an applied setting), or content students must demonstrate to be considered proficient. Competencies are based on the state’s content and performance standards.

²² Higher-order thinking requires students to demonstrate understanding, analyze, evaluate, or otherwise use the academic content they have learned beyond recalling discrete facts or applying simple memorized algorithms.

²³ In addition to the academic skills identified in each state’s academic content standards, many other skills are considered important for students to graduate from high school ready for jobs or college. (See, for example, skills discussed throughout OESE, 2016b.) These skills (e.g., problem-solving,

communication, collaboration, and creativity) are commonly referred to as “21st century learning skills” (Partnership for 21st Century Learning, 2019).

²⁴ Performance tasks require that students demonstrate their learning by performing contextually relevant activities (e.g., evaluating the levels of pollution in a pond rather than simply answering questions about pollution).

²⁵ One example of a district-required competency is the following: “Students will demonstrate an understanding of single-variable linear equations by applying analytic methods to simplify, to perform arithmetic operations, and to solve problems.” This competency requires students to demonstrate a mastery of components present in 13 different state academic content standards.

²⁶ An operational assessment requires that students participate in all IADA testing components within a given grade and subject, and that IADA results can be used to determine whether they are proficient against the state’s academic standards.

²⁷ The different paths for participation are: Full Implementation (districts implement NH PACE in ELA, math, and science, and use it for federal accountability), Partial Implementation I (districts implement NH PACE in one or all grades, and use it for federal accountability), and Partial Implementation II (districts implement NH PACE in one content in one grade until internal capacity and processes demonstrate readiness for Partial Implementation I).

²⁸ Lower-level skills primarily require memorization or familiarization (e.g., matching vocabulary words with definitions).

²⁹ “Standards” refer to academic content standards developed by a state. ESEA (2015), Sec. 1111(b)(1)(C) requires states to develop and adopt standards in math, English language arts, science, and “any other subject determined by the State.”

³⁰ LEAP Humanities high school assessments (Humanities I and Humanities II) will be administered to students enrolled in English I and English II courses, respectively.

³¹ Technology-enhanced items are computer-administered items that allow for interaction. They are more complex than multiple-choice/multiple-select items. For example, some technology-enhanced items allow students to respond by dragging and dropping answers from a list, ordering items, or highlighting text to identify a central idea.

³² Test security includes processes and procedures to limit potential cheating and/or unauthorized access and use of test content (e.g., ensuring test items are not copied and posted on the internet).

³³ The Georgia legislature allowed for up to 10 Georgia districts or consortia of districts to develop and implement their own locally developed innovative assessment systems. Georgia’s application for IADA included three locally developed assessment systems, but only two systems (the GMAP through-year assessment and Navvy) were approved for IADA. In addition to pursuing IADA, Georgia policymakers were interested in pursuing maximum flexibility allowed under ESEA through the use of a locally selected, nationally recognized test for high school achievement, if comparability with current end-of-course assessments could be established.

³⁴ After the IADA demonstration period, the state of Georgia plans to select one model for statewide use based on an external evaluation of nationally recognized professional and technical standards required for assessment systems used for federal accountability.

³⁵ MAP Growth is a product of NWEA and was previously called Measures of Academic Progress, or MAP. The Georgia application did not specify for how long the GMAP districts have used the MAP Growth assessments before IADA.

³⁶ The GMAP through-year assessment will repeatedly assess the breadth of the content standards across the interim administrations. That is, the content is not tied explicitly to content recently taught. The depth (e.g., cognitive rigor, content, and context of the items that are associated with item difficulty) will vary with the student's ability level (Georgia Department of Education, 2019, p.11).

³⁷ The Georgia legislature allowed for up to 10 Georgia districts or consortia of districts to develop and implement their own locally developed innovative assessment systems. Georgia's application for IADA included three locally developed assessment systems, but only two systems (the GMAP through-year assessment and Navvy) were approved for IADA. In addition to pursuing IADA, Georgia policymakers were interested in pursuing maximum flexibility allowed under ESEA through the use of a locally selected, nationally recognized test for high school achievement, if comparability with current end-of-course assessments could be established.

³⁸ After the IADA demonstration period, the state of Georgia plans to select one model for statewide use based on an external evaluation of nationally recognized professional and technical standards required for assessment systems used for federal accountability.

³⁹ Georgia requires comparability with the statewide assessment. As a result, the consortium will have to administer their IADA assessments for at least a year to generate evidence of comparability before they can administer an operational assessment under IADA instead of the regular statewide assessment.

⁴⁰ Application requirements are specified in IADA (2017), §200.105. Among the requirements, applicants must discuss how their IADA assessment system does or will meet the following technical requirements: meet the same technical requirements for statewide assessments required under 1111(b)(2)(B) of ESEA (2015) except that the IADA assessment need not be the same assessment administered to all students in the state and need not be administered annually in each of the required grades and subjects as the statewide assessments; align with challenging state content standards; measure a student's academic proficiency and growth using items above or below the student's grade level; express student results or competencies consistent with the state standards; generate annual summative determinations that are valid, reliable, and comparable for all students and student subgroups to results from the traditional statewide assessments; plan to annually determine comparability with the statewide assessment in one of five specified ways; provide for the participation of all students; be accessible to all students; provide appropriate accommodations; annually measure in each participating school progress of at least 95 percent of all students and 95 percent of each student subgroup; generate an annual summative determination of achievement using the data from the IADA assessment for each student in participating schools that describes student mastery of standards or alternate standards for students with the most significant cognitive disabilities; provide disaggregated results by each student subgroup; and provide an unbiased, rational, and consistent determination of progress toward state's long-term goals for academic achievement under ESEA so the state may validly and reliably aggregate data from the system for meeting accountability and reporting requirements under ESEA. (See IADA, 2017, §200.105(b) for more details on these requirements.)

⁴¹ See IADA (2017), §200.105 for more details on these requirements.

⁴² The IADA program office directly shared the APRs with the study team as states submitted them. The APRs are now available on the IADA program [website](#).

⁴³ Georgia submitted one APR that included information on the GMAP through-year assessment system and the Navvy assessment system.

⁴⁴ There were fewer clarifying questions related to the APRs after the 2018-19 APR because the IADA program office incorporated additional details from the study team's initial clarifying questions into subsequent APR forms.

⁴⁵ Although challenges were a major focus, the interviews also addressed lessons from system leaders' IADA experience through the 2019-20 school year, including whether participating in IADA had helped develop the kind of assessment system the state or consortium wanted. Findings on these topics may be incorporated into the study's next report.

⁴⁶ The six activities were based on major topics included in *Operational Best Practices for Statewide Large-Scale Assessment Programs* (Council of Chief State School Officers & Association of Test Publishers, 2013).

⁴⁷ Whether a system completed an activity is based on state or district consortium reports of activities associated with their first assessment(s) planned for IADA. Some systems planned to stagger the administration of the full set of their IADA assessments, choosing to focus first on particular grades or subjects (e.g., grade 4 math and grade 7 ELA/reading) then add grades or subjects over the course of the demonstration period.

⁴⁸ The framework for the indicators was based on the five areas of progress identified in the ESEA for this IADA Progress Report: the extent to which (1) the state has solicited feedback from key stakeholders about their satisfaction with the IADA system; (2) educators have demonstrated a commitment and capacity to implement or continue to implement the IADA system; (3) the IADA system meets technical requirements for assessments; (4) the same innovative assessment was used to measure the achievement of all participating students; and (5) there were similar student participation rates in the IADA and traditional state assessments (ESEA 2015, Sec. 1204[c][2]). The indicators were then grouped into six major assessment development and administration activities for presentation: initial planning, preparation for the first operational assessment, operational assessment administered, post-administration activities, continuous improvement activities, and scale-up.

⁴⁹ Forty-eight of the 50 indicators fall into five areas identified in ESEA for this report (see prior endnote). Two indicators were added to: (1) determine if the system had administered an operational assessment, and (2) if the system had increased the number of participating districts. While the legislation does not explicitly require this report to assess progress on these two elements, the study team elected to include them because administering an operational assessment and scaling up during the demonstration period are key program expectations.

⁵⁰ See: Council of Chief State School Officers and the Association of Test Publishers (2013).

⁵¹ See: American Educational Research Association, American Psychological Association, and National Council on Measurement in Education (2014).

⁵² The study team used the U.S. Department of Education's 2018 assessment review process guidance to inform indicator development and their minimum thresholds and modified that guidance to account for the nature and development of the IADA systems (OESE, 2018).

⁵³ Although New Hampshire submitted an APR for 2020-21, it reported no NH PACE activities during that year. New Hampshire's ratings for 2020-21 therefore reflect information from the prior APRs.

REFERENCES

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
https://www.testingstandards.net/uploads/7/6/6/4/76643089/standards_2014edition.pdf
- Applications for New Authorities; Innovative Assessment Demonstration Authority, 83 F.R. 46924. (2018, September 17). <https://www.govinfo.gov/content/pkg/FR-2018-09-17/pdf/2018-20152.pdf>
- Applications for New Awards; Competitive Grants for State Assessments Program, 85 F.R. 25422. (2020, May 1). <https://www.govinfo.gov/content/pkg/FR-2020-05-01/pdf/2020-09336.pdf>
- Council of Chief State School Officers & Association of Test Publishers. (2013). *Operational best practices for statewide large-scale assessment programs*. Washington, DC: Author.
- Elementary and Secondary Education Act, 20 U.S.C. § 6301. (2015).
<https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf>
- Georgia Department of Education. (2019, May 31). *Georgia's application for the Innovative Assessment Demonstration Authority—Addendum*. U.S. Department of Education, Office of Elementary and Secondary Education. <https://www2.ed.gov/admins/lead/account/iada/gaiadaaddendum.pdf>
- Innovative Assessment Demonstration Authority, 34 C.F.R. § 200. (2017).
<https://www.ecfr.gov/current/title-34/subtitle-B/chapter-II/part-200/subpart-E/subject-group-ECFR9277b2b0db822d9>
- Konstantopoulus, S., Li, W., Miller, S.R., & van der Ploeg, A. (2016). Effects of interim assessments across the achievement distribution. *Educational and Psychological Measurement*, 76(4), 587-608.
- Office of Elementary and Secondary Education. (2016a, December 7). *Dear colleague letter regarding Every Student Succeeds Act (ESSA) and assessments* (OESE-2016-29). U.S. Department of Education. <https://oese.ed.gov/files/2020/07/essaassessmentdcltr1207.pdf>
- Office of Elementary and Secondary Education. (2016b, October). *Non-regulatory guidance student support and academic enrichment grants*. U.S. Department of Education.
<https://www2.ed.gov/policy/elsec/leg/essa/essassaegrantguid10212016.pdf>
- Office of Elementary and Secondary Education. (2018, September). *A state's guide to the U.S. Department of Education's assessment peer-review process*. U. S. Department of Education.
<https://www2.ed.gov/admins/lead/account/saa/assessmentpeerreview.pdf>
- Office of Elementary and Secondary Education. (2020). *Application package for Innovative Assessment Demonstration Authority*. U.S. Department of Education.
<https://www2.ed.gov/admins/lead/account/iada/iadaapplication2020.pdf>
- Office of Elementary and Secondary Education. (2022, August 26). *Awards: Competitive Grants for State Assessments*. U.S. Department of Education. <https://oese.ed.gov/offices/office-of-formula-grants/school-support-and-accountability/competitive-grants-for-state-assessments/awards/>

Partnership for 21st Century Learning. (2019). *Framework for 21st century learning*. Battelle for Kids.
https://static.battelleforkids.org/documents/p21/P21_Framework_Brief.pdf

Perie, M., Marion, S., Gong, B., & Wurtzel, J. (2007). *The role of interim assessments in a comprehensive assessment system*. The Aspen Institute.
<https://www.achieve.org/files/TheRoleofInterimAssessments.pdf>