STUDENT ASSESSMENTS AND ASSOCIATED GROWTH MODELS FOR TEACHER AND PRINCIPAL EVALUATION

FORM *C*

PUBLICLY AVAILABLE SERVICES SUMMARY

This form will be posted on the New York State Education Department's Web site and distributed through other means for all applications that are approved in conjunction with this RFQ to allow districts and BOCES to understand proposed offerings in advance of directly contacting Assessment Providers regarding potential further procurements.

Assessment Provider Information	
Name of Assessment Provider:	NWEA®
Assessment Provider Contact Information:	Jennifer Schreiber, Proposal Manager Email: proposals@nwea.org Phone: (503) 624-1951
Name of Assessment:	MAP® Growth™ English Language Arts (ELA), MAP Growth Science and MAP Growth Course-Specific Assessments
Nature of Assessment:	☒ ASSESSMENT FOR USE WITH STUDENT LEARNING OBJECTIVES WITH A TARGET SETTING MODEL; OR ☐ SUPPLEMENTAL ASSESSMENT WITH AN ASSOCIATED GROWTH MODEL: ☐ GAIN SCORE MODEL ☐ GROWTH-TO-PROFICIENCY MODEL ☐ STUDENT GROWTH PERCENTILES ☐ PROJECTION MODELS ☐ VALUE-ADDED MODELS ☐ OTHER:
What are the grade(s) for which the assessment can be used to generate a 0-20 APPR score?	MAP Growth Course-Specific Mathematics: K-12 MAP Growth Science: Grades 3–8 MAP Growth ELA (Reading and Language Usage) for high school. This is in addition to our currently approved MAP Growth ELA (Reading and Language Usage) for Grades K–10
What are the subject area(s) for which the assessment can be used to generate a 0-20 APPR score?	▪ MAP Growth Course-Specific Mathematics in Algebra 1, Algebra 2, and Geometry ▪ MAP Growth Science ▪ MAP Growth ELA (Reading and Language Usage)
What are the technology requirements associated with the assessment?	MAP Growth assessments are a fully hosted software-as-a-service solution housed on a single platform, the MAP Administration and Reporting Center (MARC). From the MARC, users can perform the following tasks: ▪ Manage user, student, organization, program, and test data ▪ View MAP Growth reports ▪ Create testing sessions and administer tests Although MAP Growth does not require any downloading or installation of software, NWEA recommends students use our secure testing browser. Our secure testing browser provides a secure testing environment by prohibiting access to the desktop, internet, and other applications and programs during test administration.

	The secure testing browser will prevent students from initiating other browser sessions and from having access to other content on the testing device, unless they exit the test. In the event students exit the test during test administration, they will need to log back in and have the proctor reactivate the test before proceeding. As an additional security feature, when a student logs back in after exiting the test, our adaptive test engine will present a new item in place of any item that was presented before the student exited the test. This alleviates potential item exposure the student had, both from a test security and test validity perspective. MAP Growth assessments can be administered with a variety of devices, operating systems, browsers, and apps. Computer equipment must meet the minimum requirements specified by the manufacturers of the operating system and browser or app in use. Supported devices and software are provided and regularly updated online at https://teach.mapnwea.org/impl/maphelp/Content/MAPSetup/Technology/SystemRequirements.htm. NWEA provides, at no additional cost, all necessary updates and enhancements to our assessments. Our assessment system typically has monthly releases of new features, bug fixes, and changes to improve system performance and stability.
Is the assessment available, either for free or through purchase, to other districts or BOCES in New York State?	☒ YES ☐ NO

Please provide an overview of the assessment for districts and BOCES. Please include:

- **A description of the assessment;**
- **A description of how the assessment is administered;**
- **A description of how scores are reported (include links to sample reports as appropriate);**
- **A description of how the Assessment Provider supports implementation of the assessment, including any technical assistance. (3 pages max)**

NWEA is pleased to offer our MAP Growth Science, Course-Specific Mathematics, and ELA (Reading and Language Usage) assessments to be included on the New York State Department of Education (NYSED) website as approved Teacher and Principal Evaluation assessments for New York districts and Board of Cooperative Educational Services (BOCES) organizations. MAP Growth assessments in reading, language usage, and mathematics are already included on this list for students in grades K–10, and this application is intended to add new content areas and expand the grade levels of our currently approved assessments through high school.

Along with our web-based, item-level adaptive interim assessments, we offer our experience, expertise, and research-based assessments to districts and BOCES in New York to empower educators to accurately measure student achievement and growth within the school year and across years. The precise data MAP Growth assessments provide will allow educators to make the kinds of immediate instructional decisions that can affect positive change for every student.

With over forty years of experience in providing adaptive assessments, we currently partner with over 10,000 education organizations worldwide to provide assessment solutions, reports, instructional resources, professional learning, and research services. These partnerships include state departments of education; school districts of various sizes; private schools and charter schools; foundations; international schools; and national education organizations such as the Bureau of Indian Education. Since we began partnering with education organizations, we have continued to expand globally, reaching over ten million students throughout the United States and in 140 foreign countries. MAP Growth assessments are used throughout New York in districts such as New York City Public Schools, Rochester City School District, Rockville Centre Union Free School District, and Mineola Public Schools, not only for teacher and principal evaluation under the Annual Professional Performance Review (APPR), but also to inform instruction and support student achievement and growth.

Overview of MAP Growth Assessments

Education organizations use MAP Growth for multiple purposes:

- As a universal screener in a response to intervention framework
- To measure student growth and inform instruction
- To differentiate and personalize learning
- To measure the performance of programs or core instruction
- To predict performance on summative and college readiness assessments

MAP Growth assessments measure what students know and inform educators and parents about what they are ready to learn next. By dynamically adjusting to each student's answers, the computer adaptive tests — calibrated to an equal-interval scale for each subject — create a personalized experience that accurately measures performance, regardless of demographics or test-taking ability.

MAP Growth reports transform data into timely insights. Teachers use the information to differentiate instruction and pinpoint needs of students or sub-groups of students. Higher-level reports give administrators the context to drive improvement across schools and districts.

MAP Growth Science

MAP Growth Science is a computer adaptive test that assesses achievement according to standards-aligned content. Scores from repeated administrations are used to measure growth. A single test covers: Life Sciences, Earth and Space Sciences, and Physical Sciences. MAP Growth Science can be administered three times per school year, with an optional summer administration.

MAP Growth Course-Specific Mathematics

Our Course-Specific Mathematics tests are computer adaptive, aligned to standards, and designed to assess course-specific mathematics content, measuring student growth over one academic year of instruction, from fall to spring. Scores provide one indicator of whether the student is ready to move to the next mathematics course. The Course-Specific Mathematics tests can be taken two to three times a year and are available in Algebra 1, Algebra 2, and Geometry.

MAP Growth ELA (Reading and Language Usage)

MAP Growth assessments in reading and language usage are computer adaptive, aligned to standards, and report results using a cross-grade vertical scale that assesses achievement according to standards-aligned content. MAP Growth in reading and language usage can be administered three times per school year, with an optional summer administration. Scores from repeated administrations are used to measure growth over time.

MAP Growth Assessment Administration

MAP Growth assessments can be administered to a whole class, a group of students, or one student at a time. The system is accessed from a single platform, which contains administration access, reports, and other data management and reporting tools. To access the system, educators enter their username and password in the login page. After educators log in, they will see the dashboard. From this dashboard, users can perform the following tasks, depending on assigned role: manage user, student, organization, program, and test data; view MAP Growth reports; and create testing sessions and administer tests

It is simple for users of all technical abilities to administer the assessments. Our partners' experience of the MAP Growth assessment system is that it is intuitive, easy to operate, and engaging for both students and educators. The MAP Growth system has a visually appealing interface that is simple to navigate, providing users with embedded, page-specific online help, guides, and tutorials on-demand. All of this supports our ability to keep educators' time invested in learning the application to a minimum, while maximizing the ability to obtain useful and actionable information from the data. The MAP Growth Test Administration Manual with details on how the tests are administered is provided as Attachment A.

Fast, Informative Scoring and Reporting

MAP Growth assessments are adaptive and employ a common scoring algorithm. During the assessment, a Bayesian scoring algorithm is used to inform item selection. This prevents the artificially dramatic fluctuations in student achievement at the beginning of the test that can occur with other scoring algorithms. In addition, we use classical item analysis statistics to monitor keys and to confirm students are being scored correctly.

Although the Bayesian scoring works well as a procedure for selecting items during test administration, Bayesian scores are not appropriate for the calculation of final student achievement scores. This is because Bayesian scoring uses information other than the student's responses to items (e.g., past performance) to calculate the achievement estimate. Since only the student's performance on the day he/she takes a test should be used to give the student's current score, a maximum-likelihood algorithm is used to calculate a student's actual score at the completion of the test.

NWEA uses the Rasch item response theory model to create its RIT scales. MAP Growth results, reported as RIT scores, relate directly to the curriculum scale in each subject area. There is one RIT scale for each subject. Using the RIT scale to report test results makes it possible to follow a student's educational growth from testing season to testing season and year to year. Each subject area has a unique alignment to the RIT scale, meaning that scores cannot be compared across subject areas. By using item response theory to create the scales and anchoring item difficulty estimates to them, the RIT scales are comparable from one set of items to another and from one set of examinees to another. This enables comparisons of the scores from different students, or from the same student at different times, even though different sets of test items are administered. This also allows longitudinal comparison of student performance.

MAP Growth Reports

Upon completion of a MAP Growth test, the assessment software calculates each student's score and immediately displays the score for the subject and instructional areas via the end-of-test screen. While student reports are available immediately after testing, student, classroom, school, and grade-level reports, as well as data exports, are updated every twenty-four hours. All reports are available online using the MARC and can be accessed from any location with an Internet connection. An overview of the reports available for MAP Growth assessments can be found here: https://www.nwea.org/resource-library/brochures/map-growth-reports-portfolio-4.

Implementation and Technical Support Services

In conjunction with our organizational mission — *Partnering to help all kids learn*® — NWEA is pleased to offer its support services to New York school districts and BOCES. We are committed to building and sustaining long-term relationships with our partners, which continue to deepen after implementation is complete. We provide each New York partner a team to support the use of MAP Growth assessments. This team consists of an account team; an Implementation Specialist through the first testing season; Professional Learning Consultants; and the ongoing assistance of our Partner Support team.

Implementation Services

NWEA will assign an Implementation Specialist to coordinate access to the MAP Growth assessments. The Implementation Specialist will support and guide New York school districts and BOCES staff members as they prepare to use the assessments and learn to access and retrieve reports. The Implementation Specialist will be available to answer questions throughout that time. The availability of an Implementation Specialist can be negotiated to extend beyond the first academic term, should stakeholders need or desire further implementation support.

Account Management

To build and sustain long-term relationships with our partners, we provide an account management team to support New York school districts and BOCES. Our account management team works closely with the school district or BOCES from initiation of the partnership, through enrollment and initial test administration, and then continues to grow and develop that partnership through deeper training and professional learning to support the effective use of our assessment solutions in support of school and district improvement. This team will help school and district leaders understand instructional use of assessment data, monitor student growth, evaluate school performance, and target instructional resources to improve learning.

Partner Support

Since our inception, the NWEA Partner Support team provides timely, knowledgeable, and courteous support to numerous partners across the country and throughout the world. We are proud of the relationships that we have built with our partners and are humbled by the positive feedback we have received.

New York school districts and BOCES may reach Partner Support via a toll-free telephone number, email, and our chat support platform. Our representatives take ownership of support requests, triage accordingly, and monitor from inception to resolution. Our first responders resolve over 90 percent of support requests during their initial contact with the partner.

Our Partner Support team is available by phone and email from 7:00 a.m. to 5:00 p.m. local time, Monday through Friday, excluding federally recognized holidays observed by NWEA.

Help Center

Our assessments are online and easy to use for administrators and educators of all technical abilities, including staff members with limited technology experience. Your staff members also have ongoing personalized support from the Account Management team, described above. Our assessments also have embedded help resources and troubleshooting support built into the system. These materials include step-by-step training videos and guides for proctors, educators, and administrators. These resources are available online in the Help Center, an easy-to-navigate page that provides users with resources and training materials.

Helpdesk Staff

For larger districts and BOCES within New York, we propose working closely with them to train district helpdesk staff to serve as initial technical response. This has enabled deeper knowledge of our products within the districts and schools where it is used, while also facilitating the most rapid response and resolution for end users. Any unresolved calls can be escalated to NWEA for a second tier of support and resolution.

Please provide an overview of the student-level growth model or target setting model for SLOs for districts and BOCES, along with how student-level growth scores are aggregated to create teacher-level scores, and how those teacher-level scores are converted to New York State's 0-20 metric.

For this application, we are adding new content areas and expanding the grade levels of our currently approved MAP Growth assessments through high school. Consistent with the model used for our currently approved assessments to measure educator effectiveness based on student growth on MAP Growth assessments, NWEA is partnering with Education Analytics Inc. (EdAnalytics) to help New York school districts implement and calculate high-quality student learning objectives (SLOs). Under this approach, student growth — as measured from fall to spring each year — will be compared to the NWEA student growth norms. NWEA has growth norms for students in the subject areas of mathematics, reading, language usage, and science, as well as course-specific mathematics norms (Algebra 1, Algebra 2, and Geometry).

These nationally representative norms in mathematics, reading, language usage, and science provide information about mean levels of growth for students based on a student's grade, subject area, starting achievement level, and the number of instructional weeks between test events. These fall-to-spring growth norms will serve as the basis for measuring one year's worth of growth, with the standard error of growth taken into consideration when interpreting student growth relative to the growth norms.

For grades in which growth norms are available, the student growth norms will be available to all schools that use MAP Growth assessments, so educators will know at the start of the school year what the year of growth expectations are for each of their students. Educators will have access to growth projections for each student after fall testing; these growth projections are based on the student norms and further ensure that educators know early in the school year's instructional cycle what constitutes a year of growth for each student. EdAnalytics and NWEA will work with partner districts to help educators better understand the magnitude of their fall-to-spring growth goals, and will provide guidance around how to interpret their students' growth relative to the year of growth expectation and the role that the standard error of growth plays in that comparison.

Specifically, student growth goals — and, in turn, teacher SLOs — will be based on measured student growth compared to the student growth norms, factoring in the standard error of measure for the growth norms. **With that consideration in mind, a student whose growth is equal to or greater than his or her student-specific growth norm minus 1.6 standard errors of growth will be classified as meeting the goal of one year of growth.** The standard error band was included to ensure that students whose growth was not statistically different than expectations set via the growth norms (at a 90 percent probability level) would still be counted as meeting their growth goals. Taking the standard error of growth into consideration also provides better differentiation across HEDI categories than if student growth was simply compared to the growth norms (which would result in only approximately 30 percent of teachers receiving a rating of Developing or higher and the majority of teachers receiving an Ineffective rating).

Using this proposed SLO approach, EdAnalytics will calculate the percentage of an educator's students whose fall-to-spring growth meets or exceeds their growth goals. To ensure accurate linkage between students and educators, EdAnalytics will use official district Student Information Repository System (SIRS) data for SLO calculations. Participating districts can provide EdAnalytics with these data directly, or EdAnalytics can collect these data from a district's Region Information Centers/BOCES.

Based on the student-educator linkages from district SIRS data, EdAnalytics will convert the percentage of an educator's students who meet or exceed their growth goals onto New York State's 0-20 HEDI scale. These conversions will be based on the following proposed translation table:

Effectiveness Rating	Points Awarded	Percentage of Students Meeting Growth Targets
Highly Effective	20 points	97% to 100%
	19 points	93% to 96%
	18 points	90% to 92%
Effective	17 points	85% to 89%
	16 points	80% to 84%
	15 points	75% to 79%

Developing	14 points	67% to 74%
	13 points	60% to 66%
Ineffective	12 points	55% to 59%
	11 points	49% to 54%
	10 points	44% to 48%
	9 points	39% to 43%
	8 points	34% to 38%
	7 points	29% to 33%
	6 points	25% to 28%
	5 points	21% to 24%
	4 points	17% to 20%
	3 points	13% to 16%
	2 points	9% to 12%
	1 point	5% to 8%
	0 points	0% to 4%

This proposed translation was designed to result in a distribution of teachers to HEDI categories consistent with the distribution of teachers in the Ineffective and Developing categories under New York State's own growth model. Based on classroom-level data from a subset of New York districts from the 2014-2015 school year, 3.99 percent of classrooms within those districts would have received an Ineffective rating, while 11.56 percent, 42.35 percent, and 42.10 percent of classrooms would have received ratings of Developing, Effective, and Highly Effective, respectively.

At the conclusion of the school year, EdAnalytics will calculate the percentage of an educator's students whose growth meets or exceeds their growth goals. EdAnalytics will provide participating districts with a summary of this percentage for each educator evaluated using this SLO approach, and will convert these percentages to New York's 0-20 HEDI scale based on the aforementioned proposed translation table. EdAnalytics will only include students in an educator's SLO calculation if the students have valid test results from the fall and spring and the educator is the teacher of record for these students. The teacher of record will be determined based on educational best practices and in discussions with participating districts.

Additionally, students will not be included in an educator's SLO for the following reasons (based on data from a district's official SIRS data files):

- Teacher is not teaching the student in a course aligned to the tested area
- Teacher has an "instructional responsibility weight" of 0
- Student has been flagged with an "exclude from evaluation" indicator
- Teacher was only associated with a student outside of the period of time between pre-test and post-test

Student scores will also not be included in an educator's SLO if those scores have been invalidated for any reason (test scores are outside the valid range, the test duration was less than six minutes, the standard error of measurement is outside of acceptable limits, etc.).

While the NWEA student growth norms do take into account a student's starting achievement level, they do not explicitly control for student factors that may impact student growth that are outside of an educator's direct control, such as a student's socio-economic or English language learner status, or if a student is eligible for special education services, etc. NWEA and EdAnalytics are also partnering to provide value-added analyses as a student-level growth model option to any interested New York districts that use MAP Growth assessments. The MAP Growth Technical Manual is provided as Attachment B.

New York State Next Generation Assessment Priorities Please provide detail on how the proposed supplemental assessment or assessment to be used with SLOs addresses each of the Next Generation Assessment Priorities below.	
Characteristics of Good ELA and Math Assessments (only applicable to ELA and math assessments):	MAP Growth assessments are computer adaptive and have items of varied difficulty to cover the expected ability distribution and to support students with diverse needs, including students with disabilities, English language learners, and those performing above and below grade-level expectations. We adhere to the *Standards for Educational and Psychological Testing* in our item development processes. The MAP Growth item bank contains more than 42,000 items at varying difficulty levels, all calibrated to a point on our vertical RIT scale. The quality and depth of our item pools enable our assessments to measure every student precisely while meeting the requirements of test specifications. All NWEA items go through a rigorous item development and review process. Our process yields items with strong alignment to the breadth and depth of the standards. To achieve this, we have developed a deep understanding of the standards and use a variety of approaches and item types to assess them. Items are developed and reviewed through multiple lenses, including how they align to their targeted standard and grade level, how they adhere to the principles of universal design, and whether they are free from potential bias and sensitivity issues. Items are initially reviewed for: ▪ Content alignment ▪ Instructional relevance ▪ Currency ▪ Face validity ▪ Item construction ▪ Bias, sensitivity, and fairness ▪ Permissions ▪ Plagiarism The item then undergoes a check by the item owner and content confirmation reviews. Two NWEA Content Specialists assign and review depth of knowledge (DOK) level and other content-related metadata to the item, and perform a thorough and rigorous content and alignment review to confirm the item content: ▪ Is instructionally relevant ▪ Has clear face validity ▪ Is free of sensitivity and fairness issues ▪ Is sound in terms of item construction ▪ Is grade-level-appropriate

	After the content reviews, each item goes through an item publishing quality review where NWEA Content Production Specialists perform: ▪ A thorough copy edit to check syntax, grammar usage, spelling, and punctuation ▪ A comprehensive display review to confirm the item will properly display ▪ A formatting and style review for item consistency ▪ An accessibility review to add alternative descriptive text to item elements ▪ A final bias, sensitivity, and fairness review Upon completion of the item reviews, items are published and field-tested. Student response data from field-test items, which are presented within a set of calibrated items, are used to analyze the item's performance and to calibrate the difficulty estimate for each new item to the existing measurement scale. Successfully calibrated items are added to the operational item banks. It is important to note that only operational items contribute to student test scores. At any point during the reviews, an item can be sent back to a previous stage or rejected if it does not meet our strict standards. **Cognitive Demand and Depth of Knowledge** NWEA Content Specialists create item specifications, derived from unpacking the standards for the fullest understanding of the intention, scope, and focus of the standards. From this, we provide specific guidance to item writers regarding the content, context, and cognitive complexity of items that will assess a standard. The goal is to ensure that the MAP Growth assessments draw from a deep pool of items that span a full range of cognitive levels and skills. **Embedded Field Testing** Item development and field testing for MAP Growth assessments occurs continually to enhance and deepen the item pool. Field testing is required to maintain the item bank, as existing items are retired or removed due to changes in standards or item parameter drift. An embedded field-test design is used so that field-test items can be integrated seamlessly into the operational test. Often, in an embedded field-test design, field-test items are stored in a field-test item pool, and a predetermined number are administered to each examinee. When the assessment is scored, only the responses to the active items are used, so a student's score is not influenced by the presence of field-test items. This approach provides advantages, such as: ▪ Preserving testing mode ▪ Obtaining response data in an efficient manner ▪ Reducing the impact of motivation and representativeness concerns of the field test The assessment algorithm holds a fixed number of slots for field-test items for all students. Therefore, the sample of

	students taking the field test-items is representative of the student population. NWEA uses a procedure called “adaptive item calibration.” This procedure essentially divides item calibration into two stages: ▪ The first stage administers field-test items based on pre-specified criteria. ▪ The second stage estimates item parameters. Specifically, to start the field-testing process, each field-test item needs to have an initial item difficulty estimate, established by content experts based on item characteristics such as item length and complexity or difficulties of existing items with similar content. This initial item difficulty estimate is treated as if it represents the true item difficulty and is used to collect responses for calibration. Once a pre-specified number of responses is collected, a field-test item goes through a calibration process and its difficulty estimate is updated. The updated item difficulty estimate is then used to collect the next specified number of responses and its item difficulty estimate is updated again by all accumulated responses. This process is repeated in an iterative manner until either a pre-specified level of item parameter estimate precision is obtained or item parameter estimate is stabilized. This adaptive item calibration approach allows items to be presented to students that closely match their estimated achievement level. This helps optimize the use of testing time by presenting items that are neither too difficult nor too easy for a student. To make sure that the quality of the data is high, field-test items are administered only in the grade range suggested by the item author. Thus, the sample of students taking any field-test item is reflective of the sample of students who will be taking the item after it becomes active. A field-test item must have a minimum of 1,000 responses to be calibrated and become operational, enough to establish the stability and precision of item parameter estimates. This number is the minimum and is generally higher. The analysis of the field-test items includes distractor analysis and review against a series of statistical indices such as: ▪ Percent correct ▪ Point-biserial ▪ Mean squared error ▪ Z chi-square ▪ Rasch fit and unfit statistics ▪ Correlation between expected and actual percent correct score NWEA field-test items are calibrated with rigor due to our use of multiple statistical indices. **Item Bias and Sensitivity** We are committed to developing engaging, authentic, rigorous, and culturally diverse assessments that effectively measure

	the full range of the standards. Therefore, it is vital that we address a wide variety of texts in a balanced, respectful way that does not upset, distract, or exclude any student populations. Item writers employ careful consideration and sound judgment while crafting items, considering each item from a variety of angles regarding bias and sensitivity, in accordance with the NWEA Sensitivity, Fairness, and Accessibility Guidelines. A well-constructed item serves to activate and focus a student's thought process on the task presented in the item. To meet our high expectation of fairness to all students, every item is thoroughly examined at multiple points in the item development process, undergoing specific bias and sensitivity reviews. Sensitivity in this context means an awareness of the different things that can distract a student during assessment. Fairness in this context relates to giving each student equal opportunity to answer the item correctly based solely on their knowledge of the item content. Any sensitivity and fairness issues found in items are eliminated in revision or rejection of the item during development. Each item is evaluated against a set of criteria and is flagged if it: ▪ Requires prior knowledge other than the skill/concept being assessed ▪ Requires construct-irrelevant or specialized knowledge ▪ Has cultural bias ▪ Has linguistic bias ▪ Has socioeconomic bias ▪ Has religious bias ▪ Has geographic bias ▪ Has color-blind bias ▪ Has gender bias ▪ Favors students who have no visual impairments ▪ Favors students who have no disabilities ▪ Inappropriately employs idiomatic English ▪ Offensively stereotypes a group of people ▪ Mentions body/weight issues ▪ Contains inappropriate or sensitive topics ▪ Distracts, upsets, or confuses in any way ▪ Has other bias issues **Alignment to Standards** For a video explaining how we align items to specific state standards, please visit https://vimeo.com/166551954. (Please note that the video uses the terminology "goals" and "sub-goals," which we now refer to as the "instructional areas" and "sub-areas" of MAP Growth tests.) As part of the test development process, our Content Specialists review the standards and organize them into reporting frameworks for MAP Growth tests. This work consists of linking the grade-level expectations across grades

	to create a two-tier framework consisting of instructional areas and sub-areas. Creating tests in this manner means that they align tightly to standards and provide an accurate measure of student achievement.
Assessments Woven Tightly into the Curriculum:	NWEA believes that each and every student matters, and we offer assessments designed to help guide meaningful classroom instruction. MAP Growth assessments offer a personalized experience for students by adapting to each student's learning level – precisely measuring student progress and growth for each individual. Assessments are designed to be completed within a short amount of time (forty to sixty minutes per domain) and to provide teachers with valuable information about what each student knows and is ready to learn, which can be used to inform classroom instruction. MAP Growth assessments provide teachers with a means to measure the growth and progress of every student over time, regardless of whether a student is performing on, above, or below grade level. In addition, the assessments compare and predict student achievement and growth over time via NWEA achievement and growth norms. These data can be used by teachers to personalize instruction quickly for individual, small group, or class activities. Teachers can also use the data to support efforts to engage students in achieving personalized learning goals and progress through student and family goal-setting activities. MAP Growth provides summary reports that align student results to the applicable standards. Data in these reports are used to guide teacher-led instruction and help educators design lesson plans based on the readiness level of each student, groups of students, or an entire class. MAP Growth assessments include our interactive tool for teachers, the Learning Continuum. Teachers can use the Learning Continuum to filter by subject and grade level, then group students by standard. The Learning Continuum identifies skills and concepts each student is ready to learn by connecting RIT ranges to standards-aligned Learning Statements. This helps educators bridge the gap between MAP Growth test data and a district's chosen curriculum. Through the Learning Continuum reports, Test and Class Views supply global and student-specific information for tailoring instruction in which RIT scores are connected to skills and concepts students are ready to learn, helping to identify learning goals and targets to be shared with students and parents, as well as support efforts to create more personalized lesson plans.
Performance Assessment:	MAP Growth assessments engage students with rich content that covers DOK levels 1, 2, and 3 and includes technology-enhanced items. Results are provided quickly and accurately, with valuable data that can be used to inform, support, and teach. **Depth of Knowledge of Assessment Items** MAP Growth assessments draw from an item bank containing more than 42,000 items, including technology-enhanced items

and items at DOK levels 1, 2, and 3. Our high-quality item pools have been widely acknowledged to play a critical role in the measurement precision and efficiency of computer adaptive tests. The item pools cover all instructional areas and difficulty levels across the full range of the RIT scale.

The DOK framework, a system used to describe content complexity, was developed by Dr. Norman L. Webb in the late 1990s. DOK is now a widely accepted common language that teachers, administrators, curriculum and test developers, and others use to understand and apply content complexity.

MAP Growth test items span a full range of cognitive levels and skills, further supporting the alignment of the tests to the New York Learning Standards and Common Core State Standards. Each item in the pool is evaluated and tagged with a Bloom's cognitive process dimension and DOK level.

Variety of Item Types

MAP Growth items are presented in multiple response formats, such as multiselect, and include technology-enhanced items. Item types included in MAP Growth, MAP Growth Science and Course-Specific Mathematics assessments are:

- **Selection** (student selects answer option(s))
  - **Multiple Choice:** student chooses one response from multiple options.
  - **Multiple Select/Multiselect (Choice Multiple):** student chooses two or more responses from multiple options.
  - **Selectable Text (Hot Text):** students select a response from within a piece of text or a table of information (e.g., word, section of a passage, number, symbol, or equation).
- **Construction** (student constructs the response using provided options)
  - **Drag-and-Drop:** students select an option or options in an area called the toolbar and move or "drag" these options (e.g., words, phrases, symbols, numbers, or graphic elements) to designated containers on the screen.
  - **Click-and-Pop:** students move options (e.g., words, phrases, symbols, numbers, or graphic elements) from the area called the toolbar to designated container(s) on the screen by selecting an option; the option then "pops" into the container on screen.
- **Generation** (student generates the response with no answer options available)
  - **Text Entry (Short Constructed-Response):** Students use the keyboard to type their response directly onto the screen in response to a question or prompt.
- **Item Delivery Mechanism** (ways items are presented in addition to standalone)

- **Item Set:** students are presented with a set of items that all focus on a single passage or a narrowly defined topic. (Currently used only in MAP Growth Reading and Science. Not used in K–2.)
- **Composite Items:** students interact with multiple interaction types included within a single item.

Multidimensional Science Items

The MAP Growth Science assessments include multidimensional items that align to the dimensions of their particular standard. All items are rated for their degree of alignment to the three dimensions. Well-aligned items must address at least two of these dimensions in the context of a phenomenon or problem, rather than assessing isolated skills or knowledge. Some items assess all dimensions of the items' aligned performance expectations, and others assess different combinations of the dimensions.

More information about our content development process, including item types, is provided in our responses above.

Superior Precision of Adaptive MAP Growth Tests

MAP Growth assessment scoring is automatic and timely. Individual scores are available immediately in the end-of-test screen, which displays immediate score information including overall RIT score and instructional area scores. Classroom, school, and district reports are updated every twenty-four hours. We also provide information about measurement precision, in the form of the standard error of measure (SEM) of an individual score. Our adaptive assessments give educators data to establish baseline measures for student placement and to immediately identify and target instruction to individual student needs.

Item response theory models define a specific statistical relationship between test-takers and items. The item response theory framework has proven to be a powerful tool for test construction, scaling, equating, and score reporting. One important advantage of item response theory-based scores is that the precision of each score can be estimated using the SEM.

SEM can be used to judge the level of confidence that can be placed in a score — the lower the SEM, the greater the confidence. Using an item response theory model with adaptive item selection provides a powerful tool to reduce SEM and precisely measure students' ability and achievement. This translates into reliable measures of student growth, empowering teachers to make better-informed instructional decisions.

SEM is a function of the relationship among item parameters, the ability of the student, and the number of items administered. Measurement precision is best when students are given items closely matched to their abilities. Our adaptive

	test algorithm selects the best items for each student, producing a significantly lower SEM than fixed-form tests. Because MAP Growth assessments are computer adaptive and supported by high-quality item pools anchored to vertical scales, educators will see the following two major benefits: ▪ Tests that provide scores with similar precision across the achievement range ▪ Tests that provide superior precision over fixed-form tests
Efficient Time-Saving Assessments:	MAP Growth assessments are efficient, flexible, and easy to use, allowing all students — with or without accommodations — to test at various times and days during the test window. They were developed to be administered within a classroom setting, one-on-one, in small groups, or to an entire class. Not all students must test at the same time, giving educators the freedom to choose what works best in their classrooms, although all students participate in the assessments within a designated testing window. While group administration provides consistency and preserves instructional time and teacher resources, none of the administration options has adverse effects on the validity or reliability of MAP Growth tests. MAP Growth tests are intended to be completed in a class period. Because MAP Growth tests measure how much a student knows, not how much he or she can read and respond to in a given period of time, they are always untimed. Students receive items within their individual range of ability and have as much time as needed to work, so MAP Growth becomes part of the learning process rather than merely a test event. While they are untimed, each assessment takes, on average, forty to sixty minutes. MAP Growth assessments also provide flexibility for even shorter single administration times: The assessments can be paused at any point and resumed within thirty minutes, or suspended and resumed within twenty-eight days, allowing for a range of administration times and multiple test sessions. MAP Growth assessments produce multiple data points that give educators information for multiple purposes, providing an efficient way to view a student's full performance. The data points include: ▪ **Overall Scale Score (RIT Score):** A RIT (for Rasch Unit) score is an objective estimate of a student's overall achievement level in a subject. RIT scores are equal interval in nature, meaning that the distance between 150 RITs and 151 RITs is the same as the distance between 230 RITs and 231 RITs. ▪ **Observed or RIT Growth:** The growth in RIT points made between two terms in a growth comparison period. ▪ **Projected Growth:** The amount the student's RIT score is predicted to change, based on student growth norms. The student's initial score plus projected growth equals projected RIT. ▪ **Standard Error of Measurement (SEM):** The SEM is a measure of the precision of the RIT score. RIT scores that are more precise have a smaller SEM. MAP Growth reports for individual student performance routinely include

	the SEMs at the overall test level, at the instructional area level, as well as for change (growth). Score ranges are also included for all individual scores and are based on ±1 SEM. The SEM of a MAP Growth assessment is calculated using maximum-likelihood procedures. Although the SEM can theoretically range from zero to infinity, typical values for the test fall between 2.9 and 3.5 RIT points. Adaptive tests such as MAP Growth assessments minimize measurement error by using items of difficulty that best match a student's performance level. ▪ **RIT Range:** A range of RIT scores defined by the student's RIT score ±1 SEM. If the student took the test again relatively soon, one could expect his or her score to fall within this range about 68 percent of the time. ▪ **Instructional Areas:** The students' performance in the instructional area strands tested in a subject. Data will display either by instructional area strand RIT ranges or descriptors if students took a MAP Growth test. ▪ **National Percentage Rank (Percentile Rank):** The percentile rank is a normative statistic that indicates how well a student performed in comparison to students in the norm group. The most recent norm sample used a pool of approximately ten million students. A student's percentile rank indicates that the student scored as well as, or better than, the percent of students in the norm group. ▪ **Mean RIT:** The RIT score that indicates the average achievement status of a group. An individual student's RIT can be compared to the overall Mean RIT of a class, grade level, school, or district.
Technology:	Our technology requirements are described beginning on page 19 of this application. With computer adaptive tests such as MAP Growth, each student experiences a unique test based on his or her responses to each question. This adaptivity supports students with diverse needs, including students with disabilities, English language learners, and those performing outside of grade-level expectations. MAP Growth assessments require students to answer every question presented, instead of giving them the option to skip, so the difficulty level of the assessment is accurate and precise. The assessment begins by delivering a question based on known information about that student — grade level the first time tested, and previous score after that. If the student answers the question correctly, he or she receives a more difficult question. An incorrect response prompts an easier question. A MAP Growth test ends when the measurement precision or the maximum test length is reached. The algorithms used to deliver a unique test to each student are based upon Rasch item difficulty calibrations, where items are delivered so that students will likely respond correctly 50 percent of the time. As a result, students at all levels of learning stay engaged with our assessment. Struggling

	students who might otherwise get frustrated and stop trying and high-achieving students who might get bored by strictly grade-level assessments will remain interested, as subsequent questions adapt to their abilities. **Test Security** MAP Growth assessments includes protocols and tools designed to preserve test security, including forms of cheating. These protocols and tools — including computer adaptivity, item exposure control, high-quality item pools, and the secure testing browser — not only help preserve test security but also confirm the measurement accuracy and precision of MAP Growth. ▪ **Computer Adaptivity:** The adaptive test engine of MAP Growth assessments selects items based on an individual test-taker's ability. This gives each student a completely unique test and prevents cheating among students. ▪ **Item Exposure Control:** This procedure helps prevent items from being overused with MAP Growth tests. Test items that are exposed too often can create a security problem that can threaten test validity. Our item-exposure restrictions are among the most conservative in the field of adaptive testing. ▪ **Item Pools:** The unmatched, high-quality item pools that comprise MAP Growth assessments not only provide highly accurate and precise measurement, but they also prevent students from seeing repeated items, which would compromise the validity of scores. ▪ **Secure Testing Browser**: Our secure testing browser is designed to restrict access to the desktop and internet, providing a secure testing environment by prohibiting access to other applications and programs during test administration. The browser will prevent students from initiating other browser sessions and from having access to other content on the testing device, unless they exit the test. In instances when a student exits the test and has been allowed to log back in, as an additional security feature, our adaptive test engine will present a new item in place of any item that was presented before the student exited the test. This alleviates any item exposure the student had, both from a test security and test validity perspective. Additional benefits of the adaptive technology of the MAP Growth assessment system include: ▪ **Precise Data Faster**: Adaptive tests provide more accuracy in determining each student's achievement level using fewer items than a traditional fixed-form test, leading to shorter testing time. Upon completion of a MAP Growth test, the assessment software calculates each student's score and immediately displays the score for the subject and instructional areas via the end-of-test screen. ▪ **Flexible Reporting**: NWEA offers a robust suite of reports at the student, class, school, and district level in the MARC. The assessment software calculates each student's score and displays an overall RIT score via the end-of-test screen. MAP Growth reports and instructional

	resources are student-centric, research-based, and data-driven. The reports also provide data needed to inform instruction, evaluate programs, justify budget decisions, and help educators make key decisions. ▪ **Increased Student Confidence**: With adaptive testing, students gain confidence as they demonstrate what they are capable of doing without being bound by the restrictions inherent to a fixed-grade level instrument. ▪ **Broader Spectrum of Measurement**: Tests adapt to each student's instructional level independent of grade level, providing a greater depth of performance analysis. ▪ **Improved Security**: Each student develops a unique version of the test based on their performance, thereby reducing the likelihood that students may observe and use the answer of another student.
Degree to which the growth model must differentiate across New York State's four levels of teacher effectiveness (only applicable to supplemental assessments):	N/A; NWEA is not proposing a supplemental assessment with a corresponding growth model.

STUDENT ASSESSMENTS FOR TEACHER AND PRINCIPAL EVALUATION

FORM *H*

APPLICANT CERTIFICATION FORM

ASSESSMENTS FOR USE WITH STUDENT LEARNING OBJECTIVES

Please read each of the items below and check the corresponding box to ensure the fulfillment of the technical criteria.

PLEASE SUBMIT ONE "FORM H" FOR EACH APPLICANT. CO-APPLICANTS SHOULD SUBMIT SEPARATE FORMS.

The Applicant makes the following assurances:

Assurance	Check each box:
The assessment is rigorous, meaning that it is aligned to the New York State learning standards or, in instances where there are no such learning standards that apply to a subject/grade level, alignment to research-based learning standards.	☒
To the extent practicable, the assessment must be valid and reliable as defined by the Standards of Educational and Psychological Testing.	☒
The assessment can be used to measure one year's expected growth for individual students.	☒
For K-2 assessments, the assessment is not a "Traditional Standardized Assessment" as defined in Section 1.3 of this RFQ.	N/A ☐
For assessments previously used under Education Law §3012-c, the assessment results in differentiated student-level performance. If the assessment has not produced differentiated results in prior school years, the applicant assures that the lack of differentiation is justified by equivalently consistent student results based on other measures of student achievement.	N/A ☐
For assessments not previously used in teacher/principal evaluation, the applicant has a plan for collecting evidence of differentiated student results such that the evidence will be available by the end of each school year.	☒
At the end of each school year, the applicant will collect evidence demonstrating that the assessment has produced differentiated student-level results and will provide such evidence to the Department upon request.[3]	☒

[3] Please note, pursuant to Section 2.3 of this RFQ, an assessment may be removed from the approved list if such assessment does not comply with one or more of the criteria for approval set forth in this RFQ

To be completed by the Copyright Owner/Assessment Representative of the assessment being proposed and, where necessary, the co-applicant LEA:

NWEA 1. Name of Organization (PLEASE PRINT/TYPE)	4. Signature of Authorized Representative (PLEASE USE **BLUE** INK)
Geri Cohen 2. Name of Authorized Representative (PLEASE PRINT/TYPE)	6/5/19 5. Date Signed
Executive VP & CFO 3. Title of Authorized Representative (PLEASE PRINT/TYPE)	

1. Name of **LEA** (PLEASE PRINT/TYPE)	4. Signature of School Representative (PLEASE USE **BLUE** INK)
2. School Representative's Name (PLEASE PRINT/TYPE)	5. Date Signed
3. Title of School Representative (PLEASE PRINT/TYPE)	