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:	Data Recognition Corporation
Assessment Provider Contact Information:	Robert Starr, Assessment Solutions Consultant 631-757-6464 RStarr@datarecognitioncorp.com
Name of Assessment:	*TerraNova,* Common Core
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?	Grades 3 – 8
What are the subject area(s) for which the assessment can be used to generate a 0-20 APPR score?	Reading, English Language Arts, Mathematics
What are the technology requirements associated with the assessment?	*TerraNova* Common Core is an assessment administered through paper and pencil. DRC provides services for scanning and scoring of answer documents, including the handscoring of constructed-response items.
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)**

TERRANOVA COMMON CORE ASSESSMENT DESCRIPTION

At DRC, we know that results matter. We also know that there are proactive steps schools can take to help teachers provide more targeted instruction and help students raise their achievement levels. Educators, districts, dioceses, and individual schools rely on DRC year after year for their testing needs. DRC is a stable leader in the testing industry and has 35 years of experience in providing highly accurate and dependable results that translate into valuable, informed decision-making in the classroom. Great design, innovation, technical excellence, and curricula-based content are all reflected in *TerraNova* Common Core—one of the most respected, proven assessment solutions for today's educators.

TerraNova Common Core is norm-referenced assessment which measures student growth as a component of evaluating teacher effectiveness for grades 3–8 in Reading, English Language Arts, and Mathematics.

TerraNova Common Core offers 2011 norms and features item types that go beyond traditional multiple-choice tests. On the same scale and all in the same test book, the test includes:

- Integrated Performance Tasks (PT)
- Constructed-Response (CR) items
- Extended Constructed-Response (ECR) items
- Multiple-Choice

Using one test educators can compare student results on national and common core standards across grades and ability levels. All items have been developed and field-tested offering the most reliable and valid measure of any national achievement test in the market.

The Common Core State Standards provide a consistent, clear understanding of what students are expected to learn, so teachers and parents know what they need to do to help them. The standards are designed to be robust and relevant to the real world, reflecting the knowledge and skills that our young people need for success in college and careers. The Common Core promotes critical thinking, problem solving, and attention to 21st century skills. They are aligned to college and work expectations.

Assessments that measure these new standards must look and feel different from traditional multiple-choice assessments. Students must be provided opportunities to demonstrate their understanding of the new standards through writing, producing work, and explaining how they arrived at the answer.

Items measuring the Common Core can ask students to perform a series of tasks that measure multiple standards. Gone are the days when test items map to a only a single standard and a multiple-choice item is sufficient for measuring what level of mastery a student has on a specific standard. The depth, rigor, and complexity of the Common Core move beyond the measurement ability of a multiple-choice test alone.

TerraNova Common Core breaks the limitations of a multiple-choice test by offering unique item types such as CR, ECR, and PT items. No other national achievement test offers this combination of items all on the same scale, all in one test book.

ADMINISTERING TERRANOVA COMMON CORE

TerraNova Common Core is administered via paper and pencil format to students in grades 3–8 in Reading, English Language Arts, and Mathematics. Scoring options include handscoring or machine scoring. If DRC provides scoring services test results are available in seven days or less and reports are designed to show administrators, students, and teachers where they stand on both national and the Common Core standards today and over time. *TerraNova* Common Core is the only field tested, valid, and authentic measure of the Common Core available to districts today. DRC is the only publisher that uses 2-parameter IRT scoring to offer students partial credit for partial understanding. Objective and item level reports showcase student mastery on both national and Common Core standards. Report data can be accessed both online and in print for flexible report assembly. The Premier Home report includes links to helpful instructional activities and is available in 10 languages.

TERRANOVA SCORES ARE REPORTED

TerraNova Common Core has been designed for rapid scoring and reporting. You'll receive data back in 7 days or less and the data is immediately actionable:

- Objective and item level reports showcase student mastery on both national and College and Career Readiness standards
- Report data is available online and in traditional paper formats so you can assemble reports in the ways that work best for you as your team interprets the information and communicates with parents and other stakeholders
- Reports pinpoint where, within a comprehensive performance task, the student needs more work, by highlighting partial credit results
- Reports enable you to view all items by item type so you can see all parts of a performance task to gain deeper insight into what students know and where knowledge and skill gaps may remain

TerraNova Common Core includes the Premier Home Report—an easy-to-understand snapshot of how an individual student is progressing toward mastery on specific College and Career Readiness learning objectives. The report also includes links to instructional activities that will help parents and teachers individualize instruction.

The Premier Home Report offers dynamic translation in the top 10 languages, including Arabic, Cantonese, Filipino, French Creole, Hmong, Korean, Punjabi, Russian, Spanish, and Vietnamese. Prescribed individual activities can also be translated, helping teachers and parents help their children.

HOW DRC SUPPORTS IMPLEMENTATION AND PROVIDES TECHNICAL ASSISTANCE

DRC offers a full array of implementation and technical support services. Customer Support includes three tiers of customer service and technical support, which creates a timely start to the assessment program. Our team provides appropriate training for successful implementation of the *TerraNova* assessments. Another element of training involves working with educators on how to use the data from *TerraNova* to identify instructionally actionable information in order to

improve their classroom instruction and student learning. Sessions will provide participants with a strong foundation to interpret student test results and apply that information to instruction for students. Participants will review a variety of methods and exercises to be used in the classroom and will engage in large and small group activities, including the development of lessons while practicing differentiation across grade levels and content areas.

DRC has a dedicated team of resources for responding to any and all customer inquiries. Your team includes Assessment Solutions Consultant, Robert Starr, and National Sales Manager, John Reginald, who provide hands-on, consultative direction.

All New York LEA's will also have access to our Customer Care team with members from Customer Service, Scoring, Research, Technology, Publishing, and Product Management. Working together, this team ensures that every customer receives personal, helpful, and timely responses regardless of whether service is requested though the website, email, or through a phone call. For general support including scoring services representatives are available from 9:00 AM to 7:00 PM Eastern Time. For help with online/software products technical support staff are available from 7:30 AM to 5:00 PM Eastern time.

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 the create teacher-level scores, and how those teacher-level scores are converted to New York State's 0-20 metric.

Students' normative growth can be calculated using students' normative scores. Summary statistics on the students' normative growth at class level for each teacher can be calculated and compared. The following is an example how to use the national percentile rank scores for this purpose.

Each student taking *TerraNova* Common Core at two time points will receive two scale scores and two normative scores, national percentile ranks (NPR). The NPR represents how the student performed at each time point relative to his peers in the nation. Growth can be calculated either using the scale scores or the NPRs (for normative growth). For a class of students, the mean of the scale scores and the median national percentile rank of the class can be calculated and used as a summary measure of the class performance. The comparison of the two mean scale scores or the two median national percentile ranks provides information whether the class performance has improved between the two test administrations.

The following is a possible crosswalk, from the native growth scale (median difference of student national percentile ranks) to the 0-20 evaluation metric. This crosswalk is provided here as a demonstrating example only. The actual crosswalk will be conducted upon consultation with districts and will take into account information provided by studies such as the Contrasting Groups study described above and/or information of class sizes.

Difference of Median Percentile Metric	0-20 Evaluation Metric
<=(-24)	0
(-23)-(-22)	1
(-21)-(-19)	2
(-18)-(-17)	3
(-16)-(-14)	4
(-13)-(-12)	5
(-11)-(-9)	6
(-8)-(-7)	7
(-6)-(-4)	8
(-3)-(-2)	9
(-1)- 1	10
2- 3	11
4- 6	12
7- 8	13
9- 11	14
12- 13	15
14- 16	16
17- 18	17
19- 21	18
22- 23	19
>= 24	20

New York State Next Generation Assessment Priorities Please provide detail on how the proposed supplemental assessment l 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):	DRC is committed to high-quality, error-free assessments that are valid and reliable for all students. The *TerraNova* Common Core Reading, English Language Arts, and Mathematics assessments measure the New York State Learning Standards with statistical reliability and content validity. The tests meet the requirements for a quality assessment. To achieve high quality assessments, content-area test development specialists and Research specialists reviewed items, for technical and statistical quality; for a match to standards; for bias, fairness, and sensitivity; for depth of knowledge; for estimated difficulty; and for adherence to the Principles of Universal Design. After norm-referenced data were gathered across the United States for the field test study, the best items were selected and reviewed by our test development and psychometric specialists for both content excellence and technical, statistical quality. In addition, our content specialists reviewed the test content to make sure it was in compliance with industry guidelines for clarity, style, accuracy, and appropriateness for all students. ***Overview of Stimuli: Passages, Graphics, and Scenarios*** The *TerraNova* Common Core Reading test uses stimuli like passages, graphics, and scenarios, that are accessible to the diverse student population and that are consistent with the content standards. Content specialists examined a wide array of primary source materials for a range of authentic topic areas. In addition, we considered the criteria in the following figure when evaluating content for its appropriateness for inclusion in the *TerraNova* Common Core assessment. **Criteria for Evaluating Content** ■ Have interest value for students. ■ Are grade appropriate in terms of vocabulary and language characteristics. ■ Are free of bias, fairness, and sensitivity issues. ■ Represent different cultures. ■ Are able to stand the test of time. ■ Sufficiently “rich” to generate a variety of items. ■ Avoid dated or specific subject matter unless a relevant historical context is provided. (For passages, students should not have to have extensive background knowledge in a particular discipline or area to understand a text.) All stimuli identified were reviewed by members of the content committee.

Item Development Reviews

Content committees reviewed the items using the following criteria:

- **Alignment—**Does the content of the item align with the Standard? Writers were trained to consider the degree to which the item is, in fact, aligned with the indicated eligible content. In making this judgment, writers considered whether the content was aligned (e.g., did the eligible content and the item both deal with fractions) and whether the required performance was aligned (e.g., if the eligible content called for a comparison to be made, was this reflected in the item).
- **Grade-level Appropriateness—**Is the item grade-level appropriate? Is the content consistent with the experiences of a student at the grade level assessed? Is the challenge level appropriate for the grade?
- **Correct Keys—**Is there one clear, correct answer? There should be no other answer that "could" be correct. Note: This does not mean that "good" distractors are unfair.
- **Difficulty—**Item Difficulty is indicated as Easy, Medium, and Hard. Is your rating in agreement with the difficulty rating on the Item Card?
- **Source of Challenge—**Is the source of challenge appropriately targeted to the content? The hardest part of the item (i.e., source of challenge) should be the content that is targeted. For example, in mathematics, the mathematics should be the major source of challenge rather than the wording or graphic. Students should not give an incorrect answer to a mathematics item because the reading level is too high or a graphic is flawed or excessively complex. Conversely, students should not give correct answers for reasons such as prior knowledge that make the answer to the item obvious (e.g., if the item asks which country has the largest population and students are to read a graph that includes China, there is no need to read the graph to answer the item).
- **Distractors—**Are distractors fair and appropriate? Distractors that are appropriate offer students reasonable choices that can be arrived at by making common errors. There should be no distractors that make no sense at all. It should be possible to examine each option and to reason how a student with some deficiency in knowledge or skill could choose it. The distractors should be formatted according to acceptable standards of test construction (e.g., a phrase that is common to each distractor should be placed in the stem).

- **Universal Design**
  - **Language Demand:** Is the language clear, well-formatted, and precise? Does the item use correct terminology for the content area? In order for all students to enter into the items of the assessment, they must be able to understand them. If the items are formatted poorly, use unnecessarily complex words or phrases, or use figures or layouts that are difficult to understand, some students will give incorrect answers due to these factors rather than the content is being assessed.
  - **Bias:** Is the item free of bias? All students will not be able to enter into the assessment if bias considerations are not resolved. Does the item contain clear bias problems? *A thorough, independent bias review was completed for all items*.
- **Depth of Knowledge—**Depth of Knowledge is based on the alignment work of Dr. Norman Webb. Rate each item based on the cognitive demand, using the following levels:
  - **Recall:** Recall of a fact, information, or procedure.
  - **Basic Application of Skill or Concept:** Use of information, conceptual knowledge, procedures, two or more steps.
  - **Strategic Thinking:** Requires reasoning, developing a plan or sequence of steps; has some complexity; more than one possible answer.
  - **Extended Thinking:** Requires an investigation, time to think and process multiple conditions of the problem or task, and more than 10 minutes to do non-routine manipulations. (This level is generally not assessed in on-demand assessments.)

Item Development Considerations

A strong assessment system is built upon sound assessment items that are instructionally sensitive and which align to content standards. The *TerraNova* Common Core items provide clear, focused expectations for grade-level performance.

Item Development Considerations

- Alignment to the Assessment Anchors and Eligible Content
- Grade-level appropriateness (reading/interest level, etc.)
- Depth of knowledge
- Cognitive level
- Item/task level of complexity
- Estimated difficulty level
- Performance Level Descriptor
- Relevancy of context
- Rationale for distractors
- Accuracy
- Style
- Correct terminology

Conformity with Professional Standards

The development of *TerraNova* Common Core followed the *Standards for Educational and Psychological Testing* (AERA, APA, NCME, 2014) and was designed to produce reliable and instructionally valid tests that reflect the range of cognitive ability articulated in the standards. Furthermore, the item development adhered to the Principles of Universal Design, and reflected attention to the accessibility by diverse groups of students and function appropriately across a broad range of test administration accommodations.

In addition to the committee review process, psychometric procedures for detecting bias were implemented. Both the *Code of Fair Testing Practices in Education* (Joint Committee on Testing Practices 1988) and the *Standards for Educational and Psychological Testing* (AERA, APA, and NCME 2014) assert that test items must be free from construct-irrelevant sources of differential difficulty. It is important that subgroup differences in performance be examined when sample sizes permit, and actions should be taken to ensure that differences in performance are due to factors that are construct-relevant, rather than construct-irrelevant. As part of the effort to identify construct-irrelevant differences in performance, assessment items will be evaluated by means of differential item functioning (DIF) analysis procedures.

Universal Design

The development of *TerraNova* Common Core was informed by the elements of universal design that characterize sound assessment practice. The Principles of Universal Design were created to ensure accessible environments for all people through equitable use, simple and intuitive design, effective communication, tolerance for variability, and minimal fatigue.

Their application is defended by research that links them to higher performance for all students.

The Frameworks for Universal Design for Computer-Based Testing (UD-CBT) and Universal Design for Learning specify how digital technologies can create tests that more accurately assess students who possess a diverse range of physical, sensory, and cognitive abilities and challenges.

The Development team used the Universal Design checklist in all stages of building the test.

Universal Design Checklist

- Items measure what they are intended to measure.
- Items respect the diversity of the assessment population.
- Items have a clear format for text.
- Stimuli and items have clear pictures and graphics.
- Items have concise and readable text.
- Items allow changes to other formats, such as Unified English Braille, without changing meaning or difficulty.
- The arrangement of the items on the test has an overall appearance that is clean and well organized.

Alignment, Rigor, and Cognitive Complexity

The cognitive complexity of the *TerraNova* Common Core items was classified using Dr. Norman Webb's depth-of-knowledge framework (Webb, N.L. 1997, 1999, 2007). The methodology Dr. Webb (1999) developed offers a comprehensive model that can be applied to a wide variety of contexts. With regard to the alignment between standards and assessment instruments, Dr. Webb's criteria include five categories, one of which deals with content. Within the content category is a useful set of levels for evaluating depth of knowledge (DOK) According to Dr. Webb (1999), "dependence of knowledge consistency between standards and assessments indicates alignment if what is elicited from students on the assessment is as demanding cognitively as what students are expected to know and do as stated in the standards" (p. 7-8). The four levels of cognitive complexity (i.e., depths of knowledge) are as follows:

Level 1: Recall

Level 2: Application of a Skill/Concept

Level 3: Strategic Thinking

Level 4: Extended Thinking

Stimulus Passage Development, Searching, and the Use of Copyrighted Materials

The passages in *TerraNova* Common Core are a balance of commissioned, public domain, and previously published works. When copyright materials were used, we secured all permissions for use of such material in *TerraNova* Common Core.

Commissioned passages represent a variety of topics to include, but not limited to, the following: science, biography, technology, how-tos, and other informational topics as well as poetry, dramas, and narratives for literary passages.

General Guidelines for Passage Selection and Writing

Passages will have:

- The appropriate length (for given grade and use) and complexity for the designated grade level
- Text that is rich enough to allow for the needed number of items to be generated
- Relevancy of context
- Text that will appeal to students at the designated grade level
- Appropriate subjects for the designated grade level
- Grade-appropriate vocabulary
- Text structures that will be familiar for the designated grade level
- Text that is written in Standard American English. Fiction passages may contain colloquial expressions in dialogue, but these expressions should be grade-level appropriate

Sources for Passages

The sources for the *TerraNova* Common Core passages include non-copyrighted and copyrighted newspapers; novels; trade books; anthologies of literature and poetry; short story collections; and children's, young adult, and general magazines. Test developers avoided sources such as Newbery Medal-winning book titles, Caldecott Medal-winning book titles, federal government forms, and selections from any basal reading series or from textbooks. Passage finders avoided selecting passages that are older and contain outdated information, especially when collecting nonfictional materials and passages that are too popular or may have been used in reading/English curricula. In addition, reading passages (literature and informational) and graphics (e.g., drawings, timelines, photographs, graphs) were reviewed for any bias and sensitivity issues.

Assessments Woven Tightly Into the Curriculum:	The *TerraNova* Common Core Reading, Language, and Mathematics assessments measure skills found within the New York State Learning Standards and taught at each grade, or at the adjacent grade, to offer an appropriate range of difficulty. For example, the **Reading** tests measure Literature: Key Ideas and Details (3-8), Craft and Structure (3-8); Informational Text: Key Ideas and Details (3-8), Craft and Structure (3-8), Integration of Knowledge and Ideas (3-8); Vocabulary Acquisition (3-8). The **Language** tests measure Conventions of Standard English (3-8); Knowledge of Language (3-8); Text Types and Purposes (3-8); Production and Distribution of Writing (3-8); Research to Build and Present Knowledge (6). The **Mathematics** tests measure the following skills: Measurement and Data (3-5), Geometry (4, 5, 7, 8), Number and Operations in Base Ten (1-5), Number and Operations – Fractions (3-4), Operations and Algebraic Thinking (3-5), Ratios and Proportional Relationships (Grades 6-8), Expressions and Equations (Grades 6-8), Number System (Grades 6-8), and Statistics and Probability (Grades 6, 7). In addition to the norm-referenced selected response items that measure the New York State Learning Standards, the constructed response items within the performance task sections measure the application and integration of information, increasing the depth of knowledge level and giving student an opportunity to demonstrate their knowledge, skills, and abilities that meet the Reading, Language, and Mathematics content standards.
Performance Assessment:	*TerraNova* Common Core's mix of response formats offers many of the benefits of full performance assessments. *TerraNova* Common Core features item types that go beyond traditional multiple-choice tests. On the same scale and all in the same test book, the test includes: ■ Integrated Performance Tasks (PT) ■ Constructed-Response (CR) items ■ Extended Constructed-Response (ECR) items ■ Multiple-Choice DRC is the only publisher that uses 2-parameter IRT scoring to offer students partial credit for partial understanding. This is an industry first which benefits the student because every item contributes to a student's score profile.
Efficient Time-Saving Assessments:	*TerraNova* Common Core offers time-saving formats that move beyond traditional standardized tests. The *TerraNova* Common Core assessment takes 3 hours 50 minutes to administer all sections. Even greater time savings is possible with *TerraNova's* Survey edition, a short form selected-response assessment that can be used interchangeably with other *TerraNova* editions because of the development of a common scale.

Technology:	Additional next-generation innovations are being developed with unparalleled resources of DRC. These include second generation computer adaptive tests and application of Artificial Intelligence in scoring written responses. A robust online reporting system is available to offer helpful pinpointing of student information.
Degree to which the growth model must differentiate across New York State's four levels of teacher effectiveness (only applicable to supplemental assessments):	The proposed growth model uses native growth scale (median difference of student national percentile ranks) to evaluate teacher's effectiveness. It provides a numerical measure of teacher's effectiveness. Based on the native growth scale the four levels of teacher effectiveness can be defined.

STUDENT ASSESSMENTS FOR TEACHER AND PRINCIPAL EVALUATION

FORM *G*

ATTESTATION OF TECHNICAL CRITERIA – SUPPLEMENTAL ASSESSMENTS WITH CORRESPONDING GROWTH MODELS

Please read each of the items below and check the corresponding box to ensure the fulfillment of the technical criteria outlined in the Technical Application on "FORM B-2".

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

COMPLETE THIS SECTION:

2.2(A) Narrative Overview of Proposed Supplemental Assessment and Associated Growth Model	
This application contains a short overview of the assessment being proposed, including the intended purpose of the assessment, and how the assessment is administered.	☑
For supplemental assessments, this application contains a description of the growth model and how it is used in conjunction with the assessment.	☑ ☐ N/A
For K-2 assessments, this application contains evidence that the proposed assessment is consistent with this RFQ's requirement that the assessment not be a "Traditional Standardized Assessment" as defined above in the section "Definitions of Key Terms Used in this RFQ."	☐ ☑ N/A
2.2(B) Evidence of Capability	
This application provides an overview of services provided by the Assessment Provider, including a description of the range of support / technical assistance that the Assessment Provider would provide to an LEA if selected by an LEA for this service.	☑
This application contains information as to whether the Applicant or Assessment Provider has been denied approval as a provider of assessment services in another state(s) and the reason(s) for such denial. If denied within New York State, the location and reason are indicated.	☑ ☐ N/A
2.2(C): Evidence of Copyright Owner/Assessment Representative History of Assessment Development	
This application contains evidence that the Copyright Owner/Assessment Representative has a history of developing assessments of student learning (achievement or growth) for the purpose of making defensible judgments about educator effectiveness.	☑ ☐ N/A

2.2(D)-i: Technical Documentation Related to Assessment and Student Growth Score Properties: RELIABILITY ***Both "minimum" and "desired" qualifications are listed. For the purposes of this RFQ, applications will only be rated against the "minimum" qualifications; however, NYSED's aspirational "desired" qualifications are also listed to identify possible future requirements for assessments and associated growth models.***	
For supplemental assessments used in conjunction with growth models: This application contains evidence of the *minimum* criteria for reliability: • Student test scores have adequate levels of reliability (e.g., coefficient alpha > 0.75). This application contains evidence of the *desired* criteria for reliability: • Standard errors provided for students growth scores. • Student growth classifications have adequate decision consistency. • Teacher effectiveness classifications demonstrate adequate consistency. *Examples include agreement statistics (e.g., kappa coefficients) based on simulation studies.*	**Check all that apply:** ☑ ☑ ☑ ☐
2.2(D)-ii: Technical Documentation Related to Assessment and Student Growth Score Properties: VALIDITY – ALIGNMENT ***Both "minimum" and "desired" qualifications are listed. For the purposes of this RFQ, applications will only be rated against the "minimum" qualifications; however, NYSED's aspirational "desired" qualifications are also listed to identify possible future requirements for assessments and associated growth models.***	
For supplemental assessments used in conjunction with growth models: This application contains evidence of the *minimum* criteria for alignment validity: • Evidence that test content is sufficiently aligned with New York State Learning Standards and covers a range of measurable standards. Documentation that demonstrates that: (a) at least 80% of the test measures content aligned with NYS learning standards, (b) no more than 20% of test content is aligned with other learning standards or objectives, and (c) a range of content from the NYS learning standards is measured *Note: Other relevant standards can be proposed if NYS Learning Standards do not apply to subject area.* This application contains evidence of the *desired* criteria for alignment validity: • 100% alignment between NYS Learning Standards and assessment.	**Check all that apply:** ☑ ☐
2.2(D)-iii: Technical Documentation Related to Assessment and Student Growth Score Properties: VALIDITY – RELATIONS TO OTHER VARIABLES ***Both "minimum" and "desired" qualifications are listed. For the purposes of this RFQ, applications will only be rated against the "minimum" qualifications; however, NYSED's aspirational "desired" qualifications are also listed to identify possible future requirements for assessments and associated growth models.***	
For supplemental assessments used in conjunction with growth models: This application contains evidence of the *minimum* criteria for validity in relation to other variables: • Evidence students' growth scores are correlated with other measures of student progress (e.g., r > .5 with measures such as the number of objectives mastered by a student over the course of the year, teachers' ratings of students' progress, or scores from other assessments).	**Check all that apply:** ☑

This application contains evidence of the *desired* criteria for validity in relation to other variables: • Evidence teacher effectiveness ratings are positively correlated (e.g., r > .5) with other measures of teaching effectiveness.	☐
2.2(D)-iv: Technical Documentation Related to Assessment and Student Growth Score Properties: VALIDITY – INTERNAL STRUCTURE ***Both "minimum" and "desired" qualifications are listed. For the purposes of this RFQ, applications will only be rated against the "minimum" qualifications; however, NYSED's aspirational "desired" qualifications are also listed to identify possible future requirements for assessments and associated growth models.***	
For supplemental assessments used in conjunction with growth models: This application contains evidence of the *minimum* criteria for validity of internal structure: • Scale properties appropriate for growth model used (*see notes*). Total scores and subscores on student assessments should be supported by dimensionality analyses (e.g., IRT residual analyses, factor analyses). This application contains evidence of the *desired* criteria for validity of internal structure: • Evidence students' scores are on an interval scale. **Notes: If **gain score model** is used, evidence is needed that students' pretest and posttest scores are on the same scale. If **student growth percentile** model used, justification for the number of years included in the model should be provided. If **growth-to-proficiency**, projection, or **value-added models** are used, evidence is needed that the model explains a significant amount of variability in student achievement. Also, models should demonstrate robustness to missing data.*	**Check all that apply:** ☑ ☑
2.2(D)-v: Technical Documentation Related to Assessment and Student Growth Score Properties: UTILITY AND COMPREHENSIBILITY ***Both "minimum" and "desired" qualifications are listed. For the purposes of this RFQ, applications will only be rated against the "minimum" qualifications; however, NYSED's aspirational "desired" qualifications are also listed to identify possible future requirements for assessments and associated growth models.***	
For supplemental assessments used in conjunction with growth models: This application contains evidence of the *minimum* criteria for utility and comprehensibility: • Technical documentation that describes how student growth and educator effectiveness are calculated. This application contains evidence of the *desired* criteria for utility and comprehensibility: • Student growth reports support instructional improvement. Resources and supporting materials available to the field.	**Check all that apply:** ☑ ☑
2.2(E)-i: Technical Documentation Related to Aggregating Student-Level Growth Scores to Teacher-Level Scores: CREATION OF TEACHER LEVEL SCORES	
For supplemental assessments used in conjunction with growth models: This application includes a narrative description of how student-level scores are aggregated to create a single teacher-level score for each teacher.	☑ ☐ N/A

2.2(E)-ii: Technical Documentation Related to Aggregating Student-Level Growth Scores to Teacher-Level Scores: EXCLUSION RULES	
This application includes a description of any exclusion rules that remove students associated with a given teacher from the teacher's teacher-level score (either through a growth model or in conjunction with an SLO).	☑ ☐ N/A
2.2(F): Technical Documentation Related to Converting Teacher-Level Growth Score to New York State's 0-20 APPR Scale	
This application includes a crosswalk that maps scores on the assessment's aggregated teacher-level growth score to the required New York State teacher and principal evaluation metric, which ranges from 0-20.	☑
This application includes procedures for converting teacher-level growth scores to the 0-20 APPR scale comply with the New York Standards for each evaluation rating category, which are based on the following definitions.	☑
For supplemental assessments used in conjunction with growth models: This application includes an explanation of the assignment of HEDI rating categories based on the following ranges: • Highly Effective: results are well-above State average* for similar students • Effective: results meet State average* for similar students • Developing: results are below State average* for similar students • Ineffective: Results are well-below State average* for similar students	☑ ☐ N/A
2.2(G)-i: Technical Documentation Related to Fairness: TEST TAKERS Consistent with the new Testing Standards (2014), there is an increased focus in the industry on fairness of assessments and their uses. Please provide evidence of fairness for both the proposed assessment and, if applicable, the proposed growth model.	
This application includes evidence that the proposed assessments are fair to all test takers (e.g., Differential Item Functioning [DIF] / bias information, fairness evaluation / sensitivity review plan.)	☑
2.2(G)-ii: Technical Documentation Related to Fairness: TEACHER GROWTH SCORES	
This application includes evidence of fairness of the proposed aggregated teacher growth scores (e.g., lack of correlation between aggregated teacher growth scores and student demographics).	☑
The evidence of fairness of the proposed aggregated teacher growth scores includes an explanation of how the growth model incorporates (a) prior academic history, (b) poverty, (c) students with disabilities, and (d) English language learners.	☑ ☐ N/A

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

1. Name of Organization (PLEASE PRINT/TYPE) Data Recognition Corporation	4. Signature of Authorized Representative (PLEASE USE BLUE INK) 
2. Name of Authorized Representative (PLEASE PRINT/TYPE) Susan S. Engeleiter	5. Date Signed February 16, 2016
3. Title of Authorized Representative (PLEASE PRINT/TYPE) Chief Executive Officer and President	

1. Name of LEA (please print/type) N/A	4. Signature of School Representative (Please use Blue ink) N/A
2. School Representative's Name (please print/type) N/A	5. Date Signed N/A
3. Title of School Representative (please print/type) N/A	