edreports.org

Center for Assessment

Review Criteria
Interim Assessment
Mathematics
Grades 3-8

v1.0

Gateway 1

Alignment, Fairness, & Accessibility

The assessment is designed to measure student knowledge and skills in alignment with the expectations of college- and career-ready standards.

The process for the development of items and test forms/events is clearly-articulated, intentional, and seeks to measure the depth and breadth of the standards.

The assessment is designed for fairness and accessibility; all students are provided opportunities to have their learning and achievement accurately assessed.

Gateway 1 Overview	Available Points
Criterion 1.1: Test Development Alignment Indicators 1.1.a-1.1.b Assessment design specifications align to the expectations of college-and career-ready (CCR) standards.	8
Criterion 1.2: Item and Form Alignment Indicators 1.2.a-1.2.g Assessment items and resulting test forms align to the expectations of the mathematics domains as outlined by college- and career-ready (CCR) standards.	16
Criterion 1.3: Fairness and Accessibility Indicators 1.3.a-1.3.c The assessment is fair and accessible for all students in the intended test-taking population.	12

▸ Criterion 1.1: Test Development	Assessment design specifications align to the expectations of college-and career-ready (CCR) standards.

Indicators	Rating
1.1.a Assessment design specifications provide clear expectations and detailed guidance to support the development of high-quality, CCR standards-aligned materials. • Assessment rationale explains the design of the assessment, the benefits of the assessment, and a research foundation grounding the assessment process. • Item development documentation is sufficiently robust to support the writing and review of items measuring CCR standards. • Across all item types, test program documentation provides clear scoring information and/or rubrics to evaluate students' levels of understanding with respect to the standards. • Item development documentation includes a description of processes used to ensure items are content-accurate and without technical or editorial flaws. • The suggested ranges of cognitive demand reflected in assessment design specifications are sufficient to measure the depth of the standards.	0 2 4
1.1.b Test blueprints and/or assessment design specifications focus strongly on the content that is most important for students to master by reflecting an appropriate expected distribution of content and related score points. • The expected distribution of score points across the K-8 content blueprints and/or assessment design specifications focuses strongly on the major work as established in CCR standards. • The expected distribution of score points across the HS course content blueprints and/or assessment design specifications focuses on the content and skills students need to be successful in college and careers as established in CCR standards.	0 2 4

Total Available Points	8	Meets: 8 Partially Meets: 6 Does Not Meet: <6

▸ Criterion 1.2: Item and Form Alignment	Assessment items and resulting test forms align to the expectations of the mathematics standards as outlined by college- and career-ready (CCR) standards.

Indicators	Rating
1.2.a Test forms/events delivered to students reflect an appropriate distribution of content and related score points and item types within forms/events. ● The test forms/events delivered to students focus strongly on the major and supporting clusters as established in CCSS standards. ● The test forms/events delivered to students focus on the content and skills students need to be successful in college and careers as established in CCR standards.	0 2 4
1.2.b Test items are written to elicit evidence of learning relative to one or more CCR standard/s and aligned to assessment design specifications. ● Test items can be clearly identified as measuring one or more CCR standard/s without formally measuring knowledge and skills that are not included with CCR standards. ● Test items align to the assessment design specifications. ● Items are content-accurate, reflecting no technical or editorial flaws.	0 1 2
1.2.c The range of item types and cognitive demand among test events is sufficient to strategically assess the full intent and complexity of CCR standards being addressed and is aligned to blueprints or assessment design specifications. ● The items are constructed to reach the depth and complexity of CCR standards and cover the expected range of content measured by any assessment sub-scores. ● There is an appropriate distribution and/or range of cognitive demand exercised among test events submitted for review. ● The range of item types and cognitive demand among test events align to blueprints or assessment design specifications.	0 1 2
1.2.d The assessment is aligned to the procedural skill and fluency expectations of CCR standards. ● The item development documentation and distributions of points from the assessment design specifications that directly address standards requiring procedural skill and fluency are reflected in the assessment items and resulting forms/test events.	0 1 2
1.2.e The assessment is aligned to the conceptual understanding expectations of CCR standards. ● The item development documentation and distributions of points from the assessment design specifications that directly address standards requiring conceptual understanding are reflected in the assessment items and resulting forms/test events.	0 1 2
1.2.f The assessment is aligned to the application expectations of CCR standards. ● The item development documentation and distributions of points from the assessment design specifications that directly address standards requiring application are reflected in the assessment items and resulting forms/test events. ● For high school assessments, items that attend to the full intent of the modeling process are administered to all students.	0 1 2

1.2.g The assessment includes mathematical practices as described in CCR standards. • The assessment design specifications addressing mathematical practices are reflected in the assessment items and resulting forms/test events. • Items including mathematical practices should be connected to CCR-aligned mathematics. • If alignments are provided, items aligned to targeted mathematical practice(s) require them to receive full credit. • Test forms/events reflect the distribution of mathematical practices as outlined in the assessment design specifications.	0 1 2

Total Available Points	16	Meets: 13-16 Partially Meets: 8-12 Does Not Meet: <8

▸ Criterion 1.3: Fairness and Accessibility	The assessment is fair and accessible for all students in the intended test-taking population.

Indicators	Rating
1.3.a Items and test events are developed and reviewed using procedures that ensure fairness. • Item development documentation/procedures clearly demonstrate adherence to the principles of universal design. • Item rendering specifications clearly reflect the principles of universal design. • Item review processes are designed to minimize construct-irrelevant variance. • Items go through a content bias/sensitivity review to make sure they are appropriate and fair for all relevant student groups. • Procedures are in place to evaluate the technical quality and appropriateness of items and test events for student subgroups and students utilizing different accommodations.	0 2 4
1.3.b Appropriate accommodations and supports are in place to ensure the assessment is accessible to all students in the intended test-taking population, including special populations of students and English Learners. • The test-taking population for which the assessment was/was not designed to support is clearly documented. • The list of accommodations is aligned to the vendor's definition of the assessment's intended uses. • The list of accommodations is sufficient to serve the needs of the full population of intended test takers. • Evidence is available to support the validity and fairness of the intended interpretations and uses for those students who access the exam using the supported accommodations. • Evidence is available that supports the quality and appropriateness of provided accommodations. • The administration manual is clearly worded and supports teachers and other educational personnel in providing an appropriate testing experience for all students. • Sample forms or released test items are available to stakeholders at each grade level.	0 2 4
1.3.c The range and types of technology provided within the assessment support the validity of assessment outcomes. • Guidance is provided to support accessibility to the assessment system on a variety of platforms. • Auditory supports present stimuli and items in a natural voice and at a cadence that can be adjusted to accommodate the learner. • Overall visual design, including digital tools (e.g., dictionaries, calculators, sticky notes, and highlighters) enhances the test-taking experience, does not distract or clutter the digital workspace, and can be easily navigated by students.	0 2 4

Total Available Points	12	Meets: 10-12 Partially Meets: 6-9 Does Not Meet: <6

Gateway 2

Technical Quality

The assessment supports valid score interpretations related to student achievement and other reported information such as predicted student performance on future tests, areas of student strengths and need, and progress on student learning since previous test administrations.

Gateway 2 Overview	Available Points
Criterion 2.1: Overall Achievement Indicators 2.1.a-2.1.d The interim assessment provides for valid inferences about a student's current overall achievement in the target content domain.	8
Criterion 2.2: Predicted Student Performance Indicators 2.2.a-2.2.d The interim assessment provides valid information regarding predicted student performance on a state's summative assessment or other intended criterion measure(s).	Points may vary based on indicators that are claimed by the publisher to be assessed.
Criterion 2.3: Sub-scores Indicators 2.3.a-2.3.d The interim assessment provides for valid inferences about a student's specific areas of strength and need (e.g., at the reportable category, content strand or objective level).	Points may vary based on indicators that are claimed by the publisher to be assessed..
Criterion 2.4: Student Progress Indicators 2.4.a-2.4.d The interim assessment provides valid information regarding student progress in the content domain.	Points may vary based on indicators that are claimed by the publisher to be assessed.

▸ Criterion 2.1: Overall Achievement	The interim assessment provides for valid inferences about a student's current overall achievement in the target content domain.

Indicators	Rating
2.1.a Item and form development procedures result in high-quality test events. • Item development, review, and piloting procedures and materials are designed to ensure all newly developed items meet technical quality standards. • Assessment design specifications and test development and review procedures ensure test events meet content and statistical quality criteria.	0 1 2
2.1.b Achievement scores are reliable. • Item/test development and review procedures facilitate the reliability of test scores. • Procedures for calculating and evaluating reliability are well-documented and appropriate. • Obtained reliability indices and estimates of precision are at an appropriate level to support the use of results as intended.	0 1 2
2.1.c Achievement scores support intended interpretations of student performance. • Evidence is provided to support the intended interpretations of student achievement. • Equating/linking procedures supporting the comparability of achievement scores and score-based inferences across test events/administrations are described and reasonable. • Item development specifications, task models, and scoring rubrics include enough detail to support consistency in the presentation, format, and degree of scaffolding observed in items and associated stimuli across test events. • There is empirical evidence and an active research agenda supporting the validity of achievement scores as measures of the intended knowledge and skills.	0 1 2
2.1.d Achievement scores are appropriate for supporting their intended uses. • The intended uses for the achievement scores are clearly and consistently articulated. • There is sufficient theoretical and empirical evidence supporting the intended uses of achievement scores.	0 1 2

Total Available Points	8	Meets: Claim-Dependent Partially Meets: Claim-Dependent Does Not Meet: Claim-Dependent

▸ Criterion 2.2: Predicted Student Performance	The interim assessment provides valid information regarding predicted student performance on a state's summative assessment or other intended criterion measure(s).

Indicators	Rating
*2.2.a The design of the interim assessment supports its use in predicting performance on one or more external measures. ● Sufficient information is provided to evaluate the degree to which the construct or content domain targeted by the interim assessment is similar to that assessed by the criterion measure(s). ● The intended use of the interim assessment does not invalidate or contradict its appropriateness for predicting performance on the intended criterion measure(s). ● If an interim assessment was designed to predict performance on specific assessments (e.g., ACT, SAT), evidence supporting that claim is provided.	0 1 2 N/C
*2.2.b Predicted results are reliable. ● Procedures used for calculating and evaluating the reliability of predicted scores/classifications are well documented and appropriate. ● The reliability of the predicted result is calculated in a manner that is consistent with the inferences they were designed to support (e.g., CCR). ● The predictions demonstrate sufficient reliability to support their intended uses.	0 1 2 N/C
*2.2.c Predicted results (e.g., expected scaled scores, performance levels, passing status, etc.) reflect a student's likely performance on the state summative assessment or other intended criterion measure(s). ● The data and procedures used to establish and evaluate the predictive relationship for a given test-taking sample are documented and reasonable. ● The procedures used to support intended interpretations are clearly articulated. ● Studies support the appropriateness of the predicted result as a measure of future performance.	0 1 2 N/C
*2.2.d Predicted results are appropriate for supporting their intended uses. ● The intended uses for the predicted results are clearly and consistently articulated. ● There is sufficient theoretical and empirical evidence to support the appropriateness of the intended uses of predicted results.	0 1 2 N/C

**Note: Indicators and Criteria should be considered not claimed (N/C) if the assessment was not intentionally designed to be predictive.*

Total Available Points	Points may vary based on indicators that are claimed by the publisher to be assessed.	Meets: Claim-Dependent Partially Meets: Claim-Dependent Does Not Meet: Claim-Dependent

▸ Criterion 2.3: Sub-scores	The interim assessment provides for valid inferences about a student's specific areas of strength and need (e.g., at the reportable category, content strand or objective level).

Indicators	Rating
*2.3.a Test events are designed to provide specific information about a student's areas of strength and need in the content domain. ● The assessment design supports the reporting of sub-scores at each level of granularity for which they are provided ● The assessment design supports interpretations of students' areas of strength and need in the content domain.	0 1 2 N/C
*2.3.b Reported sub-scores are reliable. ● Estimates of reliability/precision are provided for all reported sub-scores. ● Procedures for calculating reliability indices and precision for the sub-score results are defensible and well documented. ● The calculated reliability and precision indices indicate adequate support for the intended interpretations and uses.	0 1 2 N/C
*2.3.c Reported sub-scores support intended interpretations of student performance in defined sub-skill areas. ● Evidence is provided to support intended interpretations of all reported sub-scores. ● Empirical data suggest sub-scores represent distinct sub-domains and should be reported separately.	0 1 2 N/C
*2.3.d Reported sub-scores are appropriate for supporting their intended uses. ● The intended uses for the sub-scores are clearly and consistently articulated. ● There is sufficient theoretical and empirical evidence supporting the intended uses for the sub-scores.	0 1 2 N/C

**Note: Indicators and Criteria should be considered not claimed (N/C) if the assessment was not intentionally designed with the use of subscores.*

Total Available Points	Points may vary based on indicators that are claimed by the publisher to be assessed.	Meets: Claim-Dependent Partially Meets: Claim-Dependent Does Not Meet: Claim-Dependent

▸ Criterion 2.4: Student Progress

The interim assessment provides valid information regarding student progress in the content domain.

Indicators	Rating
*2.4.a The interim assessment is designed to support measures of growth. • Test design and content specifications (within and across grades) support the use of assessment results as a means of evaluating growth in the manner specified by the vendor. • The technical characteristics of the test and reportable scale support the reported growth measure.	0 1 2 N/C
*2.4.b Student growth scores are reliable. • Procedures for estimating standard errors around the growth estimates are appropriate and well documented. • The reliability of the growth scores have been evaluated for students at different places along the ability scale. • The calculated reliability and precision indices indicate adequate support for the intended uses of the reported growth scores.	0 1 2 N/C
*2.4.c Student growth scores support the intended interpretations. • The procedures and measures for calculating student growth are clearly documented and appropriate. • If significant modifications are made to the interim assessment that might break the trend line (i.e., test design changes, rescaling, and shifts in performance standards), empirical evidence is provided to support the intended interpretations and uses of growth scores. • Empirical evidence confirms that growth scores provide for valid intended inferences about student learning in the content domain.	0 1 2 N/C
*2.4.d Student growth scores are appropriate for supporting the intended uses. • The intended uses for the growth scores are clearly and consistently articulated. • There is sufficient theoretical and empirical evidence supporting the intended uses for the growth scores.	0 1 2 N/C

**Note: Indicators and Criteria should be considered not claimed (N/C) if the assessment was not intentionally designed to support student progress.*

Total Points Available	Points may vary based on indicators that are claimed by the publisher to be assessed.	Meets: Claim-Dependent Partially Meets: Claim-Dependent Does Not Meet: Claim-Dependent

Gateway 3

Score Reports and Interpretive Guides

Score reports provide sound information for stakeholders and support to assure assessment data is interpreted correctly and appropriately for use in multiple contexts.

Gateway 3 Overview	Available Points
Criterion 3.1: Overall Achievement Indicators 3.1.a-3.1.c Score reports and other resources (e.g., user manuals, interpretive guides, instructional or curricular resources) are appropriate for facilitating the intended interpretations and uses of overall achievement results.	10
Criterion 3.2: Predicted Student Performance Indicators 3.2.a-3.2.c Score reports and other resources (e.g., user manuals, interpretive guides, instructional or curricular resources) are appropriate for facilitating the intended interpretations and uses of predicted student performance.	Points may vary based on indicators that are claimed by the publisher to be assessed.
Criterion 3.3: Sub-scores Indicators 3.3.a-3.3.c Score reports and other resources (e.g., user manuals, interpretive guides, instructional or curricular resources) are appropriate for facilitating the intended interpretations and uses of sub-scores.	Points may vary based on indicators that are claimed by the publisher to be assessed.
Criterion 3.4: Student Progress Indicators 3.4.a-3.4.c Score reports and other resources (e.g., user manuals, interpretive guides, instructional or curricular resources) are appropriate for facilitating the intended interpretations and uses of student growth or progress results.	Points may vary based on indicators that are claimed by the publisher to be assessed.

▸ Criterion 3.1: Overall Achievement	Score reports and other resources (e.g., user manuals, interpretive guides, instructional or curricular resources) are appropriate for facilitating the intended interpretations and uses of overall achievement results.

Indicators	Rating
3.1.a The design of the score reports and supporting materials (e.g., user manuals and interpretive guides) and the types of information provided are consistent with the intended interpretations and uses for specific users (e.g., educators, parents, students, or administrators). ● Score reports effectively represent the intended interpretations and uses of overall achievement results. ● The type and grain size of the information reported is appropriate for effectively serving the intended interpretations and uses. ● Evidence shows that there was attention to the audience and specific users in the design process, including user-specific versions of reports when needed. ● Evidence (e.g., studies, focus groups) is provided that users are able to effectively interpret and use reports in the manner intended. ● The documentation should include warnings of potential or common misuses of the results that may result in negative, unintended consequences. ● Reports identify and flag students for whom the integrity of the test interpretations may be compromised (e.g., student clicks through rapidly). ○ The conditions which bring about a flag are articulated on reports and/or in interpretive guides.	0 2 4
3.1.b Score reports include information about the degree of error associated with the achievement score. ● For example, confidence intervals, error bands, or probability statements are provided to represent potential score variability. ● Supports (e.g., illustrative examples, informational text) are provided to facilitate accurate interpretations of error estimates and clarify the practical implications of error on score use.	0 1 2
3.1.c Sufficient and appropriate guidance (e.g., instructional or curricular supports) is provided to support the intended interpretations and uses, when needed. ● Guidance is aligned to the use. ● Any guidance provided has a basis in research and/or was created in consultation with educators experienced in using educational data. ● Guidance is provided to support appropriate use for students scoring at the full range of performance outcomes.	0 2 4

Total Available Points	10	Meets: 8-10 Partially Meets: 5-7 Does Not Meet: <5

▸ Criterion 3.2: Predicted Student Performance	Score reports and other resources (e.g., user manuals, interpretive guides, instructional or curricular resources) are appropriate for facilitating the intended interpretations and uses of predicted student performance.

Indicators	Rating
*3.2.a The design of the score reports and supporting materials (e.g., user manuals and interpretive guides) and the types of information provided are consistent with the intended interpretations and uses for specific users (e.g., educators, parents, students, or administrators). ● Score reports effectively represent the intended interpretations and uses of overall achievement results. ● The type and grain size of the information reported is appropriate for effectively serving the intended interpretations and uses. ● Evidence shows that there was attention to the audience and specific users in the design process, including user-specific versions of reports when needed. ● Evidence (e.g., studies, focus groups) is provided that users are able to effectively interpret and use reports in the manner intended. ● The documentation should include warnings of potential or common misuses of the results that may result in negative, unintended consequences. ● Reports identify and flag students for whom the integrity of the test interpretations may be compromised (e.g., student clicks through rapidly). ○ The conditions which bring about a flag are articulated on reports and/or in interpretive guides.	0 2 4 N/C
*3.2.b Score reports include information about the degree of error associated with the predicted performance score. ● For example, confidence intervals, error bands, or probability statements are provided to represent potential score variability. ● Supports (e.g., illustrative examples, informational text) are provided to facilitate accurate interpretations of error estimates and clarify the practical implications of error on score use.	0 1 2 N/C
*3.2.c Sufficient and appropriate guidance (e.g., instructional or curricular supports) is provided to support the intended interpretations and uses, when needed. ● Guidance is aligned to the use. ● Any guidance provided has a basis in research and/or was created in consultation with educators experienced in using educational data. ● Guidance is provided to support appropriate use for students scoring at the full range of performance outcomes.	0 2 4 N/C

**Note: These claims should match claims made and evaluated in Gateway 2. Indicators and Criteria should be considered not claimed (N/C) if the assessment was not intentionally designed to be predictive.*

Total Points Available	Points may vary based on indicators that are claimed by the publisher to be assessed.	Meets: Claim-Dependent Partially Meets: Claim-Dependent Does Not Meet: Claim-Dependent

▸ Criterion 3.3: Sub-scores	Score reports and other resources (e.g., user manuals, interpretive guides, instructional or curricular resources) are appropriate for facilitating the intended interpretations and uses of sub-scores.

Indicators	Rating
*3.3.a The design of the score reports and supporting materials (e.g., user manuals and interpretive guides) and the types of information provided are consistent with the intended interpretations and uses for specific users (e.g., educators, parents, students, or administrators). ● Score reports effectively represent the intended interpretations and uses of overall achievement results. ● The type and grain size of the information reported is appropriate for effectively serving the intended interpretations and uses. ● Evidence shows that there was attention to the audience and specific users in the design process, including user-specific versions of reports when needed. ● Evidence (e.g., studies, focus groups) is provided that users are able to effectively interpret and use reports in the manner intended. ● The documentation should include warnings of potential or common misuses of the results that may result in negative, unintended consequences. ● Reports identify and flag students for whom the integrity of the test interpretations may be compromised (e.g., student clicks through rapidly). ○ The conditions which bring about a flag are articulated on reports and/or in interpretive guides.	0 2 4 N/C
*3.3.b Score reports include information about the degree of error associated with sub-scores. ● For example, confidence intervals, error bands, or probability statements are provided to represent potential score variability. ● Supports (e.g., illustrative examples, informational text) are provided to facilitate accurate interpretations of error estimates and clarify the practical implications of error on score use.	0 1 2 N/C
*3.3.c Sufficient and appropriate guidance (e.g., instructional or curricular supports) is provided to support the intended interpretations and uses, when needed. ● Guidance is aligned to the use. ● Any guidance provided has a basis in research and/or was created in consultation with educators experienced in using educational data. ● Guidance is provided to support appropriate use for students scoring at the full range of performance outcomes.	0 2 4 N/C

**Note: These claims should match claims made and evaluated in Gateway 2. Indicators and Criteria should be considered not claimed (N/C) if the assessment was not intentionally designed with the use of subscores.*

Total Points Available	Points may vary based on indicators that are claimed by the publisher to be assessed.	Meets: Claim-Dependent Partially Meets: Claim-Dependent Does Not Meet: Claim-Dependent

▸ Criterion 3.4: Student Progress	Score reports and other resources (e.g., user manuals, interpretive guides, instructional or curricular resources) are appropriate for facilitating the intended interpretations and uses of student growth or progress results.

Indicators	Rating
*3.4.a The design of the score reports and supporting materials (e.g., user manuals and interpretive guides) and the types of information provided are consistent with the intended interpretations and uses for specific users (e.g., educators, parents, students, or administrators). ● Score reports effectively represent the intended interpretations and uses of overall achievement results. ● The type and grain size of the information reported is appropriate for effectively serving the intended interpretations and uses. ● Evidence shows that there was attention to the audience and specific users in the design process, including user-specific versions of reports when needed. ● Evidence (e.g., studies, focus groups) is provided that users are able to effectively interpret and use reports in the manner intended. ● The documentation should include warnings of potential or common misuses of the results that may result in negative, unintended consequences. ● Reports identify and flag students for whom the integrity of the test interpretations may be compromised (e.g., student clicks through rapidly). ○ The conditions which bring about a flag are articulated on reports and/or in interpretive guides.	0 2 4 N/C
*3.4.b Score reports include information about the degree of error associated with the student progress score. ● For example, confidence intervals, error bands, or probability statements are provided to represent potential score variability. ● Supports (e.g., illustrative examples, informational text) are provided to facilitate accurate interpretations of error estimates and clarify the practical implications of error on score use.	0 1 2 N/C
*3.4.c Sufficient and appropriate guidance (e.g., instructional or curricular supports) is provided to support the intended interpretations and uses, when needed. ● Guidance is aligned to the use. ● Any guidance provided has a basis in research and/or was created in consultation with educators experienced in using educational data. ● Guidance is provided to support appropriate use for students scoring at the full range of performance outcomes.	0 2 4 N/C

**Note: These claims should match claims made and evaluated in Gateway 2. Indicators and Criteria should be considered not claimed (N/C) if the assessment was not intentionally designed to support student progress.*

Total Points Available	Points may vary based on indicators that are claimed by the publisher to be assessed.	Meets: Claim-Dependent Partially Meets: Claim-Dependent Does Not Meet: Claim-Dependent