

ITEMS THAT REQUIRE ADDITIONAL INFORMATION OR REVISION IN NORTH CAROLINA’S INNOVATIVE ASSESSMENT DEMONSTRATION AUTHORITY PLAN

April 2, 2019

CONTENTS

- Regulatory Requirements 3
 - Consultation 3
 - Innovative Assessment System (2) and (3) 6
 - Innovative Assessment System (4)..... 8
 - Innovative Assessment System (6)..... 15
 - Innovative Assessment System (7) and (9)..... 16
 - Initial Implementation in a Subset of LEAS or Schools..... 18
- Application Selection Criteria 19
 - (a)(3) 19
 - (b)(2) 21
 - (c)(1) 23
 - (c)(2) 30
 - (d)(1) 32
 - (d)(2)..... 33
 - (e)(1)..... 34
- Appendix A..... 37

Note:

North Carolina has revised the proposed model detailed in the application. Rather than administer three interim assessments which are combined for an aggregate academic achievement level, the model being proposed will consist of two through-grade assessments, NCPAT 1 and NCPAT 2, that provide data to determine which mini cluster of items a student takes for the end-of-year NCPAT 3.

REGULATORY REQUIREMENTS

CONSULTATION

Regulatory Requirement	Required information from the SEA
<p>Evidence that the SEA or consortium has developed an innovative assessment system in collaboration with--</p> <p>(1) Experts in the planning, development, implementation, and evaluation of innovative assessment systems, which may include external partners; and</p> <p>(2) Affected stakeholders in the State, or in each State in the consortium, including—</p> <ul style="list-style-type: none"> i. Those representing the interests of children with disabilities, English learners, and other subgroups of students described in section 1111(c)(2) of the Act; ii. Teachers, principals, and other school leaders; iii. Local educational agencies (LEAs); iv. Representatives of Indian tribes located in the State; v. Students and parents, including parents of children described in paragraph (a)(2)(i) of this section; and vi. Civil rights organizations. 	<p>Evidence that the SEA developed the innovative assessment system through consultation with the following stakeholders:</p> <ul style="list-style-type: none"> 1. Those representing the interests of children with disabilities, English learners and other subgroups of students; 2. Representatives of Indian tribes located in the State; 3. Students; 4. Parents; and 5. Civil rights organizations.

North Carolina’s Response:

The foundation of the North Carolina Personalized Assessment Tool (NCPAT) is the NC Check-Ins, an interim/formative assessment in grades 3–8 that was developed in response to input from the Task Force on Summative Assessment. Throughout its work, which began in January 2014, the task force had input from multiple stakeholders representing parent and teacher organizations, education advocacy groups, and professional education organizations. These organizations, in addition to the diversity of the task force, focused the discussions on the development of an assessment system that supports learning throughout the school year so all students, including all subgroups of students, may achieve academic success. As NC Check-Ins was developed, additional input was gathered periodically via surveys of teachers. The survey responses included teachers of all student subgroups: economically disadvantaged, English learners, students with disabilities, and ethnicities.

During the design phase of NCPAT, the North Carolina Department of Public Instruction held a stakeholder’s meeting on December 3, 2018. The meeting participants included representatives of the North Carolina Association of Educators, the North Carolina National Network of State Teachers of the Year, the North Carolina PTA, and the Office of the Governor. These participants provided perspectives cognizant of the needs of all students and all subgroups of students. To further gain input, the Testing and Growth Advisory webinar held on December 4, 2018 was open to any interested persons. Of those participating on the webinar, there was representation from the students with disabilities’ and English learners’ communities. Feedback and questions included the following:

- Whether the NCPAT design is appropriate for all students and subgroups of students, particularly those in grades 3–5
- For students who may not perform well on the through-grade assessments, will there be an opportunity to demonstrate subsequent learning on the content standards such as additional test sessions
- The importance of usable data to guide instructional decisions
- Clarification of the content being measured throughout the school year
- The impact on transitory students

With respect to input from the State’s Indian Tribes, Southside Ashpole in Robeson County has committed to participating in the IADA (see Application Appendix B Letter of Support from the Innovative School District, Southside Ashpole Elementary School). With an American Indian population of 33 percent of the total student population, this school’s teachers, parents, and students will provide valuable input on the development of the NCPAT throughout the granted authority period.

As with the NC Check-Ins’ development, on-going input and targeted opportunities for feedback will be integrated into the design and development of the NCPAT. Stakeholder input is not a singular event as evidenced in the following list of stakeholder engagement opportunities that have occurred since the submission of the IADA application in December 2018:

1. March 1, 2019: Central Carolina Regional Education Service Alliance; presentation and discussion with the Curriculum and Instruction Council, included members with expertise with English learners and students with disabilities.
2. March 28, 2019: Superintendents’ Quarterly Meeting; presentation and discussion with district superintendents
3. April 11, 2019: The Governor’s Commission on Access to Sound Basic Education; presentation and discussion on the history of assessment and accountability and the proposed innovation assessment model; members represent school and district administrators, local school boards, county commissioners, the North Carolina Association of Educators, higher education

including Duke's Sanford School of Public Policy, the judicial system, the North Carolina Justice Center, the Downeast Partnership for Children, Turning Point Workforce Development Board and the business community.

4. May 2, 2019: North Carolina State Board of Education; presentation on the history of assessment and accountability and the proposed innovation assessment model; part of a broader look at standards, assessments, and accountability at the national, state, and district levels. In addition to members of the Board, participants included the Teacher of the Year, the Principal of the Year, and the Superintendent of the Year.

INNOVATIVE ASSESSMENT SYSTEM (2) AND (3)

Regulatory Requirement	Required information from the SEA
<p>(2)(i) Align with the challenging State academic content standards under section 1111(b)(1) of the Act, including the depth and breadth of such standards, for the grade in which a student is enrolled; and</p> <p>(ii) May measure a student's academic proficiency and growth using items above or below the student's grade level so long as, for purposes of meeting the requirements for reporting and school accountability under sections 1111(c) and 1111(h) of the Act and paragraphs (b)(3) and (b)(7)–(9) of this section, the State measures each student's academic proficiency based on the challenging State academic standards for the grade in which the student is enrolled;</p> <p>(3) Express student results or competencies consistent with the challenging State academic achievement standards under section 1111(b)(1) of the Act and identify which students are not making sufficient progress toward, and attaining, grade-level proficiency on such standards;</p>	<p>Clarification regarding:</p> <ol style="list-style-type: none"> 1. The process of establishing achievement standards for the innovative assessment, the North Carolina Personalized Assessment Tool (NCPAT), that are consistent and comparable with the State academic achievement standards. In particular, clarification regarding the timeline for establishing these achievement standards is needed. 2. Information regarding the processes and procedures of the multiple event administration design, in order to ensure that all students who participate in the pilot assessments are assessed against all of the State's academic content standards (e.g., what are the procedures in the event of a student absence from one of the scheduled testing administrations).

North Carolina’s Response:

Following feedback from multiple stakeholders and technical reviews, North Carolina has revised the design of NCPAT and will not aggregate score data from all three administrations to determine students final summative score. In this revised design, the purpose of NCPAT 1 and NCPAT 2 are to:

1. Provide educators, students and stakeholders with immediate and detailed feedback on grade level specific content standards at the end of the first and second trimester so classroom instruction may be tailored to individual student's needs.
2. Serve as an indicator to determine an appropriate starting point for NCPAT 3

Likewise, the design and purposes of NCPAT 3 has been modified from the application submitted on December 14, 2019. The NCPAT 3 will be administered towards the end of the last trimester. Based on current North Carolina state statute, NCPAT 3 will be administered during the last 10 days of the school year. The scope of NCPAT 3 will be broad to include items from the entire grade level content standards but emphasis will be given to content standards covered during the last trimester for mathematics. The English language arts/reading content standards are spiraled so all content standards will be assessed on each NCPAT test.

The purposes of NCPAT 3 are to:

1. Provide educators, students and stakeholders with immediate and detailed feedback on grade level specific content standards covered during the third trimester so future instruction may be tailored to individual student's needs
2. Provide a scale score and an academic achievement level on student performance based on the entire grade level content standards for general accountability uses.

NCPAT 3 will be designed as a multistage assessment with two or more mini clusters of items where information from NCPAT 1 and NCPAT 2 will be used to route students to the cluster that best maximizes their measurement precision while exposing them to a shorter test. Matrix sampling will be used to design these mini clusters to ensure they are aligned to grade level test specifications, and across all NCPAT 3 clusters, the full breadth and depth of grade level content standards will be assessed. This design also ensures students participating in NCPAT 1 through NCPAT 3 will be assessed on the full breadth and depth of grade level content standards.

For score interpretations and uses to be valid, all students must have the opportunity to learn content standards being assessed. The ELA/reading assessed standards are a group of common standards that differs by the reading genre (information and literature). The ELA/reading standards are taught together; therefore, it is the expectation that all students would have the opportunity to learn before the end of the first trimester. The difference in terms of content coverage between NCPAT 1 and NCPAT 2 will be in the depth and complexity of the reading genre.

The mathematics content standards are more discrete, and students access them differently depending on the local curriculum. North Carolina is a local control state; there is no designated state curriculum and pacing. As a result, the design of NCPAT 1 and NCPAT 2 will require more planning. Building from our experiences with NC Check-Ins, for mathematics the plan is to find consensus from across the state on common standards that all schools are expected to cover by the end of trimester 1. Only these standards will be assessed in NCPAT 1. NCPAT 2 will only assess standards for which there is consensus that all districts across the state have covered by the time the window closes. NCPAT 3 will assess all grade level content standards for the entire school year and the testing window is the final 10 days of the school year.

During the pilot years, to ensure all students participating in the innovative assessment also receive consistent and comparable grade level state academic achievement standards, a subset of items from the regular end-of-grade assessments will be embedded into NCPAT 3 and used to link the NCPAT 3 scale with the end-of-grade scale.

INNOVATIVE ASSESSMENT SYSTEM (4)

Regulatory Requirement	Required information from the SEA
<p>(i) Generate results, including annual summative determinations as defined in paragraph (b)(7) of this section, that are valid, reliable, and comparable for all students and for each subgroup of students described in 34 CFR 200.2(b)(11)(i)(A)–(1) and sections 1111(b)(2)(B)(xi) and 1111(h)(1)(C)(ii) of the Act, to the results generated by the State academic assessments described in 34 CFR 200.2(a)(1) and section 111(b)(2) of the Act for such students.</p> <p>Consistent with the SEA’s or consortium’s evaluation plan under 34 CFR 200.106(e), the SEA must plan to annually determine comparability during each year of its demonstration authority in one of the following ways:</p> <p>(A) Administering full assessments from both the innovation and statewide assessment systems to all students enrolled in participating schools, such that at least once in any grade span (i.e., 3–5, 6–8, 9–12) and subject for which there is an innovative assessment, a statewide assessment in the same subject would also be administered to all such students. As part of this determination, the innovative assessment and statewide assessment need not be administered to an individual student in the same school year.</p> <p>(B) Administering full assessments from both the innovative and statewide assessment systems to a demographically representative sample of all students and subgroups of students described in section 1111(c)(2) of the Act, from among those students enrolled in participating schools, such that at least once in any grade span (i.e., 3–5, 6–8, 9–12) and subject for which there is an innovative assessment, a statewide assessment in the same subject would also be administered in the same school year to all students included in the sample.</p> <p>(C) Including, as a significant portion of the innovative assessment system in each required grade and subject in which both an innovative and statewide assessment are administered, items or performance tasks from the statewide assessment system that, at a minimum, have been previously pilot tested or field tested for use in the statewide assessment system.</p> <p>(D) Including, as a significant portion of the statewide assessment system in each required grade and subject in which both an innovative and statewide assessment are administered, items or performance tasks from the innovative assessment system that, at a minimum, have been previously pilot tested or field tested for use in the innovative assessment</p>	<p>NCDPI must clarify:</p> <ol style="list-style-type: none"> 1. How the SEA will ensure that all students who participate in the NCPAT pilot tests will receive an annual summative determination of proficiency from either the NCPAT OR from the statewide assessment. 2. How the SEA will ensure that all students will have had an opportunity to learn the standards being assessed at each point of administration of the NCPAT throughout the year. 3. How the SEA will aggregate data from the three interim assessments to generate results that are valid, reliable, and comparable to the state summative assessment, for all students and for each subgroup of students? 4. How the SEA will address whether valid and reliable summative scores for the innovative assessment can be established by having students answer the same linking test items multiple times (both on the innovative assessment and on the statewide assessment); OR by clarifying the procedures used when students take both the innovative and statewide assessment in the same school year.

<p>system.</p> <p>(E) An alternative method for demonstrating comparability that an SEA can demonstrate will provide for an equally rigorous and statistically valid comparison between student performance on the innovative assessment and the statewide assessment, including for each subgroup of students described in 34 CFR 200.2(b)(11)(i)(A)–(I) and sections 1111(b)(2)(B)(xi) and 1111(h)(l)(C)(ii) of the Act;</p> <p>(ii) Generate results, including annual summative determinations as defined in paragraph (b)(7) of this section, that are valid, reliable, and comparable, for all students and for each subgroup of students described in 34 CFR 200.2(b)(11)(i)(A)–(I) and sections 1111(b)(2)(B)(xi) and 1111(h)(l)(C)(ii) of the Act, among participating schools and LEAs in the innovative assessment demonstration authority. Consistent with the SEA's or consortium's evaluation plan under 34 CFR 200.106(e), the SEA must plan to annually determine comparability during each year of its demonstration authority period;</p>	
---	--

North Carolina’s Response:

IADA Assessment Revised Design Proposal

The proposed plan is to move towards a balanced assessment system that will provide granular data for immediate feedback about students’ performance throughout the year and summative data at the end of the year for general accountability purposes. The goals of this proposal are:

1. To design and evaluate an assessment system which will be used to measure and provide immediate feedback on what students know during the school year, “acquisition of knowledge,” and is also able to provide summative estimate of what students’ know based on overall grade level content standards at the end of the year, “maintenance of knowledge”.
2. To design an assessment system that seeks to strike a balance between providing formative feedback data to educators and reduce the impact of test time and test anxiety for students and schools. Each assessment in the system will be designed so it can be effectively administered within 2 hours.

The North Carolina Personalized Assessment Tool (NCPAT) will consist of 3 assessments that will each be administered at the end of each trimester during the school year. Each NCPAT will be designed to provide educators and students with immediate and detailed feedback about students estimated understanding of selected grade level content standards and skills. To ensure students have had the opportunity to learn content being assessed particularly for mathematics, the scope of NCPAT 1 and NCPAT 2 will be limited to

content standards that are expected to have been covered by the end of each trimester. This will allow Local Education Agencies to still maintain local control of their individual curriculum.

The main purposes of NCPAT 1 and NCPAT 2 are:

1. Provide educators, students and stakeholders with immediate and detailed feedback on grade level specific content standards at the end of the first and second trimester so classroom instruction may be tailored to individual student's needs.
2. Serve as an indicator to determine an appropriate start point for NCPAT 3

The design and purposes of NCPAT 3 will be slightly different from NCPAT1 and NCPAT 2. NCPAT 3 will be administered towards the end of the last trimester. Based on the current North Carolina state law, NCPAT 3 will be administered during the last 10 days of the school year. The scope of NCPAT 3 will be broad to include sample items from the entire grade level content standards but significant emphasis will be given to content standards covered during the last trimester for math.

The purposes of NCPAT 3 are to:

1. Provide educators, students and stakeholders with immediate and detailed feedback on grade level specific content standards covered during the third trimester so future instructions could be tailored to individual student's needs.
2. Provide a scale score and an academic achievement level on student performance based on the entire grade level content standards for general accountability uses.

The proposed plan is to use a multistage test design for NCPAT 3 to ensure all essential content and psychometric properties of breadth and depth of content coverage and precision/reliability are satisfied so valid inferences may be made about test scores. Each NCPAT 3 will be constructed to have a minimum of two mini item cluster sets. Matrix sampling will be used to sample and place grade level items on each cluster to match statistical target and test blueprint. Each NCPAT 3 cluster will be designed to maximize measurement precision around a critical decision point on the performance scale. For example, the test information function for Cluster 1 items will be designed to maximize precision at the 2/3 performance cut. The content scope for each NCPAT 3 Cluster will include standards covered during the final trimester as well as an overlap of content standards previously covered during the first and second trimesters. Each NCPAT 3 Cluster will span the entire performance scale so every student is given an unrestricted opportunity to demonstrate what they can do irrespective of the NCPAT 3 Cluster they are being assigned.

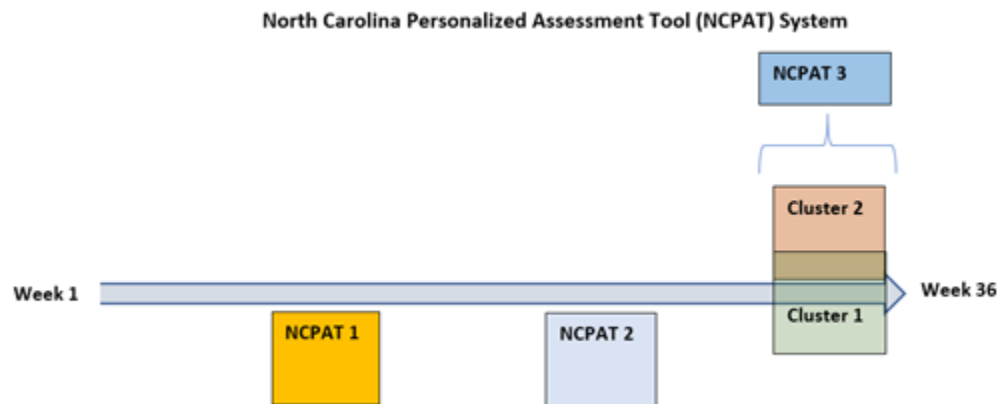
Important advantages of this design are:

- Precise Measurement: At the end of the year, students will be assigned to a Cluster that is best suited for them based on information gathered throughout the year from NCPAT 1 and NCPAT 2. Student performance from NCPAT 1 and NCPAT 2 will be used as a locator to route students to the most appropriate NCPAT 3 Cluster that will maximize measurement precision conditional to their expected ability level.
- Balanced Assessment: Sampling techniques will be used to ensure all grade level content standards are assessed throughout the year without the need for a longer end-of-year test. Students will have the opportunity to learn and receive feedback throughout the year on content covered.
- Shorter End-of-Grade Test: This proposed design can be viewed as a hybrid multi-stage adaptive design with a very reliable and robust examinee starting ability estimate. Test data from NCPAT 1 and NCPAT 2 will be used to determine student starting point for NCPAT 3. At the end of the year, all students will be assessed reliably on all grade level content standards with a shorter but more targeted NCPAT 3 designed to maximize precision.

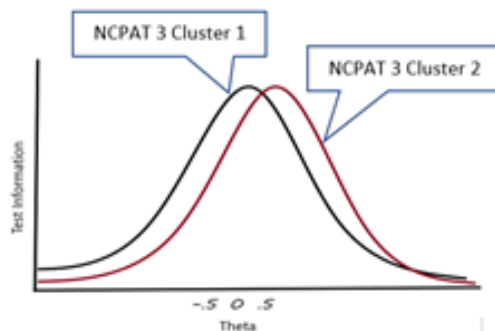
Design Features and Business Rules:

1. To ensure all students within the system have had the opportunity to learn contents being assessed, the scope for NCPAT 1 and NCPAT 2 will only include standards that all students are expected to have covered by the time each assessment is administered. This will be a strong requirement for mathematics NCPAT 1 and NCPAT 2.
2. NCPAT 1 and NCPAT 2 are considered low stake assessments with minimal test security and administration policies. Teachers will continue to administer them within a relaxed testing window when their classroom is ready based on their local curriculum and pacing.
3. All NCPAT assessments will be designed so they may be administered within two hours for all students.
4. For NCPAT 1 and NCPAT 2, teachers will get an immediate detailed report on student performance by standard. Actual test items may not be available for review, but additional reports will include granular references about each item and sample items for teachers to review.
5. For NCPAT 3, educators will get an immediate detailed report on student performance by content standard covered during the third trimester.

6. To provide teachers with formative data, items (selected response and technology-enhanced) will be developed with a descriptive summary of what is being measured and a rationale for each response option.
7. Students will be divided into groups using data from NCPAT 1 and NCPAT 2. Each group will be assigned an NCPAT 3 Cluster that is best targeted to maximize measurement precision for their ability range.
8. NCPAT 3 may include up to three (3) Clusters and each student is only expected to complete one cluster.
9. During the pilot, students in participating schools with no information from NCPAT 1 or NCPAT 2 will be administered the current EOG test.
10. The content scope for NCPAT 3 (mathematics) will over sample from standards covered in the final trimester. Matrix sampling will be used to also include standards from the entire grade level content standard. This will ensure the NCPAT system covers the entire breadth and depth of grade level content standards.
11. Each Cluster in NCPAT 3 will be designed to span the entire performance range to give all students an opportunity to demonstrate what they know and are able to do.
12. NCPAT 3 Clusters will have common items to allow for statistical linking.



NCPAT 3 Clusters Test Information Functions



Proposed Scaling Plan

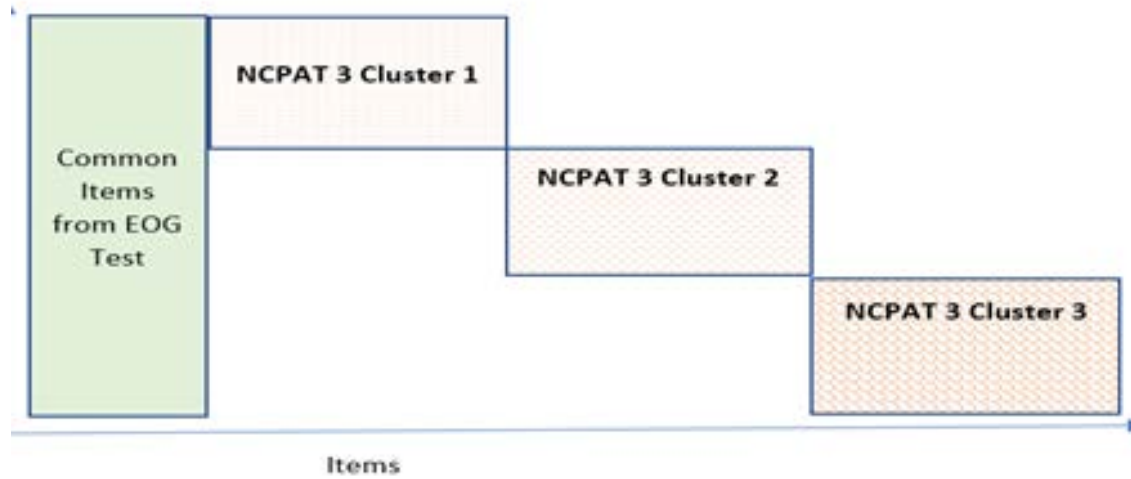
During the demonstration authority period of the NCPAT, common item non-equivalent groups data collection design will be used to ensure students participating in the NCPAT pilot will continue to have scale scores and academic achievement levels reported on the same grade level scale as non-participating students. A representative set of common items from the statewide assessment will be included on the NCPAT 3. Data from the common item set will be used to transform the NCPAT 3 scale onto the end-of-grade scale using methodology consistent with IRT scale transformation.

North Carolina currently uses a random group item pre-equated design to place parallel end-of-grade forms onto a common scale. Under the current design, parallel forms embedded with newly developed items are administered to random groups of students across the state. These new field test items are then calibrated and placed onto the common grade level IRT scale. New parallel forms are then created with pre-equated raw-to-scale tables for subsequent administration.

During the NCPAT pilot years, North Carolina will use a common item design to transform the scale of NCPAT 3 onto the end-of-grade scale. The rationale is that North Carolina will recruit a small proportion of students who might not be statistically representative of the entire grade level population to participate in NCPAT. The common item set will also be used to place the

NCPAT 3 Clusters onto a single common scale. All students participating in the IADA pilot will take the common item sets in addition to items in their assigned cluster.

Matrix Design for Common Item Equating NCPAT 3.



INNOVATIVE ASSESSMENT SYSTEM (6)

Regulatory Requirement	Required information from the SEA
<p>For purposes of the State accountability system consistent with section 1111(c)(4)(E) of the Act, annually measure in each participating school progress on the Academic Achievement indicator under section 1111(c)(4)(B) of the Act of at least 95 percent of all students, and 95 percent of students in each subgroup of students described in sections 1111(c)(2) of the Act, who are required to take such assessments consistent with paragraph (b)(1)(iii) of this section;</p>	<p>Clarification from the SEA that at least 95 percent of students in each subgroup described in section 1111(c)(2) of the Act will participate in either the state assessment or the innovative assessment.</p>

North Carolina’s Response:

Throughout each year of the pilot, North Carolina will assess at least 95 percent of all students, and 95 percent of students in each subgroup, on the North Carolina Personalized Assessment Tool (NCPAT), End-of-Grade (EOG), or NCEXTEND1 summative assessments. Results from all three assessments will be included in the state accountability model.

INNOVATIVE ASSESSMENT SYSTEM (7) AND (9)

Regulatory Requirement	Required information from the SEA
<p>(7) Generate an annual summative determination of achievement, using the annual data from the innovative assessment, for each student in a participating school in the demonstration authority that describes--</p> <ul style="list-style-type: none"> (i) The student's mastery of the challenging State academic standards under section 1111(b)(1) of the Act for the grade in which the student is enrolled; or (ii) In the case of a student with the most significant cognitive disabilities assessed with an alternate assessment aligned with alternate academic achievement standards under section 1111(b)(1)(E) of the Act, the student's mastery of those standards; 	<p>NCDPI must clarify:</p> <ol style="list-style-type: none"> 1. A description of how the innovative assessments will be combined to generate an annual summative determination of achievement. 2. The assessment that will be provided to students who require an alternate assessment.
<p>(9) Provide an unbiased, rational, and consistent determination of progress toward the State's long-term goals for academic achievement under section 1111(c)(4)(A) of the Act for all students and each subgroup of students described in section 1111(c)(2) of the Act and a comparable measure of student performance on the Academic Achievement indicator under section 1111(c)(4)(B) of the Act for participating schools relative to non-participating schools so that the SEA may validly and reliably aggregate data from the system for purposes of meeting requirements for--</p> <ul style="list-style-type: none"> (i) Accountability under sections 1003 and 1111(c) and (d) of the Act, including how the SEA will identify participating and non-participating schools in a consistent manner for comprehensive and targeted support and improvement under section 1111(c)(4)(D) of the Act; and (ii) Reporting on State and LEA report cards under section 1111 (h) of the Act. 	<p>Evidence provided to address the concerns noted in section (7) will also address this section.</p>

North Carolina's Response:

North Carolina has consulted with its technical advisory committee and other research partners to investigate psychometrically reliable and valid methods to generate summative scores from multiple administrations of interim assessments. As part of its effort in 2017, North Carolina partnered with The National Center for the Improvement of Educational Assessment Inc. to conduct an empirical study. The purpose of this exploratory research study was to examine how such a single summative score might be created and evaluated within the context of North Carolina's interim assessments. Consistent with research in this area, three approaches to derive a single summative score were explored (Wise, 2011): The maximum of the interim assessment sum scores, the mean of the interim

assessment sum scores, and the weighted mean of interim assessment sum scores. Results from these multiple approaches did not vary sustainably from one another, which was not surprising given (i) the relatively high correlations between interim assessments, and (ii) that each approach is like one another in terms of the algebraic formula.

However, as captured in the discussion section of the report and in reviewing these results with our technical advisory committee it was apparent there are several significant interpretative and practical issues with this design. Most notable are what summative interpretative claims should be made about students. For example, should the claim be about average student performance in a fashion similar to the way in which course grades or grade point averages are defined? Should the claim be about a student's best performance, similar to the way in which a student's best work is selected for a portfolio? Or should the claim be about a composite that weights each assessment according to some value judgment, similar to the way different kinds of work contribute more or less to a student's course grade? Also, this design did not account for student growth throughout the year and a consistent procedure to handle missing data and transient students.

Following these feedbacks, North Carolina has revised its design and will no longer attempt to combine scores from all three assessments into a single summative score. The revised plan is to use information from NCPAT 1 and NCPAT 2 to classify students into performance groups and route each group to the NCPAT 3 cluster sets that will maximize their measurement precision.

Students identified to take the alternate assessment will continue to participate in the NCEXTEND1 assessment.

See Appendix A: Combining Information across Interim Assessments

INITIAL IMPLEMENTATION IN A SUBSET OF LEAS OR SCHOOLS

Regulatory Requirement	Required information from the SEA
<p>If the innovative assessment system will initially be administered in a subset of LEAs or schools in a State--</p> <ol style="list-style-type: none"> 1. A description of each LEA, and each of its participating schools, that will initially participate, including demographic information and its most recent LEA report card under section 1111(h)(2) of the Act; and 2. An assurance from each participating LEA, for each year that the LEA is participating, that the LEA will comply with all requirements of this section.; 	<p>Demographic information and report cards for the initial participating schools, or a plan for when that can be provided.</p>

North Carolina’s Response:

Since the submission of the application on December 14, 2019, additional districts have indicated interest in participating in the IADA. The stakeholder opportunities cited in the Consultation section and the ensuing discussion of the innovative assessment design have led additional districts and charter schools to commit to participation. These discussions will continue with the goal of having firm commitments for participation by late fall 2019. Such commitments are typically at the district level, but during the planning year, specific school participation within a district will be determined with the purpose of reflecting the demographics of the State. The demographic information for each school and the NC Report Card for each school will be provided by January 2020 to the U.S. Department of Education.

APPLICATION SELECTION CRITERIA

(A)(3)

Application Selection Criteria	Required information from the SEA
<p>If the system will initially be administered in a subset of schools or LEAs in a State--</p> <ul style="list-style-type: none"> (i) The strategies the SEA, including each SEA in a consortium, will use to scale the innovative assessment to all schools statewide, with a rationale for selecting those strategies; (ii) The strength of the SEA's or consortium's criteria that will be used to determine LEAs and schools that will initially participate and when to approve additional LEAs and schools, if applicable, to participate during the requested demonstration authority period; and (iii) The SEA's plan, including each SEA in a consortium, for how it will ensure that, during the demonstration authority period, the inclusion of additional LEAs and schools continues to reflect high-quality and consistent implementation across demographically diverse LEAs and schools, or contributes to progress toward achieving such implementation across demographically diverse LEAs and schools, including diversity based on enrollment of subgroups of students described in section 1111(c)(2) of the Act and student achievement. The plan must also include annual benchmarks toward achieving high-quality and consistent implementation across participating schools that are, as a group, demographically similar to the State as a whole during the demonstration authority period, using the demographics of initially participating schools as a baseline. 	<p>Additional information about the State's strategy to scale statewide is needed, including:</p> <ol style="list-style-type: none"> 1. Strategies for encouraging or requiring participation in order to scale the innovative assessment system statewide. 2. A plan for ensuring demographic diversity, including diversity of subgroups, among participating LEAs and schools, and ensuring that participating schools are, as a group, demographically similar to the State. 3. Annual benchmarks toward achieving high-quality and consistent implementation across participating schools.

North Carolina's Response:

The NCDPI will require up to 30% of the student population for each grade level/content area to participate in the IADA beginning with Year 2 and continuing through Year 4 of the demonstration authority period. With the release of the State's application and subsequent discussions with stakeholders, much of the feedback from districts and schools has been concern with the initial plan to require participating students take both the NCPAT and the current end-of-grade test. As stated in this response to the USED, the NCDPI is utilizing a linking design for the NCPAT 3 and the end-of-grade assessment so the participating students will only take one

assessment, not two. This modification of the design of the NCPAT during the IADA period will remove a primary barrier to voluntary participation.

North Carolina has the support of state statute if it becomes necessary to require participation to meet the 30% requirement or to meet the demographic diversity, including diversity of subgroups, requirement. However, it is preferred the requirements are met through voluntary participation, and without the requirement that students participate in two assessments, this is very doable. To accomplish this, the NCDPI will communicate the following to LEA and schools:

- The NCPAT's design will provide student-level information throughout the school year so additional instruction may be provided.
- The tests will be shorter in length and the NCPAT 3 will provide a more precise estimate of a student's performance on the end-of-year assessment.
- Participating districts and schools will be at the table with the NCDPI team in the development and design of communication documents, score reports, individual student reports, etc.
- At the conclusion of each school year, participating districts and schools will have the opportunity to provide input on what went well and what did not, with respect to logistics and policy so subsequent changes can be implemented that will optimize the assessment for all students and subgroups of students.

This communication will be shared in multiple ways: (1) monthly regional trainings led by our Regional Accountability Coordinators, (2) periodic webinars open to all district and charter school staff, (3) presentations at the Regional Support Team meetings, and (4) conference sessions across the state.

In addition to reviewing the historical and the NCPAT assessment data for all participating districts and schools, the NCDPI will monitor selected schools to affirm consistency in the administration of the assessments. Monitoring teams will conduct on-site visits on test day to document the adherence to policies and practices and to gain insight on possible changes for the administrations. As part of the monitoring visits, the NCDPI team will seek input from the teachers and administrators on the usability of the NCPAT 1 and NCPAT 2 for improving instructional outcomes and on the costs/benefits of the system.

(B)(2)

Application Selection Criteria	Required information from the SEA
<p>The extent and depth of SEA, including each SEA in a consortium, and LEA capacity to implement the innovative assessment system considering the availability of technological infrastructure; State and local laws; dedicated and sufficient staff, expertise, and resources; and other relevant factors. An SEA or consortium may also describe how it plans to enhance its capacity by collaborating with external partners that will be participating in or supporting its demonstration authority. In evaluating the extent and depth of capacity, the Secretary considers—</p> <ul style="list-style-type: none">(i) The SEA's analysis of how capacity influenced the success of prior efforts to develop and implement innovative assessments or innovative assessment items; and(ii) The strategies the SEA is using, or will use, to mitigate risks, including those identified in its analysis, and support successful implementation of the innovative assessment.	<p>A description of the strategies NCDPI is using, or will use, to mitigate risks, including those identified in its analysis, and support successful implementation of the innovative assessment.</p>

North Carolina's Response:

The lesson learned from designing, developing and implementing the NC Check-Ins was to mitigate risk. It is necessary (1) to have an extensive communication plan that includes repeated opportunities for feedback and (2) to be willing to listen to such input and make needed changes. With our statewide assessment, typically we gather input from the test specification panelists, but that is a one-time event. Likewise, we may receive periodic feedback and change policies or procedures, but such feedback is not as intentionally gathered as it was with NC Check-Ins. With the NC Check-Ins, training on the purpose of the assessment occurred at the beginning of the school year, and the training was not limited to our train-the-trainer model. The trainings were designed for teachers and co-hosted by our accountability team and our Standards, Curriculum and Instruction colleagues in the Department. This approach placed the emphasis on the classroom instruction: The purpose of the NC Check-Ins is to provide timely data to adjust instruction. So, as all testing experts repeatedly stress, it is all about the purpose of the test.

Another lesson learned from NC Check-Ins is that useful data is most valued. With NC Check-Ins, the NCDPI provided student-level data on each item. This information included the content standard assessed, the Depth of Knowledge, and the answer choice selected by the student. Teachers and curriculum planners used this information to plan subsequent instruction.

Applying these lessons to the development of NCPAT is critical to its potential for success. Thus, to mitigate risks, beginning with the planning year, the NCDPI will do the following:

- For each year of the IADA period, the NCDPI will host three webinars with teachers, principals, and district-level staff to share the purpose of the NCPAT and to listen to the participants' needs.
- The webinars will be co-hosted with our colleagues in the Standards, Curriculum and Instruction Division, the Exceptional Children's Division, and the English Learners' Section.
- Across all administrations, the NCDPI will monitor the administrations at schools and use the opportunity to gather verbal feedback from teachers and principals and if allowed, students.
- After each testing event in each year, the NCDPI will collect survey data from teachers, principals and district-level staff for feedback, particularly on the use of the data at the classroom-level.
- The NCDPI will monitor the Online Testing Irregularity System to identify areas that require action such as security breaches.

With these strategies in place, the NCDPI will ensure the development of a system that changes the use of assessment data in schools.

(C)(1)

Application Selection Criteria	Required information from the SEA
The extent to which the timeline reasonably demonstrates that each SEA will implement the system statewide by the end of the requested demonstration authority period, including a description of-- (i) The activities to occur in each year of the requested demonstration authority period; (ii) The parties responsible for each activity; and (iii) If applicable, how a consortium's member SEAs will implement activities at different paces and how the consortium will implement interdependent activities, so long as each non-affiliate member SEA begins using the innovative assessment in the same school year consistent with 34 CFR part 200.104(b)(2);	An amended timeline that includes training for teachers in item development, communications with schools and parents such as score reporting, and revisions to the assessment system.

North Carolina's Response:

Year One: 2019–20 School Year

Planning Year

Develop: Grade 4 Mathematics and Grade 7 ELA/Reading

	Task/Deliverable	Responsible
1	Convene teacher panels for test specifications for: Grade 4 Mathematics and Grade 7 ELA/Reading	NCDPI/NC State
2	Finalize test specifications	NCDPI
3	Finalize Analyses Plan for Demonstration Period, including standard setting	NCDPI w/Technical Advisors
4	Contract with teachers to write and review items. Process for item writing and training already in place.	NCDPI/NC State
5	Embed items in operational end-of-grade assessments	NCDPI/NC State
6	Review online delivery system for innovative assessment	NCDPI/NC State
7	Develop professional development materials for schools and conduct training: Teachers and Administrators: • Assessment Literacy- purpose, design, and data reports of NCPAT	NCDPI

	Task/Deliverable	Responsible
	<ul style="list-style-type: none"> • Data Literacy- how to read and use the NCPAT data reports (Item level, standard level, teacher level, school level) • Test Administration- how to administer NCPATs Students: <ul style="list-style-type: none"> • Online tutorials to become familiar with the testing platform and item types used in NCPATs 	
8	Develop training materials for state regional support teams and regional accountability coordinators and conduct: <ul style="list-style-type: none"> • Assessment Literacy- purpose, design, and data reports of NCPAT • Data Literacy- how to read and use the NCPAT data reports (Item level, standard level, teacher level, school level) • Test Administration- how to administer NCPATs 	NCDPI
9	Administer survey to teachers, administrators, district/charter school staff (on-going)	NCDPI
10	Conduct focus groups on alignment	NCDPI
11	Develop Communication Plan	NCDPI
12	Disseminate parent communication about innovative pilot (on-going)	NCDPI
13	Assess 95% or more of all students in NC on the End-of-Grade or NCEXTEND1 summative assessments and include these results in the state accountability model.	NCDPI

Year Two: 2020–21 School Year

Assess: Grade 4 Mathematics and Grade 7 ELA/Reading

Develop: Grade 4 & 6 ELA/Reading, Grade 6 & 7 Mathematics

	Task/Deliverable	Responsible
1	Build test forms from embedded items in 2019–20: Grade 4 Mathematics and Grade 7 ELA/Reading.	NCDPI/NC State
2	Select and embed 15 anchor items from the EOG onto the NCPAT 3.	NCDPI/NC State
3	Administer test forms (NCPAT 1 ~ week 12, NCPAT 2 ~ week 24, and NCPAT 3 during the last 10 days of the school year.)	NCDPI
4	Conduct administration observations, accommodations monitoring, and cognitive labs.	

	Task/Deliverable	Responsible
5	Analyze data from each administration, including growth.	NCDPI and SAS
6	Post equate the NCPAT 3 scale on to the EOG scale to assign achievement levels for students participating in NCPAT.	NCDPI
7	Update/modify training materials as needed and conduct professional development for: Teachers: <ul style="list-style-type: none"> • Assessment Literacy- purpose, design, and data reports of NCPAT • Data Literacy- how to read and use the NCPAT data reports (Item level, standard level, teacher level, school level) • Test Administration- how to administer NCPATs Students: <ul style="list-style-type: none"> • Online tutorials to become familiar with the online platform and item types used in NCPATs 	NCDPI
8	Update/modify training material as needed and conduct training for state regional support teams and regional accountability coordinators: <ul style="list-style-type: none"> • Assessment Literacy- purpose, design, and data reports of NCPAT • Data Literacy- how to read and use the NCPAT data reports (Item level, standard level, teacher level, school level) • Test Administration- how to administer NCPATs 	NCDPI
9	Administer survey to teachers, district/charter school staff	NCDPI
10	Conduct focus groups	NCDPI
11	Assess 95% or more of all students and 95% of students in subgroups, on the NCPAT 3, End-of-Grade, or NCEXTEND1 summative assessments and include these results in the state accountability model.	NCDPI
12	Convene teacher panels for test specifications for: Grade 4 & 6 ELA/Reading, Grade 6 & 7 Mathematics	NCDPI/NC State
13	Finalize test specifications	NCDPI
14	Contract with teachers to write and review items. Process for item writing and training already in place.	NCDPI/NC State
15	Select and embed 15 anchor items from the EOG onto the NCPAT 3.	NCDPI/NC State
16	Review online delivery system for innovative assessment	NCDPI/NC State

Year Three: 2021–22 School Year

Assess: ELA/Reading and mathematics in Grade 4, 6 & 7

Develop: ELA/Reading and mathematics Grades 3, 5 & 8

	Task/Deliverable	Responsible
1	Build test forms from embedded items in 2020–21: Grade 4 & 6 ELA/Reading, Grade 6 & 7 Mathematics	NCDPI/NC State
2	Evaluate linking relationship between NCPAT and EOG scale and check anchor items for any potential drift.	NCDPI/NC State
3	Administer test forms: ELA/Reading and mathematics in Grade 4, 6 & 7	NCDPI
4	Conduct administration observations, accommodations monitoring, and cognitive labs	NCDPI
5	Analyze data from each administration, including growth.	NCDPI
6	Evaluate pre-equated raw-to-scale tables for NCPAT 3 and reliability of equating procedure.	NCDPI
7	Update/modify training material as needed and conduct training for state regional support teams and regional accountability coordinators: <ul style="list-style-type: none"> • Assessment Literacy- purpose, design, and data reports of NCPAT • Data Literacy- how to read and use the NCPAT data reports (Item level, standard level, teacher level, school level) • Test Administration- how to administer NCPATs 	NCDPI
8	Update/modify training materials as needed and conduct professional development for: <p>Teachers:</p> <ul style="list-style-type: none"> • Assessment Literacy- purpose, design, and data reports of NCPAT • Data Literacy- how to read and use the NCPAT data reports (Item level, standard level, teacher level, school level) • Test Administration- how to administer NCPATs <p>Students:</p> <ul style="list-style-type: none"> • Online tutorials to become familiar with the online platform and item types used in NCPATs 	NCDPI
9	Administer survey to teachers, district/charter school staff	NCDPI
10	Conduct focus groups	NCDPI

	Task/Deliverable	Responsible
11	Assess 95% or more of all students in NC on the End-of-Grade or NCEXTEND1 summative assessments and include these results in the state accountability model.	NCDPI
12	Convene teacher panels for test specifications for: ELA/Reading and mathematics Grades 3, 5 & 8	NCDPI/NC State
13	Finalize test specifications	NCDPI
14	Contract with teachers to write and review items. Process for item writing and training already in place.	NCDPI/NC State

Year Four: 2022–23 School Year

Assess: ELA/Reading and mathematics in Grade 4, 6 & 7

Develop: ELA/Reading and mathematics Grades 3, 5 & 8

	Task/Deliverable	Responsible
1	Select and embed 15 anchor items from the EOG onto the NCPAT 3.	NCDPI/NC State
2	Administer test forms: ELA/Reading and mathematics in Grade 4, 6 & 7	NCDPI
3	Conduct administration observations, accommodations monitoring, and cognitive labs	
4	Analyze data from each administration, including growth.	NCDPI
5	Evaluate pre-equating design and reliability of raw-to scale tables	NCDPI
6	Update/modify training materials as needed and conduct professional development for: Teachers: <ul style="list-style-type: none"> • Assessment Literacy- purpose, design, and data reports of NCPAT • Data Literacy- how to read and use the NCPAT data reports (Item level, standard level, teacher level, school level) • Test Administration- how to administer NCPATs Students: <ul style="list-style-type: none"> • Online tutorials to become familiar with the online platform and item types used in NCPATs 	NCDPI
7	Update/modify training material as needed and conduct training for state regional support teams and regional accountability coordinators: <ul style="list-style-type: none"> • Assessment Literacy- purpose, design, and data reports of NCPAT 	NCDPI

	Task/Deliverable	Responsible
	<ul style="list-style-type: none"> • Data Literacy- how to read and use the NCPAT data reports (Item level, standard level, teacher level, school level) • Test Administration- how to administer NCPATs 	
8	Administer survey to teachers, district/charter school staff	NCDPI
9	Conduct focus groups	
10	Assess 95% or more of all students in NC on the End-of-Grade or NCEXTEND1 summative assessments and include these results in the state accountability model.	NCDPI
11	Contract with teachers to write and review items. Process for item writing and training already in place.	NCDPI/NC State

Year Five: 2023–24 School Year

Assess: Grades 3–8 Mathematics and Grades 3–8 ELA/Reading Statewide

	Task/Deliverable	Responsible
1	Build test forms: ELA/Reading and mathematics Grades 3, 5 & 8	NCDPI/NC State
3	Administer test forms: Grades 3–8 Mathematics and Grades 3–8 ELA/Reading	NCDPI
4	Conduct administration observations, accommodations monitoring, and cognitive labs	
5	Analyze data from each administration, including growth.	NCDPI
6	Conduct standard setting for academic achievement standards	NCDPI/External Vendor
7	Update/modify training materials as needed and conduct professional development for: Teachers: <ul style="list-style-type: none"> • Assessment Literacy- purpose, design, and data reports of NCPAT • Data Literacy- how to read and use the NCPAT data reports (Item level, standard level, teacher level, school level) • Test Administration- how to administer NCPATs Students: <ul style="list-style-type: none"> • Online tutorials to become familiar with the online platform and item types used in NCPATs 	NCDPI

	Task/Deliverable	Responsible
8	Update/modify training material as needed and conduct training for state regional support teams and regional accountability coordinators: <ul style="list-style-type: none"> • Assessment Literacy- purpose, design, and data reports of NCPAT • Data Literacy- how to read and use the NCPAT data reports (Item level, standard level, teacher level, school level) • Test Administration- how to administer NCPATs 	NCDPI
9	Administer survey to teachers, administrators, and district/charter school staff (on-going)	NCDPI
10	Conduct focus groups (on-going)	NCDPI

(C)(2)

Application Selection Criteria	Required information from the SEA
<p>The adequacy of the project budget for the duration of the requested demonstration authority period, including Federal, State, local, and non-public sources of funds to support and sustain, as applicable, the activities in the timeline under paragraph (c)(1) of this section, including-</p> <ul style="list-style-type: none">(i) How the budget will be sufficient to meet the expected costs at each phase of the SEA's planned expansion of its innovative assessment system; and(ii) The degree to which funding in the project budget is contingent upon future appropriations at the State or local level or additional commitments from non-public sources of funds.	<p>An estimated budget for the innovative assessment system at each phase of NC's planned expansion of its innovative assessment system.</p>

North Carolina's Response:

The estimated budget for the NCPAT is based on the current costs for the end-of-grade assessments. With the linking model not requiring students to take both the NCPAT and the end-of-grade assessment, the development of the NCPAT is not an additional cost.

The work required for the development of the NCPAT will be fulfilled by existing staff at the NCDPI and North Carolina State University. The allocation of responsibilities will shift as needed. For example, the Test Measurement Specialists will oversee the test specifications panels, the reviewing of field test items, the selection of operational items, and the quality control checks of all delivered test forms. These responsibilities are the same as their responsibilities for the current assessments. As the IADA is implemented, more time will be allocated for it as opposed to the current assessments, so the amount of time spend on the current assessments will decrease and the time spent on the IADA will increase.

The estimated cost for the current mathematics end-of-grade test is \$6.44 per student and the estimated cost for the current English language arts/reading end-of-grade test is \$6.42 per student. With 30% of the students at each grade level participating in the NCPAT, the estimated cost for each year is as shown in the chart below:

Year	Number of Students	Mathematics Cost (\$6.44 per Student)	ELA/Reading Cost (\$6.42 per Student)
1 (Planning)	N/A	N/A	N/A
2 (Grade 4 Mathematics and 7 ELA/Reading)	228,992	\$1,474,708	1,470,129
3 (Grades 4, 6, and 7 Mathematics and ELA/Reading)	342,336 each content area	\$2,204,644	\$2,197,797
4 (Grades 4, 6, and 7 Mathematics and ELA/Reading)	342,336 each content area	\$2,204,644	\$2,197,797
5 (Statewide 3–8 Mathematics and ELA/Reading)	681,543 each content area	\$4,389,137	\$4,375,506
Total		\$10,273,133	\$10,241,229

(D)(1)

Application Selection Criteria	Required information from the SEA
The extent to which the SEA or consortium has developed, provided, and will continue to provide training to LEA and school staff, including teachers, principals, and other school leaders, that will familiarize them with the innovative assessment system and develop teacher capacity to implement instruction that is informed by the innovative assessment system and its results;	A detailed description of the training that will be provided to LEA and school staff.

North Carolina’s Response:

The NCDPI currently has a statewide train the trainer model in place. Consultants at the department develop trainings then train our Regional Accountability Coordinators (RACs) and our Regional Computing Consultants (RCCs). The RACs and RCCs train the LEA test coordinators, the LEA test coordinators train the school testing coordinators, and the school coordinators train the teachers. For the first 3 years, the NCDPI consultants will partner with the RACs and RCCs to provide direct training to our pilot schools. In year 5, we will move towards the training the trainer model already in place.

The following trainings will be developed and disseminated through webinars and in person trainings:

- **Assessment Literacy**- purpose, design, and data reports of NCPATs
- **Data Literacy**- how to read and use the NCPAT data reports (item level, standard level, domain level, teacher level, school level)
- **Test Administration training**- how to administer NCPATs

Also please see information in (B)(2).

(D)(2)

Application Selection Criteria	Required information from the SEA
The strategies the SEA or consortium has developed and will use to familiarize students and parents with the innovative assessment system;	A description of the strategies NC will use to familiarize students with the innovative assessment system.

North Carolina’s Response:

The NCPAT assessment system is modeled on the NC Check-Ins which are currently administered to approximately 60% of students in grades 3–8. The NCDPI will build on the existing familiarity with the NC Check-Ins framework to ensure students and teachers understand the system. Additionally, the NCDPI will work with Technical Outreach for Public Schools (TOPS) at North Carolina State University to develop online tutorials and practice activities for students to become familiar with the online platform and item types used in NCPATs. The NCDPI will also release test specifications and items for parents, teachers, and students to help familiarize them with the format of the test, item type, and rigor of questions. To help teachers, communicate with their students, student friendly information regarding the purpose, design, and types of reports generated from the assessments will be shared during trainings.

(E)(1)

Application Selection Criteria	Required information from the SEA
The strength of the proposed evaluation of the innovative assessment system included in the application, including whether the evaluation will be conducted by an independent, experienced third party, and the likelihood that the evaluation will sufficiently determine the system's validity, reliability, and comparability to the statewide assessment system consistent with the requirements of 34 CFR part200.105(b)(4) and (9);	Additional information about the State's strategy to evaluate the innovative assessment is needed, including: <ol style="list-style-type: none">1. Evidence of a plan for evaluation of the innovative assessment system, particularly how the evaluation will determine the system's:<ol style="list-style-type: none">A. ReliabilityB. ValidityC. Comparability to the statewide assessment system2. A timeline for the evaluation plan proposed throughout the period of the authority.

North Carolina's Response:

1. Additional plan for measuring reliability and validity evidences of the proposed North Carolina Innovative Assessment Pilot tests and comparability to the statewide assessment system.

The resulting scores from the innovative assessment system will be used for instructional adjustments and accountability purposes. A plan for gathering additional information on reliability and validity evidences of the North Carolina innovative assessment system are described below. These evidences will be collected throughout the years and over administrations.

- Internal consistency measured by Cronbach alpha (Cronbach, 1951) is one of the methods generally used to estimate test reliability. Cronbach alpha reliability estimate will be computed and documented for all NCPAT assessments to ensure they meet or exceed industry standards. Reliability estimates will also be computed and compared for sub-groups (gender, ethnicity, students with disability and economically disadvantaged students).
- Reliability also refers to consistency of constructs or Unidimensionality of NCPAT assessments over years. The NCDPI will conduct dimensionality analysis for each NCPAT and continue to make improvements as needed to ensure each NCPAT is designed to be unidimensional.
- “Validity refers to the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests.” (Standards, 2014). Validity process therefore involves collecting evidences to support proposed score interpretations.

- Content related evidence: NCPAT tests will be constructed using the same content and statistical specifications as the grade level end-of-grade (EOG) tests. The NCDPI will gather content related evidence to show NCPAT are aligned to North Carolina Standard Course of Study (NCSCS). During the test development, the NCDPI Psychometricians and Test Measurement Specialists (TMSs) will focus on closely matching the classical and item response theory-based parameters including mean P-value, Point-Biserials, IRT parameters, Test Characteristic Curves (TCCs), and Test Information Functions (TIFs) for a given test over years.
- Item and Test Alignment evidences: North Carolina will continue to gather all related alignment evidence throughout the demonstrated pilot. Teachers will continue to write items for NCPAT and only items that have gone through all test development checks and have been field tested will be used for NCPAT. In our current design, every new item must undergo three separate independent alignment evaluations before it is administered to students. Finally, the plan is to conduct an independent test alignment study on all NCPAT forms in year 5.
- Opportunity to learn is an important concept to ensure fair and valid interpretation of scores. North Carolina is a local control state which allow local school districts to adopt their individual curriculum. The content scope of NCPAT 1 and NCPAT 2 for mathematics will be limited to content standards that have been commonly agreed all students have had the opportunity to learn. North Carolina is going to survey participating teachers and students to ensure there are no threats to opportunity to learn.
- Evidence based on relations to other variables: if a test measures what it is intended to measure, then the results must agree (correlate reasonably) with results from other tests external to the given test. Students performance from NCPAT are expected to be correlated with other related measures such as previous year test score or other related measures.

2. Timeline for the evaluation plan proposed throughout the period of the authority

Table below lists timelines and tasks to collect evidences for the program evaluation.

Year	Tasks
Year 1 (2019–20)	1. Year 1 is a planning year. The NCDPI will develop test specifications, develop items by embedding in the operational tests, review online testing platform, select sample schools for grade 4 mathematics and grade 7 ELA, and prepare for professional development and training.

Year 2 (2020–21)	<p>1. In Year 2, grade 4 mathematics and grade 7 ELA tests will be developed and administered.</p> <p>The NCDPI will:</p> <ol style="list-style-type: none"> a. Document and compare classical (p-value, Pbis, and alpha) and IRT statistics (parameters, TCCs and TIFs) of the interim, clusters, and EOG tests. b. Perform correlation and dimensionality analysis. c. Conduct survey of teachers and administrators to facilitate evaluation of the pilot. d. Assess usability of the online delivery system. e. Evaluate calibration and equating results and monitor performance of items including anchor items. f. Evaluate classical, IRT, and survey results; usability of the online delivery results; teacher and administrator survey results; results comparing NCPAT3 and EOG; and comparing the pilot study sample results from clusters with current as well as previous year’s EOG results. g. Produce technical report documenting the evaluation results. <p>2. The sample identified in Year 2 for grade 4 math and grade 7 ELA will continue to take the pilot tests in Year 3. Additional samples will be selected for grade 4 ELA, grade 6 ELA/mathematics, and grade 7 mathematics.</p>
Year 3 (2021–22)	<p>1. In Year 3, the NCDPI will:</p> <ol style="list-style-type: none"> a. Continue to administer both ELA and mathematics tests for grades 4, 6, and 7 samples. b. Repeat the same analysis and evaluation processes that was used in Year 2 and monitor the results. c. Compare results between Year 3 and Year 2. d. Produce technical report documenting the evaluation results.
Year 4 (2022–23)	<p>1. In Year 4, the NCDPI will:</p> <ol style="list-style-type: none"> a. Continue to administer both ELA and mathematics tests for the same grades and contents. b. Repeat the same analysis and evaluation processes that was performed in Year 3 and monitor results. c. Compare results between Year 4, Year 3, and Year 2. d. Produce technical report documenting the evaluation results. e. Evaluate the trend across administrations. Based on the results from the pilot study, decide whether to continue or abandon the innovative assessments.
Year 5 (2023–24)	<ol style="list-style-type: none"> 1. If decided to continue and use the model for future testing, the NCPAT pilot design will be rolled to statewide implementation for grades 3–8 in mathematics and ELA. 2. A standard setting will be conducted in the Summer of 2024. 3. The results will be submitted to NCSBE for approval on August 2024 meeting.

Combining Information across Interim Assessments: Considering a Single Summative Score based on North Carolina Interim Assessments

DRAFT–9/19/17

Nathan Dadey & Brian Gong

The National Center for the Improvement of Educational Assessment, Inc.

Introduction

North Carolina’s Proof of Concept (POC) Interim Assessment Study provides a unique opportunity to investigate questions key to the interpretation and use of interim assessment scores. One question that has been brought to the fore by the Every Student Succeeds Act (ESSA) is how the results of multiple interim assessments can be combined to produce a “single summative score”. The purpose of this exploratory research study¹ is to examine how such a single summative score might be created and evaluated within the context of North Carolina’s interim assessments.

Context

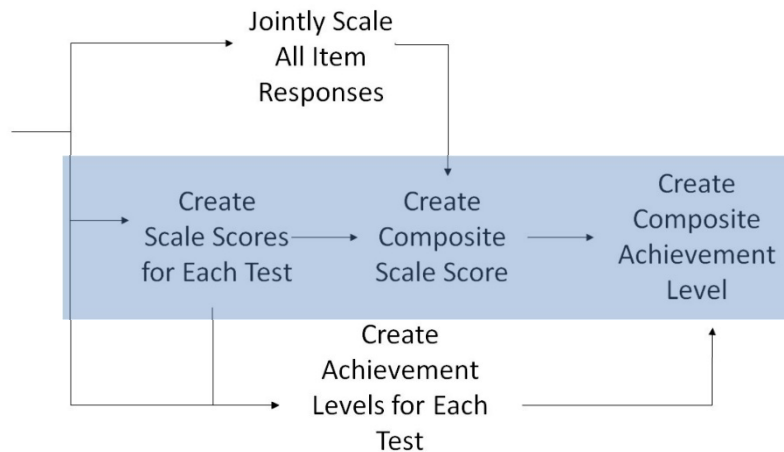
ESSA includes a provision a state’s accountability assessments may “be administered through multiple statewide interim assessments that result in a single summative score that provides valid, reliable, and transparent information on student achievement or growth” (ESSA, §1111(b)(2)(B)(viii)). The implication is that this single summative score will be used as the indicator of achievement within a given state’s accountability system. Therefore, this single summative score will need to fulfill the role that the state’s current indicator of academic achievement plays (or these roles will need to be revised). At bare minimum then, this single summative score is a proficiency or achievement level classification – in other words, a single summative achievement level. In addition, given most states’ context, a scale score will also need to be produced – a single summative scale score.

There are a several key decision points to be made in the creation of a single summative score. Specifically, decisions made around the appropriate **level of aggregation** and the **aggregation method** will interact with the **design of the interim assessments**, potentially resulting in different judgements about students. By level of aggregation, we mean the unit at which results are combined into a single summative score. The level of aggregation could be at the item level, in which all of the item responses to all assessments are treated as if they came from the same test and scaled accordingly, or at the test level, in which scale scores or achievement levels from each interim assessment are combined into a single summative scale score or achievement level. The choices involved with the level of aggregation are shown in Figure 1, which illustrates that there are multiple ways in which the level of aggregation

¹For considerations of other issues key related to the implementation of an ESSA compliant system of interim assessments, see the brief entitled *Using interim assessments in place of summative assessments? Consideration of an ESSA Option*, available at http://www.ccsso.org/Resources/Publications/Using_Interim_Assessments_in_Place_of_Summative_Assessments_-_Consideration_of_an_ESSA_Option.html

may be addressed. For example, a scale score could be created for each interim assessment, then aggregated to create the single summative score, and finally that single summative score could be used to create a single summative achievement level. Alternatively, all of the items could be used to create the single summative score, through methods like item response theory – skipping the creation of scale scores for each test.

Figure 1. Flowchart illustrating the Levels of Aggregation for Interim Assessment Results. Note: the blue box indicates the approaches examined within this



The above flowchart illustrates the levels at which aggregation may take place, but does not define the methods by which the results will be aggregated. While there are numerous aggregation methods, in the context of a simulation study, Wise (2011) explored three different approaches to aggregating the results of multiple interim assessments - taking the maximum, mean or weighted mean of sum scores. In addition, Wise also suggested that, in some cases, the results from interim assessments can be treated as “if they were different sections of the same test” (Wise, 2011, p. 11) and thus combined through psychometric scaling (e.g., subjecting all of the item response data to one item response theory calibration) or through the aggregation of each assessment score into a single composite index or scale (e.g., adding the scores from each interim together, as is often done with assessment batteries).

The case Wise is referring to is when the blueprints from each interim assessment differ in terms of their coverage of the standards. Under this “modular” or “differing” blueprint design, each interim assessment blueprint covers a specific subset of the standards with minimal overlap. In this case, the results must be combined not only to satisfy ESSA, but also to represent the full set of content standards. In addition, the results can, potentially, be combined as if they were different sections of the same test². The key conceptual problem under the modular blueprint design is determining how to combine the assessment scores to most appropriately represent the content standards. North Carolina’s POC mathematics assessments generally follow the modular design. In contrast to the modular design, under the “mini-summative” or “same” blueprint design, each interim assessment covers the same or

²Wise introduces this model in terms of assessments that are tied to specific units of instruction and administered directly after instruction.

similar set of standards. The key conceptual problem for the mini-summative design is summarizing student performance that is likely changing over time. That is, it is generally expected that student performance on a set of standards changes over the course of an academic year, hopefully for the better. The summative score under this design must therefore determine how to summarize across these changes. North Carolina's POC English Language Arts (ELA) assessments generally follow the mini-summative design.

Methods

In this section we first summarize North Carolina's POC interim assessments and then define our approach to creating single summative scores, based on subset of the multiple approaches defined in the prior section.

Data

The data come from a set of interim assessments administered by North Carolina's Department of Public Instruction (NCDPI) during the 2015-2016 school year. Three interim assessments in fifth grade mathematics and three assessments in sixth grade ELA were administered to stratified random sample of students. The fifth-grade mathematics data sets contained 3,964 students and the sixth grade ELA assessment contained 3,961 students. The assessments were administered as pencil and paper forms within NCDPI defined assessment windows: the window for the first interim was October 1–30, 2015, for the second, December 8, 2015–January 22, 2016 and for the third, March 3–31, 2016. In addition to taking the interim assessments, students within the sample also took a shorted version of North Carolina's End of Grade (EOG) assessments, which were created by omitting field test items from the regular version.

Each interim assessment in math and ELA were administered in sessions with a maximum time of 90 minutes. Each math interim contained 21 multiple choice items and 4 gridded response items. The first ELA interim contained 20 multiple choice items, and the second and third interims contained 19 multiple choice items and 1 constructed response item. There was no overlap in the items across interim assessments. The interims were design with the blueprint of the summative assessment in mind, as can be seen in the Summary Report: 2015–16 Proof of Concept Study Grade 5 Mathematics and Grade 6 English Language Arts/Reading (North Carolina Department of Public Instruction, 201, p. 28-29) However, the interim items were pulled from the summative assessment item pool – suggesting that the interims could, potentially, be placed onto the summative scale by fixing the item parameters to their values based on the scaling of the summative assessment.

The Pearson correlations between the assessments are shown in Table 1 below. The associations for ELA are slightly weaker than those for math. In general, the associations with the summative assessment are lower for the first interim, but very similar for the second and third interim – indicating that approaches to aggregation that weight the latter assessments will most likely better associate with the summative.

Table 1. Pearson Correlations between and among Interim and Summative Assessments

	Grade 5 Math				Grade 6 ELA			
	Interim 1	Interim 2	Interim 3	EOG Summative	Interim 1	Interim 2	Interim 3	EOG Summative
Interim 1								
Interim 2	0.768				0.707			
Interim 3	0.758	0.837			0.667	0.718		
EOC Summative	0.782	0.848	0.849		0.740	0.785	0.760	

Methods

We consider³ the following types of single summative scale scores:

- The maximum of the interim assessment sum scores,
- The mean of the interim assessment sum scores,
- The weighted mean of interim assessment sum scores, where the weights are roughly proportional to the amount of instruction students receive before taking the assessment (0.33 for the first interim, 0.50 for the second interim and 0.78 for the third interim), and

We then compare these scores to the EOG assessment summative score. Following the flowchart in Figure 1, we also produce proficiency classifications (i.e., dichotomized achievement level classifications, where a level 3 and above is proficient and level 2 and below is not proficient) for each of the summative scale scores. We do so by conducting logistic regression to find a cutpoint on the interim single summative scale score that best predicts student classifications on the summative assessment. As with the comparison to the EOG summative score, we do so based on the premise that the results of the EOG summative assessment are, if not the target of inference, a worthwhile point of comparison.

³We are also considering the Rasch scaling of all of the item responses from all of the items, ignoring assessments, however the estimation runs are ongoing at the time of this paper.

Results

Table 2. *Pearson Correlations between and among Single Summative Scale Scores and EOC Scale Score.*

	Grade 5 Math			Grade 6 ELA				
	Mean	Weighted Mean	Max	EOG Summative	Mean	Weighted Mean	Max	EOG Summative
Mean								
Weighted Mean	0.995				0.993			
Max	0.962	0.947			0.952	0.943		
EOG Summative	0.893	0.893	0.860		0.853	0.848	0.821	

Table 3. *Comparison of Proficiency Classifications between Single Summative Proficiency Levels and EOG Proficiency Levels.*

	Grade 5 Math			Grade 6 ELA		
	% Same Classification	Different Classification		% Same Classification	Different Classification	
		Identified as Proficient by Interim	Identified as Not Proficient by Interim		Identified as Proficient by Interim	Identified as Not Proficient by Interim
Mean	87.6%	5.8%	6.6%	85.0%	7.9%	7.1%
Weighted Mean	87.8%	6.0%	6.3%	84.7%	8.6%	6.7%
Max	85.1%	7.6%	7.2%	83.2%	8.9%	7.9%

The multiple approaches do not vary sustainably from one another, which is not surprising given (i) the relatively high correlations between interim assessments, and (ii) that each approach is similar to one another in terms of the algebraic formula. We suspect that the Rasch theta estimates, which are currently being estimated, will be similar to the mean, as these estimates are monotonically to the sum score of all assessment items.

Discussion

To be clear, **the choice of summative score used should be rooted in the claims to be made about students.** For example, should the claim be about average student performance in a fashion similar to the way in which course grades or grade point averages are defined? Should the claim be about a student’s best performance, similar to the way in which a student’s best work is selected for a portfolio? Or should the claim be about a composite that weights each assessment according to some value judgment, similar to the way different kinds of work contribute more or less to a student’s course

grade? These claims align to particular types of summative scores and should be carefully considered by the state.

In addition, there are a number of considerations we have not yet examined. First among these considerations is reliability. We are currently working on estimating reliability for the various composite scores, but expect that the reliabilities of the mean and weighted scores to be the highest. In addition, this report has been agnostic to North Carolina's accountability system. While we have considered the classifications of students identified as proficient, there are likely other ways in which summative scores are used within the system, and shifts in summative scores could affect such uses.

References

Every Student Succeeds Act §1111, S.1177. (December, 10, 2015). 114th Congress. Retrieved from:
<https://www.congress.gov/bill/114th-congress/senate-bill/1177/text>

Wise, L. (2011). Picking up the pieces: Aggregating results from through-course assessments. Paper presented at the Invitational Research Symposium on Through-Course Summative Assessments, Atlanta, GA.