Committee on Standards, Design and Methodology

December 7, 2020
3:00 – 5:30 pm ET
Webex

AGENDA

3:00 – 3:05 pm	**Welcome from New Committee Leadership** **Welcome to New Members** *Gregory Cizek, Chair* *Carey Wright, Vice Chair*	
3:05 – 4:20 pm	**Design of Studies to Review and Revise NAEP Achievement Level Descriptions (ALDs) in Mathematics and Reading** *Sharyn Rosenberg, Assistant Director for Psychometrics* *Eric Moyer, Pearson*	Attachment A
4:20 – 5:00 pm	**Discussion of Strategic Vision** *Gregory Cizek*	Attachment B
5:00 – 5:30 pm	**Feedback on Report Relating Michigan Students' Performance on NAEP to Postsecondary Performance** *Gregory Cizek* *Sharyn Rosenberg*	Attachment C

Studies to Review and Revise NAEP Achievement Level Descriptions (ALDs) for Mathematics, Reading, and Other Subjects

Background

On September 24, 2020, the National Assessment Governing Board (Governing Board) awarded contract# 91995920C0004 to Pearson (as a result of a competitive bidding process) for conducting studies to review and revise NAEP achievement level descriptions (ALDs) in mathematics and reading using the 2019 NAEP assessments at grades 4, 8, and 12[1]. This work is intended to address the first recommendation of the evaluation of NAEP achievement levels that was conducted by the National Academies of Sciences, Engineering, and Medicine:

Recommendation #1: Alignment among the frameworks, the item pools, the achievement-level descriptors, and the cut scores is fundamental to the validity of inferences about student achievement. In 2009, alignment was evaluated for all grades in reading and for grade 12 in mathematics, and changes were made to the achievement-level descriptors, as needed. Similar research is needed to evaluate alignment for the grade 4 and grade 8 mathematics assessments and to revise them as needed to ensure that they represent the knowledge and skills of students at each achievement level. Moreover, additional work to verify alignment for grade 4 reading and grade 12 mathematics is needed.

The Board committed to conducting studies to review and revise the NAEP ALDs in its initial response to the evaluation that was formally adopted and sent to the Secretary of Education and Congress in December 2016. The Board's Achievement Levels Work Plan, adopted in March 2020, further describes the intention for this work: "Addressing Recommendation #1 should focus on the current reporting ALDs for mathematics and reading at grades 4, 8, and 12. The methodology will be similar to what was done to evaluate the alignment and revise the 2009 NAEP Reading ALDs for grades 4, 8, and 12 (Donohue, Pitoniak, & Beaulieu, 2010) and the 2009 NAEP Mathematics ALDs for grade 12 (Pitoniak, Dion, & Garber, 2010). This process will generate new reporting ALDs that comply with the revised Board policy statement" (p. 3).

According to Principle 1a of the Board policy on Developing Student Achievement Levels for NAEP, "Content achievement level descriptions translate the policy definitions into specific

[1] The base year of this contract includes the review and revision of ALDs in mathematics and reading at grades 4, 8, and 12; in addition, an option may be exercised for a second year of the contract focusing on review and revision of ALDs in U.S. history, civics, science, technology and engineering literacy (TEL) at grade 8 based on data from the most recent administrations of those assessments in 2018 and 2019.

expectations about student knowledge and skills in a particular content area, at each achievement level, for each subject and grade. Content ALDs provide descriptions of specific expected knowledge, skills, or abilities of students performing at each achievement level. They reflect the range of performance that items and tasks should measure. When setting achievement levels, the content ALDs provide consistency and specificity for panelist interpretations of policy definitions for a given assessment. During reporting, content ALDs communicate the specific knowledge and skills represented by *NAEP Basic*, *NAEP Proficient*, and *NAEP Advanced* for a given assessment" (p. 5).

Principles 3g and 4a of the Board policy apply specifically to this project of reviewing and revising the current ALDs and creating reporting ALDs (based on empirical data) that indicate what students at each achievement level *do* know and *can* do rather than what they *should* know and *should* be able to do[2]. Additional details for carrying out the work described by principles 3g and 4a are included in the Achievement Levels Procedures Manual.

The basis for the evaluation of NAEP achievement levels (and subsequently for this project) is the *existing* NAEP frameworks and item pools, not the new NAEP Mathematics Framework scheduled for implementation in 2025 or the NAEP Reading Framework that is currently under development and consideration by the Board. In accordance with principle 4b of the Board policy, the achievement levels and/or ALDs will need to be reviewed again once the new frameworks are implemented. Such work is beyond the scope of this project.

Project Overview

Dr. Eric Moyer is the project director at Pearson and Dr. Jennifer Galindo is the assistant project director at Pearson. Pearson will conduct a pilot study and an operational meeting using scale anchoring studies where panels of content experts judge the alignment of the current mathematics and reading ALDs and produce a set of recommended reporting ALDs for the Governing Board to consider in reporting the results from the next regular administration of the NAEP reading and mathematics assessments at grades 4, 8, and 12. The Governing Board is expected to take action on the reporting ALDs for mathematics and reading at grades 4, 8, and 12 in August 2021.

Based on careful review of the history of ALD development, review, and revisions for NAEP mathematics and reading, a model-based anchored approach for reviewing the alignment of the ALDs for NAEP mathematics and reading will be used. The methodology for this alignment review study is based on that of previous studies, including the ALD development and review meeting held in 2009. The methodology was specified by the Board's Achievement Levels Work Plan and was selected to reduce the potential for possible inconsistencies from the use of different methods. The process of the model-based anchored approach will result in organizing

[2] According to the Board policy, ALDs will continue to describe what students *should* know and *should* be able to do for the purposes of item development and standard setting; only the reporting ALDs will be written in terms of what students *do* know and *can* do.

specific NAEP items by achievement level, which will serve as a key referent for panelists in reviewing and revising the current ALDs.

The model-based anchored approach includes three stages. The first stage will involve conducting statistical analyses to determine the items from the subject and grade that are anchored to a level corresponding to the score range within cut scores set to represent the achievement level descriptors (ALDs). The second stage relies on panels of content experts for each individual assessment. The panelists individually review the items that are anchored to each performance level and create summary descriptions of what students in each level are expected to know and be able to demonstrate based on the knowledge and skills measured by the items. In the final stage, the panelists compare the current ALDs for the respective assessment with their summary descriptions. The panelists note the similarities and differences, to make a recommendation regarding whether the current ALDs accurately describe what students in each level are expected to know and be able to demonstrate or if revisions to the current ALDs are needed to improve alignment. The final alignment judgment will be used to report whether the panels determined that there exists alignment between the current ALDs and student expectations. The final panel summary descriptions will be used to revise the current ALDs to create reporting ALDs that indicate what students at each achievement level do know and can do.

There is a technical advisory committee (TAC) consisting of the following experts in ALDs:

Dr. Karla Egan (Principal, EdMetric)

Dr. Ellen Forte (CEO and Chief Scientist, edCount)

Dr. Susan Loomis (Independent Consultant)

Dr. Marianne Perie (President, Measurement in Practice)

Dr. Mark Reckase (University Distinguished Professor Emeritus, Michigan State University)

Dr. Lauress Wise (Principal Scientist, Human Resources Research Organization)

The TAC is scheduled to meet for a total of 80 hours between November 2020 and August 2021 to provide technical advice on all aspects of the project; this is intended to help ensure that all procedures, materials, and reports are carried out in accordance with current best practices, providing additional validity evidence for the process and results. In addition to frequent meetings and reviews of materials, two TAC members will attend the pilot and operational meetings to observe and provide feedback on the process.

Project Update (November 2020)

Since the contract was awarded at the end of September, project staff have been working with the Board's Contracting Officer's Representative (COR), Dr. Sharyn Rosenberg, to put together detailed plans for the design and implementation of the studies. In addition, project staff have been working with NCES via the COR to obtain the data and materials that are necessary for conducting the studies.

The TAC began meeting (virtually) on November 2 and 6 to discuss and provide advice on the proposed study design, including how to best conduct this work in the midst of a pandemic. Additional TAC discussions are planned for December 3 and 15.

The COSDAM discussion on Monday, December 7 will include an overview of the project and the proposed design for the studies (attached). As noted previously, the design is intended to be similar to the studies conducted on the 2009 reading assessment at grades 4, 8 and 12 and the 2009 mathematics assessment at grade 12. The primary ways in which the proposed design differs from the previous studies are summarized below:

- There is a TAC to provide technical advice on all aspects of the project, imparting additional validity evidence for the process and results.
- There will be replicate panels for each subject and grade, allowing for a comparison of results across groups.
- Panelists are intended to be nationally representative and will not be limited to individuals who worked extensively on NAEP frameworks and item development.
- Reporting ALDs will be written in terms of what students *do* know and *can* do rather than what students *should* know and *should* be able to do, to reflect the current Board policy.
- Panelists will examine items that anchor below *NAEP Basic* to try to differentiate between the knowledge and skills demonstrated by students in the *NAEP Basic* range as compared to students in the upper region of below *NAEP Basic.*
- The review of items and other relevant materials will occur using Pearson's standard setting software on computers rather than on paper, since the 2019 NAEP assessments were digitally-based.
- The panel meetings will need to be conducted virtually rather than in person due to the COVID-19 pandemic.

National Assessment Governing Board

Developing Achievement Level Descriptions for Mathematics and Reading

Submitted: November 2020

NAEP ALD Development Design Document

Submitted to:
National Assessment Governing Board
800 North Capitol Street, NW, Suite 825
Washington, DC 20002-4233

This study was funded by the
National Assessment Governing Board under
Contract 91995920R0004.

Submitted by:
NCS Pearson, Inc.
2510 N. Dodge Street
Iowa City, IA 52245-9945
Phone: 319.354.9200

Table of Contents

Executive Summary

The National Assessment of Educational Progress (NAEP), known as the "The Nation's Report Card," provides information on what students in the United States know and can do in various subject areas. As part of its legislative mandate for overseeing and setting policy for NAEP, the National Assessment Governing Board (Governing Board) develops achievement levels that further define expectations of what students should know and be able to do. Achievement on all NAEP assessments is reported using the following achievement levels, in accordance with the Board policy on Developing Student Achievement Levels for NAEP:

NAEP Basic – This level denotes partial mastery of prerequisite knowledge and skills that are fundamental for performance at the NAEP Proficient level.

NAEP Proficient – This level represents solid academic performance for each NAEP assessment. Students reaching this level have demonstrated competency over challenging subject matter, including subject-matter knowledge, application of such knowledge to real world situations, and analytical skills appropriate to the subject matter.

NAEP Advanced – This level signifies superior performance beyond NAEP Proficient.

With the achievement levels, content achievement level descriptions (ALDs) are developed to define the expected knowledge and skills for student at each achievement level. The policy ALDs apply to all NAEP assessments, regardless of subject and grade.

The National Academies of Sciences, Engineering, and Medicine completed an evaluation of NAEP achievement levels in November 2016, which included seven recommendations. Two of the recommendations from the report were related to NAEP ALDs, recommendations #1 and #3.

Recommendation #1: Alignment among the framework, the item pools, the achievement-level descriptors, and the cut scores is fundamental to the validity of inferences about student achievement. In 2009, alignment was evaluated for all grades in reading and for grade 12 in mathematics, and changes were made to the achievement-level descriptors, as needed. Similar research is needed to evaluate

alignment for the grade 4 and grade 8 mathematics assessments and to revise them as needed to ensure that they represent the knowledge and skills of students at each achievement level. Moreover, additional work to verify alignment for grade 4 reading and grade 12 mathematics is needed.

Recommendation #3: To maintain the validity and usefulness of achievement levels, there should be regular recurring reviews of the achievement-level descriptors, with updates as needed, to ensure they reflect both the frameworks and the incorporation of those frameworks in NAEP assessments.

In response to this evaluation of the NAEP achievement levels, the Governing Board developed and approved an Achievement Levels Work Plan in March 2020. To address the recommendations regarding the ALDs from the evaluation, the Governing Board issued a contract to Pearson in September 2020 to develop and conduct anchoring studies using NAEP 2019 data to review and revise the ALDs for grades 4, 8, and 12.

On behalf of the Governing Board, Pearson has developed this Design Document, which describes in detail the activities for the anchoring studies for grades 4, 8, and 12 NAEP mathematics and reading. This document is intended to provide the foundation for all ALD alignment review and revision activities. The Design Document will guide all aspects of the ALD review and revision process, including (1) a set of judgments about the alignment between the policy descriptions and current ALDs and expectations of what students should know and be able to do, based on the items that anchor to each achievement level for each assessment and (2) a set of revised ALDs that could be used as reporting ALDs, describing what students do know and can do.

For the ALD review and revision studies, Pearson plans to use a model-based approach for reviewing the alignment of the ALDs for NAEP mathematics and reading. The Board's Achievement Levels Work Plan indicated that the methodology for these studies should be similar to previous ALD development and review meetings held in 2009 (Donahue, Pitoniak, & Beaulieu, 2010; Pitoniak, Dion, & Garber, 2010) to reduce the potential for possible inconsistencies from the use of different methods.

The model-based approach includes three stages. The first stage involves conducting statistical analysis to determine the items from the subject and grade that are anchored to each achievement level. The second stage relies on panels of content experts for each individual assessment. Each panelist will review the items that are anchored to each achievement level to identify the knowledge and skills needed to respond to the items associated with a level. The panels will then work together to develop common summary descriptions of the knowledge and skills that students in each level are expected to know and be able to do. In the third stage, the panelists complete several alignment judgment rounds, comparing the current ALDs for the respective assessment with the summary descriptions. During the alignment judgment rounds, the panelists rate the degree of alignment and take notes regarding the similarities and differences between the ALDs and summary descriptions. The goal of this process is for the committee to make a recommendation regarding whether the current ALDs accurately describe what students in each level are expected to know and be able to demonstrate or if revisions to the current ALDs are needed to improve alignment and to provide revised ALDs that accurately represent what students within each achievement level are expected to know and be able to do.

To facilitate this process, Pearson will use computers during both the pilot and operational ALD review meetings. Using computers along with the online interface in the Pearson Standard Setting website will increase the efficiency of the activities the panelists will need to complete for this study. Additionally, this method will allow these meetings to be conducted virtually, to address current logistical concerns with holding in-person meetings in the midst of the COVID-19 pandemic.

Below is a summary of what each section in the Design Document includes.

Section 1: Achievement Level Descriptions Review Panels describes the panelist identification and recruitment plan designed to obtain broadly representative and well-qualified panelist groups for all studies.

Section 2: Briefing Materials describes the briefing materials sent to panelist prior to each panel study in the ALD review process (pilot study and operational ALD review meeting).

Section 3: Pilot Study describes the pilot study designed to incorporate the exact procedures planned for the operational ALD review meeting, including the use of a virtual meeting.

Section 4: Achievement Level Descriptions Review Tasks and Procedures describes the ALD review tasks, the nature of the tasks, and the procedures to be implemented prior to and as a part of the operational ALD review meeting—including how panelists are trained and supported in implementing all activities.

Section 5: External Feedback describes the process of obtaining feedback on the results and recommendations from the operational ALD review meeting.

Section 1: Achievement Level Descriptions Review Panels

Pearson will implement a multi-step panelist recruitment plan for the pilot study and operational ALS meeting. Panelist for the pilot study and operational panels will be recruited from across the nation. The objective of the recruitment plan is to produce well-qualified panels with a high level of content expertise, consisting of classroom teachers and content experts that will be in the best position to provide the necessary judgments for the respective subject and grade. For each meeting, there will be six panels convened, with each panel focused on a single subject and grade.

Grade 4 Mathematics

Grade 8 Mathematics

Grade 12 Mathematics

Grade 4 Reading

Grade 8 Reading

Grade 12 Reading

A maximum of 48 panelists will be recruited for the operational study, with up to eight panelists recruited for each panel. Additionally, a maximum of 48 panelists with similar background distribution will be recruited for the pilot study, with up to eight panelists recruited for each panel. Panels will have a minimum of six panelists, in the event that one or two panelists have to drop out prior to the meetings.

Panelists in each panel will be assigned to one of two replicate groups, allowing for a comparison of results across groups.

Panels for the ALD alignment review and revision operational and pilot studies will reflect an overall balance of gender, race/ethnicity, geographic location, and urbanicity, i.e., no more than 75% homogenous. Classroom teachers currently engaged in instruction in the respective grade and subject area will compose at least five of the panelists in each panel. At least two of the panelists will be non-classroom educators with curriculum experience within the respective subject, such as state or local curriculum coordinators or higher-education faculty teaching education courses associated with the respective subject and level.

Our goal for this study is to obtain panelists with a high level of content expertise that would enable them to engage in the ALD review process and provide meaningful recommendations and judgments. Pearson will work with staff from the Governing Board along with allied organizations in recruiting panelists.

Panelist Recruiting Process

A multi-phase process will be used to identify panelists:

Phase 1: Identify nominators through allied organization and state departments of education; contact nominators and ask them to nominate qualified classroom and non-classroom educators using an online nomination form (e.g., name, contact information, and basic qualifications). Nominators will be asked to indicate if they judge the nominee to be "outstanding" in terms of their professional performance and to briefly describe the rationale for that judgment when provided.

Phase 2: Notify nominees; request résumés and completed panelist form (e.g., background in instruction with respective subject and grade, professional achievements, experience with students). Nominees will be asked to specify training and experience that makes them an outstanding candidate for panel selection.

Phase 3: Evaluate nominated candidates based on their background and experience; select the most qualified panelists and assign them to panel groups with respect to gender, race/ ethnicity, geographic location, instruction experience, type of institutional affiliation, and urbanicity.

Prior to finalizing the selection of candidates to participate in the studies, the list of prospective panelists will be prepared and presented to the Governing Board for review and approval.

To optimize recruitment, we will use email and phone calls to communicate with prospective panelists. An honorarium will be paid to panelists for the ALD alignment review and revision operational and pilot studies. Substitute teacher costs will be reimbursed directly to schools based on actual school costs for substitute teacher payments.

Identification of Panelist Nominators

Panelist nominators will be recruited using multiple sources. One source for nominators will be professional organizations that have a strong background in providing professional development in mathematics education or reading and literacy education. To recruit panelists for the mathematics panels, some of allied organizations may include:

National Council of Teachers of Mathematics (NCTM)

National Council of Supervisors of Mathematics (NCSM)

Association of Mathematics Teacher Educators (AMTE)

To recruit panelists for the reading panels, some of allied organizations may include:

National Council of Teachers of English (NCTE)

International Literacy Association (ILA)

American Literacy Corporation (ALC)

In addition to these organizations, state education organizations in mathematics or reading and literacy, state superintendents and departments of education, school board presidents, and district and school administrators of public and private education entities will be contacted in the four NAEP regions to

propose qualified nominators across both panelist type (classroom teacher and non-classroom teacher educator).

Based on previous experience in recruiting NAEP panelists for achievement level setting meetings, Pearson estimates that 20 percent of the nominators will respond by submitting at least one nominee for consideration. We further estimate that no more than 20 percent of the nominees would meet the qualifications, satisfy the requirements for representation, and agree to serve on the panel. For the pilot and operational studies, an estimated 2400 nominators must be identified to yield at least 480 active nominators, resulting in at least 480 nominees. Assuming that 20 percent of those nominees will be eligible, meet the distribution requirements for representation on the panels, and be available and agree to serve as panelists, the yield would be the target of 96 panelists, with a target of 48 panelists for the pilot study and 48 panelists for the operational meeting. Pearson will supplement the number of nominators, as needed, to attain the panel targets.

Selection of Panelists

Nominees will be asked to complete an online questionnaire regarding their qualifications and experiences for serving on the panel. Candidates that present with the credentials required of panelists will be contacted by phone to collect any missing information, verify the information provided, and confirm the willingness of the candidate to serve on the panel, if selected. The goal is to select the most qualified panelists who are knowledgeable about the related subject at the appropriate level, while maintaining the goal to recruit 75% classroom teachers and 25% non-classroom teacher educators to compose each of the panels.

Panelists nominated in each panel must meet the following minimal qualifications.

Classroom Teacher Qualifications:

The nominee must meet all of the following qualifications:

At least five years of overall teaching experience

At least two years of experience teaching the related subject and the grade

Judged to be “outstanding” in their professional performance by a nominator

Non-Classroom Teacher Educator Qualifications:

The nominee must meet one of the following qualifications:

Non-teacher educational staff within school or district with education experience in the related subject and grade

Curriculum director or content specialist serving schools at the related level or state department of education with education experience in the related subject and grade

Postsecondary teacher education faculty teaching courses in the related subject and level

The credentials of panelists will be evaluated based on the number and importance of the credentials that are presented. Nominees having no distinguishing credentials will be scored low. Nominees having extensive credentials, including having been named outstanding teacher/teacher of the year and/or being actively engaged at the national level in professional activities within the specific subject and level, will score high. The scoring scheme differs for each panelist type (classroom teacher and non-classroom teacher educator). Nominees with the highest scores are given top priority by placing the best-qualified candidates at the beginning of the candidate list. The selection process then selects panelists to reach the targets for representation listed above, with nominees having the highest qualifications being the first selected each time. All panels will be selected to have approximately equal proportions of males and females and equal proportions of representation from each of the four NAEP regions. We will also attempt to draw panels so that at least 20 percent of the panelists self-identify as a minority.

Each panelist for the pilot study and operational ALD review and revision meeting will be given an honorarium. Pearson acknowledges that the funds available to offer panelists are not commensurate with their contribution. They will emphasize that panelists’ participation in the ALD review and revision project represents an exceptional contribution to education in the United States.

Section 2: Briefing Materials

Pearson will send access to a set of briefing materials to each confirmed panelist for their review and familiarization prior to the relevant panel meeting (pilot study/operational). The Pearson Standard Setting website, customized specifically for the project, will provide panelists with secure online access. The first time the panelist logs in to the website, he or she must read and electronically sign a nondisclosure agreement. Once signed, the panelist will be guided through a brief online training for using the website before having access to the non-secure advanced materials designated for the particular committee through links on the website. Panelists will use an online checklist to guide them through the online materials and will be able to check off each document after it has been reviewed. Documents will include the following:

- Confidentiality agreement
- Purpose and overview of the meeting
- Meeting roles and responsibilities
- Request for reimbursement form
- Meeting agenda
- NAEP framework for the relevant assessment
- Current NAEP ALDs
- Other materials identified as appropriate

Communication with panelists will encourage them to engage with the briefing materials as those materials are intended to serve as a foundation for successfully carrying out the process designed for each panel. Pearson staff will be able to monitor the panelist activity on the website to determine which materials a panelist has accessed.

Section 3: Pilot Study

Pearson will conduct a pilot study of 4-5 days to implement the exact virtual meeting procedure planned for the operational ALD alignment review and revision meeting in late February of 2021. Conducting the pilot study at this time offers an opportunity to preview, revise, and resolve issues prior to the operational ALD alignment review and revision meeting in May of 2021. To maintain uniformity of conditions, the pilot study and operational ALD alignment review and revision meetings will be held using the same virtual process, including the same agenda of activities, and all steps in the pilot study will be the same as those planned for the operational ALD alignment review and revision meeting. By fully replicating the process of the operational study, the pilot will provide the information needed to determine whether any modifications are needed for the operational study.

The Technical Advisory Committee (TAC) will be consulted for advice and recommendations regarding details of the design prior to and after the pilot study. Up to two members of the TAC will be asked to observe the pilot study. All pilot and operational study materials will be based on data from the 2019 operational administration of the NAEP mathematics and reading assessments. The pilot study has the following goals:

- Determine whether modifications for training, instructions, materials, timing, and logistics will be needed for the operational ALD alignment review and revision study.
- Provide an opportunity for facilitators to practice the process before moving to the operational ALD alignment review and revision meeting.

Given that all steps in the pilot study represent those planned for the operational ALD alignment review meeting, details on the process for both the pilot study and the operational meeting are provided under Section 4 of this document which clearly describes the operational ALD alignment review meeting.

A pilot study report will be prepared no later than March 2021, for presentation to COSDAM during a webinar or conference call. The timing of the pilot study allows for preparation of the report

and review of the report by the COR, COSDAM and TAC so that improvements to the process can be made in advance of the operational ALD review meeting.

Section 4: ALD Alignment Review Tasks and Procedures

Pearson will conduct an operational ALD alignment review meeting in May 2021. To maintain uniformity of conditions, the operational ALD alignment review meeting will be held with the same agenda of activities used for the pilot study, with the exception of adjustments made based on improvement to the pilot study process. This section describes the ALD alignment review procedures and tasks that Pearson will implement during both the pilot study and operational ALD alignment review meeting and includes information about the configuration of panels and materials, training of panelists, the collection of panelists' ratings, and the feedback given to panelists. The operational meeting will involve 48 panelists, with approximately eight per subject/grade. The panelists will be assigned to two groups with approximately similar representation, which will be used as replicate panels during the meeting, since they will complete the same process under the same facilitation. The TAC will be consulted prior to the operational meeting, and up to two members of the TAC will be asked to observe the meeting. All operational materials will be based on data from the most recent 2019 administration of the grade 4, grade 8, and grade 12 NAEP mathematics and reading assessments.

As previously indicated, Pearson will implement a method-based anchor approach and use the Pearson Standard Setting website platform as well as Zoom to facilitate key aspects of ALD alignment review process, including panelist training, review of ALDs, housing and reviewing selected anchor item sets, recording of panelists' feedback and ratings during each round of review, provision of feedback, and evaluation of the ALD alignment review process. The Pearson Standard Setting website will be set up to guide panelists through the steps of the ALD review process, with facilitators having the ability to restrict or provide panelists with access to sections of the site or activities, as needed. The night before the first day of the pilot study and the operational meeting, Pearson will reset the panelists' passwords for all panelists to a common meeting password. Additionally, access to the sites is controlled by Pearson administrators. The sites are deactivated overnight and during any extended

breaks during the day. The Zoom meetings platform will be used to host the meeting, screen-share facilitator materials, and create breakout groups.

The model-based anchored approach includes three stages. The first stage will involve conducting statistical analyses to determine the items that anchor to each achievement level (*NAEP Basic, NAEP Proficient,* and *NAEP Advanced*). Additionally, an anchor set will be developed for items that map to the higher end of the region below *NAEP Basic*. The second stage relies on panels of content experts for each individual assessment. The panelists in each committee individually review the items that are anchored to each performance level. In table groups, they create summary descriptions of what students in each level are expected to know and be able to demonstrate based on the knowledge and skills measured by the items. In the final stage, the panelists compare the current ALDs for the respective assessment with the summary descriptions, noting the similarities and differences, to make a recommendation regarding whether the current ALDs accurately describe what students in each level are expected to know and be able to demonstrate or if revisions to the current ALDs are needed to improve alignment. If it is determined that revisions are recommended, the panel provides recommended modifications to the ALDs.

Computer Use and Software

Pearson will use two computers during the pilot study and the operational ALD review meeting. The use of computers and an online interface will reduce the time required for panelists to complete most steps in the ALD review activities. In addition, the use of computers will allow the panelists to interact with the items as students did, such that panelists are better able to understand what examinees would have to know or be able to do.

The computer will be a laptop computer (provided by Pearson) that is used by panelists to access the online Pearson standard setting website interface. As a part of supporting both efficiency and accuracy of the ALD review process, Pearson will use computers with access to the Pearson Standard Setting website, with a section specifically designed for the NAEP ALD review meeting. The NAEP

website can be accessed only through a user identification code (ID) assigned by Pearson. Permissions will be set up for each user ID so each panelist can access only the materials he/she will be using.

The use of the website will reduce the time required for panelists to complete multiple steps in the process, since they will be able to access materials and complete activities within the website interface. As a website-based system, all materials and data will be stored, organized, and accessed through the website, which will ease the demands of development of materials and ensure the consistency of the materials with which the panelists interact. The use of the website will also ensure the security of the materials during the standard setting meeting, since the system will be used throughout the ALD review process and requires a secure login by all users. Pearson designed the interface for the NAEP ALD review process to have the following features:

Simultaneous access by multiple users, with each individual user assigned a profile which defines their level of access to the site, including the materials the user can access and the specific functionalities available. Facilitators and site administrators can use conditional access features based on date, activity completion, or custom settings to define access to materials and functionalities of the site.

User access prior to the ALD review meeting, so users can interact with materials and experience training that will prepare them for the meeting and the ALD review process.

Consistent user interface throughout the entire ALD review process, from pre-meeting work to post-meeting feedback.

Management of storage and access to materials through the website, as a single access point for the meeting.

Assignment of panelists to materials and items, based on, based on panelists grouping.

Embedded quality control features through conditional settings within activity that limit the range of responses to only valid entries and ensure no blank entries are accepted.

Export of panelists' item review notes, as a group or individually. Facilitators will be able to view the panelists' notes within the website, to verify panelists' responses.

Secure storage of all panelist information and judgments within the website, with access restricted to facilitators and site administrators.

Preparing for the ALD Alignment Review Meeting

This section describes key activities Pearson will complete before the pilot and operational ALD alignment review meetings that contribute to the success of the meetings.

Development of Anchor Item Sets.

The development of the anchor item sets starts by grouping individual students from the most recent (2019) administration of the grade 4, grade 8, and grade 12 NAEP mathematics and reading assessments into achievement levels. The achievement level classification for each student is based on their NAEP "plausible value" and the relationship to the boundaries of the achievement levels for the respective assessment. The average plausible value will be used for each student. This approach will use all students in the NAEP sample from the most recent administration in 2019 to ensure that there are sufficient students associated with each achievement level for the analysis to determine each anchor item set. This is an approach has been utilized in previous NAEP alignment studies.

After students are assigned to an achievement level, the conditional *p*-value, or probability of each student in that achievement level answering each item correctly, will be calculated using the IRT statistics from the most recent administration of the assessments. The conditional *p*-value for students across a given level will be averaged to derive the anchoring probability for that item or score point for multi-point items. Each item or score point will be assigned four conditional *p*-values, one each for below *NAEP Basic, NAEP Basic, NAEP Proficient,* and *NAEP Advanced*, which represent the average performance on the item of the typical student within the four achievement levels. In addition to the conditional *p*-values, item discrimination values will be calculated for each item. The item discrimination is defined as the difference between the item's conditional *p*-value for the anchor achievement level and the conditional *p*-value at the next lower level. Each item or score level will be assigned four discrimination values, one each for achievement levels *NAEP Basic, NAEP Proficient*, and

NAEP Advanced and one for just below the *NAEP Basic* level. To provide a criterion for the discrimination values for each achievement level, the 40th percentile of the distribution of the item discrimination value for the respective achievement level will be calculated. The conditional *p*-values and discrimination values are then used to analyze all items from the most recent NAEP administration to determine which items map into the given achievement levels. Items qualify as anchoring at an achievement level if the conditional *p*-value equals or exceeds 0.67, which is the same criterion used for prior ALD alignment studies. When items meet the conditional *p*-value criterion for more than one achievement level, the item is classified as mapping onto the lowest achievement level at which the conditional *p*-value criterion is met.

Additionally, an item also needs to meet the discrimination value criterion, of being greater than the 40th percentile of discrimination values, for the achievement level to be classified as anchoring to the achievement level. Based on these criteria, items will be classified into one of six categories: (1) high below the *NAEP Basic* level, (2) *NAEP Basic* level, (3) *NAEP Proficient* level, (4) *NAEP Advanced* level, (5) does not discriminate, or (6) does not anchor. The items in the anchor item sets for the respective assessment will be grouped by content area. By reviewing the items within a content area, across all achievement levels, the panelists will be able to maintain a consistent focus on the knowledge and skills associated with the content area. The items associated with a content area will be ordered by achievement level from below the *NAEP Basic* level to the *NAEP Basic* level, to the *NAEP Proficient* level, and then finally the *NAEP Advanced* level. Within an achievement level, the items are in decreasing order of conditional *p*-value, so the easiest item associated with the achievement level is first and the most difficult item is last. In this way, panelists will see a progression in what students know and are able to demonstrate while working through the set of items that anchor to that achievement level. The item set with achievement level anchor classifications will be reviewed with the TAC and the content staff to determine a final anchor set.

Division of Panelists into Replicate Panels. In order to assess the degree of internal validity, two replicate groups (table groups) will be created within a committee. Approximately three to four panelists will be in each table group. Pearson will assign panelists to a table group with the intent of creating groups that are as equivalent as possible to one another. The purpose of the replication committee is to be able to assess the degree of internal validity. The summary statistics from the table group rounds of individual comparisons will be calculated and compared. In addition, a qualitative analysis will be conducted to assess the degree to which the summary descriptions created by each table group vary.

Provision of Advanced Briefing Materials. As discussed in Section 3 Briefing Materials, panelists will have access to advanced materials through the Pearson Standard Setting website. Pearson will send each confirmed panelist access to the materials for their review and familiarization prior to the operational meeting, including information to log into the website and change the assigned password. The first time the panelist logs in to the website, he or she must read and electronically sign a nondisclosure agreement. Once signed, the panelist will be guided through a brief online training for using the website before having access to the non-secure advanced materials. Panelists will then use an online checklist to progress through the review and to ensure that each document is reviewed. Panelists will also have access to an orientation activity that includes an overview video as well as multiple engagement check-ins to gauge interaction and preparedness.

Training of Facilitators. The ALD alignment review and revision study will involve two content facilitators as well as a process facilitator per committee. The content facilitators are selected for their expertise and experience in instruction in the related subject and grade as the framework they bring. The process facilitator is selected for the expertise and experience conducting meetings they have. Recognizing that facilitators may introduce individual difference that can result in slightly different instructions, content and process facilitators will be properly trained to implement the process uniformly and as intended. Pearson will prepare the PowerPoint presentations that facilitators

will use during the meeting. In addition, facilitator handbooks will include the tables and graphs, a script for providing instructions, a description of the activities and an explanation of the feedback. Facilitators will attend a one-day training prior to both the pilot study and the operational meeting. The project director overseeing the activities will lead the training. In addition, the facilitators and project director will do a walkthrough of the entire meeting the day before the pilot study and the day before the operational meeting.

Preparation of Pearson Laptops. Pearson will be providing each of the panelists a Pearson laptop that has been configured to have the appropriate software needed to access the NAEP items as well as the Pearson Standard Setting website. The laptops will be shipped to each panelist along with a return shipping label.

Technology Check-in. Approximately two weeks prior to the ALD alignment review meeting, Pearson will host a technology check-in to ensure all panelists have been able to connect the laptops to the internet, have access to the Pearson Standard Setting website and to Zoom, and know how to navigate in Pearson's website and in Zoom.

Preparing Panelists for the ALD Alignment Review Process

Provide an Orientation. The operational meeting will have 48 panelists, with approximately eight panelists assigned to each committee. The six committees, along with the process and content facilitators, will begin the meeting with introductions and a description of the panelist recruitment process.

Overview the Alignment Review and Revision Process. Facilitators will provide an overview of the purpose of the ALD alignment review and revision in general and description of the process that will be used. Panelists will receive training in the key components of the NAEP framework for the related assessment and the rationale supporting these components. The process facilitator will describe the process that was used to create the anchor item set and an orientation to the information that is provided for each item. The panelists will also be trained in how the items are scored with the item key

or scoring rubrics. We will also orient the panelist in the structure of the alignment judgment process, including the anchoring of items to specific achievement levels and the ordering of the items within achievement levels.

Individual Item Review and Summary Descriptions

Item Review. During the individual item review process, panelists will review all items within the anchor item set. Prior to the panelists beginning the item review process, in the table groups the facilitator will demonstrate the use of the website to review each item, view item information, and collect individual panelists' notes about each item using a practice anchor set. Additionally, the content facilitator will model to the table group how to approach the items in the anchor set and the process of developing item descriptions. The facilitator will also show how one might convert the item description into a summary statement about what students know and can do for an achievement level. The panelists will work independently to review each item and create a description of the knowledge and skills demonstrated by students who answered the item correctly or who received a particular score for constructed-response items. During this step, panelists will review items within the same content area, grouped by achievement level (including the items anchored to just below the *NAEP Basic* level) and by conditional *p*-value, easiest to hardest. When they have completed the review of the anchor items for an achievement level, they will be asked to write a single summary statement about what students within that level know and can do based on the item descriptions. Once they have completed the review of all the items across the achievement levels, they will be asked to compare what students can do in one achievement level compared to the adjacent achievement level(s). They will complete the entire process for a content area before progressing to the next content area.

Summary Descriptions. After individually writing descriptions for each item for a content area, the panelist will convene in their table groups to review the collected set of knowledge and skills associated with the items within an achievement level and construct a summary of what students performing in that content area and achievement level (as well as just below the *NAEP Basic* level) are

expected to know and can do. Prior to beginning the table group development of summary descriptions, the facilitator will model the process of using the item descriptions to develop summary descriptions by achievement level. This process will start with the just below the *NAEP Basic* level, then the *NAEP Basic* level, then the *NAEP Proficient* level, then the *NAEP Advanced* level. They will create their summary descriptions for each achievement level, within a content area, in the website using a shared document. The summary descriptions will be captured by the content facilitator. Once all panelists complete the summary descriptions for a content area, they will then begin individual work for the next content area. After each content area of individual review, they will meet in the table groups to create the summary descriptions for that content area. When all content areas are completed, the panelists will progress to the alignment judgment rounds with table group discussions, whole group discussions, and ALDs revisions, if needed.

Alignment Judgment Rounds with Group Discussions

The process of creating and reviewing ALDs is an iterative process, where individuals provide individual alignment judgments, followed by table group discussions and whole group discussions that are informed with results from the alignment judgments, resulting in more informed judgments in subsequent rounds. There will be three alignment judgment rounds. Prior to the first alignment judgment round, the content facilitator will provide the panelists training about the meaning of alignment between the summary statements and the policy ALDs as well as the current content ALDs.

Alignment Judgment Rounds. The replicate panels or table groups will be a significant part of the review and judgment process. After the panelists complete their individual item review the panelists will then meet in their assigned table groups to discuss their individual summary statements and to develop a table group set of summary statements for that content area. They will toggle between individual review of items by content area and table group development of summary descriptions until all content areas have been addressed. To identify areas of alignment and lack of alignment, panelists will then work independently to make comparisons between the table group summary

descriptions with the policy descriptions and current ALDs. Panelists will be asked to assess the degree of alignment of the table group summary descriptions by content area and achievement level to the policy descriptions and ALDs. If they indicate weak to moderate alignment, they will be prompted to provide suggestions/comments for why the alignment is weak and what modifications could be made. Additionally, they will make a judgment on the overall alignment of the summary descriptions as a whole to the policy descriptions and current ALDs.

After the first round of individual alignment judgments, panelist will meet back in their table groups and discuss why they made their specific alignment judgments. If necessary, they can make adjustments to the summary descriptions to clarify or modify the language to better indicate the original intended meaning. The panelists will then proceed with another round of individual alignment judgments. They will then meet back as a whole group and review summary statistics of the alignment judgments made and have a whole group discussion similar to the table group discussion. Where necessary, they will visit the summary descriptions to evaluate if the difference in the judgments relates to differences in the summary descriptions across the table groups. Panelists will be shown the summary descriptions from each table group side-by-side to evaluate the degree to which they are similar or different and develop a set of whole group summary descriptions. The panelists will have one more opportunity to make individual alignment judgments based on the whole group summary descriptions.

After the third round of individual alignment judgments, a whole group discussion will take place. For the final whole group discussion, panelists will be asked to focus on any sub-categories that still indicate greater than 50% panelist agreement of weak alignment or greater than 66% panelist agreement of moderate or weak agreement with the current ALDs. The panelist will review the ALDs for these subcategories and have the opportunity to suggest edits to the ALDs to improve alignment.

Cross-grade Review. Before the end of the meeting, the panelists will be brought together to review the reporting ALDs developed by each committee, to review the terminology used across the achievement levels and grades to ensure that they clearly delineate progression of skills across grades

and levels. This discussion will be led by one of the process and content facilitator teams. Any suggested revisions to the reporting ALDs will be determined by the panelists that constructed them, since they are most familiar with the content and discussions, but recommended changes may be provided by any panelists. The recommended ALDs will be the result of this final meeting.

Process Evaluations. Procedural evidence refers to the appropriateness of the procedures and how well those procedures were implemented. Evidence for procedural validity may come from a number of sources, including criteria for selecting panelists, the justification for the method, the quality of the implementation of the procedure, and the completeness of the documentation of the process. As another source of evidence of procedural validity, panelists will be asked to complete evaluation forms after each major activity of the process. Evaluations will include both selected-response and open-ended questions that address the panelists understanding of the process and confidence in the results. No key-entry by staff is required because panelists will use the secure website interface to complete their evaluations. Panelist entries will be available for viewing using the facilitator login to the website. Facilitators will scan written responses for possible problems as they are collected during each day. Summary statistics will be computed for all ratings items and written responses. These analyses will be reviewed in real time throughout each day, and any sources of confusion will be identified for clarification with individual panelists or the panel as a whole. The operational meeting must be completed in time for the recommended cut scores to be used for reporting the next NAEP mathematics and reading assessments in grades 4, 8, and 12.

Section 5: External Feedback

Pearson understands the potential contribution of external feedback to the review and revision of ALDs for the NAEP mathematics and reading assessments. Pearson has designed a strategy for collecting and using feedback from content experts and NAEP stakeholders for the panel recommendations for revising the ALDs.

Pearson plans to implement a vigorous and targeted approach to soliciting feedback regarding the panel recommendations through personal contact with key leaders and members of stakeholder groups, and through the use of a simplified and directed format for reviewer response. Pearson will create a website to obtain external feedback on the panel recommendations for revising the ALDs resulting from the operational ALD review meeting. The website will provide a means for stakeholders and the public to find information about the study design and the panel recommendations and to provide feedback. Prior to opening the site to the public, Pearson will submit the site to Governing Board staff for review and approval.

Pearson will solicit comments from content persons, technical persons, and persons in education-related organizations that are known to use NAEP. When organizations are engaged, a personal call from project staff will establish the initial contact and explain the importance of the request. For collection of comments from members of organizations, Pearson will ask the organization leadership to communicate the request for feedback to the membership and to encourage members' cooperation and response. In all cases, a personalized email message will be sent to explain the purpose of the request for comment and to provide a link to the website having the Design Document, instructions, and questions for respondents to consider. Reviewer comments can be recorded through the link, downloaded by Pearson staff, and saved for analysis. Follow-up emails will be sent to encourage responses and to contact individuals for clarification and additional information, should that be necessary. The feedback will be summarized and presented to member of the panel meetings during a virtual meeting, to allow

the panelists to make any revisions to their recommendations before the ALDs are presented to the Governing Board.

Discussion of Strategic Vision

During a special Board meeting on September 29, the Board adopted its second strategic vision, Strategic Vision 2025, which is intended to guide the Board's work over the next four years. Strategic Vision 2025 (attached) includes three pillars (Inform, Innovate, and Engage) and eight priorities. Each strategic priority may include activities across several different standing committees, but it is anticipated that COSDAM will have primary responsibility for the following priorities:

Link NAEP resources with external data sources and disseminate what is learned from these sources so that NAEP can inform policy and practice in understandable and actionable ways

Develop a body of evidence to improve the interpretation and communication of NAEP achievement levels to ensure that they are reasonable, valid, and informative to the public

In addition, COSDAM may lead activities that fall under any of the other strategic priorities.

As discussed by the Executive Committee on November 16, Board staff are developing proposed work plans for committees to discuss in conjunction with the March 2021 quarterly Board meeting.

During the upcoming COSDAM meeting, Committee members will discuss initial guidance for developing COSDAM activities for the strategic priorities.

Strategic Vision 2025

The National Assessment of Educational Progress (NAEP), also known as The Nation's Report Card, was developed in 1969 to answer the important question: "How are our nation's students doing?" In 1988, Congress created the independent, bipartisan National Assessment Governing Board to set policy for The Nation's Report Card, that is, deciding what subjects to assess, approving test and survey questions, determining achievement levels, and improving the reporting and use of results. Since that time, the Governing Board and its partner, the National Center for Education Statistics (NCES), have worked to safeguard NAEP as the "gold standard" assessment of educational achievement in the United States.

The Governing Board established its first Strategic Vision in 2016 with two major goals – to inform and to innovate – for the Board's work on behalf of policymakers, educators, researchers, business leaders, and the general public. The Governing Board's second Strategic Vision, which will guide the Board's efforts through 2024, adds a third goal, "to engage." This new goal in the Board's vision highlights the importance of not only reporting results accurately, but also in promoting their use:

- To serve as a catalyst for action to improve student achievement;
- To inspire improvement in the quality of assessments and standards; and
- To tell the stories of American achievement for all, over time and in context.

The Governing Board's new Strategic Vision comes at a time of worldwide disruption in education, with the COVID-19 pandemic forcing educators to rethink long-held practices in teaching, learning, and assessment. Educators and policymakers alike are focused on educational equity with renewed commitment. The Governing Board will continue to partner with NCES to ensure NAEP remains the gold standard in assessment, providing objective information about the status and progress of American education at a time of significant disruption to our nation's schools and students.

The pandemic has heightened the need for the Governing Board to adopt a Strategic Vision that fulfills its Congressional mandate and preserves NAEP as a trusted, valuable resource that endures over time. Over the next four years, the Strategic Vision will guide the Governing Board as it strives to inform efforts to improve our nation's schools, innovate in carrying out its mandate, and engage stakeholders in understanding and using the results of The Nation's Report Card.

INFORM

The National Assessment Governing Board will disseminate NAEP resources to inform the work of numerous education stakeholders and to promote high-quality uses of The Nation's Report Card that support improvements in policy and practice. NAEP resources include results; focused studies; assessment questions and tasks; assessment innovations; and contextual variables about the educational experiences of students, teachers, and schools. The Governing Board will:

- **Identify** the needs of stakeholders and refine resources to promote sustained use of NAEP data, enabling educators, researchers, advocates, and policymakers to understand and improve student achievement.
- **Elevate** high-quality uses of NAEP resources to demonstrate NAEP's utility and to highlight the unique value of the Nation's Report Card to inform education policy and practice.
- **Link** NAEP resources with external data sources and disseminate what is learned from these sources so that NAEP can inform policy and practice in understandable and actionable ways.

INNOVATE

The National Assessment Governing Board will ensure The Nation's Report Card remains at the forefront of assessment design and technology by refining design, content, and reporting, increasing relevancy for NAEP users and inspiring action to improve achievement for all. The Governing Board will:

- **Optimize** the utility, relevance, and timing of NAEP subject-area frameworks and assessment updates to measure expectations valued by the public.
- **Monitor** and make decisions about the NAEP assessment schedule based on the Board's policy priorities of utility, frequency, and efficiency to ensure NAEP results are policy-relevant.
- **Develop** a body of evidence to improve the interpretation and communication of NAEP achievement levels to ensure that they are reasonable, valid, and informative to the public.

ENGAGE

The National Assessment Governing Board will strengthen partnerships and communications with stakeholder organizations, building capacity to understand and harness the resources of The Nation's Report Card to advance policy and practice. The Governing Board will:

- **Develop**, sustain, and deepen strategic partnerships to ensure that NAEP remains a trusted, relevant, and useful resource.
- **Help** stakeholders understand how the Governing Board and NAEP can illuminate important skills for postsecondary education pathways.

National Assessment Governing Board
800 North Capitol Street, NW Suite 825 • Washington, DC 20002
Phone: 202-357-6938 • E-mail: nagb@ed.gov • www.nagb.gov

Relating Michigan Students' Performance on NAEP to Postsecondary Performance

Since 2003, the Governing Board has been interested in using NAEP 12th grade assessments in reading and mathematics to support statements about students' academic preparedness for college. During the first phase of this work several research studies were conducted between 2008 and 2013. This research included statistical relationship studies using data from students who were sampled and assessed by NAEP 12th grade reading or mathematics in 2009 and also attended public colleges in Florida: https://www.nagb.gov/focus-areas/reports/preparedness-research/docs/statistical-relationships/landing/Florida_Statistical_Study.html.

In conjunction with the administration of the 2013 NAEP 12th grade assessments in reading and mathematics, the Governing Board approached the Michigan Department of Education about the possibility of conducting statistical relationship studies with data from the state longitudinal data system. Data sharing agreements were established between Governing Board staff, NCES staff, NCES contractors Westat and ETS, and the Michigan Department of Education. The data sharing agreement limits the scope of the study to the specific research questions that were developed in conjunction with COSDAM and reported here. A similar study is still underway using data provided by the Massachusetts Department of Education and is expected to be available for COSDAM review in spring 2021.

In addition to bearing on the Board's longstanding interest in the knowledge and skills relevant for postsecondary preparedness, this work is an example of how NAEP can be linked to external data sources with the potential to inform policy and practice, as described in the Board's new Strategic Vision.

During the upcoming COSDAM meeting, COSDAM members will discuss the draft report and provide feedback to be used in finalizing the report and closing out this study.

Draft for COSDAM Review

NAEP Grade 12 Academic Preparedness Research: Analyses Relating Michigan Students' Performance on NAEP to Postsecondary Performance

Nuo Xi

Laura Jerry

David Freund

Mei-Jang Lin

NCES Project Officer: Bill Tirre, Senior Technical Advisor

National Assessment Governing Board Staff: Sharyn Rosenberg, Assistant Director for Psychometrics

Prepared by Educational Testing Service for the National Center for Education Statistics under contract ED-IES-12-R-0021, Task 9.3.1 at the request of the National Assessment Governing Board.

This research result used data collected and maintained by the Michigan Department of Education (MDE) and/or Michigan's Center for Educational Performance and Information (CEPI). Results, information and opinions solely represent the analysis, information and opinions of the author(s) and are not endorsed by, or reflect the views or positions of, grantors, MDE and/or CEPI or any employee thereof.

Introduction

Starting in early 2003, the National Assessment Governing Board (Governing Board) embarked on an ambitious mission to redesign grade 12 assessments and reporting as recommended by the National Commission on 12th Grade Assessment and Reporting. The commission recommended that a state program similar to the 4th and 8th grades should be implemented at grade 12 and that NAEP should start reporting on the readiness of 12th graders for college, training for employment, and entrance into the military. As a result of the second recommendation, a number of studies were conducted to assess whether and in what ways NAEP could report on *academic preparedness*.

The Phase I of this research endeavor was concluded with more than 30 studies using the 2009 NAEP grade 12 Reading and Math data. The Phase I research findings supported inferences about NAEP performance and academic preparedness for college (Governing Board preparedness research website[1]) and identified potential benchmarks on the 12th grade NAEP Reading and Math scales to indicate a reasonable probability of being academically prepared (Fields, 2014). According to the Governing Board definition, to be "academically prepared for college," 12th graders should have the knowledge and skills in Reading and Math to qualify for placement into entry-level college credit courses that meet general education requirements, without the need for remedial coursework in either subject. A scale score of 302 on the NAEP grade 12 Reading assessment (equivalent to the *NAEP Proficient* cut score) and 163 on the NAEP grade 12 Math assessment (between the *NAEP Basic* cut score of 141 and the *NAEP Proficient* cut score of 176) were identified to project a reasonable probability of being academically prepared for college. As a result, the percentage of 12th grade students in the U.S. who were academically prepared for college has been estimated and reported beginning with the 2013 assessments in Reading and Math. Extensive details about the Phase I research work can be found in Fields (2014) and the Governing Board preparedness research website.

As part of the Phase I research efforts, Florida participated in the research by providing longitudinal data that could be linked to the Florida students' performance on the 2009 NAEP grade 12 Reading and Math. These data were a critical component for the validity evaluation of the benchmarks identified by the statistical linking study conducted in Phase I. The Florida longitudinal study results can be found in Moran, Freund, and Oranje (2012).

In the second phase of the Governing Board's academic preparedness research, additional state partners agreed to provide longitudinal data that can be linked to the 2013 NAEP Reading and Math assessments at grades 8 and 12. Michigan, as one of the state partners, participated in the state-level statistical linking research connecting NAEP and ACT scores and provided data on students who were part of the NAEP grade 12 sample during the 2012-2013 school year. The statistical linking study connecting Michigan students' performance on the NAEP and ACT assessments was completed and documented in a separate report (Xi, Lin, Jerry, Freund, and Oranje, 2016). The current report describes the relationship between the performance on NAEP and postsecondary outcomes observed from the Michigan longitudinal data.

---

[1] Governing Board preparedness research website: https://www.nagb.gov/focus-areas/reports/preparedness-research.html

Data

This study used data from Michigan public school students who participated in the 2013 NAEP grade 12 Reading and Math assessments. Whereas grade 12 NAEP assessments in Reading and Math are in general administered to only nationally representative samples, 13 states (Arkansas, Connecticut, Florida, Idaho, Illinois, Iowa, Massachusetts, Michigan, New Hampshire, New Jersey, South Dakota, Tennessee, and West Virginia) volunteered to participate in a twelfth-grade state pilot program in 2013. From late January to early March 2013, NAEP assessments in Reading and Math were administered nationwide. In Michigan, about 2,900 and 3,100 students at grade 12 were assessed in Reading and Math, respectively. Sample sizes were rounded to the nearest hundred as required by the NCES Statistical Standards.

The process of matching longitudinal data to 2013 NAEP participants in Michigan was carried out through an agreement between the National Assessment Governing Board and the National Center for Education Statistics to have NAEP contractors Westat and ETS conduct the preparedness research work. In addition, data confidentiality agreements were established between all parties involved and the Michigan Department of Education. The Michigan Department of Education provided six years of longitudinal data that cover year 1 through year 6 out of high school for those 12th graders who took grade 12 NAEP assessments in 2013. Information provided by the longitudinal data include college enrollment, remedial course taking, GPA, and degrees obtained. Confidentiality of the students and student data was assured through the assignment of a pseudo ID for students. The pseudo ID was employed as a way to transfer student data to ETS *without* the need to include Personally Identifiable Information (PII) such as names and birthdates. Similarly, the pseudo ID was appended to the NAEP files by Westat who then provided that file to ETS, again *without* any PII. Via the pseudo ID, ETS subsequently matched longitudinal data to NAEP files. The overall matching rate between the NAEP sample and the longitudinal data is 98%.

Research Questions

The intention of the current study is to examine the relationship between performance on NAEP and postsecondary outcome for the Michigan 12th graders who took NAEP Reading and Math assessments in 2013. The longitudinal information was collected to pursue the following four research questions in accordance with the data sharing agreement:

1. What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students who enroll in remedial and non-remedial courses?
2. What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students with a first-year GPA of B- or above?
3. What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students who remain in college after each year?
4. What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students who graduate from college within 6 years?

The next section will provide details on how the longitudinal data were used to answer each research question, along with plots and summaries that help readers interpret the results.

Analysis Results

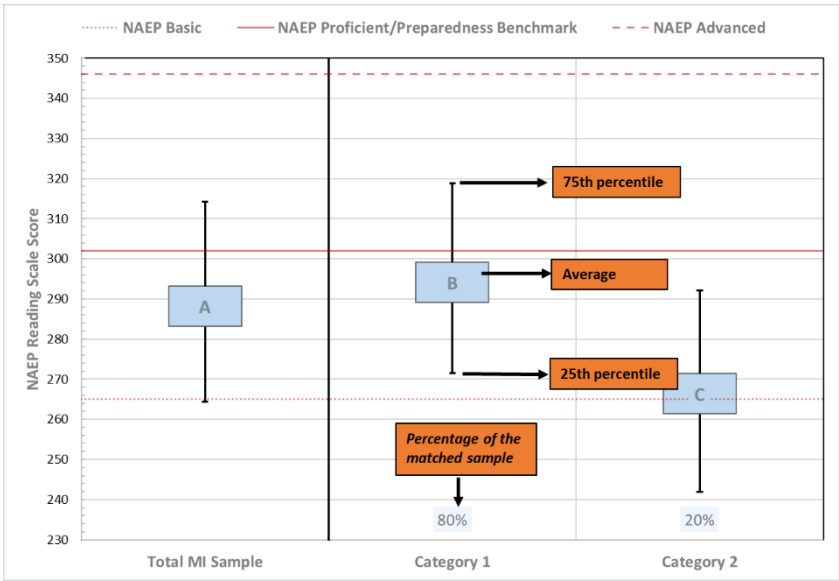
Interpretation of Figures

The relationships between grade 12 NAEP performance and postsecondary outcomes are displayed in boxplots, which is a standardized way to show distribution of data based on summary statistics such as the average, 5th, 25th, 75th, and 95th percentiles. The current report concerns the four research questions that request information on average and IQR. Therefore, the boxplots showed in this report will be constructed specifically to display the average, 25th, and 75th percentile of considered distributions.

Figure 1 presents an example of boxplot used in the current report. The NAEP sample matched to the longitudinal data will first be classified into categories of interest based on each research question. The average NAEP scale score, 25th, and 75th percentiles of each category will then be computed and displayed in a boxplot. Taking "Category 1" in Figure 1 as an example. The number in the box represents the average NAEP scale score of this category, while the ends of the whiskers are the 25th and 75th percentiles of the same category. The number above "Category 1" indicates the weighted percentage of this category out of the matched sample. As a reference, the distributional information of the total Michigan NAEP sample before matching (i.e., "Total MI Sample" in Figure 1), is also shown in each figure, but separated from the matched sample by a solid black line. The four horizontal lines indicate the *NAEP Basic*, *NAEP Proficient,* and *NAEP Advanced* cut scores of the corresponding subject, as well as the Preparedness Benchmark identified by the Phase I research studies (Fields, 2014). Notice that for NAEP Reading, the Preparedness Benchmark coincides with the *NAEP Proficient* cut score, and therefore only three red horizontal lines are shown in the Reading plots (Preparedness Benchmark line is shown in blue for Math).

To understand the relationship between the grade 12 NAEP performance and each postsecondary outcome variable, one needs to compare the boxplots between different categories to see whether the average NAEP scale score changes from one category to another and if so whether the difference is in a direction one would expect.

FIGURE 1: ILLUSTRATION OF BOXPLOT



RQ1: What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students who enroll in remedial and non-remedial courses?

The longitudinal data provided by the Michigan Department of Education include an indicator of college enrollment status for each academic year starting from the 2013-2014 through the 2018-2019 academic year. If this indicator is "No" for a student throughout these six years, he/she will be classified as *never enrolled in college*. Otherwise, he/she will be further categorized as *took at least 1 remedial course* or *never took remedial course* depending on his/her remedial course enrollment indicators.

Figure 2 and **Figure 3** depict the relationship between remedial course enrollment status and grade 12 NAEP Reading and Math performance, respectively. The NAEP performance of the overall grade 12 Michigan sample (before matching) is shown on the left of these two figures as a reference. These two figures show that between 67% and 68% of the matched Michigan 12th graders have never taken any remedial courses, about 13% have enrolled in at least one remedial course, while the other 20% never enrolled in college. Students who have never taken any remedial courses performed better on the NAEP assessments, with an average score close to (but still lower than) the corresponding *NAEP Proficient* cut score.

The average NAEP Reading scale score of those who have never taken any remedial courses is 298, very close to the *NAEP Proficient* cut score/Preparedness Benchmark identified for Reading (302) and the corresponding 25th and 75th percentiles are 276 and 322 (IQR=46). In comparison, the average NAEP Reading scale score of those who have taken at least one remedial course is 274, and the corresponding 25th and 75th percentiles are 255 and 295 (IQR=40).

The average NAEP Math scale score of those who have never taken any remedial courses is 164, below the *NAEP Proficient* cut score for Math (176) but slightly above the Preparedness Benchmark identified for Math (163). The corresponding 25th and 75th percentiles are 146 and 185 (IQR=39). In comparison, the average NAEP Math scale score of those who have taken at least one remedial course is 136, and the corresponding 25th and 75th percentiles are 122 and 152 (IQR=30).

FIGURE 2: NAEP GRADE 12 READING PERFORMANCE BY STUDENT ENROLLED IN REMEDIAL COURSES

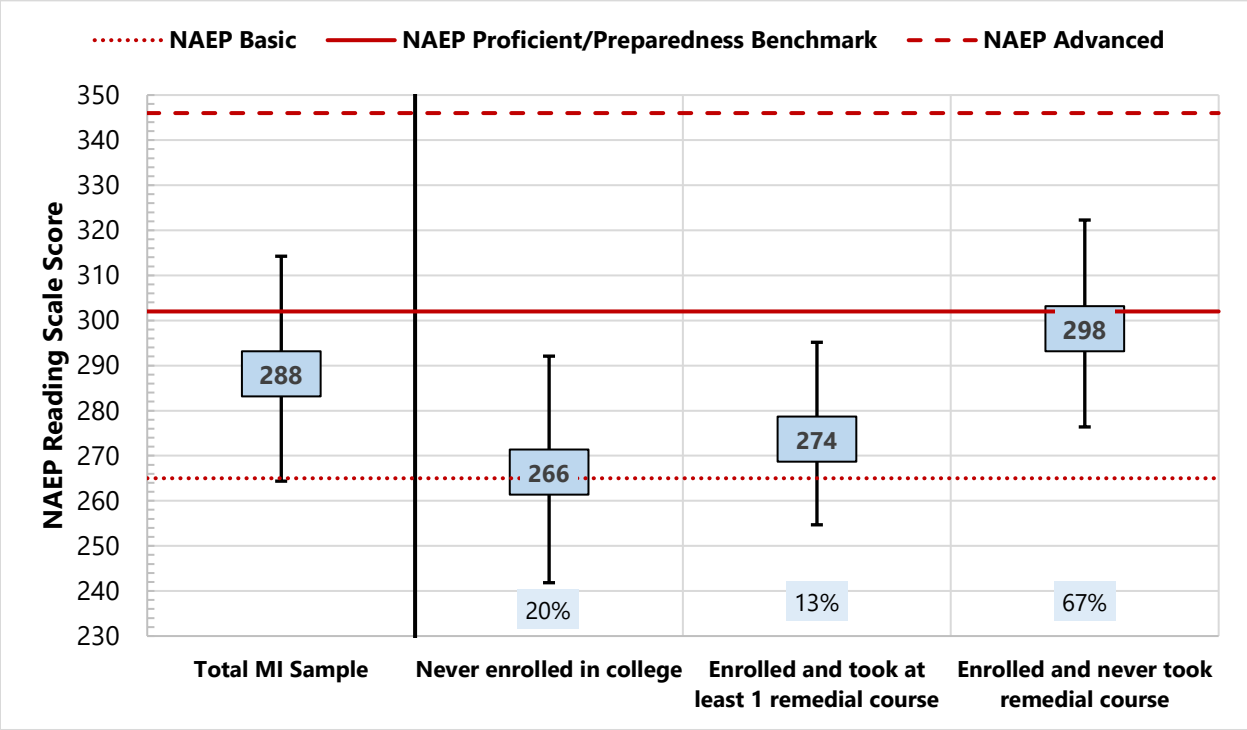
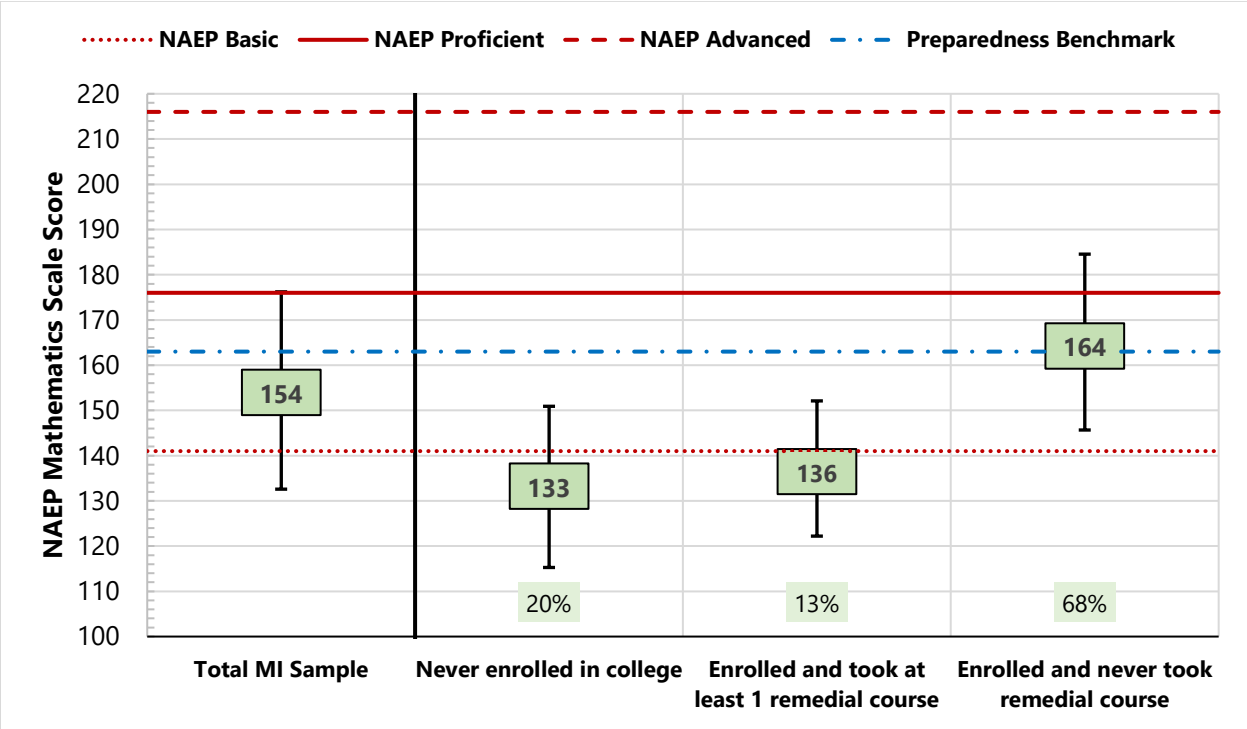


FIGURE 3: NAEP GRADE 12 MATH PERFORMANCE BY STUDENT ENROLLED IN REMEDIAL COURSES



RQ2: What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students with a first-year GPA of B- or above?

Based on the matched sample's enrollment status during the 2013-2014 academic year and their GPA record, students were categorized into the following four categories:

- Not enrolled in college 2013-2014
- Enrolled 2013-2014 but no GPA available
- First-year GPA (2013-2014) < 2.5
- First-year GPA (2013-2014) >= 2.5

Figure 4 and **Figure 5** provide a graphical summary of the relationship between first-year GPA (FYGPA) and grade 12 NAEP performance. These two figures show that, of the matched grade 12 Michigan sample, between 32% and 33% had a FYGPA at B- or above[2], 15% had a FYGPA below B-, and the rest 53% were either not enrolled in the 2013-2014 academic year or enrolled but with no GPA available. Students who had a FYGPA at B- or above had an average NAEP Reading scale score of 305, above the *NAEP Proficient* cut score (302). The corresponding 25% and 75% percentiles are 285 and 327 (IQR=42). The average NAEP Math scale score for those who had a FYGPA at B- or above was 169, below the *NAEP Proficient* cut score (176) but above the Preparedness Benchmark (163). The corresponding 25% and 75% percentiles are 151 and 189 (IQR=38). Students who enrolled in the 2013-2014 academic year but with no GPA available had the second highest average NAEP scale scores among the four categories, suggesting that this category may also include some of the higher-performing students.

[2] A 2.5 grade point average (GPA) from a possible 4.0 total GPA is equal to a B- letter grade.

FIGURE 4 NAEP GRADE 12 READING PERFORMANCE BY FIRST-YEAR GPA

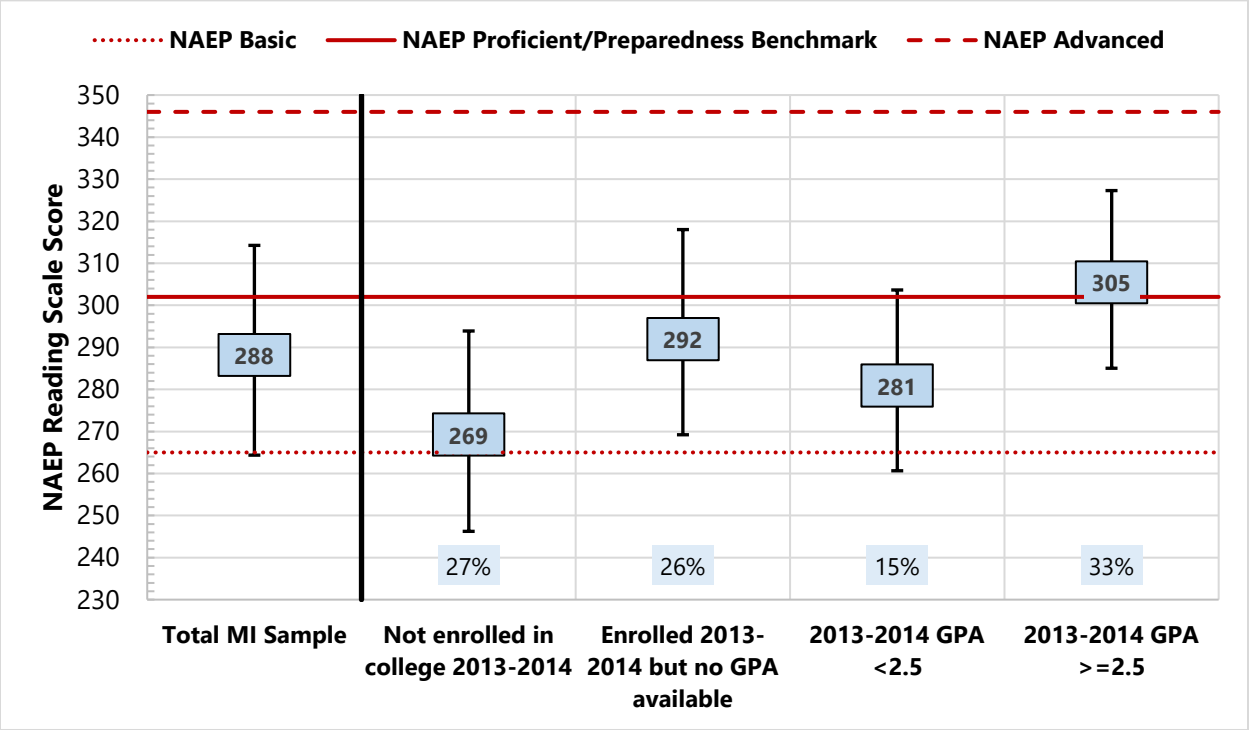
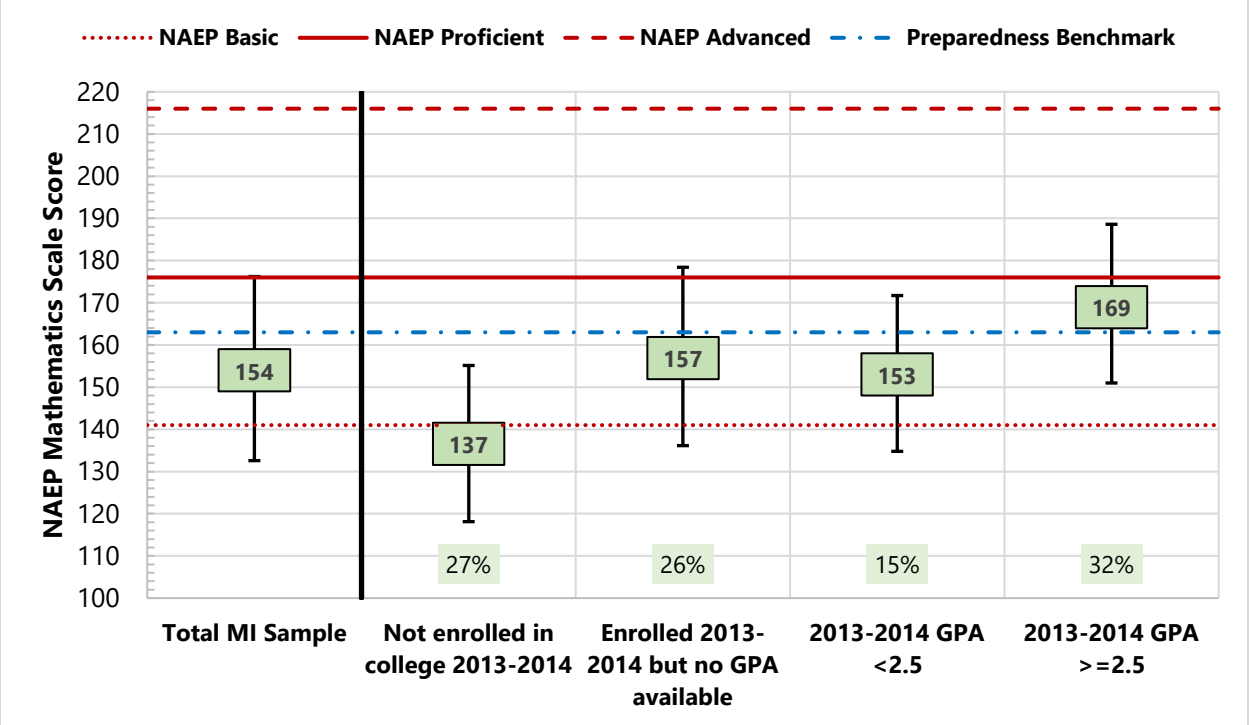


FIGURE 5 NAEP GRADE 12 MATH PERFORMANCE BY FIRST-YEAR GPA



RQ3: What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students who remain in college after each year?

To define "remain in college after each year," students were categorized beginning with the 2013-2014 academic year by consecutive years of enrollment, into the following seven categories:

- Not enrolled 2013-2014 (year 1)
- Enrolled 2013-2014 (year 1) but not 2014-2015 (year 2)
- Enrolled 2013-2015 (year 1 to year 2) but not 2015-2016 (year 3)
- Enrolled 2013-2016 (year 1 to year 3) but not 2016-2017 (year 4)
- Enrolled 2013-2017 (year 1 to year 4) but not 2017-2018 (year 5)
- Enrolled 2013-2018 (year 1 to year 5) but not 2018-2019 (year 6)
- Enrolled 2013-2019 (year 1 to year 6)

If there was a break and then resumption in enrollment, the student is kept in the lower category. For example, a student who enrolled in 2013-2014, took a break in 2014-2015, and then enrolled again in 2015-2016 is kept in the category of "Enrolled 2013-2014, but not 2014-2015". Given that there were students representing nearly every combination of enrollment years, it was determined that looking only at consecutive years of enrollment was the most straightforward and consistent approach.

Figure 6 and **Figure 7** display the relationship between the grade 12 NAEP performance on Reading and Math and consecutive years of enrollment. In general, these two figures suggest a positive relationship between grade 12 NAEP performance and consecutive years of college enrollment, with the fifth category (i.e., "Enrolled 2013-2017 but not 2017-2018") having the highest average NAEP scale score for both Reading and Math. Students who enrolled in 2013-2017 but not 2017-2018 had an average NAEP Reading scale score of 311, above the corresponding *NAEP Proficient* cut score (302). Students who enrolled in 2013-2017 but not 2017-2018 had an average NAEP Math scale score of 172, only 4 points below the corresponding *NAEP Proficient* cut score (176) and above the Preparedness Benchmark identified for NAEP Math (163). Students in the last two categories, i.e., "Enrolled 2013-2018 but not 2018-2019" and "Enrolled 2013-2019", also had average NAEP scale scores close to or above the corresponding *NAEP Proficient* cut scores, suggesting a positive relationship between grade 12 NAEP performance and perseverance.

FIGURE 6: NAEP GRADE 12 READING PERFORMANCE BY YEARS OF ENROLLMENT

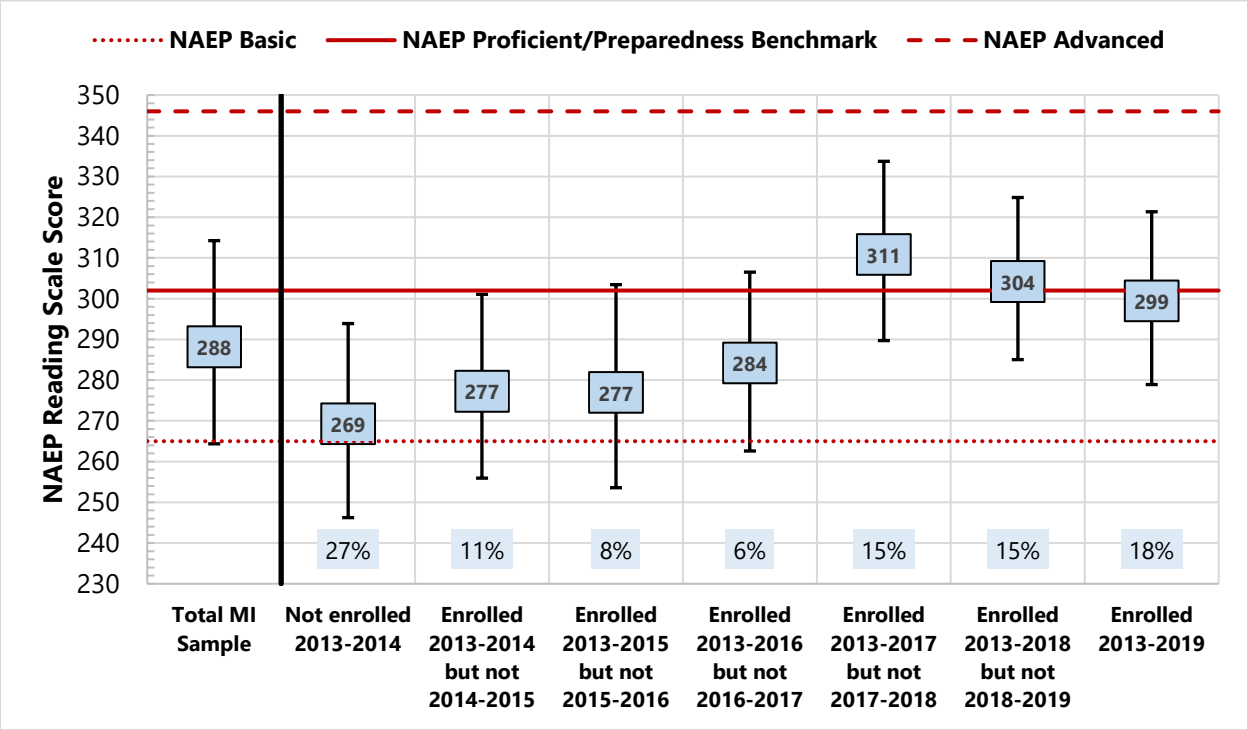
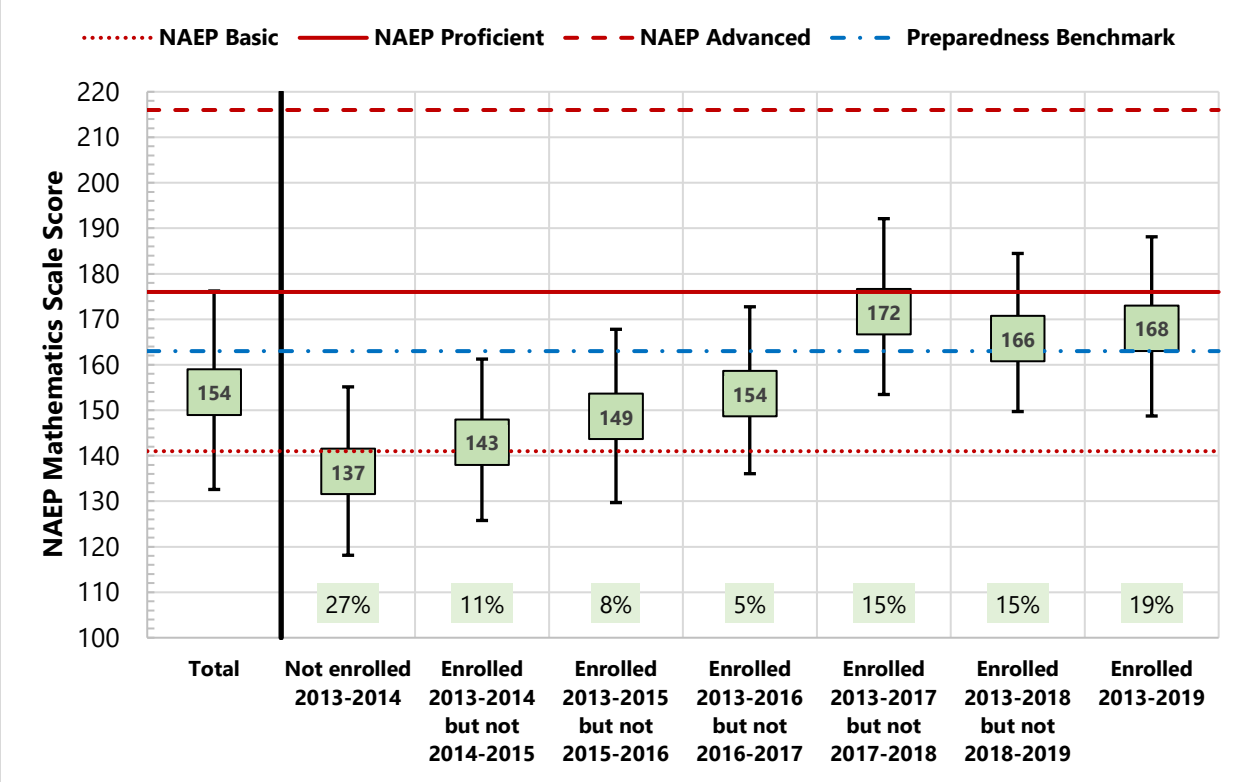


FIGURE 7: NAEP GRADE 12 MATH PERFORMANCE BY YEARS OF ENROLLMENT



RQ4: What are the average grade 12 NAEP Reading and Math scores (and the IQR) for students who graduate from college within 6 years?

By combining the college enrollment status and the information on degrees received, students are categorized into the following four categories:

- Never enrolled in college
- Enrolled but did not receive any degree/certificate
- Enrolled and achieved degree/certificate less than Bachelor's
- Received Bachelor's, Baccalaureate, or Master's degree

Students who received various certificates or Associate's degrees were put into the third category, i.e., "Enrolled and achieved degree/certificate less than Bachelor's", and students who received a Bachelor's, Baccalaureate, or Master's degree were put into the fourth category. The last two categories are of particular interest to the current research question, as the students classified in these two categories can be considered "graduate from college within 6 years".

Figure 8 and **Figure 9** provide information on the relationship between grade 12 NAEP performance and college graduation status with award level. Of the matched grade 12 Michigan sample, about 42% completed college within 6 years (depending on the matched NAEP subject), and the rest 58% of the matched sample either never enrolled in college or enrolled but did not achieve any degrees. For those who completed college with a degree/certificate less than Bachelor's, their average grade 12 NAEP Reading scale score is 290, and their average grade 12 NAEP Math scale score is 156, both are lower than the corresponding *NAEP Proficient* cut score. For those who achieved a Bachelor's, Baccalaureate, or Master's degree within 6 years, their average grade 12 NAEP Reading scale score is 311, and their average grade 12 NAEP Math scale score is 176, higher than or equal to the corresponding *NAEP Proficient* cut score. Students who never enrolled in college have the lowest average grade 12 NAEP performance, while students who enrolled but did not achieve any degree have the second lowest grade 12 NAEP performance among the four considered categories.

FIGURE 8: NAEP GRADE 12 READING PERFORMANCE BY GRADUATE WITHIN 6 YEARS

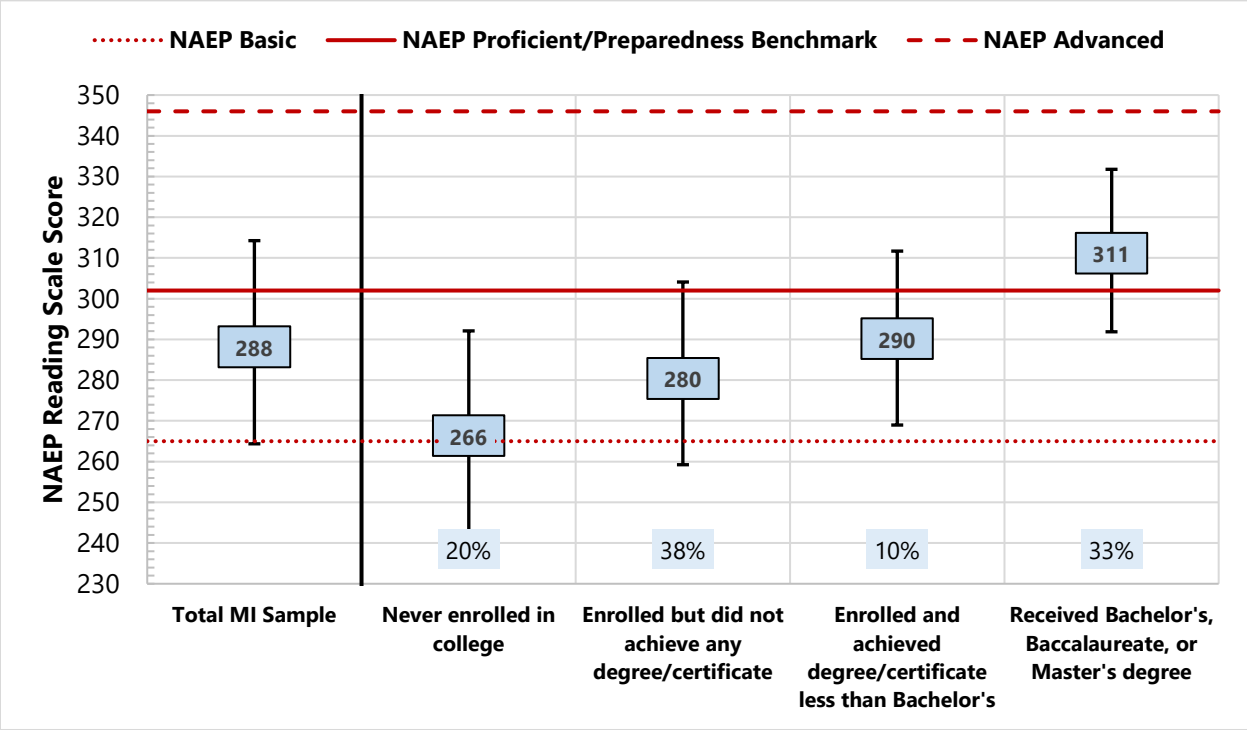
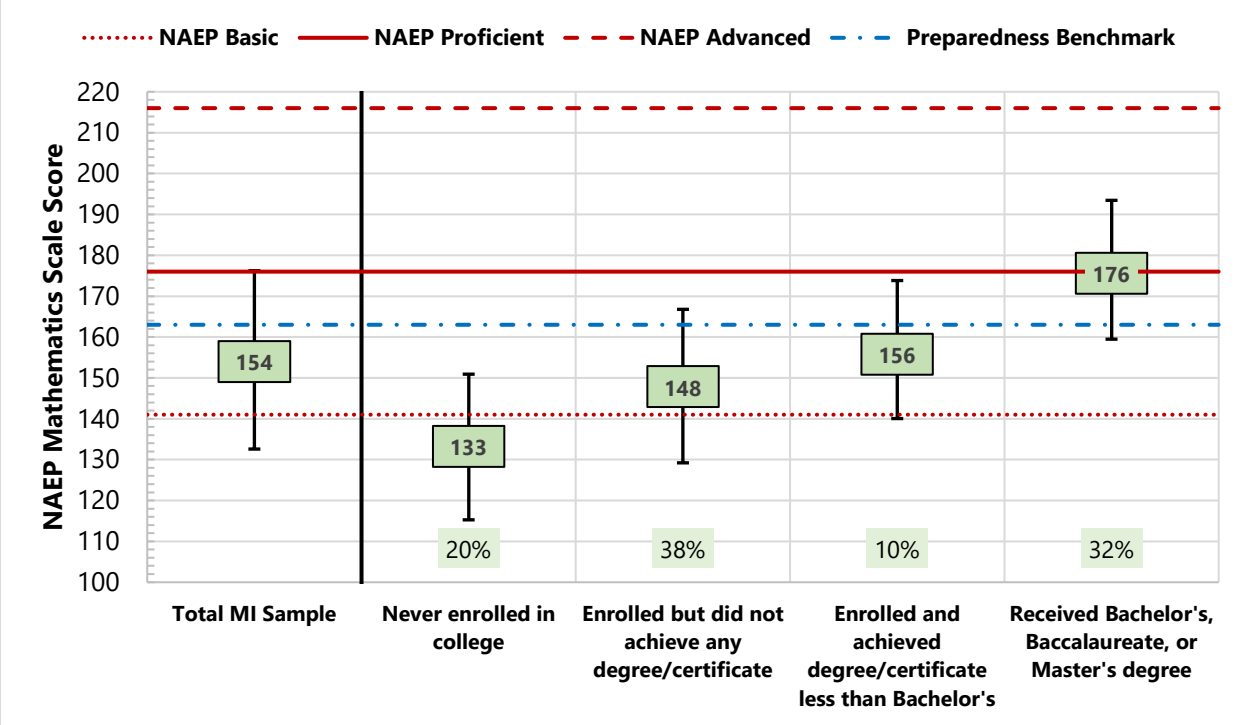


FIGURE 9: NAEP GRADE 12 MATH PERFORMANCE BY GRADUATE WITHIN 6 YEARS



Summary

The current study examines the relationship between the grade 12 NAEP Reading and Math performance and postsecondary academic outcomes for the 2013 Michigan NAEP sample. The NAEP data are matched to the longitudinal data provided by the Michigan State Department of Education through the usage of Pseudo ID, which ensures the confidentiality of student data. The overall matching rate is at 98%.

Four research questions are examined (see **Research Questions**). Analysis results along with plots and summaries are provided for each research question in the **Analysis Results** section. In general, the results suggest that for the matched Michigan 12th graders who took NAEP in 2013, their grade 12 NAEP performance is positively related to their postsecondary outcomes evaluated in this report. For instance, students who enrolled in college and never took remedial courses are associated with better performance on the grade 12 NAEP Reading and Math assessments compared to those who had taken at least one remedial course in college. In fact, the average grade 12 NAEP Reading and Math scale scores of those who never took remedial courses are very close to the Preparedness Benchmark identified for each subject. Recall that the Preparedness Benchmark was set at NAEP scale score that indicates students have required knowledge and skills to be placed into entry-level college credit courses without the need for remedial courses, using the 2009 NAEP national data. Therefore, the closeness between the identified Preparedness Benchmarks and the observed 2013 NAEP performance of the Michigan 12th graders who never took remedial courses could be seen as confirmative and validity evidence for using the Preparedness Benchmarks to indicate 12th graders' academic preparedness for college.

It should be emphasized that the longitudinal data were matched to the Michigan sample who took the 2013 NAEP grade 12 Reading and Math assessments. As the grade 12 NAEP Reading and Math assessments were administered to public school students only, interpretation of the results presented in the current report should be limited to this population (i.e., public school students).

References

Fields, R. (2014). *Towards the National Assessment of Education Progress (NAEP) as an Indicator of Academic Preparedness for College and Job Training*. Washington, DC: National Assessment Governing Board.

Moran, R., Freund, D., and Oranje, A. (2012). *NAEP 12th Grade Preparedness Research: Analyses Relating Florida Students' Performance on NAEP to Preparedness Indicators and Postsecondary Performance*. Research report prepared for National Assessment Governing Board.

Xi, N., Lin, M., Jerry, L., Freund, D., & Oranje, A. (2016). *NAEP Grade 12 Academic Preparedness Research: Establishing a Statistical Relationship between NAEP and ACT Assessments in Reading and Mathematics for Grade 12 Michigan Students*. Research report prepared for National Assessment Governing Board.