

DELAWARE ASSESSMENT INVENTORY – SUMMARY OF FINDINGS

Prepared for Cape Henlopen School District

December 2015



In the following report, Hanover Research presents summary findings from a series of projects conducted on behalf of Cape Henlopen School District assessing the usefulness of the various district- and school-mandated assessments administered to students in the district for diagnostic and evaluative purposes.



www.hanoverresearch.com

TABLE OF CONTENTS

Executive Summary and Key Findings	3
INTRODUCTION	3
KEY FINDINGS	4
Section I: Methodology	5
OVERVIEW OF DELAWARE ASSESSMENT INVENTORY PROJECT	5
Inventory Study	5
Teacher Surveys	6
Parent and Student Surveys	7
Section II: Assessment Rankings and Analysis	9
FINDINGS	9
RANKINGS BY USEFULNESS, ALIGNMENT, AND TEACHER RECOMMENDATION	11
RANKINGS BY STUDENT PERCEPTIONS	14
Section III: Results by Assessment	17
DIBELS/DAZE	17
GATES MACGINITIE	19
MATH UNIVERSAL SCREENING TOOL (MUST)	22
STAR READING	23
STAR MATH	26
PERFORMANCEPLUS	29

EXECUTIVE SUMMARY AND KEY FINDINGS

INTRODUCTION

In this report, Hanover Research presents summary findings from a series of projects conducted on behalf of Cape Henlopen School District assessing the usefulness of the various district- and school-mandated assessments administered to students in the district for diagnostic and evaluative purposes. This research is in response to a recent initiative launched by Delaware Governor Jack Markell to review “tests administered by the state, districts, and individual schools with the goal of decreasing the testing burden on students and teachers and increasing the time available for teaching.”¹ Cape Henlopen School District has partnered with Hanover Research to conduct assessment inventories and gather feedback from teachers, students, and parents, in accordance with the methodological guidelines prescribed by the Delaware Department of Education (DDOE) and Achieve, Inc., which have together managed Governor Markell’s initiative.

This summary report comprises three sections:

- **Section I** summarizes the methodology prescribed by DDOE and Achieve and executed by Hanover Research and Cape Henlopen School District. It describes the multiple projects completed over the course of four months that will inform the district in making recommendations to the DDOE.
- **Section II** presents assessment rankings and analysis across a number of metrics as measured through stakeholder surveys. Hanover Research ranks each of six assessments according to their reported usefulness, alignment to the Common Core State Standards, teachers’ recommendation that they be continued, and student perceptions.
- **Section III** presents individual results for each of the six assessments covered by the teacher surveys, noting reported usefulness, alignment to the Common Core, and teachers’ recommendations for continuation.

¹ Delaware Assessment Inventory Project Supplementary Guidelines.” Delaware Department of Education, April 2015. p. 3.

KEY FINDINGS

- **Teachers report the highest satisfaction with the Math Universal Screening Tool (MUST).** Stakeholder surveys distributed to teachers suggest that MUST is perceived as one of the most useful and most recommended assessments given to students in the district. Cape Henlopen teachers feel that MUST is useful across a wide number of areas, especially diagnostic and instructional purposes, and 62 percent recommend that the district continue its use without reservation, more than any other assessment.
- **Only 24 percent of teachers recommend that the district continue using the Gates MacGinitie reading assessment without reservations.** With the exception of PerformancePLUS, no other assessment saw fewer than 45 percent of teachers recommend it without reservations. The Gates MacGinitie was further seen as having average usefulness and relatively poor alignment to the Common Core.
- **While teachers report that the PerformancePLUS pre- and post-assessments are useful and aligned to the Common Core State Standards, just 29 percent recommend without reservations that the district continue using the tests.** However, given the assessment's role as the DPAS Component V assessment for evaluating teacher performance, this may be a reflection of opinions of the test-based evaluation framework as much as the quality of the PerformancePLUS.
- **Similarly, teachers feel that assessments in general are more useful for diagnostic and instructional purposes than for evaluative ones.** Across all assessments, teachers are more likely to report that they are more useful for diagnosing individual students' strengths and needs and informing and improving instructional practices than they are for making evaluative decisions such demonstrating teacher effectiveness, deciding whether to promote or retain a student, or factoring into course grades.
- **In general, parent respondents indicate feeling fairly positive about the tests their eldest child took last year, specifically the STAR Reading and STAR Math assessments.** Approximately half of all respondents or more agree or strongly agree with five of the six positive statements presented to them about the tests their child took last year. However, parents are not largely aware of assessments other than the STAR Reading and MATH tests.

SECTION I: METHODOLOGY

This section describes the methodology used by Hanover Research and Cape Henlopen School District to assess the usefulness and impact of common assessments used within the district. The methodology follows the requirements and recommendations of the Delaware Department of Education (DDOE) and Achieve, Inc., which jointly created the various components of the Delaware Assessment Inventory project. This summary report, to be delivered to the DDOE, is the culmination of multiple projects conducted by Hanover Research on behalf of Cape Henlopen School District to gather stakeholder feedback and input regarding student assessments in the district. The following subsections detail the three project stages completed previously by Hanover Research to determine what assessments are used at what grade levels, which to include in stakeholder surveys, and, finally, which are viewed by stakeholders as effective tools for measuring student progress and good uses of district resources.

OVERVIEW OF DELAWARE ASSESSMENT INVENTORY PROJECT

As part of Governor Markell’s initiative to streamline assessment mechanisms in Delaware schools, the DDOE and Achieve, Inc., a Washington D.C.-based education organization, designed a systematic methodology for evaluating assessment use and gathering stakeholder feedback that would be used by each district in the state to make recommendations for streamlining assessment in their district. The sequence of research projects undertaken by Hanover Research and Cape Henlopen School District adheres to the recommendations of DDOE and Achieve, as outlined in the document “Delaware Assessment Inventory Project Supplementary Guidelines,” which details a multi-stage process for evaluating school- and district-wide assessment use.² These projects include an initial inventory study, multiple teacher surveys, and a student and parent survey.

INVENTORY STUDY

Achieve and DDOE’s guidelines specify that the project begin with an “assessment inventory” that captures the full range of assessments being used throughout the district.³ The purpose of this initial project is to systematically identify which assessments are being used within the district and inform the design of evaluative stakeholder surveys. To begin the process, in July 2015, Hanover Research distributed a spreadsheet-based data collection tool to school- and district-level administrators who provided information about the various assessments taken by students within the district. Data were collected on a wide range of factors for each assessment, including the name of the assessment, which subject it tests, how often students take it, who requires it (the district or the school), how useful it is, and the assessment’s primary purpose. These were then returned to Hanover Research, compiled, and submitted to Cape Henlopen School District.

² “Delaware Assessment Inventory Project Supplementary Guidelines.” Delaware Department of Education, April 2015.

³ Ibid., p. 7.

TEACHER SURVEYS

Following the assessment inventory, the DDOE and Achieve require two distinct teacher surveys or focus groups, one focused on “taking the temperature” on assessments in the district and another on “understanding assessment use.”⁴ To this end, Hanover Research designed and administered two surveys of Cape Henlopen teachers over a two-month period.

The initial survey, designed in accordance with DDOE and Achieve specifications to “illuminat[e] teacher perspectives on the district’s assessment program as a whole to build a greater understanding of the testing environment and help build a case for action,” was launched on August 31 and remained open through September 11, 2015. It sought to determine which assessments with which teachers were most familiar and to gauge their initial opinions regarding them. The survey received 236 complete and 69 partial responses from Cape Henlopen teachers, who were queried regarding their familiarity with the 15 assessments identified in the inventory study and highlighted in Figure 1.1 below.

Figure 1.1: Assessments Analyzed in “Taking the Temperature” Teacher Survey

ACADEMIC AREA	ASSESSMENT
English Language Arts (ELA)	<ul style="list-style-type: none"> ▪ Dynamic Indicators of Basic Early Literacy Skills (DIBELS/DAZE) ▪ Gates MacGinitie ▪ STAR Reading ▪ Common Core Writing Assessments ▪ MAZE (Tier 2 and Tier 3 RTI only)
Mathematics	<ul style="list-style-type: none"> ▪ STAR Math ▪ Scholastic Math Inventory (SMI) ▪ Math Universal Screening Tool (MUST)
Enrichment	<ul style="list-style-type: none"> ▪ Read 180 R-Skills ▪ Scholastic R-Skills ▪ Diagnostic Reading Assessment DAR
Special Education	<ul style="list-style-type: none"> ▪ STAR Reading (modified for Special Ed) ▪ STAR Math (modified for Special Ed)
Additional Assessments	<ul style="list-style-type: none"> ▪ Advanced Placement (AP) Exams ▪ DPAS Pre- and Post-Test Assessments (Component V)⁵

The “taking the temperature” survey also served to inform the design of the second teacher survey, titled “understanding assessment use” by Achieve and DDOE. This second survey covered a smaller number of assessments than then the “taking the temperature” survey, but sought to gain deeper insight into each assessment’s perceived use and importance. Cape Henlopen School District and Hanover Research collaborated to choose only the most

⁴ “Listening to Teachers: Sample Focus Group and Survey Materials.” Achieve.

⁵ For subsequent surveys, this was replaced by PerformancePLUS.

relevant assessments on which to gather detailed feedback. Two criteria generally informed the decision to include or exclude an assessment:

- **Teacher familiarity** – The initial assessment inventory and first teacher survey included a wide variety of assessments, many of which are used by only a small number of teachers and taken by just a small subset of students. Given sample size considerations, Hanover Research and Cape Henlopen School District elected not to include in the second survey assessments with which the large majority of teachers were not familiar. This resulted in the exclusion of MAZE, SMI, and enrichment- and special education-focused assessments.
- **Decision-making ability** – Some tests, such as the Common Core Writing Assessments, are mandated by the state and not administered at the discretion of the district. Others, such as AP exams, are not likely to be changed by the district and were excluded for that reason. Hanover Research made an effort to include just those assessments over which the district had control and were known to teachers.

The result of this winnowing process was the selection of six widely-familiar assessments for the “understanding assessment use” survey:

- STAR Reading
- STAR Math
- Dynamic Indicators of Basic Early Literacy Skills (DIBELS/DAZE)
- Gates MacGinitie
- Math Universal Screening Tool (MUST)
- PerformancePLUS

The survey, which ran from late September through October 2015 and received 124 complete and 37 partial responses, asked teachers to help determine which common assessments are necessary for diagnostic, instructional, and accountability purposes, and which can be modified or eliminated, aiming to develop a more precise understanding of teachers’ experiences with particular school- and district-level assessments. The survey, which provided much of the information that is used in making final summary recommendations, consisted primarily of multiple-choice and ranking-scale questions but also included a small number of open-ended questions to give respondents an opportunity to provide qualitative feedback. **These questions asked respondents about the usefulness of the assessments with which they were familiar**, as well as how aligned the assessments are to the Common Core State Standards and whether they recommend their continued use. Respondents were shown the same set of questions for each of the six assessments.

PARENT AND STUDENT SURVEYS

In addition to the two teacher surveys, DDOE also recommends surveys of parents and students, termed “listening to parents” and “listening to students.” These two surveys were administered concurrently with the second teacher survey and results provided to Cape

Henlopen School District in a three-part analysis. The parent and student surveys were developed in accordance with DDOE and Achieve guidelines to gather information on student and parent perceptions of testing in the district and were the only surveys these two stakeholder groups received. The surveys largely comprised multiple-choice and ranking-scale questions. Both parents and students were asked about all assessments identified in the inventory study. Figure 1.2 below details the response totals for each survey.

Figure 1.2: Response Totals for Parent and Student Surveys

SURVEY	RESPONSES RECEIVED	RESPONSES INCLUDED IN ANALYSIS
Parent	105	83
Student	571	521

Based on the number of responses received and reported familiarity with the assessments, analysis was provided for STAR Reading and STAR Math (parents), and STAR Reading, STAR Math, DIBELS/DAZE, and MUST (students). **Because so few parents reported being familiar with tests other than the STAR assessments, survey results from parents are not included in the summary analysis presented in Section II.** While these results are useful for evaluating the STAR exams by themselves, they are not as helpful for comparing and ranking the wider array of assessments taken by Cape Henlopen students. Results of the student survey, which included four assessments with sufficient responses, are included in this report in Section II.

SECTION II: ASSESSMENT RANKINGS AND ANALYSIS

In this section, Hanover ranks and analyzes the six assessments covered in the “understanding assessment use” survey administered to Cape Henlopen School District teachers, according to usefulness, alignment to the Common Core State Standards, and teacher recommendations. Figure 2.1 details the assessments covered.

Figure 2.1: Assessments Covered in “Understanding Assessment Use”

ASSESSMENT	SUBJECT	GRADE LEVELS	DESCRIPTION
Dynamic Indicators of Basic Literacy Skills (DIBELS/DAZE)	ELA	K-5	Cognitive measure (Grades K-6) evaluating underlying reading skills.
Gates MacGintie	ELA	1,6	Cognitive measure (Grades K-12) of general reading achievement.
STAR Reading	ELA	2-12	Cognitive, adaptive measure (Grades 1-12) of reading comprehension.
STAR Math	Mathematics	K-12	Provides information about student growth and achievement in grades 1-12.
Math Universal Screening Tool (MUST)	Mathematics	K-5	Provides information about student growth in math. Can be given three times a year. ⁶
Performance Plus	ELA and Math	K-12	Used for Component V rating for teacher evaluation. Pre/Post Test used to show growth. ⁷

Source: Delaware Department of Education⁸

FINDINGS

- While **DIBELS/DAZE** had below average aggregate usefulness rankings, the assessment shone in teachers’ assessment of its utility as a diagnostic and instructional tool. DIBELS/DAZE ranked only behind MUST for its usefulness for diagnosing individual student strengths and needs (64 percent found it useful or very useful), instructional uses (48 percent), and improving classroom instruction (42 percent). Forty-five percent of teachers recommended the district continue with the assessment without reservations (fourth out of six assessments), and 31 percent considered it strongly or very strongly aligned to the Common Core.

⁶ Description provided by Cape Henlopen School District.

⁷ Description provided by Cape Henlopen School District.

⁸ Unless otherwise noted, all descriptions are drawn from “2015-15 DPAS-II Component V Student Growth Measures.” Delaware Department of Education. <http://www.doe.k12.de.us/cms/lib09/DE01922744/Centricity/Domain/375/External%20Measures%202015-16.pdf>

- Given to just 400 students annually, teachers reported generally low opinions of the **Gates MacGinitie** reading test. While considered relatively useful for diagnosing individual student strengths and needs, informing instructional practice, and predicting performance on future assessments, the Gates MacGinitie was recommended for continuation in the district by fewer teachers (24 percent) than any other assessment and had the second-lowest indication of alignment to the Common Core State Standards.
- The **STAR Reading** test is the only ELA assessment administered to students in Cape Henlopen past Grade 6, when the Gates MacGinitie is taken for the second time. A total of 59 percent of teachers recommended or strongly recommended that the district continue to use the assessment, more than any other test except for MUST. However, it received only average indications of usefulness, as approximately a third of teachers felt it was useful or very useful for instructional uses, improving classroom instruction, and informing instructional practice. Further, it was viewed as the least aligned to the Common Core of any examined assessment.
- **STAR Math** largely mirrored but fell slightly below STAR Reading on most metrics. Fifty-five percent of teachers recommended the district keep it without reservations (third of six exams), and teachers generally found it the *least* useful of any assessment, though not by a wide margin. For instance, STAR Math was ranked least or second-least useful for diagnosing individual student strengths and needs (fifth of six), instructional uses (sixth of six), informing instructional practice (sixth of six), and informing classroom instruction (sixth of six). The gap between STAR Math and the most useful test in each metric was never more than 24 percent and reached as low as 17 percent.
- The **Math Universal Screening Tool (MUST)** received outstandingly high feedback from Cape Henlopen teachers. Across eight metrics, MUST received an average ranking of 1.63 (out of six), including the highest ratings for diagnosing individual student strengths and needs, instructional uses, informing instructional practice, and improving classroom instruction. Furthermore, 62 percent of teachers (more than for any other assessment) recommended the district continue to use it, and more teachers considered it aligned to the Common Core than any other test, with the exception of PerformancePLUS.
- Teachers consistently ranked **PerformancePLUS** as among the more useful assessments, particularly for evaluating teacher effectiveness (first), factoring into course grades (first), and informing instructional practice (second). It was also considered the most aligned to Common Core Standards, outpacing all other tests (with the exception of MUST) by at least 29 points. However, only 29 percent of teachers recommended that Cape Henlopen continue to use the assessment, a mark lower than for every assessment other than Gates MacGinitie. However, this may be a reflection of teacher attitudes towards the DPAS Component V assessment-based evaluation framework as much as the quality of the PerformancePLUS pre- and post-assessments.

RANKINGS BY USEFULNESS, ALIGNMENT, AND TEACHER RECOMMENDATION

Figure 2.2 below presents Hanover’s rankings of the six assessments’ usefulness, alignment with the Common Core, and teacher recommendations to continue or discontinue the assessment, based on the survey data collected in the most recent faculty survey. In later sections, Hanover will provide further details regarding these rankings and the specific survey questions that inform them.

Figure 2.2: Assessment Rankings by Usefulness, Alignment, and Recommendation

AVERAGE USEFULNESS RANK ¹	RECOMMENDATION ²	AVERAGE ALIGNMENT SCORE ³
MUST (1.63)	MUST (62%)	PerformancePLUS (62%)
PerformancePLUS (3.38)	STAR Reading (59%)	MUST (55%)
Gates MacGinitie (3.63)	STAR Math (55%)	STAR Math (33%)
STAR Reading (3.75)	DIBELS/DAZE (45%)	DIBELS/DAZE (31%)
DIBELS/DAZE (4.13)	PerformancePLUS (29%)	Gates MacGinitie (27%)
STAR Math (4.50)	Gates MacGinitie (24%)	STAR Reading (26%)

Notes: 1) Scores closer to one indicate more useful assessments, while scores closer to six indicate less useful assessments.

2) Percentages indicate the proportion of teachers that either “recommend” or “strongly recommend” that the district continue to administer the assessment.

3) Percentages indicate the proportion of teachers that indicate that the assessment is either “aligned” or “strongly aligned” to the Common Core State Standards.

By a wide margin, **Cape Henlopen teachers considered the MUST assessment to have the greatest usefulness**, with an average ranking of 1.63 across the eight component variables (Figure 2.3).⁹ MUST was ranked as the most useful assessment (as measured by the percentage of teachers responding “useful” or “very useful”) at diagnosing individual student strengths and needs, instructional uses, informing instructional practice, and improving classroom instruction, and was also deemed the most useful assessment for deciding whether or not to promote or retain a student, though only 21 percent of teachers considered it useful for that purpose. The other assessments ranked somewhat similarly – ranging from 3.38 average rank to 4.50) – and were considered more useful in some areas than others. For instance, PerformancePLUS had the second-highest average ranking behind MUST, largely on the strength of top rankings for evaluating teacher performance and factoring into course grades, while other assessments such as DIBELS/DAZE were deemed useful in other areas such as diagnosing individual student strengths and needs and instructional uses.

In general, Cape Henlopen teachers feel that the assessments are most useful as diagnostic tools for gauging student learning needs and differentiating and informing instructional practices, rather than for evaluating and making definitive decisions regarding student and teacher performance. For instance, just 45 percent of teachers responded that the PerformancePLUS assessment was useful or very useful for diagnosing individual student strengths and needs – the lowest of the six assessments. However, only 15 percent of

⁹ Note that this analysis ranks all variables equally, computing the average rank across the eight variables.

teachers felt that the MUST assessment was useful for deciding whether to promote or retain a student – the highest scoring assessment for usefulness in this area. No more than 21 percent of teachers felt that any assessment was useful or very useful for promoting or retaining a student or factoring into course grades, and only two assessments – PerformancePLUS (35 percent) and MUST (21 percent) – were considered by more than a quarter of respondents to be useful for evaluating teacher effectiveness.

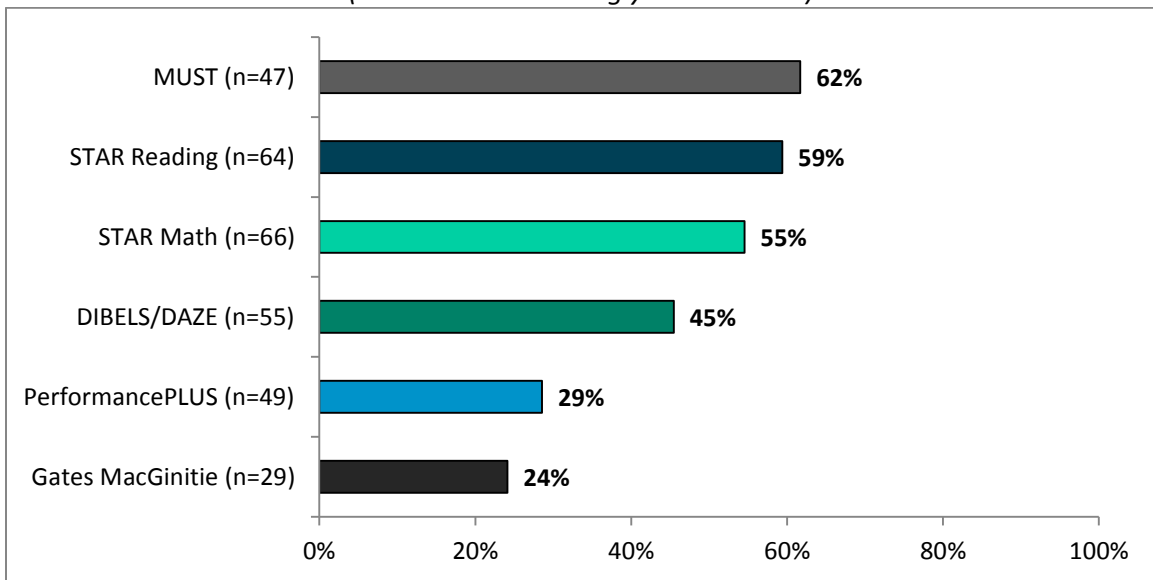
Figure 2.3: Assessment Rankings by Usefulness
(percent answering “useful” or “very useful”)

USEFULNESS CATEGORY	ASSESSMENT RANKINGS
Diagnosing individual student strengths and needs	1. MUST (68%) 2. DIBELS/DAZE (64%) 3. Gates MacGinitie (63 %) 4. STAR Reading (56%) 5. STAR Math (46 %) 6. PerformancePLUS (45%)
Instructional uses (e.g., reteaching a concept, flexible grouping)	1. MUST (58 %) 2. DIBELS/DAZE (48%) 3. PerformancePLUS (40%) 4. Gates MacGinitie (38%) 5. STAR Reading (36%) 6. STAR Math (36%)
Informing instructional practice	1. MUST (51%) 2. PerformancePLUS (43%) 3. Gates MacGinitie (42%) 4. DIBELS/DAZE (39%) 5. STAR Reading (37%) 6. STAR Math (32%)
Improving your classroom instruction	1. MUST (50%) 2. DIBELS/DAZE (42%) 3. PerformancePLUS (40%) 4. Gates MacGinitie (38 %) 5. STAR Reading (34%) 6. STAR Math (33%)
Predicting performance on a future assessment	1. Gates MacGinitie (36%) 2. STAR Math (34%) 3. STAR Reading (33%) 4. MUST (33%) 5. PerformancePLUS (29%) 6. DIBELS/DAZE (23%)
Evaluating teacher effectiveness	1. PerformancePLUS (35%) 2. MUST (25%) 3. STAR Reading (21%) 4. STAR Math (18%) 5. Gates MacGinitie (12%) 6. DIBELS/DAZE (11%)

USEFULNESS CATEGORY	ASSESSMENT RANKINGS
Promoting or retaining a student	1. MUST (21%) 2. STAR Reading (20%) 3. STAR Math (20%) 4. Gates MacGinitie (16%) 5. DIBELS/DAZE (15%) 6. PerformancePLUS (10%)
Factoring into course grades	1. PerformancePLUS (15%) 2. MUST (13%) 3. STAR Reading (12%) 4. STAR Math (11%) 5. Gates MacGinitie (8%) 6. DIBELS/DAZE (4%)

When teachers were asked to recommend whether or not the district should continue to administer each assessment, **MUST and the STAR Reading and Math assessments stood out** with 62 percent, 59 percent, and 55 percent of teachers respectively recommending or strongly recommending that they continue (Figure 2.4). Less than half of teachers recommended that DIBELS/DAZE (45 percent), PerformancePLUS (29 percent), and Gates MacGinitie (24 percent) continue to be administered. It should be noted, however, that these recommendation rankings are dependent upon the number of teachers that feel comfortable making a recommendation. For instance, 24 percent of respondents did not feel they had enough information to make a recommendation for PerformancePLUS, compared to just 9 percent for MUST.

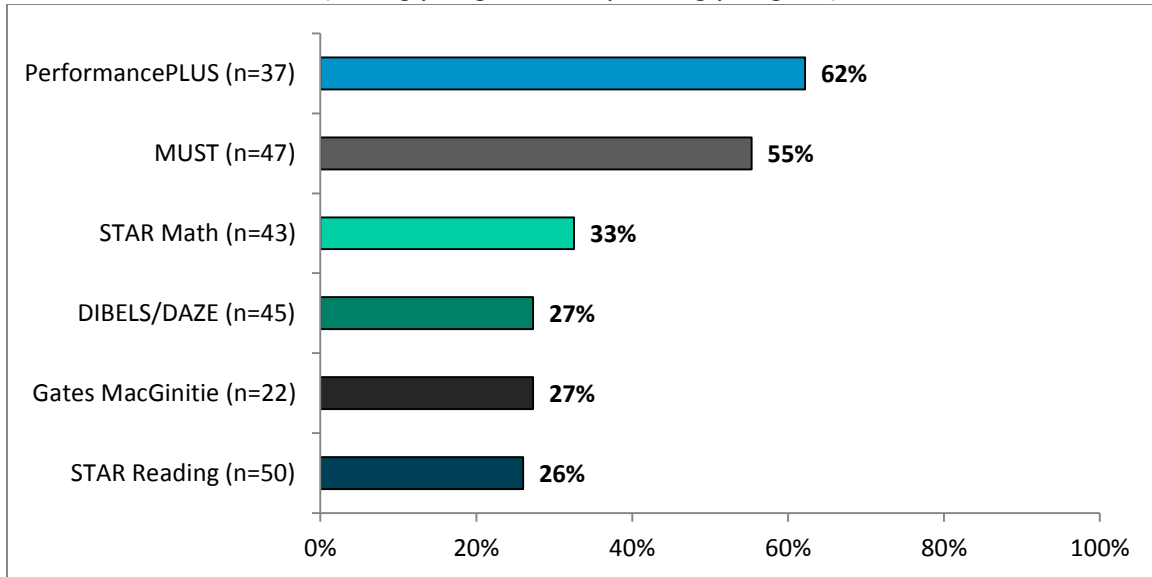
Figure 2.4: Recommend the District/School Continue to Administer the Assessment
(Recommend + Strongly Recommend)



Note: Additional options included “Recommend with reservations”, “Do not recommend”, and “Do not have enough information to make a recommendation”.

PerformancePLUS and MUST also ranked significantly higher than the other assessments for alignment to the Common Core State Standards, with 62 percent and 55 percent of teachers noting that they “strongly aligned” or “very strongly aligned” to the standards (Figure 2.5). The other four assessments ranged from 26 percent strongly or very strongly aligned (STAR Reading) to 33 percent (STAR Math).

Figure 2.5: Alignment with Common Core State Standards
(Strongly Aligned + Very Strongly Aligned)

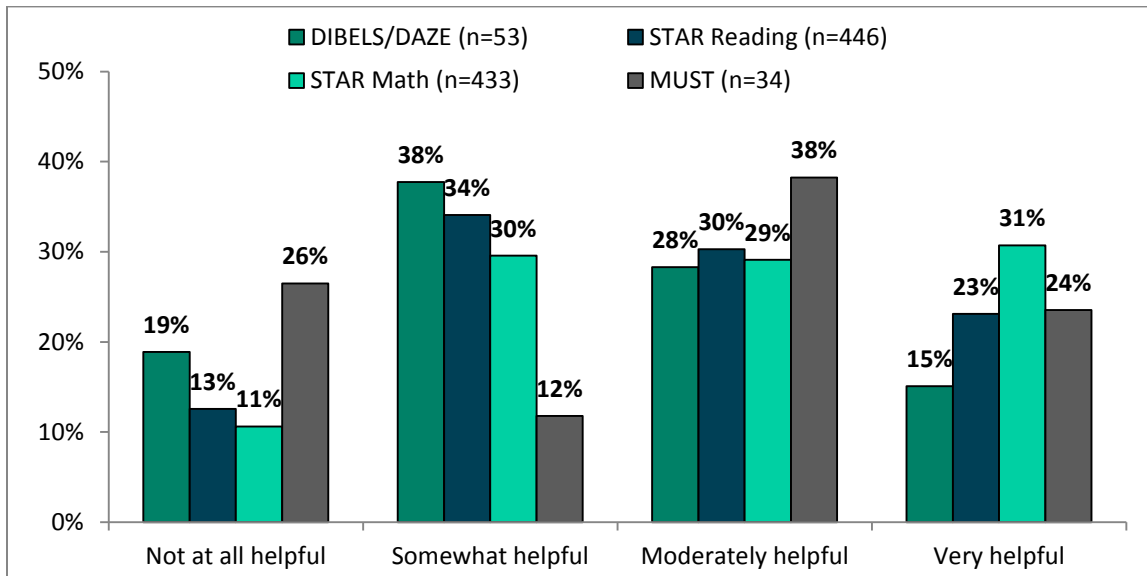


RANKINGS BY STUDENT PERCEPTIONS

This section presents students’ perceptions of the DIBELS/DAZE, MUST, and STAR Reading and MATH assessments – the tests for which the survey returned an appropriate number of responses.¹⁰ While students’ perceptions may represent a less reliable, nuanced, or experienced evaluation of an assessment’s usefulness and impact than teachers’, they may be an additional factor to consider in forming recommendations. When asked how helpful various assessments were at helping them improve in relevant subject areas, students indicated that MUST (62 percent of students responding very helpful or moderately helpful) was generally the most helpful, followed by STAR Math (60 percent), STAR Reading (53 percent), and DIBELS/DAZE (43 percent) (Figure 2.6). However, more students (26 percent) also noted that MUST was not at all helpful than any other assessment.

¹⁰ Student perceptions of Gates MacGinitie and PerformancePLUS were not measured. The survey received over 430 responses each for STAR Reading and Math but only 34 for MUST and 53 for DIBELS/DAZE. It is also important to note that students answering regarding MUST and DIBELS/DAZE are elementary school students and may have less nuanced evaluation of the assessments’ usefulness and appropriateness.

Figure 2.6: How Useful Tests Were in Helping Respondents Improve in Relevant Subject Area

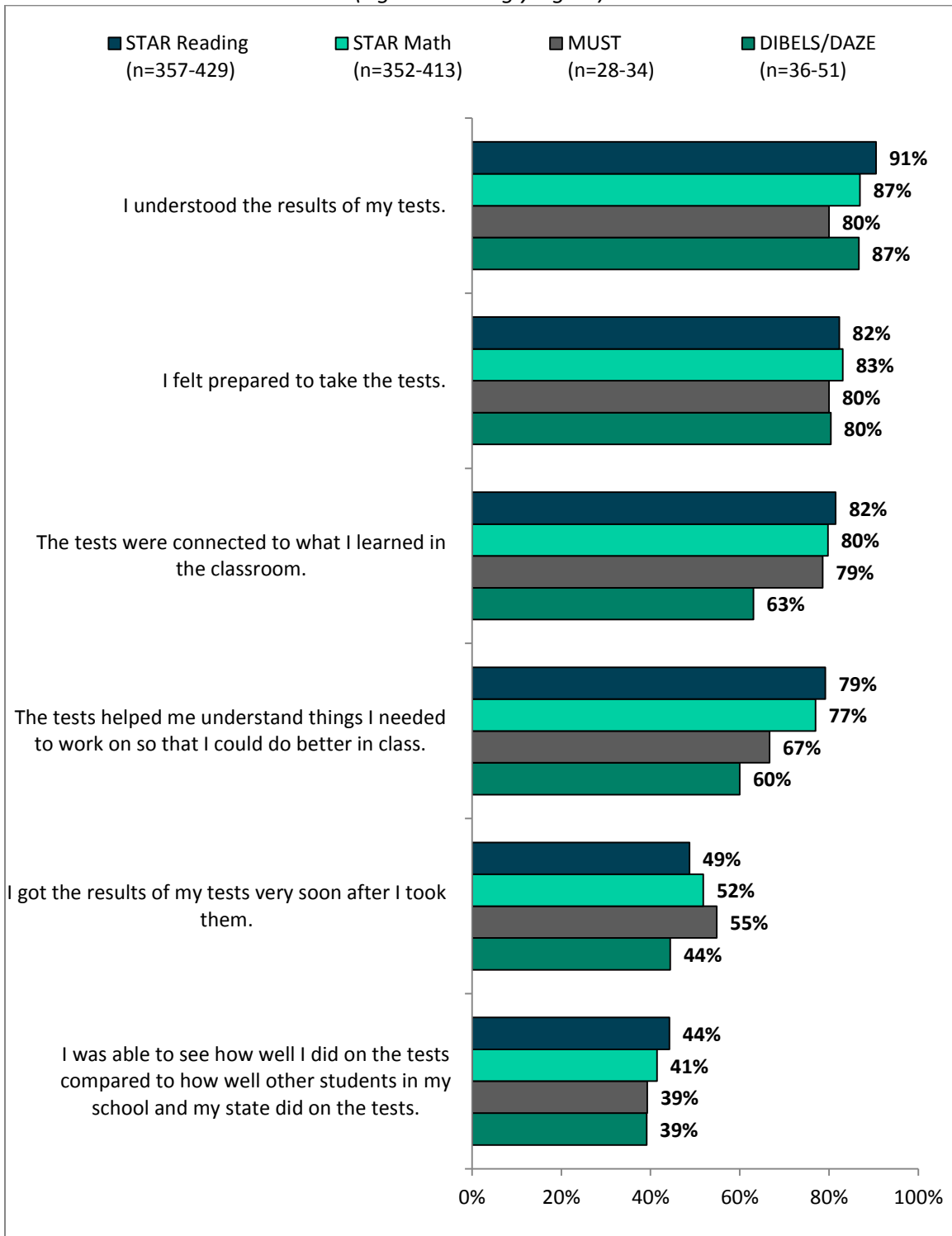


Note: Sample sizes are substantially smaller for the DIBELS/DAZE and MUST tests than for the STAR Reading and STAR Math tests.

Survey questions assessing students’ agreement with various statements about assessments revealed only a small number of differences between the four tests. For instance, the number of students agreeing or strongly agreeing that they understood the results of their tests ranged from 80 percent (MUST) to 91 percent (STAR Reading) (Figure 2.7). There was also little difference between the different tests for students’ preparedness to take them, the speed of receiving results, and ability to compare performance to peers. However, outliers included the following:

- Only 63 percent of students felt that the DIBELS/DAZE was connected to what they learned in the classroom, compared to at least 79 percent for every other test.
- The STAR Reading (79 percent) and STAR Math (77 percent) rated higher than MUST (67 percent) and DIBELS/DAZE (60 percent) for helping students understand things they needed to work on so they could do better in class.

Figure 2.7: Agreement with Various Statements Regarding Tests Student Took Last Year
(Agree + Strongly Agree)

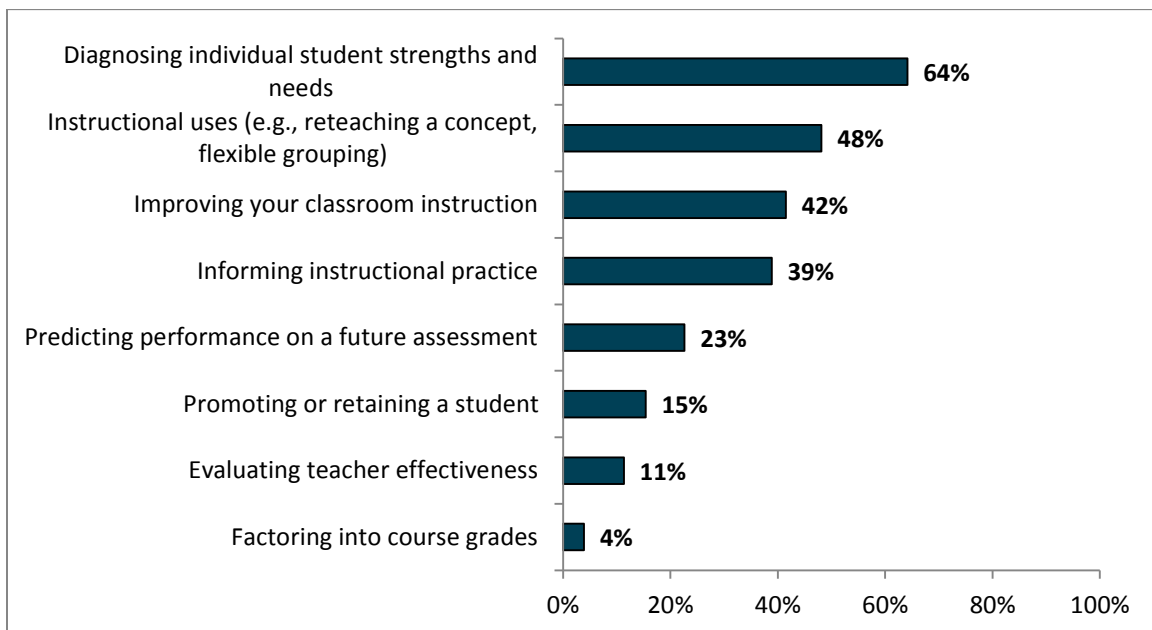


SECTION III: RESULTS BY ASSESSMENT

DIBELS/DAZE

The Dynamic Indicators of Early Literacy Skills (DIBELS/DAZE) assessment is a “cognitive measure evaluating underlying reading skills” in early grade students and is administered to Cape Henlopen students from Kindergarten through Grade 5.¹¹ Teachers generally feel that the DIBELS/DAZE assessment is a useful tool in differentiating instruction and enhancing instructional practices, yet relatively few feel that it has considerable use for evaluation purposes (a theme across found across most assessments in this analysis) (Figure 3.1). For instance, **while close to or more than half of teachers felt that the assessment is useful or very useful for diagnosing individual student strengths and needs (64 percent) and informing and improving instructional practice (between 39 and 48 percent), only a small number felt it was useful or very useful for evaluating teacher effectiveness (11 percent) or deciding whether to promote or retain a student (15 percent).** In open-ended responses to the teacher survey, numerous respondents noted that DIBELS/DAZE is most helpful for students in lower grades (e.g., K-2) rather than for older students. A small number of other responses indicated that the assessment should focus more on comprehension rather than speed and fluency. Still others noted that they wished the assessment were optional and was best for struggling students.

Figure 3.1: How useful is the DIBELS/DAZE in each of the following areas?
(percent answering “useful” or “very useful”)

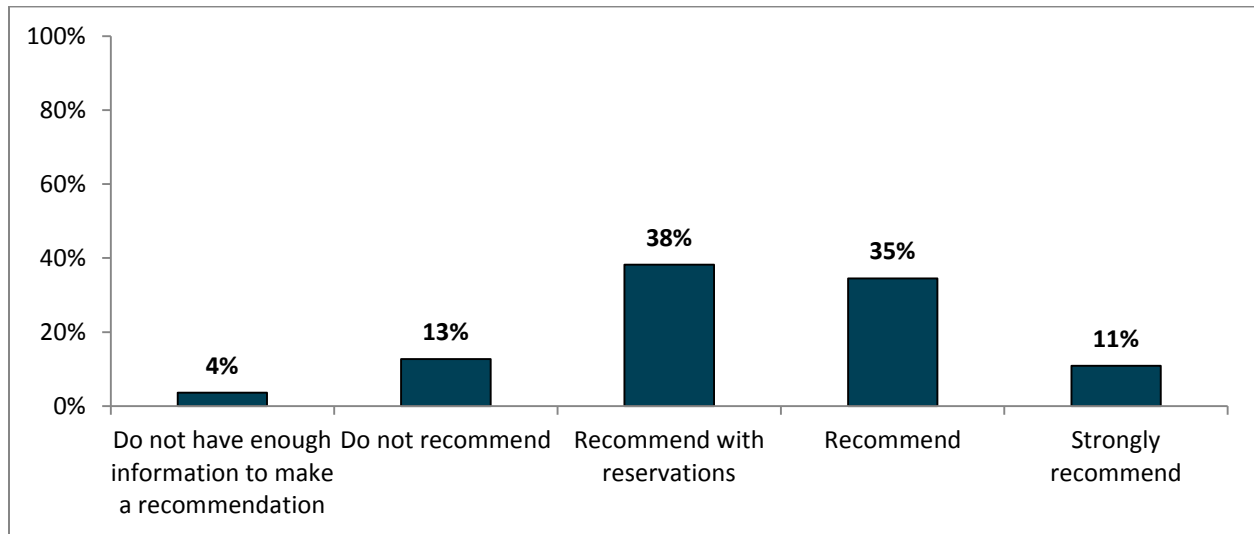


N=51-54

¹¹ “2015-15 DPAS-II Component V Student Growth Measures,” Op. cit.

Of the 55 teachers, nearly all in elementary school, who answered questions regarding DIBELS/DAZE, the large majority (96 percent) felt they had sufficient information to recommend whether or not the district should continue using the assessment (Figure 3.2). Only 13 percent of teachers recommended that the district not keep using DIBELS/DAZE, but an additional 35 percent note they would recommend but with reservations. **A combined 46 percent recommended or strongly recommended the continued use of the assessment.**

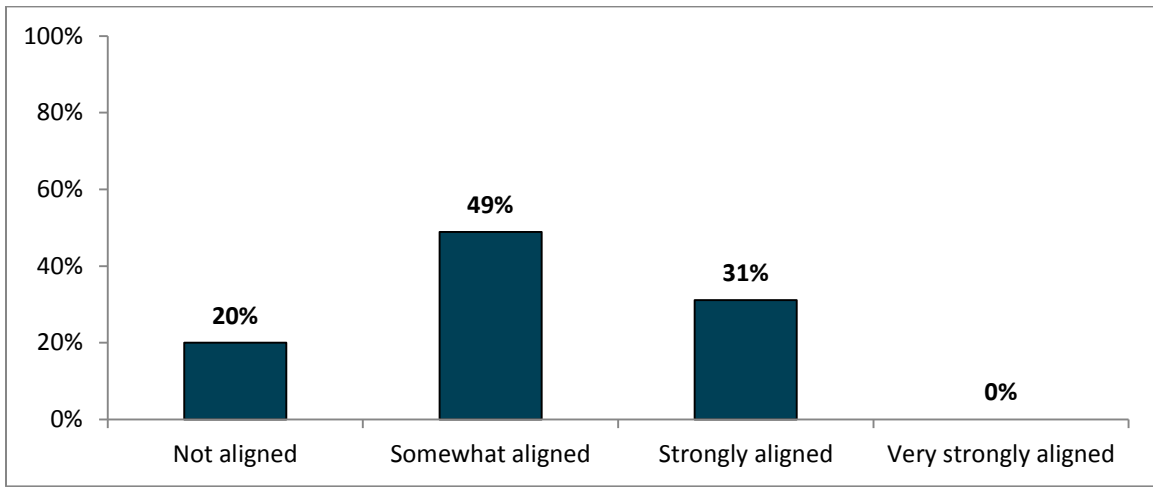
Figure 3.2: Would you recommend that the district keep using DIBELS/DAZE?



N=55

DIBELS/DAZE ranks fourth among all assessments in its perceived alignment to the Common Core State Standards (Figure 3.3). While 20 percent of teachers felt that it was not aligned at all to the Common Core, 49 percent felt it was somewhat aligned and 31 percent that it was strongly aligned. Out of the 45 respondents, none felt DIBEL/DAZE was very strongly aligned to the standards

Figure 3.3: How strong is the alignment between the DIBELS/DAZE and the Common Core State Standards?



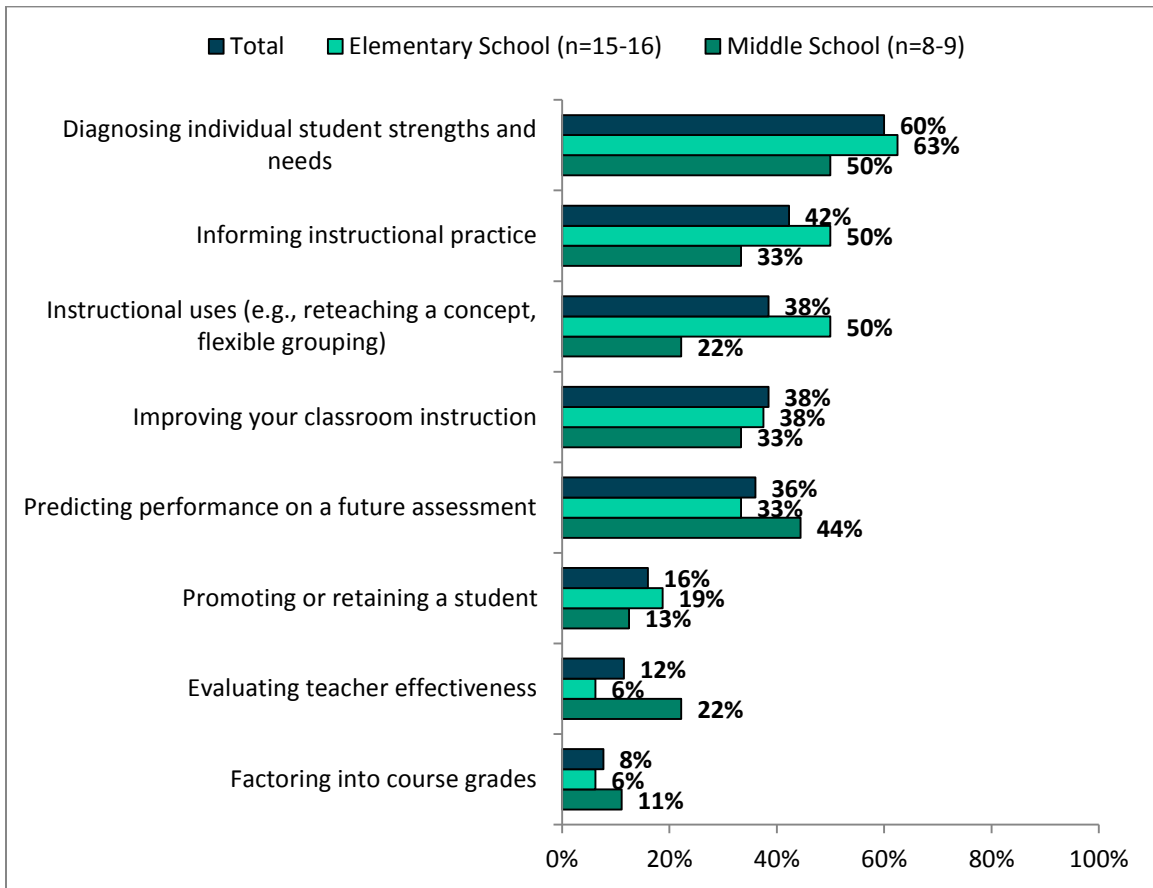
N=45

GATES MACGINITIE

The Gates MacGinitie assessment is a “cognitive measure of general reading achievement” designed for students in Grades K through 12 that is administered to Cape Henlopen students in Grade 1 and Grade 6.¹² As such, responses for this analysis were primarily drawn from elementary and middle school teachers, with one response from a high school teacher. Similar to the DIBELS/DAZE, teachers felt that the Gates MacGinitie was more useful for instructional than evaluative purposes. However, the only metric for which over half of teachers felt the assessment was useful or very useful was in diagnosing individual student strengths and needs (Figure 3.4). Less than half deemed it useful for informing instructional practice (42 percent), instructional uses (22 percent), improving classroom instruction (38 percent), and predicting performance on a future assessment (44 percent). Only 16, 12, and 11 percent of teachers respectively felt it was useful for promoting or retaining a student, evaluating teacher effectiveness, or factoring into course grades. In general, responses did not vary significantly between elementary and middle school teachers.

¹² Ibid.

Figure 3.4: How useful is the Gates MacGinitie reading test in each of the following areas?
(percent answering “useful” or “very useful”)

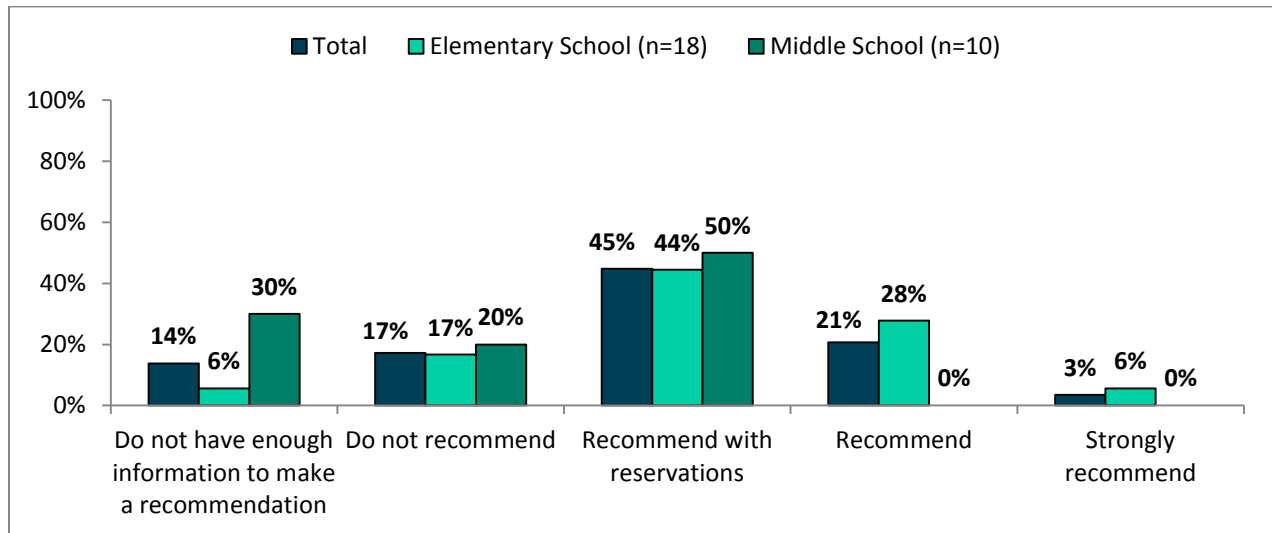


N=24-26

Note: The question received just two responses from high school teachers, so high school is excluded as a category but included in the total calculations.

Slightly less than half of teachers (45 percent) responded that they would recommend the district continue to administer the Gates MacGinitie assessment with reservations, and only a combined 24 percent would recommend or strongly recommend continuing to do so (Figure 3.5). Though the sample size is small, it is notable that no middle school teachers would recommend continuing without reservations.

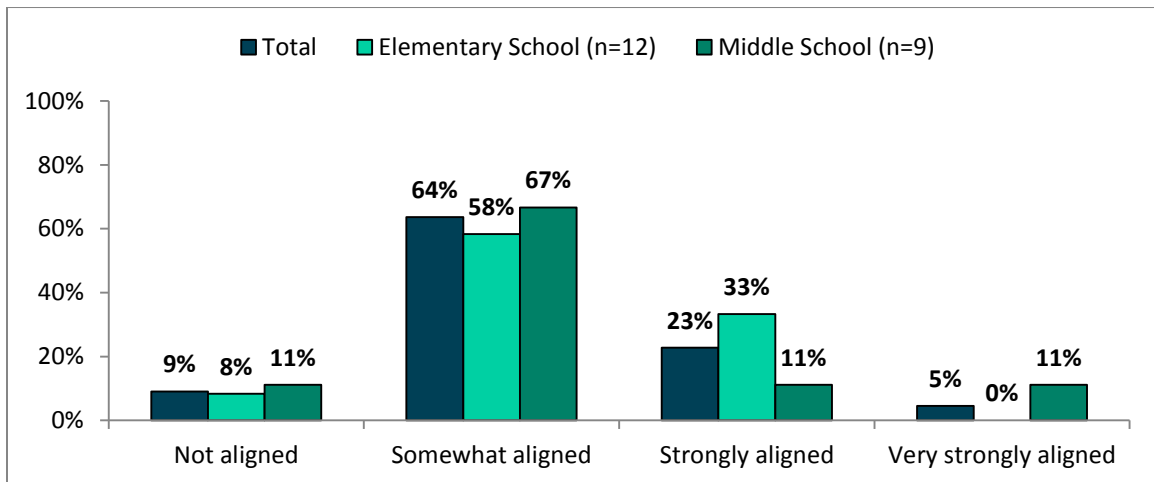
Figure 3.5: Would you recommend that the district/school continue to administer the Gates MacGinitie reading test?



N=29

Approximately two-thirds (64 percent) of respondents felt that the Gates MacGinitie was somewhat aligned to the Common Core State Standards, while a combined 28 percent felt it was strongly or very strongly aligned (Figure 3.6). Teachers were slightly less likely to feel that the Gates MacGinitie was not aligned to the standards than for the DIBELs/DAZE. In general, there were no particularly significant differences between elementary and middle school teachers.

Figure 3.6: How strong is the alignment between the Gates MacGinitie and the Common Core State Standards?

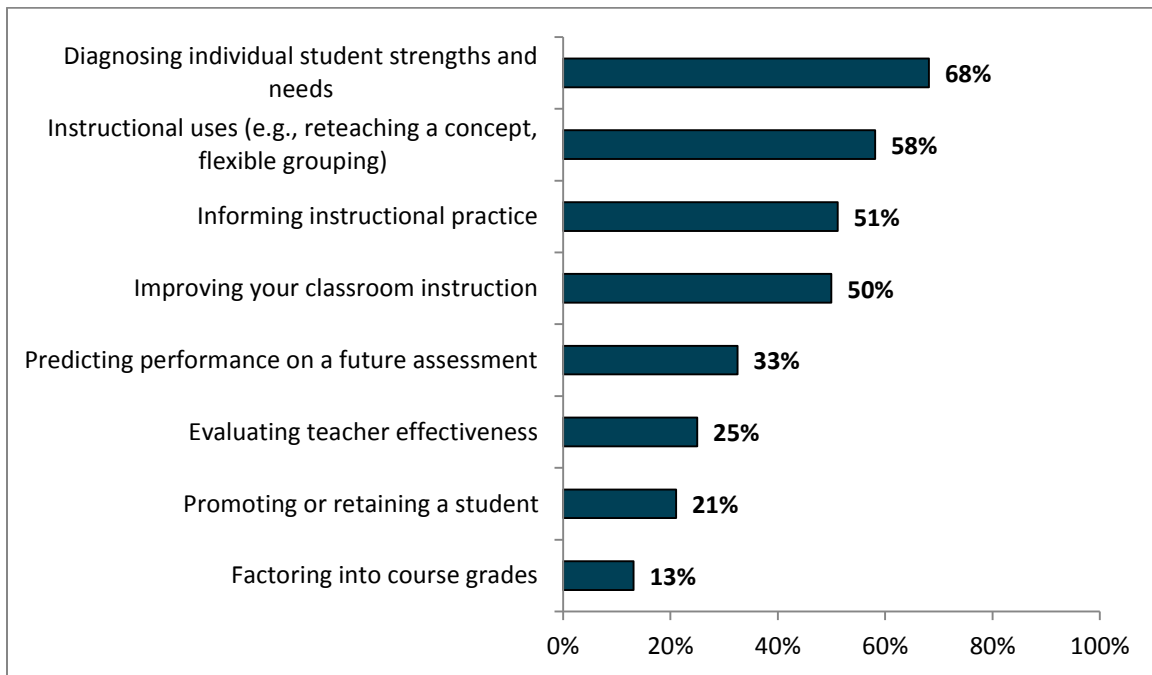


N=22

MATH UNIVERSAL SCREENING TOOL (MUST)

The Math Universal Screening Tool (MUST) is a math assessment mechanism developed by the state of Delaware and administered to approximately 1,500 Cape Henlopen students annually in Grades K through 5. More than any other assessment analyzed in this report, Cape Henlopen teachers feel that MUST is useful across a wide number of areas, especially diagnostic and instructional purposes (Figure 3.7). Over half of surveyed teachers felt that MUST was useful or very useful for diagnosing individual student strengths and needs (68 percent), instructional uses (58 percent), informing instructional practice (51 percent), and improving classroom instruction (50 percent). Similar to other assessments, teachers noted that MUST was not as effective at predicting future student performance, evaluating teacher effectiveness, promoting or retaining students, or factoring into course grades.

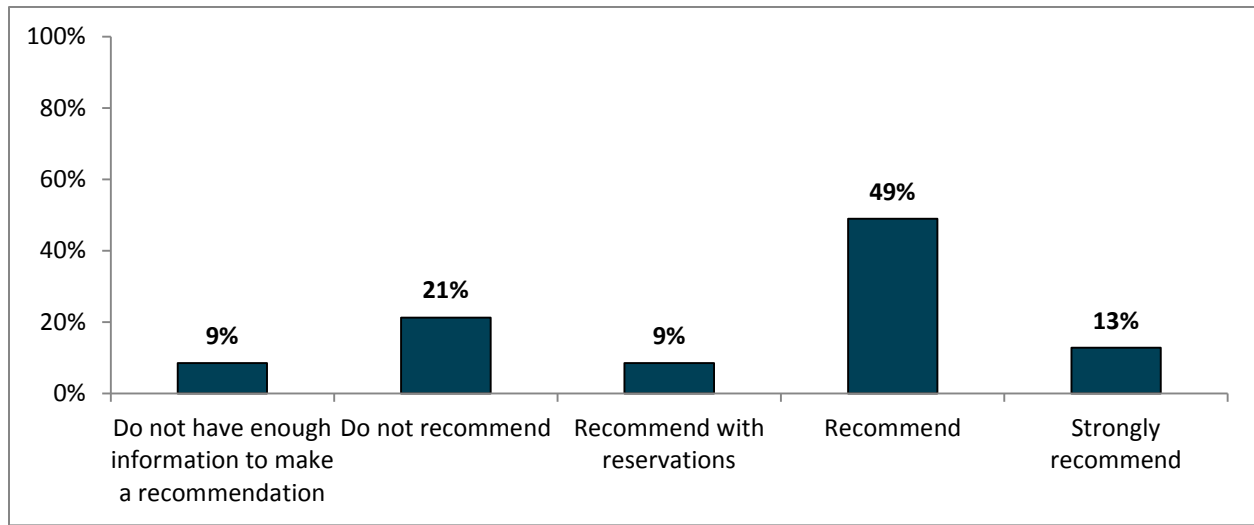
Figure 3.7: How useful is MUST in each of the following areas?
(percent answering “useful” or “very useful”)



N=38-44

Cape Henlopen teachers recommended that the district continue using MUST more than any other assessment (Figure 3.8). A combined 62 percent of teachers recommended (49 percent) or strongly recommended (13 percent) that the district keep using MUST. A further nine percent recommended with reservations, while 21 percent did not recommend and nine percent lacked sufficient information to make a recommendation.

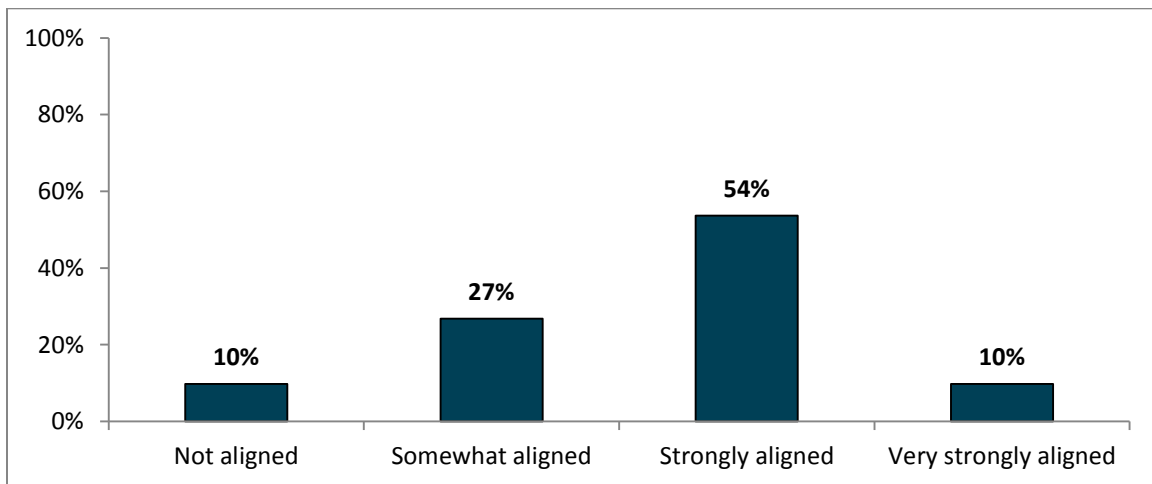
Figure 3.8: Would you recommend that the district keep using MUST?



N=47

After PerformancePLUS, teachers felt MUST was the second-most aligned assessment to the Common Core State Standards, with a combined 64 percent responding that it is strongly (54 percent) or very strongly (10 percent) aligned (Figure 3.9). Just 10 percent of respondents indicated that they felt MUST was not aligned to the standards.

Figure 3.9: How strong is the alignment between MUST and the Common Core State Standards?



N=41

STAR READING

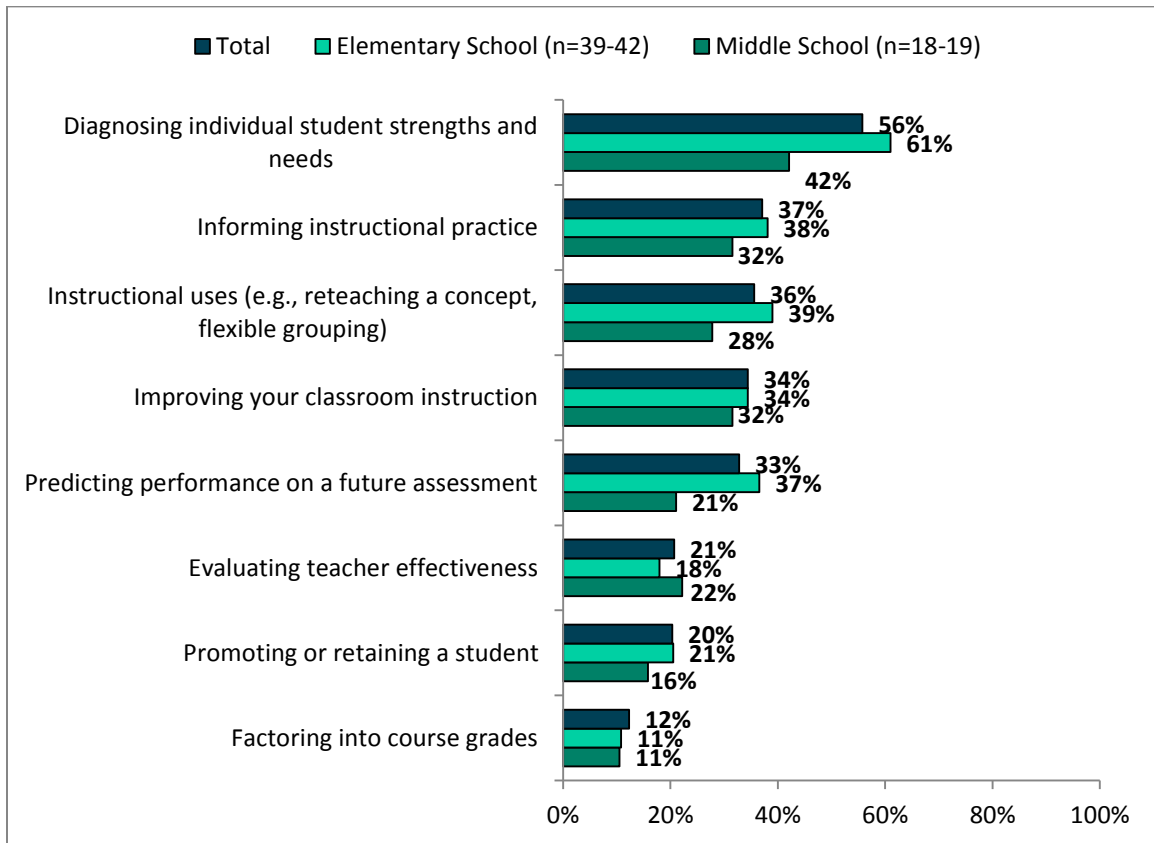
Along with the STAR Math assessment, STAR Reading is the most widely-administered assessment within the district. A “cognitive, adaptive measure of reading comprehension,” the assessment is taken by approximately 5,000 Cape Henlopen students in Grade 2 through Grade 12 annually.¹³ STAR Reading’s average usefulness rank of 3.75 places it fourth out of

¹³ Ibid.

the six assessments covered in this analysis across the eight areas of usefulness. Slightly over half (56 percent) of teachers felt that the assessment was useful or very useful for diagnosing individual student strengths and needs, but no other measure saw more than 37 percent support (Figure 2.10). Approximately a third of teachers felt that STAR Reading was useful or very useful for informing instructional practices (37 percent), instructional uses (36 percent), improving classroom instruction (34 percent), and predicting performance on a future assessment (33 percent), while fewer felt it had utility for evaluating teacher effectiveness (21 percent), promoting or retaining a student (20 percent), or factoring into course grades (12 percent).

In general, elementary school teachers often indicated higher opinions of STAR Reading’s usefulness than their middle school counterparts (Figure 3.10). In open-ended responses teachers commonly noted that they needed more training on how to use and interpret the assessment data and that the information could be broken down in more detail (e.g., by skills).

Figure 3.10: How useful is the STAR Reading test in each of the following areas?
(percent answering “useful” or “very useful”)

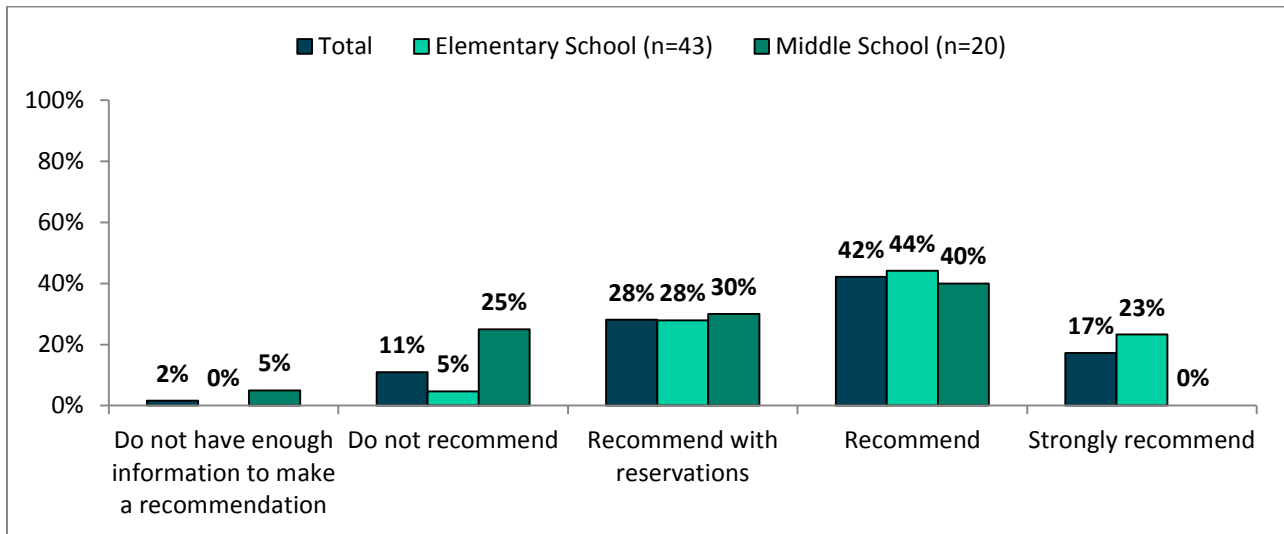


N=57-62

Note: The question received just one response from a high school teacher, so high school is excluded as a category but included in the total calculations.

While STAR Reading received only average indications of usefulness from teachers, teachers recommended that Cape Henlopen continue to use the test more than any other assessment, with the exception of MUST (Figure 3.11). In total, 59 percent of teachers recommend (42 percent) or strongly recommend (17 percent) continuing its use, while just 11 percent did not recommend and 28 percent recommended with reservations. Similar, to measures of usefulness, elementary school teachers were more likely to recommend continuing to use STAR Reading than middle school teachers.

Figure 3.11: Would you recommend that the district keep using the Star Reading test?

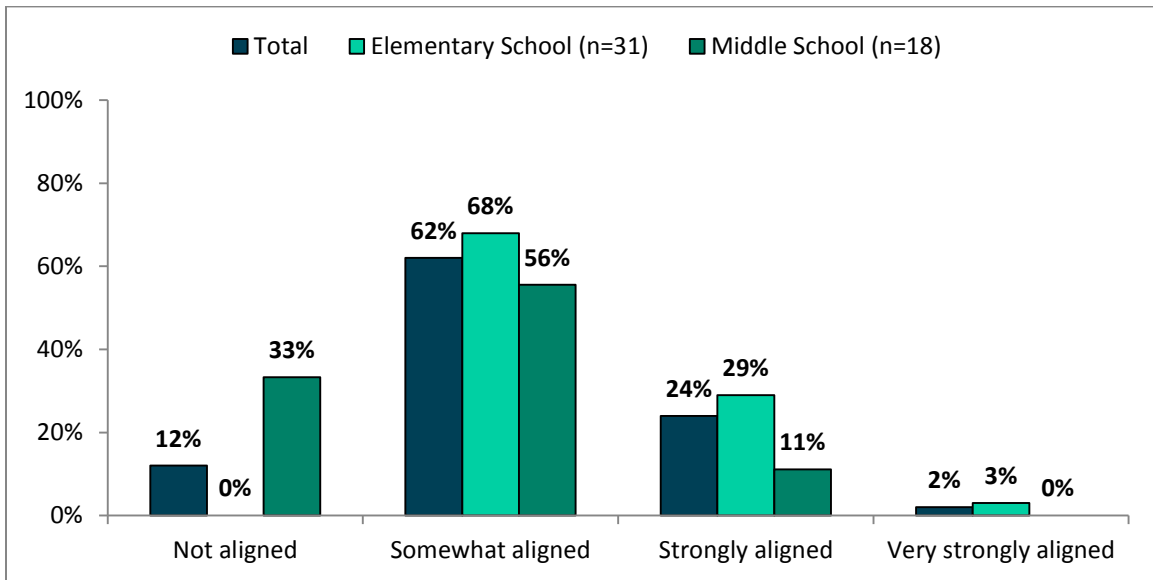


N=64

Note: The question received just one response from a high school teacher, so high school is excluded as a category but included in the total calculations.

Slightly less than two out of three teachers (62 percent) felt that STAR Reading was only somewhat aligned to Common Core State Standards, compared to 24 percent and 2 percent that thought it was strongly or very strongly aligned (Figure 3.12). Of all assessments, STAR Reading ranks last for teachers’ perceptions of its alignment to Common Core.

Figure 3.12: How strong is the alignment between the STAR Reading test and the Common Core State Standards?



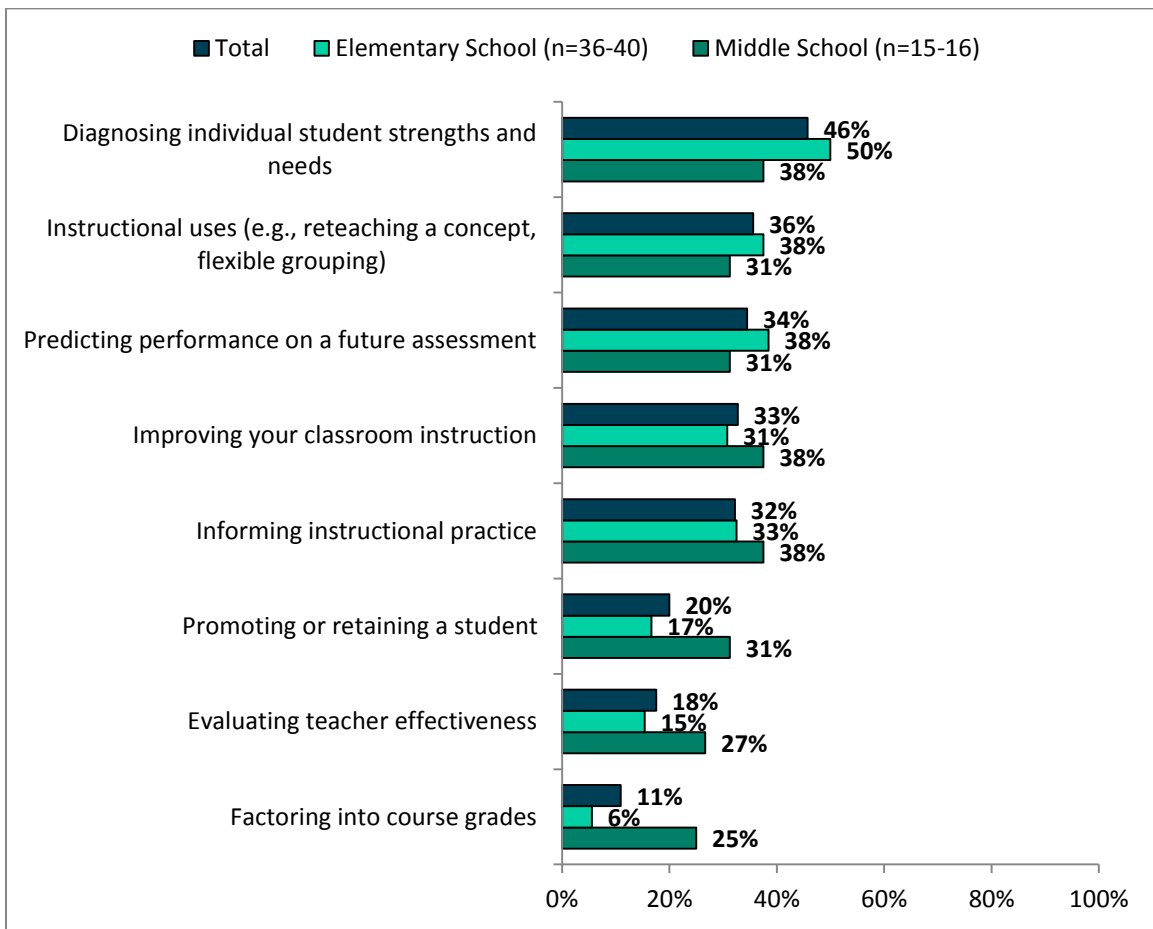
N=50

STAR MATH

Administered to approximately 5,000 students per year in the district, the STAR Math assessment “provides information about student growth and achievement” and is taken by students spanning Kindergarten through Grade 12. Less than half of Cape Henlopen teachers surveyed felt that the assessment was useful or very useful in any of the eight areas (Figure 3.13). A total of 46 percent indicated that STAR Math was useful for diagnosing individual student strengths and needs, while approximately one-third found it useful for instructional uses (36 percent), predicting performance on a future assessment (34 percent), improving classroom instruction (33 percent), and informing instructional practice (32 percent).

Of the 17 teachers that submitted open-ended responses regarding ways to make the STAR Math information more helpful, four noted that teachers need more training in using the test information and three remarked that the test was not an accurate indicator of student performance.

Figure 3.13: How useful is the STAR Math test in each of the following areas?
(percent answering “useful” or “very useful”)



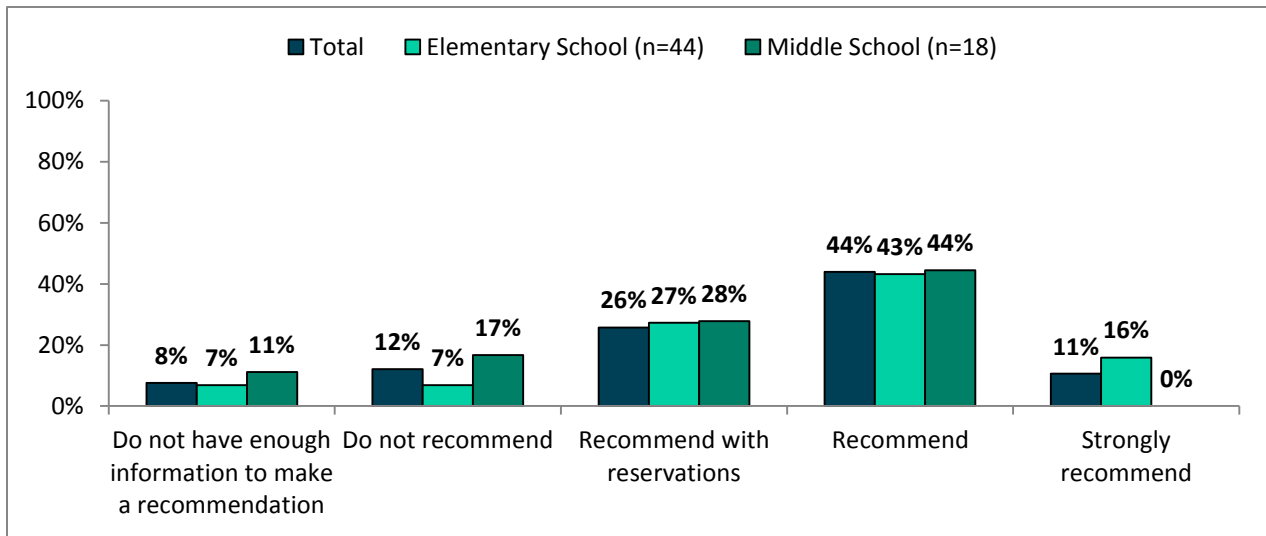
N=57-59

Note: The question received just three responses from high school teachers, so high school is excluded as a category but included in the total calculations.

The STAR Math assessment fell just behind MUST and STAR Reading in terms of teachers’ recommendations for its continued use; a total of 55 percent of teachers either recommended (44 percent) or strongly recommended (11 percent) that Cape Henlopen keep using the test, while a further 26 percent recommended with reservations (Figure 3.14). Twelve percent did not recommend continuing its use, and eight percent did not have enough information to make a recommendation.

It appears that middle school teachers were slightly less likely than elementary school teachers to strongly recommend the district keep using STAR Math and slightly more likely to not recommend its continued use.

Figure 3.14: Would you recommend that the district keep using the STAR Math test?

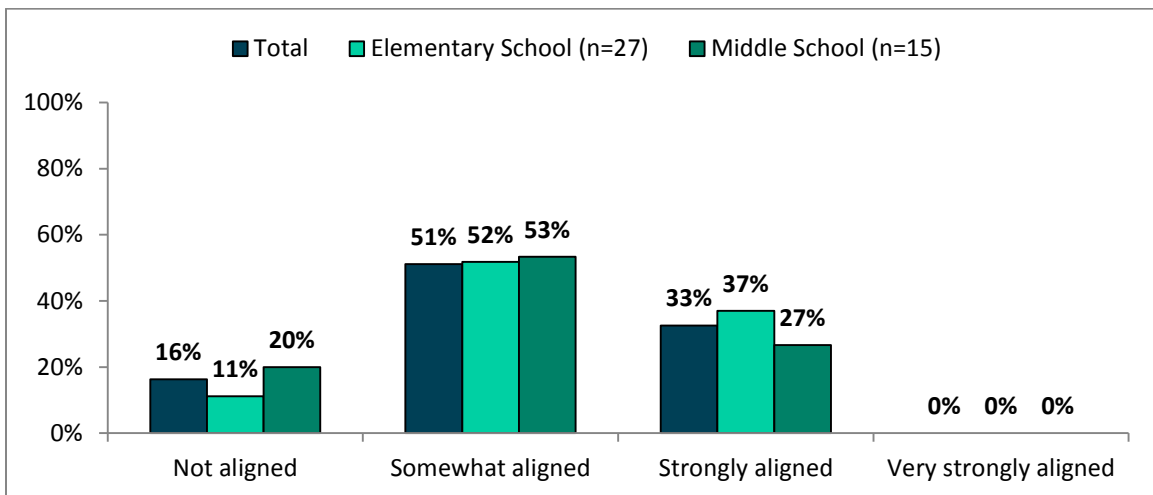


N=66

Note: The question received just four responses from high school teachers, so high school is excluded as a category but included in the total calculations.

Notably, none of the 43 teachers surveyed indicated that STAR Math was “very strongly aligned” to the Common Core State Standards, though 33 percent felt it was “strongly aligned” (Figure 3.15). Between the two categories, the STAR Math assessment ranks third among examined assessments for alignment to the Common Core, though substantially behind MUST and PerformancePLUS and only slightly ahead of DIBELS/DAZE, Gates MacGinitie, and STAR Reading. A further 51 percent felt it was slightly aligned to the standards, while 16 percent felt that it was not aligned at all.

Figure 3.15: How strong is the alignment between the STAR Math test and the Common Core State Standards?



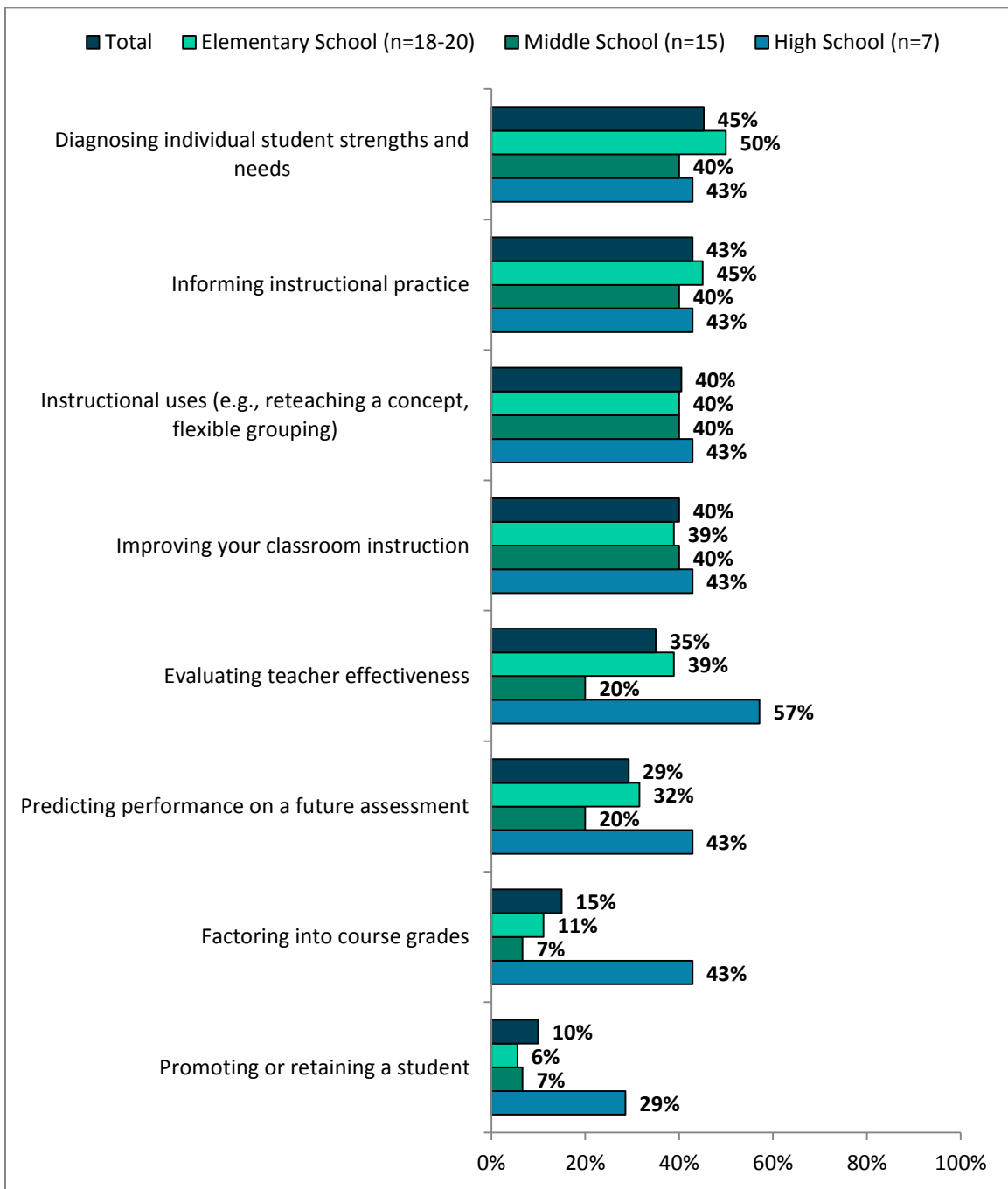
N=43

Note: The question received just one response from a high school teacher, so high school is excluded as a category but included in the total calculations.

PERFORMANCEPLUS

PerformancePLUS is the examination used by Cape Henlopen School District as the Component V assessment for evaluating teacher effectiveness and covers both ELA and mathematics topics. It is administered to approximately 3,500 students annually in Kindergarten through Grade 12. Likely in part because of its purpose to evaluate teachers, Cape Henlopen teachers rated PerformancePLUS more useful than any other assessment at evaluating teacher effectiveness, as 35 percent of teachers deemed it useful or very useful (Figure 3.16). Unlike other assessments, teachers were only slightly more likely to consider PerformancePLUS useful for diagnostic and instructional purposes than for evaluative ones. However, the usefulness of PerformancePLUS for evaluation is largely confined to evaluating teacher effectiveness, as only 15 percent and 10 percent of respondents respectively noted that the assessment was useful or very useful for factoring into course grades and promoting or retaining students (Figure 3.16). Though the number of responding high school teachers was low, it is perhaps notable that they generally felt PerformancePLUS was more effective for evaluating both students and teachers than teachers at other grade levels.

Figure 3.16: How useful is the PerformancePLUS assessment in each of the following areas?
(percent answering “useful” or “very useful”)

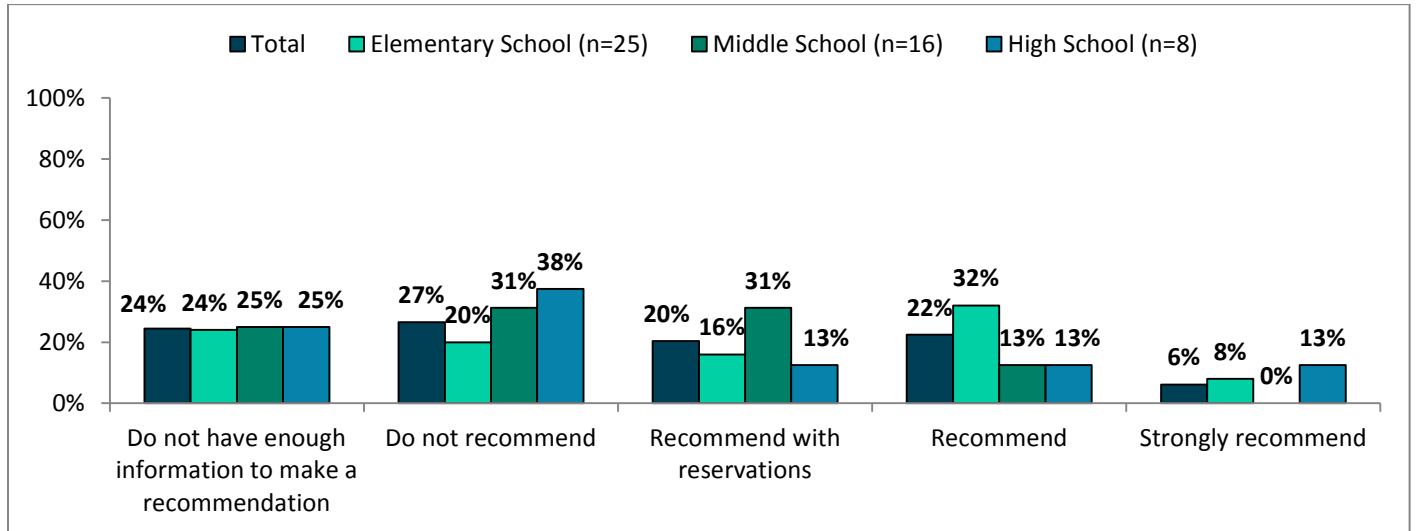


N=40-42

Teachers were relatively evenly split across response options when asked to recommend whether or not the district should continue using PerformancePLUS. Nearly a quarter of respondents noted that they did not have enough information to make a recommendation, and an additional 27 percent would not recommend that Cape Henlopen continue using the assessment (Figure 3.17). PerformancePLUS ranked behind every assessment except for

Gates MacGinitie for teacher recommendations, as just a combined 28 percent would recommend or strongly recommend keeping the tests. An additional 20 percent would recommend but with reservations.

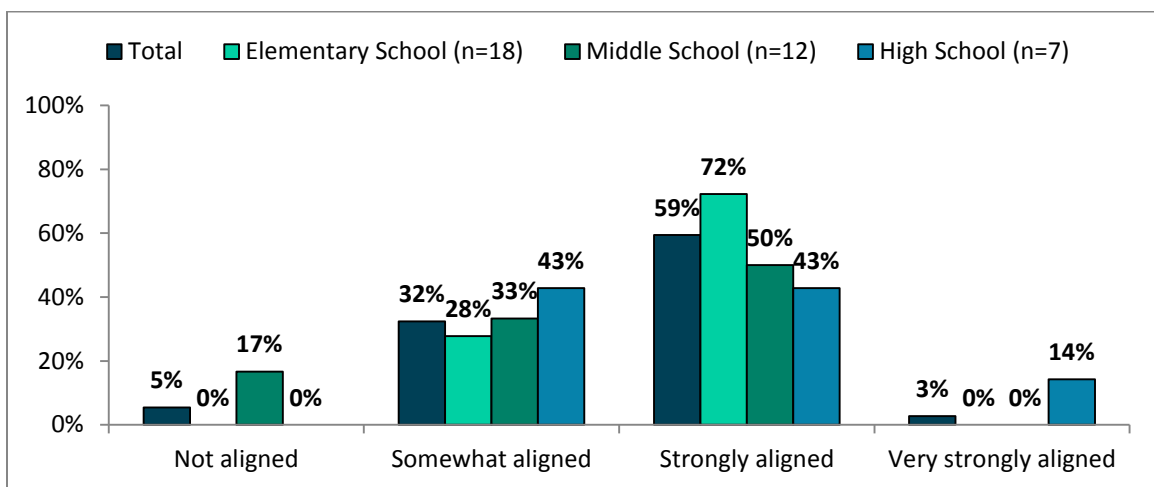
Figure 3.17: Would you recommend that the district keep using the PerformancePLUS assessment?



N=42

Teachers rated PerformancePLUS higher than any other assessment for alignment to the Common Core State Standards, with 62 percent indicating that the tests were strongly or very strongly aligned to the standards (Figure 3.18).

Figure 3.18: How strong is the alignment between the PerformancePLUS assessment and the Common Core State Standards?



N=37

PROJECT EVALUATION FORM

Hanover Research is committed to providing a work product that meets or exceeds partner expectations. In keeping with that goal, we would like to hear your opinions regarding our reports. Feedback is critically important and serves as the strongest mechanism by which we tailor our research to your organization. When you have had a chance to evaluate this report, please take a moment to fill out the following questionnaire.

<http://www.hanoverresearch.com/evaluation/index.php>

CAVEAT

The publisher and authors have used their best efforts in preparing this brief. The publisher and authors make no representations or warranties with respect to the accuracy or completeness of the contents of this brief and specifically disclaim any implied warranties of fitness for a particular purpose. There are no warranties that extend beyond the descriptions contained in this paragraph. No warranty may be created or extended by representatives of Hanover Research or its marketing materials. The accuracy and completeness of the information provided herein and the opinions stated herein are not guaranteed or warranted to produce any particular results, and the advice and strategies contained herein may not be suitable for every partner. Neither the publisher nor the authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages. Moreover, Hanover Research is not engaged in rendering legal, accounting, or other professional services. Partners requiring such services are advised to consult an appropriate professional.



4401 Wilson Boulevard, Suite 400

Arlington, VA 22203

P 202.559.0500 F 866.808.6585

www.hanoverresearch.com