



Continuous Improvement & Inclusiveness: Delaware's Student Growth Measures

January 19, 2016



Main Ideas: Delaware's Student Growth Measures

- Prior to 2011, **more than half** of the state's educators did not have formal diagnostics or summative assessments that were comparable across the state.
- Since 2011, **over 700 educators** have been gathered to develop multiple measures of student growth for every grade and subject area—this process continues to this day.
- Many LEAs are now streamlining their assessments to ensure that key assessments are utilized for **multiple purposes**.
- DDOE has a process whereby **local student growth measures** may be submitted for state approval, to build on the 300+ library/bank available to all educators.
- Over the past year, DDOE has conducted significant research with the Harvard Strategic Data Project to prove the **“validity and reliability”** of those student growth measures created by educators
- In 2015, DDOE announced a major shift in student improvement component policy: **From four required “data points” to two required “data points”**, which we believe has significantly reduced the number/time spent on student growth measures

Student Growth Measures ("Measure Bs"):

During the initial design phase with educators, the original intent of Delaware's Measure B assessments was to:

- Measure the content standards taught by Delaware educators.
- Weight/Preference the most important content standards taught within a particular subject.
- Empower/enlist educators to create assessments for use by their peers.
- Ensure all educators had more than one performance-based indicator: "Multiple Measures".
- Improve the assessment literacy of Delaware educators through the student growth measure development process.

What are Measure Bs?

External Assessments	<ul style="list-style-type: none"> • Recognized and identified by Delaware educator groups • Generally created by outside vendors • Reviewed by an outside vendor prior to approval
Internal Assessments	<ul style="list-style-type: none"> • Developed by and for groups of Delaware educators • Reviewed by an outside vendor prior to approval
Alternate Assessments	<ul style="list-style-type: none"> • Developed and submitted by a District or LEA • Reviewed by an outside vendor prior to approval

An educator, with administrator approval, may choose to use any approved measure.

Examples of Measure Bs

External Assessments	<ul style="list-style-type: none"> • 34 approved, examples include: <ul style="list-style-type: none"> • DIBELS • MAP • STAR • SRI & SMI
Internal Assessments	<ul style="list-style-type: none"> • Over 200 approved, examples include: <ul style="list-style-type: none"> • Mathematics – Grade 7 • ELA – Grade 9 • Medical Assisting I
Alternate Assessments	<ul style="list-style-type: none"> • 31 approved, examples include: <ul style="list-style-type: none"> • Allied Health III (Grades 10-12) – Appo • Visual Arts (Grades 6-8) – Brandywine • World History – Sussex Tech

An educator, with administrator approval, may choose to use any approved measure.

6

Student “Growth Goals” (“Measure Cs”)

- Developed by and for groups of Delaware educators by role and/or subject area(s)
- Reviewed by an outside vendor and approved by the Department of Education
- Based upon professional standards and position responsibilities and focused on student growth and performance
- Performance targets developed at the individual-level (generally)

Examples of Student Growth Goals (“Measure Cs”)

- Pre-K ELA
- Counselor – Middle School
- School Nurse
- School Psychologist
- Social Studies

Ongoing Refinement: “Wash Cycle”

- Four-year cycle for refinement of Measure Bs AND Measure Cs
- Led by DDOE content staff, with support from educators/content experts
- Student Growth Measures vetted for rigor, format by outside vendor
- Specialists are addressed based on demand, about 2-3 groups per year. This year the groups included are Nurses and School Counselors.

Subject	Refinement Year	Roll-Out Year
SS & Arts	14-15	15-16
CTE	15-16	16-17
Science, Math & ELA	16-17	17-18
World Languages	17-18	18-19

AN EXAMPLE OF ONGOING REFINEMENT: THE ARTS



Delaware
Department of Education

Example: Arts Educators Collaborate on Student Growth Measures

In early 2015, 30 Delaware arts educators used Model Cornerstone Assessments from the new national voluntary arts standards to refine DPAS II assessment items for Dance, Music Theatre and Visual Art.



They met over the course of two weekends to create measures of **student growth** at **high cognitive levels** that reflect current **best practices in instruction**.

An Example

**DPAS II Music Item
THEN**

- **Indicate** the type of rhythmic note used on the first beat of the third measure of the example above.

Novice I pretest **DOK 1**



DPAS II Music Item NOW

- Using the literature in your music folder, select 3 pieces and order them first, middle and last for an upcoming concert. **Create, summarize and defend** your choices. Consider personal preferences, your audience, musical criteria, **technical, historical, cultural** and expressive elements in your response.

Novice I pre-test **DOK 4**

**AN EXAMPLE OF
ONGOING REFINEMENT:
CTE**



MULTIPLE USES: “THE AMPLIFY NETWORK”



Multiple Uses: “External” SGMs

1. **Assessments-** All grade 3-12 math and ELA teachers have access to administer pre-/post assessments, interim assessments and quick checks to consistently monitor student learning throughout the year.
2. **Professional Learning Maps-** A diagnostic tool that analyzes current data use and instructional practices within the participating schools.
3. **Robust Data Reports and Access to Amplify’s “Now What” Tool-** A variety of student performance data reports at the standards level and instructional resources to address gaps in student learning.
4. **Leadership Coaching-** LEA/school leadership teams receive support to fully leverage the Amplify Network in creating a school-wide culture of data driven instruction.

Overview

Pre-/Post Assessments

- Available for Math and ELA grades 3-12
- Administered online via Amplify's Beacon Platform
- Similar look and feel to the Smarter Balanced Assessments
- Approved for use with DPAS II Component 5 evaluation ratings (Optional)
- Optional for participating schools

Interim Assessments

- Available for Math (1 IA) and ELA (2 IAs) grades 3-12
- Administered online via Amplify's Beacon Platform
- Similar look and feel to the Smarter Balanced Assessments
- Common across the network.
- Developed in collaboration with LEAs
- Required for participating schools

Quick Checks

- Available for grades K-12
- Five- to seven-item assessments, for ELA and math
- Designed after the key instructional concepts embedded in the standards
- Support short-cycle instruction
- Typically take no more than twenty minutes to administer
- Optional for participating schools

Performance Tasks

- Available for grades 2-12
- Detailed instructional scaffolding is included to support both students and teachers
- Also includes observational tasks which are developmentally appropriate tasks that assess students' understanding in the context of engaging activities for grades K-1
- Optional for participating schools

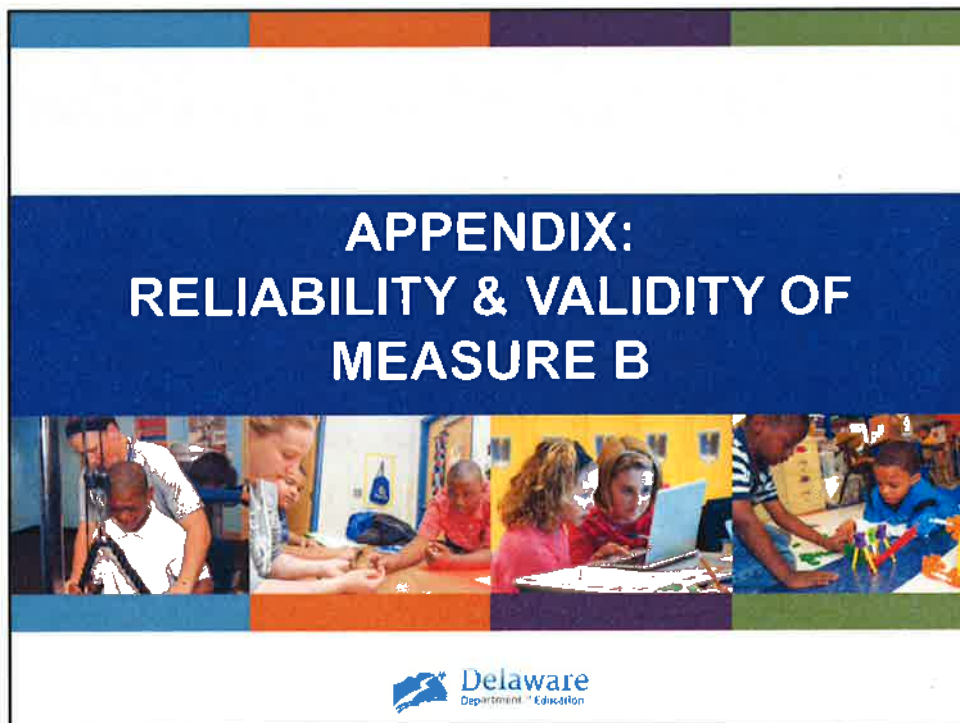
This suite of assessments can be used for: progress monitoring, RtI purposes, preparation for Smarter Balanced Assessments, and educator evaluation

15

Main Ideas: Recap Delaware's Student Growth Measures

- Prior to 2011, **more than half** of educators had very little in this area
- Since 2011, **over 700 educators** have developed Student Growth Measures (SGMs)
- Key assessments are utilized for **multiple purposes.**
- **Local student growth measures** may be submitted for approval
- The Harvard Strategic Data Project continues to prove the **"validity and reliability"** of the SGMs
- Policy shifted from **four required "data points" to two required "data points"**

16



Measure B: Technical Properties

Measure B student assessments were developed to:

- Exhibit high-content validity (alignment) by alignment with current Delaware standards.
- Provide guidance for administrative procedures for LEAs to use in administration.
- Exhibit reliability appropriate to their format and provide scoring consistency for open-items by providing detailed scoring guidelines.
- Provide transparency to educators for goal-setting conferences (that could be aligned with their context)

Measure B Analysis

- DOE maintains a database of 240+ pre/post assessments used as one of multiple measures for teacher evaluation in the student improvement component.
- While it is not required to upload assessment data, many districts/schools upload item-level data into a state-maintained system
- DOE has collected data for 2013-14 and 2014-15 and conducted analysis of assessment quality

External Measure B Review Process

The enclosed performance measure rubric is designed to examine the quality characteristics of vendor-made (external) assessments. The rubric is comprised of several technical requirements organized into eight (8) strands. Each requirement is rated on a Likert-type scale ranging from zero (not addressed) to one (fully addressed).

Rater's Task

Step 1. Review information, data, and documents associated with the creation, implementation, and refinement of the selected performance measure.

Step 2. Assign a rating to each component (TASK) within a particular strand using the following scale-

- a) (1) = fully addressed
- b) (.5) = partially addressed
- c) (0) = not addressed
- d) (N/A) = not applicable at this time

Step 3. List information, data, and document references supporting each assigned rating.

Step 4. Add notes and/or comments articulating nuances of the performance measure.

Step 5. Compile results for each strand into an overall Summary Matrix.

External Measure B Review Process (cont'd)

Strand	Points Possible	Points Earned	Strand	Points Possible	Points Earned
Design	5	3.5	Technical	9	6.5
Specifications	5	2.5	Reporting	5	5.0
Development	5	4.0	Quality	5	4.5
Administration	5	4.0	Costs	5	5.0
Total	44	35			

Most Used Post-Assessments*

Assessment Name	Number of Students
Social Studies - Post-Test 2013-14 - Grade 8	4921
Science - Biology - Post-Test 2013-14 - Grade 10	4808
Science - Post-Test 2013-14 - Grade 8	4687
Social Studies - Post-Test 2013-14 - Grade 7	4502
WLG - Spanish Level I - Post-Test 2013-14	4459
Science - Post-Test 2013-14 - Grade 7	4320
Science - Earth & Physical Sciences - Post-Test 2013-14 - Grade 9	4275
Science - Post-Test 2013-14 - Grade 6	4203
Social Studies - Post-Test 2013-14 - Grade 6	3661
Mathematics - Post-Test 2013-14 - Grade 6	3595

* Only includes data from LEAs/schools that entered data into Performance Plus system for SY2013-14 .

23

Least Used Post-Assessments*

Assessment Name	Number of Students
WLG - German Level III - Post-Test 2013-14	2
WLG - Chinese MS Exploratory - Post-Test 2013-14	10
WLG - Arabic Level III - Post-Test 2013-14	12
ESL - Post-Test 2013-14 - Grades 3-5	13
WLG - Arabic Level II - Post-Test 2013-14	13
BFM - BCM III - Post-Test 2013-14	14
BFM - Business Information Technology III - Post-Test 2013-14	14
Music - Advanced 2B - Post-Test 2013-14	14
STS - Graphic and Printing Equipment Operator I - Post-Test 2013	15
STS - Sheet Metal I - Post-Test 2013-14	15

* Only includes data from LEAs/schools that entered data into Performance Plus system for SY2013-14 .

24

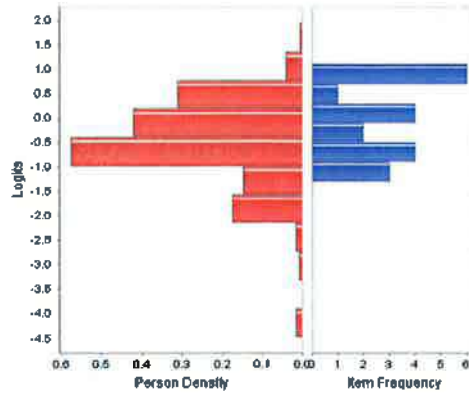
Psychometric Quality Summary

- Reliability is generally good – especially after cleaning up the scoring.
- Include linking items.
- Can clean up scores post hoc.
- Better to have more items (than simply creating items with more score points).
 - Can rescale test post-hoc
- Provide PD on scoring.
 - Ensure rubrics are very clear for all score points.
 - Eliminate superfluous score points.
 - Spanish writing has 20 – which is too difficult to distinguish unless the rubric assigns points for specific sub criteria (e.g. vocabulary words (count), grammar, organization, etc.)

Sample Item-Level Analyses

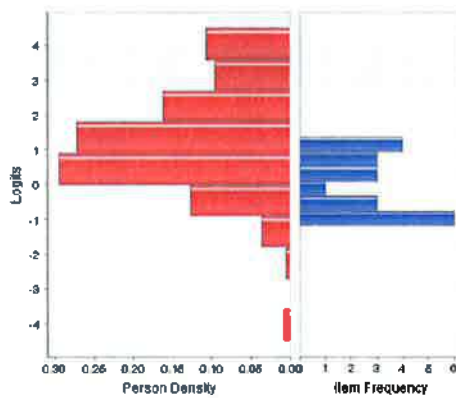
Item #	Mean Item Score	Std. Dev.	# of Observations	Item-Test Correlation	Item-Rest Correlation	Alpha if Item removed	Max Poss Score	DOK Level	Question Type
Social Studies - Post-Test 2013-14 - Grade 4									
$\alpha = 0.72$									
1	0.65	0.48	229	0.27	0.17	0.72	1	DOK 2	Multiple Choice
2	0.76	0.43	229	0.29	0.20	0.72	1	DOK 2	Multiple Choice
3	0.51	0.50	229	0.39	0.28	0.71	1	DOK 1	Multiple Choice
4	0.37	0.48	229	0.28	0.18	0.72	1	DOK 2	Multiple Choice
5	0.53	0.50	229	0.30	0.19	0.72	1	DOK 2	Multiple Choice
6	0.83	0.38	229	0.25	0.17	0.72	1	DOK 1	Multiple Choice
7	0.61	0.49	229	0.29	0.18	0.72	1	DOK 2	Multiple Choice
8	0.66	0.48	229	0.23	0.12	0.72	1	DOK 2	Multiple Choice
9	0.66	0.47	229	0.43	0.34	0.71	1	DOK 3	Multiple Choice
10	0.59	0.49	229	0.21	0.10	0.73	1	DOK 2	Multiple Choice
11	0.83	0.38	229	0.20	0.11	0.72	1	DOK 1	Multiple Choice
12	0.18	0.38	229	0.02	-0.07	0.73	1	DOK 2	Multiple Choice
13	1.05	0.93	246	0.64	0.50	0.68	2	DOK 3	Open Ended
14	1.39	0.76	246	0.56	0.42	0.69	2	DOK 3	Open Ended
15	0.95	0.84	245	0.66	0.53	0.68	2	DOK 3	Open Ended
16	1.09	0.78	246	0.68	0.56	0.67	2	DOK 3	Open Ended
17	1.03	0.88	246	0.68	0.55	0.67	2	DOK 3	Open Ended
18	1.67	0.60	246	0.50	0.37	0.70	2	DOK 3	Open Ended

Agri Science 7



Pre-test

Agri Science 7



Post - test

Reliability of Post-Assessments*

Assessment Name	Reliability
Social Studies - Grade 8	0.76
Science - Biology - Grade 10	0.82
Social Studies - Grade 7	0.79
Science - Grade 8	0.83
Science - Grade 7	0.82
WLG - Spanish Level I	0.39
Science - Grade 6	0.86
Science - Earth & Physical Sciences - Grade 9	0.83
Social Studies - Grade 6	0.81
Mathematics - Grade 6	0.72

* Only includes data from LEAs/schools that entered data into Performance Plus system for SY2013-14.

29

Reliability

Measure B	Original		Cleaned	
	Pre	Post	Pre	Post
AgriScience 7	0.52	0.87	0.55	0.75
AgriScience 8	0.60	0.87	0.62	0.95
ELA 7	0.82	0.73	0.96	0.98
ELA 8	0.85	0.70	0.95	0.98
Math 4	0.80	0.85	0.81	0.94
Math 5	0.75	0.79	0.76	0.9
Medical Asst 1	0.49	0.88	0.56	0.92
Medical Asst 2	0.52	0.26	0.67	0.65
Spanish 1	0.58	0.40	0.73	0.94
Spanish 2	0.51	0.40	0.75	0.93

Validity

- The Delaware Measure B process is substantively grounded in the expertise of educators and moderated by additional independent expertise – which is a solid basis for establishing validity.

31

APPENDIX: GOAL-SETTING RESOURCES



Measure B Assessment Reports

Calculus

Interim Measure B Assessment Report



This document provides historical student performance results from Delaware's Calculus assessment. This document is intended as a resource for educators and evaluators in preparation for and during their fall conference. The data presented here are based on data entered by districts, schools, and educators into Performance2Go over two academic years.

The maximum possible score on this assessment is 20 points.

School Year	Average Gain	Std. Dev. of Gain	Num. of Students	Pre-Test		Average Scores	
				Pre-Test Raw Score	% Score	Pre-Test Raw Score	% Score
2013-14	18.1	8.0	204	Pre	4.0	20.0%	
				Post	24.2	61.0%	
2014-15	20.2	7.5	267	Pre	5.2	13.0%	
				Post	25.4	63.5%	
Both Years	19.0	7.8	641	Pre	5.1	12.5%	
				Post	24.7	61.5%	

Average Gain
19.0 points

Table 1 shows the average gain for two academic years. It also includes the average raw score and the average score as a percent of the total possible score. You could use the average gain to set a growth target for your entire class.

All students received scores from academic years three divided into four groups (quartiles) based on their previous scores. In Figure 3, the top portion of the bars in the graph represents the average amount of gain students in each quartile made from pre-test to post-test. The labels at the bottom of the graph show the range of scores for each quartile. You could use this information to set more specific growth targets for each student based on pre-test scores.

Questions to Consider when Setting Goals & Targets

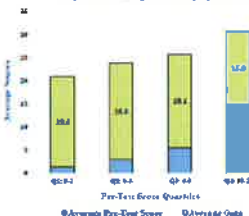
How does the performance of my students compare to statewide results? How does this inform my student improvement targets?

What approach to setting goals and targets might be best for my students? Class average? Individualized? Flexible?

What degree of improvement will my students exhibit during the instructional period/learning year? What is the appropriate amount of ambitiousness and attainability? How might this differ for "border" versus "borderless" target-achievers?

Use the local testing guidance document for more information: <http://delpet.com/More-go-guidance>
Questions? Email delstate.ed@del.gov or call 800-333-3333. Share your thoughts about this document: <http://delpet.com/delstate-feedback>

Figure 3. Average Gains by Quartile



Goal-Setting Guidance Document



TARGET-SETTING WITH DATA GUIDANCE DOCUMENT

This goal-setting guidance document contains examples of how to use historical statewide student performance data made available in the assessment reports to set high-quality goals for student growth. Educators are encouraged to use all available data in the goal-setting process—this includes data from prior years, as well as other available student performance data.

Additional goal-setting resources, including the assessment reports, can be found by visiting www.tinyurl.com/ddoe-gs-guide.

COMMONLY-USED APPROACHES TO SETTING TARGETS

This document highlights two commonly-used approaches to setting targets: 1) A class average approach—In this approach, an educator compares the class average in the fall with the class average in the spring to demonstrate student growth throughout the year and 2) A student-level approach—In this approach a target is set for each student and the percentage of students who meet their targets is calculated in the spring.

A high-quality student growth target is rigorous, ambitious, yet attainable.

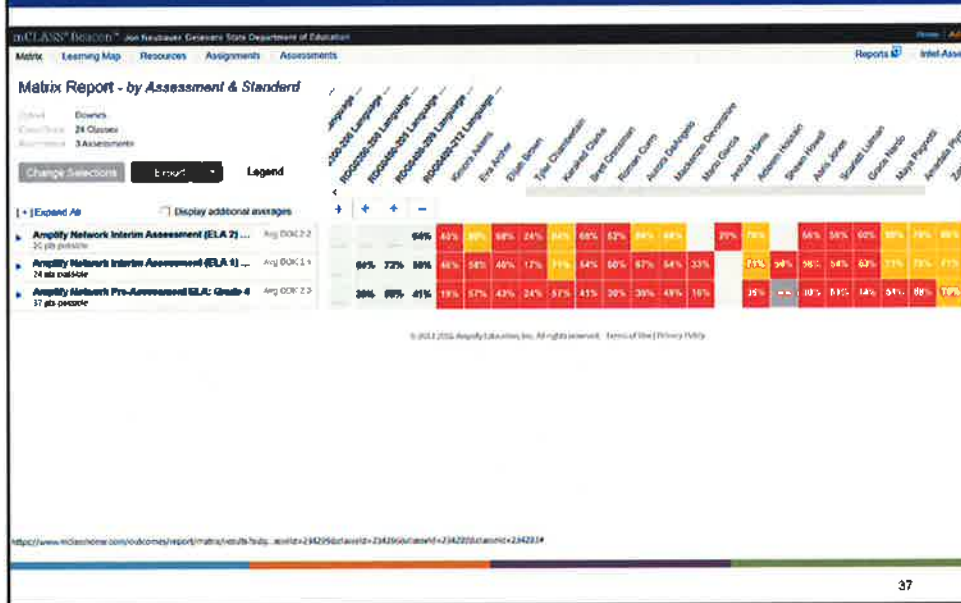
Goal-Setting Excel Templates Class Average Approach

Teacher Name:		PRIAL RATING		ADN/201
Class:				Avg. Points
Evaluation:				
Assessment used:				
Maximum Points Possible on Assessment: (Type into cell B5) →		Satisfactory Target Class Average:		
		Exceeds Target Class Average:		
	Pre-Test Average Score	Pre-Test Average as %	Post-Test Average Score	Post-Test Average as %
Roster Averages:				
ROSTER TEST SCORES				
Student Name	Pre-Test Score	Pre-Test Score as %	Post-Test Score	Post-Test Score as %

APPENDIX: AMPLIFY NETWORK



Matrix Report (Assessment View)



Matrix Report (Standards View)



Matrix Report (Item View)

Matrix Report - by Assessment & Standard

Courses: 24 Courses
Assessments: 3 Assessments

Legend

Expand All | Display statistics averages

Standard	Score	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14	Item 15	Item 16	Item 17	Item 18	Item 19	Item 20
RL.4.1	87%	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
RL.4.2	55%	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
RL.4.3	97%	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
RL.4.4	80%	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1

High Level Class View

RL.4.3

Amplify Network Interim Assessment (ELA 2) Grade 4 — Dec 18, 2015

2 pt (80%)

Item No.	Type	0 pts	1 pt	2 pt	Not Scored
1	OR	0 (0%)	3 (100%)	0 (0%)	0 (0%)

2 (100%) Not Mastered

- Tyler Chamberlain 1
- Royan Cuero 1
- Jessica Harris 1

14 (80%) Mastered

- Kimora Adams 2
- Ellynn Blevins 2
- Shelli Crossman 2
- Andi Jones 2
- Scarlett Letman 2
- Grace Harbo 2
- Eva Acker 2
- Kristina Clute 2
- Aurora Delgado 2
- Michaela Drumshire 2
- Allyson Jordan 2
- Allysa Popovich 2
- Aracelia Pizarro 2
- Tachina Wilson 2

Student Level View

RL.4.3 — K. Allens
Items Resources

RL.4.3

Describe in depth a character, setting or event in a story or drama, drawing on specific details in the text (e.g., a character's thoughts, words, or actions).

Back to Class

3 (97%) Not Mastered

- Tyler Chamberlain 1
- Roman Carlo 1
- Jessica Harris 1

14 (87%) Mastered

- Kimberly Allens 2
- Ellynn Brown 2
- Britt Crispman 2
- Aaria James 2
- Quinten Lutzman 2
- Grace Klacke 2
- Emi Archer 2
- Kristen Claffie 2
- Alexis DeLuca 2
- Madison Deveschere 2
- Adriana Hovde 2
- Mays Pagani 2

Amgily Network Initial Assessment (ELA 2 Grade 4 — Dec 18, 2014)

2-pt Items

Item No.	Type	Points earned
5	CR	2 of 2

41

Learning Map

Learning Map
Reports Initial Admins My Classro Help

Target

- CR 4 Reading (4.4) Key Details
- CR 4 Reading (4.4) Central Ideas
- CR 4 Reading (4.4) Word Meanings
- CR 4 Reading (4.4) Reasoning & Evidence
- CR 4 Reading (4.4) Analyzing What or How
- CR 4 Reading (4.4) Language Use
- CR 4 Reading (4.4) Key Details
- CR 4 Reading (4.4) Central Ideas
- CR 4 Reading (4.4) Word Meanings
- CR 4 Reading (4.4) Reasoning & Evidence
- CR 4 Reading (4.4) Language Use

Student Mastered

14 (87%) students (71%)

- Dana Dreyfus 2
- Paula Merya 2
- Julia Stewart 2
- Joan March 2
- Travis R. ... 2

42

Learning Map (Item Level)

Gr 4 Reading (Info): Word Meanings

1-4 Items

Item No.	Type	Incorrect	Correct	Not Scored	Discontinued
1	MC	1 (25%)	2 (50%)	0 (0%)	1 (25%)

The selected assessment has no items for this standard.

The selected assessment has no items for this standard.

Standards: 4-1 (25%) **Not Met/Not**, 4-2 (50%) **Met/Correct**, 4-3 (25%) **Not Scored**, 4-4 (25%) **Discontinued**

43

Learning Map (Detailed Item Level)

Gr 4 Reading (Info): Word Meanings

1-4 Items

Item No.	Type	Incorrect	Correct	Not Scored	Discontinued
1	MC	1 (25%)	2 (50%)	0 (0%)	1 (25%)

Read the sentence from paragraph 3.

What does the word **habitat** mean?

X A the other members of an animal's family

X B the health and well-being of an animal

✓ C the natural place where an animal lives

44

“Now What” Tool

The screenshot displays a web application interface for a tool called "Now What". The interface is divided into several sections:

- Header:** A dark blue banner at the top contains the text "“Now What” Tool" in white.
- Browser Bar:** Below the banner is a Chrome browser address bar showing the URL: <https://www.mindspark.com/...>
- Left Sidebar:** A sidebar on the left titled "RL.4.1" contains a list of users with a "Not Reviewed" status. The list includes:
 - Clara Stein
 - Joseph Ferguson
 - Christie Peterson
 - Tracy Johnson
 - Carolee McLeod
 - Tanya Stuy-Matt
 - Dana Bentley
 - arashan Egger
 - Rayne Dwyer
 - Nathan Ray
 - Michael Lopez
 - Carra Stevens
 - Wanda Myers
 - Arlowyn Reed
 - Hilary Vines
 - Cheryle Gura
 - Lillian Klein
- Main Content Area:** The central area displays resource cards from "Amplify insight" and "ReadWorks.org".
 - Resource 1:** "Quick Check - ELA Grade 4 (L.17) Understanding Informational Text". It is a Resource Type: Formative Assessment, dated 03/20 to 03/20. It includes a link to access the grade and subject filter.
 - Resource 2:** "Quick Check - ELA Grade 4 (L.17) Understanding Informational Text". It is a Resource Type: Formative Assessment, dated 03/20 to 03/20. It includes a link to access the grade and subject filter.
 - Resource 3:** "Read-Aloud Lesson: 14 Cows for America". It is a Resource Type: Lesson Plan, dated 01/29 to 1/30. It includes a brief description of the lesson.
- Right Sidebar:** A sidebar titled "New What?" contains a "Unsaved Plan" section with options to "Clear", "Drag a Resource", "Add RL.4.1 or drag here", "Drag a Resource", and "Print".

