

## Series 501 Teacher-Led Writing Tryouts Report

### Purpose

The purpose of the Series 501 Teacher-Led Writing Tryouts is to ensure that the Writing tasks currently in development elicit the necessary evidence to allow inferences about students' levels of English language development as defined by the WIDA ELD Standards Framework and the Writing item specifications.

The Tryouts results are used in two ways:

1. To select which tasks will continue to field testing.
2. To inform revisions to continuing tasks prior to field testing.

### Research Questions

Series 501 Teacher-Led Writing Tryouts address the following research questions:

1. Does the writing task successfully elicit language as defined in the item specification?
2. Does anything in the writing task (e.g., input, graphics, prompt, sample writing) appear to prevent students from demonstrating their written English language proficiency?
3. Do responses to the writing task lend themselves to clear and efficient scoring decisions?

### Participants

Tryouts were conducted May 10–23, 2018. A total of 30 teachers and 553 students from schools across the WIDA consortium participated in Series 501 Teacher-led Writing Tryouts. Table 1 below shows the breakdown of teacher participation by state.

| State         | # teachers |
|---------------|------------|
| Georgia       | 7          |
| Illinois      | 7          |
| Indiana       | 2          |
| Maryland      | 2          |
| Missouri      | 2          |
| New Hampshire | 1          |
| New Jersey    | 1          |
| New Mexico    | 1          |
| Pennsylvania  | 3          |
| Rhode Island  | 1          |
| Utah          | 1          |
| Virginia      | 1          |
| Wisconsin     | 1          |
| <b>Total</b>  | <b>30</b>  |

Table 2 below shows the total number of students who participated by grade level cluster and tier. Students' reported proficiency levels ranged from 1.0 through 6.0.

| Grade level cluster | Tier | # students |
|---------------------|------|------------|
| Grade 1             | A    | 48         |
|                     | BC   | 53         |
| Grades 2–3          | A    | 55         |
|                     | BC   | 88         |
| Grades 4–5          | A    | 50         |
|                     | BC   | 52         |
| Grades 6–8          | A    | 41         |
|                     | BC   | 48         |
| Grades 9–12         | A    | 74         |
|                     | BC   | 44         |
| <b>Total</b>        |      | <b>553</b> |

All writing tasks still in development for Series 501 were administered at the Teacher-Led Tryouts. Six tasks per grade cluster (three tasks per tier) were administered with the exception of Grade 1 Tier B/C which had only two tasks remaining in development, for a total of 29 tasks.

## Administration Procedures

Table 3 summarizes the procedures used to collect student data.

| Part   | Description  |
|--|--|
| 0: Background Data                                   | The teacher collected grade level and proficiency level data.  |
| 1: Scripted Introduction                             | The teacher introduced the activity and answered any student questions.  |
| 2: Writing Test Practice (Grades 4–12 students only) | The students (Grades 4–12 only) completed the new writing test practice.   |
| 3: Writing Task                                      | The students completed one writing task appropriate for their grade level and proficiency level.   |
| 4: Student Survey                                    | The students completed a checklist-style survey about the writing task with support from the teacher.  |
| 5: Teacher Survey                                    | The teacher completed a survey about the writing tasks they administered, as well as general questions about the scripting, response booklets, and tryouts procedures. |

## Data Collected

The following data was collected as part of the Tryouts:

- Student responses to writing tasks
- Student survey responses
- Students' grade levels
- Students' overall proficiency levels
- Teacher survey responses
- CAL analysis checklist responses

## Analysis Procedures

At the completion of the Tryouts, all writing responses and survey responses were compiled and reviewed. Responses for the Tryouts tasks were scored by a trained rater using the Writing Scoring Scale. The data was analyzed by CAL test development team staff members using an analysis checklist covering scoring ease, task input effectiveness, and desired response characteristics.

## Results and Recommendations

This section summarizes the results of the Tryouts for several general aspects of the writing tasks and for each writing task that was administered.

In addition, a recommendation is also provided for which tasks should continue to field testing, which could be used in the future, and which should be discontinued. For items that are recommended for field testing, any planned revisions based on the Tryouts results are also recorded. Table 4 below shows the field test plan for Series 501 Writing. Recommendations for the number of tasks to field test are based on this plan.

| Grade Level Cluster | # of field test tasks |                       |
|---------------------|-----------------------|-----------------------|
|                     | Tier A LoLA           | Tier C LoLA/LoSS (LS) |
| Grade 1             | 2                     | 1                     |
| Grades 2–3          | 1                     | 2                     |
| Grades 4–5          | 1                     | 2                     |
| Grades 6–8          | 1                     | 2                     |
| Grades 9–12         | 1                     | 2                     |

## General Observations and Feedback

### Grades 1–3: Test Administrator’s Script

Of the 30 teachers participating in the Tryouts, 14 teachers administered at least one Grades 1–3 writing task using the new Test Administrator’s Scripts. Table 5 below summarizes the teachers’ responses to survey questions about the new scripts.

| Statement  | Strongly Agree/Agree | Disagree/Strongly Disagree | Total |
|--|----------------------|----------------------------|-------|
| The Test Administrator's Script was easy to use.         | 14                   | 0                          | 14    |
| The scripting helped students understand the task input. | 13                   | 1                          | 14    |
| The scripting helped students understand what to write.  | 12                   | 2                          | 14    |

Overall, there were no concerns raised about the new script format. Feedback from teachers detailing specific concerns was related to specific tasks and was considered along with all other information about the relevant tasks.

### Grades 4–12: Stop Sign Removal

Of the 30 teachers participating in the Tryouts, 23 teachers administered at least one Grades 4–12 writing task. Table 6 below summarizes the teachers' responses to survey questions about the handwritten test administration, which did not include a stop sign screen in the test forms or in the student response booklets.

| Statement  | Strongly Agree/Agree | Disagree/Strongly Disagree | Total |
|--|----------------------|----------------------------|-------|
| Students understood <u>when to turn the page</u> in their booklet. | 20                   | 3                          | 23    |
| Students understood <u>where to write</u> in their booklet.        | 21                   | 2                          | 23    |

Overall, no major issues were observed with the handwritten administration without a stop sign screen. One teacher reported confusion due to the presence of the Writing Practice, while another reported that the confusion was due to the fact that they normally take the test online (keyboarded version) and this was their first experience writing on paper.

### Grades 4–12: New Writing Test Practice

Grades 4–12 students completed the new Writing Test Practice before beginning their writing tasks. Responses to the practice were reviewed for any issues which might necessitate revisions. Survey responses from teachers were also reviewed for any issues related to the practice.

Out of the 23 teachers who administered the practice, one teacher reported that the practice took a long time and students disliked that it was so much easier than the actual writing tasks. Another teacher reported that the overall test introduction was frustrating due to the number of times students must click, and the length of the audio on the practice screens. Because these were not widely reported concerns, we do not recommend changes be made in response.

Overall, responses to the Test Practices were as expected. Responses demonstrated understanding of the task input and of task expectations for each prompt. Some students, primarily in Tier A, produced more than the requested number of sentences in response to each prompt. However, Tier B/C students tended to produce longer sentences and use the entire response area for the specified number of sentences, so reducing the number of available lines is not recommended.

No changes to the Writing Test Practices are recommended based on the Tryouts results.

## Grade 1

## Tier A, Language of Language Arts

*Recommendation*

- **Field test: Drawing Together**
- **Field test: Making a Poster**
- Postpone: Good Morning!

*Rationale*

All Grade 1 Tier A tasks performed well at Tryouts. Based on our analysis, we believe that *Making a Poster* and *Drawing Together* are the strongest of the tasks and should be field tested. These tasks more consistently elicited higher-level discourse features, while responses to *Good Morning!* tended to be lengthy but lack organization. In addition, *Good Morning!* may pose content overlap concerns if placed on Series 501OP alongside the continuing task (*Part D: Sentences About Me*), so we believe this task and spec is best saved for the future.

*Tryouts Results*

|                          | <b>Drawing Together</b><br><i>W01A_LA_Coloring_501</i>  | <b>Making a Poster</b><br><i>W01A_LA_Poster_501</i>  | <b>Good Morning!</b><br><i>W01A_LA_Morning_501</i>  |
|--------------------------|---|--|---|
| <b>Genre</b>             | Narrative   | Procedure  | Personal Recount  |
| <b>N (students)</b>      | 14  | 21   | 13  |
| <b>N (teachers)</b>      | 3   | 3  | 4   |
| <b>Strengths</b>         | <ul style="list-style-type: none"> <li>• Topic is clearly relatable</li> <li>• Students understood what to do</li> <li>• Task elicits sufficient original language</li> </ul> | <ul style="list-style-type: none"> <li>• Task elicits clear organization</li> <li>• Students understood what to do</li> <li>• Task elicits sufficient original language</li> </ul> | <ul style="list-style-type: none"> <li>• Topic is clearly familiar</li> <li>• Students had a lot to say</li> <li>• Task elicits sufficient original language</li> </ul> |
| <b>Weaknesses</b>        | <ul style="list-style-type: none"> <li>• None</li> </ul>  | <ul style="list-style-type: none"> <li>• None</li> </ul>   | <ul style="list-style-type: none"> <li>• Task doesn't elicit well-organized discourse</li> </ul>  |
| <b>Planned Revisions</b> | <ul style="list-style-type: none"> <li>• None</li> </ul>  | <ul style="list-style-type: none"> <li>• None</li> </ul>   | <ul style="list-style-type: none"> <li>• n/a</li> </ul>   |

## Tier C, Language of Language Arts/Language of Social Studies

*Recommendation*

- **Field test: Trees**
- Postpone: What Does a Zookeeper Do?

*Rationale*

Both Grade 1 Tier C tasks performed well at Tryouts. Based on our analysis, we believe that *Trees* is the stronger of the two tasks and should be field tested. Both tasks appeared to be somewhat more challenging than previous Grade 1 Tier C narrative tasks, as students appeared less familiar with strategies for organizing this type of writing. However, both tasks were engaging and elicited ample original language. *Trees* elicited more elaboration and responses were better aligned with the targeted writing genre. *What Does a Zookeeper Do?* elicited less elaboration with some responses using more of a narrative structure, but we believe this task has potential for use in the future with revisions to the graphic layout and TA scripting.

*Tryouts Results*

|                          | <b>Trees</b><br><i>W01C_LS_TreeUse_501</i>  | <b>What Does a Zookeeper Do?</b><br><i>W01C_LS_Zookeeper_501</i>   |
|--------------------------|---|--|
| <b>Genre</b>             | Report (Resources)  | Report (Community Workers)   |
| <b>N (students)</b>      | 26  | 27   |
| <b>N (teachers)</b>      | 5   | 5  |
| <b>Strengths</b>         | <ul style="list-style-type: none"> <li>• Topic was engaging and familiar</li> <li>• Task elicits ample elaboration</li> <li>• Task elicits ample original language</li> <li>• Task elicited some organized discourse</li> </ul> | <ul style="list-style-type: none"> <li>• Topic is engaging and accessible</li> <li>• Task elicits sufficient original language</li> </ul>  |
| <b>Weaknesses</b>        | <ul style="list-style-type: none"> <li>• Some responses were list-like</li> </ul>   | <ul style="list-style-type: none"> <li>• Some responses were list-like</li> <li>• Less elaboration than <i>Trees</i></li> <li>• Layout of content may be leading some students to write a narrative</li> </ul> |
| <b>Planned Revisions</b> | <ul style="list-style-type: none"> <li>• Revise TA script to model language such as “also” and “another” to better support organization in responses</li> </ul>   | <ul style="list-style-type: none"> <li>• n/a</li> </ul>  |

## Grades 2–3

## Tier A, Language of Language Arts

*Recommendation*

- **Field Test: The Garden**
- Postpone: After School
- Postpone: Making a Card

*Rationale*

All Grades 2–3 Tier A tasks performed well at Tryouts. Based on our analysis, we believe that *The Garden* is the strongest of the tasks and should be field tested. This task consistently elicited the expected discourse features, complex sentences, and original vocabulary. *Making a Card* also performed well, eliciting organized discourse, complex sentences, and original vocabulary; however, some lower proficiency students misinterpreted the task and wrote a narrative rather than a procedure. *After School* was the weakest of the three tasks, with responses that were lengthy but tended to lack much organization.

*Tryouts Results*

|                          | <b>The Garden</b><br><i>W23A_LA_GardenSurprise_501</i>   | <b>After School</b><br><i>W23A_LA_AfterSchool_501</i>   | <b>Making a Card</b><br><i>W23A_LA_HowToMakeCards_501</i>   |
|--------------------------|--|---|---|
| <b>Genre</b>             | Narrative  | Personal Recount  | Procedure   |
| <b>N (students)</b>      | 17   | 24  | 14  |
| <b>N (teachers)</b>      | 2  | 4   | 5   |
| <b>Strengths</b>         | <ul style="list-style-type: none"> <li>• Students understood what to do</li> <li>• Students understood the graphics and were able to add details to their writing</li> <li>• Task allowed for original language</li> <li>• Task seemed engaging</li> </ul> | <ul style="list-style-type: none"> <li>• Students understood what to do</li> <li>• Topic is familiar and engaging</li> <li>• Task elicits elaborated responses and original language</li> </ul> | <ul style="list-style-type: none"> <li>• Most students understood what to do</li> <li>• Students understood the graphics and were able to describe the steps shown</li> <li>• Task allowed for original language</li> </ul> |
| <b>Weaknesses</b>        | <ul style="list-style-type: none"> <li>• None</li> </ul>   | <ul style="list-style-type: none"> <li>• Responses often lacked clear organization, with illogical sequencing at times</li> </ul>   | <ul style="list-style-type: none"> <li>• Some lower PL students wrote narratives rather than procedures</li> </ul>  |
| <b>Planned Revisions</b> | <ul style="list-style-type: none"> <li>• None</li> </ul>   | <ul style="list-style-type: none"> <li>• n/a</li> </ul>   | <ul style="list-style-type: none"> <li>• n/a</li> </ul>   |

## Tier C, Language of Language Arts/Language of Social Studies

*Recommendation*

- **Field Test: Salim Ali**
- **Field Test: Steps to Vote**
- Postpone: Visiting River City

*Rationale*

All Grades 2–3 Tier C tasks performed well at Tryouts. Based on our analysis, we believe that *Salim Ali* and *Steps to Vote* are the strongest of the tasks and should be field tested. These tasks consistently elicited extended, organized responses from higher proficiency students, with complex sentence and a variety of vocabulary. *Visiting River City* also performed well, eliciting complex sentences, and original vocabulary; however, responses showed less elaboration or less clear organization. *Visiting River City* has potential for use in the future with revisions to the task input and targeted genre.

*Tryouts Results*

|                          | <b>Salim Ali</b><br><i>W23C_LS_SalimAli_501</i>  | <b>Steps to Vote</b><br><i>W23C_LS_ClassVotingProcess_501</i>  | <b>Visiting River City</b><br><i>W23C_LS_RiverCity_501</i>  |
|--------------------------|--|--|---|
| <b>Genre</b>             | Biographical Recount   | Procedure  | Report  |
| <b>N (students)</b>      | 27   | 23   | 38  |
| <b>N (teachers)</b>      | 5  | 2  | 4   |
| <b>Strengths</b>         | <ul style="list-style-type: none"> <li>• Task was engaging</li> <li>• Task elicited organized, elaborated responses</li> <li>• Task elicited original language</li> </ul>              | <ul style="list-style-type: none"> <li>• Task elicited organized, elaborated responses</li> <li>• Students understood the graphics</li> <li>• Task elicited original language</li> </ul> | <ul style="list-style-type: none"> <li>• Task was engaging</li> <li>• Task elicited elaborated responses</li> </ul>   |
| <b>Weaknesses</b>        | <ul style="list-style-type: none"> <li>• Some responses relied on repetitive sentence structures</li> <li>• Sample writing is not effective in supporting strong beginnings</li> </ul> | <ul style="list-style-type: none"> <li>• Sample writing is not effective in supporting strong beginnings</li> </ul>  | <ul style="list-style-type: none"> <li>• Many responses relied on repetitive sentence structures</li> <li>• Many responses lacked clear organization</li> </ul> |
| <b>Planned Revisions</b> | <ul style="list-style-type: none"> <li>• Remove sample writing and add explicit direction about including an introduction and ending to the task prompt</li> </ul>                     | <ul style="list-style-type: none"> <li>• Remove sample writing and add explicit direction about including an introduction and ending to the task prompt</li> </ul>                       | <ul style="list-style-type: none"> <li>• n/a</li> </ul>   |

## Grades 4–5

## Tier A, Language of Language Arts

*Recommendation*

- **Field Test: Looking at the Stars**
- Postpone: The Best Way to Read
- Postpone: Working Together

*Rationale*

Based on the Tryouts data, *Looking at the Stars* was the strongest of the three Grades 4–5 Tier A tasks. Students understood what to do and were able to produce organized responses that included complex sentences and original vocabulary. *The Best Way to Read* worked well with students whose proficiency levels were at the higher end of Tier A, while lower proficiency students who are the target population for this task seemed to struggle. This theme appears to be better suited to a Tier C task than Tier A. *Working Together* produced scorable responses with original vocabulary and complex sentences, but responses tended to be list-like and showed some misinterpretations of the second prompt. This task could potentially be used in the future with revisions to the graphics.

*Tryouts Results*

|                          | <b>Looking at the Stars</b><br><i>W45A_LA_Stargazing_501</i>  | <b>The Best Way to Read</b><br><i>W45A_LA_WaysToRead_501</i>   | <b>Working Together</b><br><i>W45A_LA_WorkingTogether_501</i>   |
|--------------------------|---|--|---|
| <b>Genre</b>             | Narrative   | Argument   | Personal Recount  |
| <b>N (students)</b>      | 20  | 17   | 13  |
| <b>N (teachers)</b>      | 2   | 3  | 4   |
| <b>Strengths</b>         | <ul style="list-style-type: none"> <li>• Students understood what to do</li> <li>• Students were able to produce responses that reflected their reported proficiency levels</li> </ul>  | <ul style="list-style-type: none"> <li>• Topic incorporates a familiar setting and familiar vocabulary</li> <li>• Higher PL students were able to respond to the both prompts as expected</li> </ul> | <ul style="list-style-type: none"> <li>• Students were able to respond to both prompts</li> </ul>   |
| <b>Weaknesses</b>        | <ul style="list-style-type: none"> <li>• Some responses showed minor misinterpretations of graphics</li> <li>• Some responses to the second prompt began with the second picture rather than the first</li> </ul>                   | <ul style="list-style-type: none"> <li>• Lower PL students struggled with understanding their task and producing an argument</li> </ul>  | <ul style="list-style-type: none"> <li>• Responses to the second prompt weren't cohesive recounts</li> <li>• Graphic may be leading students to list times they worked together rather than describe details of one time</li> </ul> |
| <b>Planned Revisions</b> | <ul style="list-style-type: none"> <li>• Revise second prompt to include directive to begin with Picture 1</li> <li>• Add a small amount of additional scripting to the second task to support understanding of graphics</li> </ul> | <ul style="list-style-type: none"> <li>• n/a</li> </ul>  | <ul style="list-style-type: none"> <li>• n/a</li> </ul>   |

## Tier C, Language of Language Arts/Language of Social Studies

*Recommendation*

- **Field Test: Changes in Technology**
- **Field Test: Heating and Cooling**
- Postpone: Braille Writing System

*Rationale*

Based on the Tryouts data, *Changes in Technology* and *Heating and Cooling* were the strongest of the three Grades 4–5 Tier C tasks. While *Changes in Technology* wasn't interpreted exactly as expected, students were able to produce detailed arguments on this topic, including complex sentences and original vocabulary. *Heating and Cooling* also elicited organized responses that showed understanding of the task and graphics. *Braille Writing System* had mixed results with some responses as expected, but others which showed difficulty synthesizing both the information about Louis Braille and the information about how braille is used today. This task could likely be used in the future with revisions to the task input and targeted genre.

*Tryouts Results*

|                          | <b>Changes in Technology</b><br><i>W45C_LS_Industrial_501</i>   | <b>Heating and Cooling</b><br><i>W45C_LS_Heating_501</i>  | <b>Braille Writing System</b><br><i>W45C_LS_Braille_501</i>  |
|--------------------------|---|---|--|
| <b>Genre</b>             | Argument from Evidence  | Sequential Explanation  | Report   |
| <b>N (students)</b>      | 12  | 14  | 26   |
| <b>N (teachers)</b>      | 1   | 5   | 3  |
| <b>Strengths</b>         | <ul style="list-style-type: none"> <li>• Students were able to produce detailed arguments about this topic</li> <li>• Task elicited complex sentences and sufficient original language</li> </ul>                             | <ul style="list-style-type: none"> <li>• Students understood what to do</li> <li>• Students understood the graphics</li> <li>• Task elicited organized responses with original language and some elaboration</li> </ul> | <ul style="list-style-type: none"> <li>• Students were able to write detailed responses about the importance of braille, bringing in details beyond the prompt</li> <li>• Students seemed engaged in addressing the importance of braille</li> </ul> |
| <b>Weaknesses</b>        | <ul style="list-style-type: none"> <li>• Prompt wasn't interpreted as expected, with responses focusing on how technology changed between the past and the present-day, rather than how it had changed in the past</li> </ul> | <ul style="list-style-type: none"> <li>• Responses tended to lack beginnings and endings</li> </ul>   | <ul style="list-style-type: none"> <li>• Students struggled to synthesize the different parts of the task input and prompt into a cohesive response</li> </ul>   |
| <b>Planned Revisions</b> | <ul style="list-style-type: none"> <li>• Revise orientation task input to support responses focusing on changes to present-day</li> </ul>   | <ul style="list-style-type: none"> <li>• Add directive to the prompt about including an introduction and conclusion</li> </ul>  | <ul style="list-style-type: none"> <li>• n/a</li> </ul>  |

## Grades 6–8

## Tier A, Language of Language Arts

*Recommendation*

- **Field Test: Making A Decision**
- Postpone: Being Patient
- Postpone: Ways to Learn

*Rationale*

Based on the Tryouts data, *Making a Decision* was the strongest of the three Grades 6–8 Tier A tasks. Students seemed engaged and understood what to do, although some responses to the first prompt were copied. *Ways to Learn* worked well with students whose proficiency levels were at the higher end of Tier A; however, lower proficiency students who are the target population for this task seemed to struggle with producing an argument. This theme appears to be better suited to a Tier C task than Tier A. *Being Patient* produced responses that were short and lacked much detail or original language and so did not seem to allow students to fully demonstrate their language proficiency; further development of this task is not recommended.

*Tryouts Results*

|                          | <b>Making a Decision</b><br><i>W68A_LA_HourDecision_501</i>  | <b>Being Patient</b><br><i>W68A_LA_Patience_501</i>   | <b>Ways to Learn</b><br><i>W68A_LA_StudentLearning_501</i>  |
|--------------------------|--|---|---|
| <b>Genre</b>             | Critical Response  | Narrative (Personal)  | Argument  |
| <b>N (students)</b>      | 18   | 12  | 11  |
| <b>N (teachers)</b>      | 2  | 2   | 2   |
| <b>Strengths</b>         | <ul style="list-style-type: none"> <li>• Students seemed engaged with the task</li> <li>• Students were able to express their opinion and support it with details</li> </ul> | <ul style="list-style-type: none"> <li>• Students were able to complete the first prompt successfully</li> </ul>  | <ul style="list-style-type: none"> <li>• Students seemed to understand the graphics</li> <li>• Higher PL students were able to respond to both prompts as expected</li> </ul> |
| <b>Weaknesses</b>        | <ul style="list-style-type: none"> <li>• Responses to the first prompt were sometimes copied from the orientation screen</li> </ul>  | <ul style="list-style-type: none"> <li>• Responses to the second prompt tended to be about waiting for school activities, without much detail</li> <li>• Responses didn't use much original language</li> </ul> | <ul style="list-style-type: none"> <li>• Lower PL students struggled to understand and respond to the second prompt</li> </ul>  |
| <b>Planned Revisions</b> | <ul style="list-style-type: none"> <li>• Revise first prompt so that it cannot be answered by copying the orientation statement</li> </ul>                                   | <ul style="list-style-type: none"> <li>• n/a</li> </ul>   | <ul style="list-style-type: none"> <li>• n/a</li> </ul>   |

## Tier C, Language of Language Arts/Language of Social Studies

*Recommendation*

- **Field Test: Storing Grain**
- **Field Test: The Erie Canal**
- Postpone: Art Deco

*Rationale*

Based on the Tryouts data, *Storing Grain* and *The Erie Canal* were the strongest of the three Grades 6–8 Tier C tasks. Despite a small sample, *Storing Grain* performed as expected, eliciting organized arguments with complex sentences and original vocabulary. *The Erie Canal* elicited scorable responses and appeared to present a concrete and accessible topic; however, responses did not synthesize the input to address the prompt as fully as intended. Of note, however, is that both teachers who administered this task reported that their students were not very motivated to do their best due to the time of year. *Art Deco* elicited lengthy responses with some original language, but responses tended to lack coherence or did not show full understanding of the task input. This task could potentially be used in the future for Grades 9–12 with revisions to the presentation of the task input.

*Tryouts Results*

|                          | <b>Storing Grain</b><br><a href="#">W68C_LS_StoringGrain_501</a>   | <b>The Erie Canal</b><br><a href="#">W68C_LS_ErieCanal_501</a>  | <b>Art Deco</b><br><a href="#">W68C_LS_ArtDeco_501</a>   |
|--------------------------|--|---|--|
| <b>Genre</b>             | Argument from Evidence   | Historical Recount  | Factorial Explanation  |
| <b>N (students)</b>      | 7  | 18  | 23   |
| <b>N (teachers)</b>      | 1  | 2   | 1  |
| <b>Strengths</b>         | <ul style="list-style-type: none"> <li>• Task elicited organized arguments</li> <li>• Task elicited complex sentences and ample original vocabulary</li> </ul> | <ul style="list-style-type: none"> <li>• The topic appeared accessible</li> </ul>   | <ul style="list-style-type: none"> <li>• Task elicited lengthy responses</li> <li>• Task elicited some original language</li> </ul>  |
| <b>Weaknesses</b>        | <ul style="list-style-type: none"> <li>• None</li> </ul>   | <ul style="list-style-type: none"> <li>• Responses tended to rely heavily on the stimulus</li> <li>• Responses did not synthesize the different aspects of the input as expected</li> </ul> | <ul style="list-style-type: none"> <li>• Responses lacked overall coherence</li> <li>• Responses did not demonstrate a complete understanding of the task input</li> </ul> |
| <b>Planned Revisions</b> | <ul style="list-style-type: none"> <li>• None</li> </ul>   | <ul style="list-style-type: none"> <li>• Remove timeline from the task input</li> <li>• Revise prompt to help clarify students' task</li> </ul>   | <ul style="list-style-type: none"> <li>• n/a</li> </ul>  |

## Grades 9–12

## Tier A, Language of Language Arts

*Recommendations*

- **Field Test: Joon’s Story**
- Postpone: Choosing a Book
- Postpone: Spilled Paint

*Rationale*

All Grades 9–12 Tier A tasks performed well at Tryouts. Based on our analysis, we believe that *Joon’s Story* is the strongest of the tasks and should be field tested. This task consistently elicited elaborated responses with complex sentences and original vocabulary. *Choosing a Book* also performed well, eliciting organized discourse and original vocabulary; however, many students taking this task were at the high end of Tier A, and there was less evidence that this task would be successful with lower PL Tier A students. *Spilled Paint* also performed well, with responses showing some organization, complex sentences, and original vocabulary; however, some responses did show a misunderstanding of the second prompt, so this task is not recommended for field testing at this time.

*Tryouts Results*

|                          | <b>Joon’s Story</b><br><i>W91A_LA_Apples_501</i>   | <b>Choosing a Book</b><br><i>W91A_LA_BookGenres_501</i>  | <b>Spilled Paint</b><br><i>W91A_LA_PaintSpill_501</i>  |
|--------------------------|--|--|--|
| <b>Genre</b>             | Narrative (Biographical)   | Argument   | Critical Response  |
| <b>N (students)</b>      | 28   | 23   | 23   |
| <b>N (teachers)</b>      | 4  | 3  | 3  |
| <b>Strengths</b>         | <ul style="list-style-type: none"> <li>• Task elicited organized responses with complex sentences and original vocabulary</li> <li>• Responses demonstrated understanding of the graphics</li> </ul> | <ul style="list-style-type: none"> <li>• Task elicited organized responses from higher PL students</li> <li>• Task elicited original vocabulary</li> </ul> | <ul style="list-style-type: none"> <li>• Task elicited some organization, with complex sentences and original vocabulary</li> <li>• Most responses demonstrated understanding of the graphics</li> </ul> |
| <b>Weaknesses</b>        | <ul style="list-style-type: none"> <li>• Some responses used repetitive sentence structures (e.g., when...)</li> </ul>   | <ul style="list-style-type: none"> <li>• Less evidence of desired responses from lower PL students</li> </ul>  | <ul style="list-style-type: none"> <li>• Some responses showed misunderstanding of the second prompt</li> </ul>  |
| <b>Planned Revisions</b> | <ul style="list-style-type: none"> <li>• Revise sample writing for second prompt to avoid setting up “When...” sentences</li> </ul>  | <ul style="list-style-type: none"> <li>• n/a</li> </ul>  | <ul style="list-style-type: none"> <li>• n/a</li> </ul>  |

## Tier C, Language of Language Arts/Language of Social Studies

*Recommendations*

- **Field Test: Bills versus Coins**
- **Field Test: Indus Valley**
- Postpone: Highways in the United States

*Rationale*

Based on the Tryouts data, *Bills versus Coins* and *Indus Valley* were the strongest of the three Grades 9–12 Tier C tasks. *Bills versus Coins* performed as expected, eliciting strong, detailed, and well-organized arguments with complex sentences and original vocabulary. *Indus Valley* also elicited organized arguments that included complex sentences and some original vocabulary; however, responses to this task tended to rely more heavily on the input. *Highways in the United States* elicited responses with some desired characteristics; however, of the three tasks, this task seemed to be the most challenging for students. With revisions to the task input, this task could likely be used in the future.

*Tryouts Results*

|                          | <b>Bills versus Coins</b><br><i>W91C_LS_ChangingMoney_501</i>  | <b>Indus Valley</b><br><i>W91C_LS_IndusValley_501</i>  | <b>Highways in the United States</b><br><i>W91C_LS_Highways_501</i>   |
|--------------------------|--|--|---|
| <b>Genre</b>             | Argument from Evidence (Public Policy)   | Argument from Evidence (Historical)  | Consequential Explanation   |
| <b>N (students)</b>      | 14   | 15   | 15  |
| <b>N (teachers)</b>      | 4  | 4  | 4   |
| <b>Strengths</b>         | <ul style="list-style-type: none"> <li>• Task elicits great arguments with counterpoints and persuasive language</li> <li>• Task elicits ample original language</li> <li>• Students argued for both keeping the bill and switching to the coin</li> </ul> | <ul style="list-style-type: none"> <li>• Task elicits organized arguments</li> <li>• Task elicits sufficient original language</li> <li>• Students used counter-evidence to support their opinion</li> </ul> | <ul style="list-style-type: none"> <li>• Task elicits complex sentences and some original vocabulary</li> <li>• Some students were able to add details to their explanations</li> </ul> |
| <b>Weaknesses</b>        | <ul style="list-style-type: none"> <li>• None</li> </ul>   | <ul style="list-style-type: none"> <li>• Some responses relied heavily on the text input</li> <li>• All students argued for the water system; none chose the system of weights</li> </ul>                    | <ul style="list-style-type: none"> <li>• Some responses relied heavily on the text input</li> <li>• Responses tended to be less well-organized than the argument tasks</li> </ul>       |
| <b>Planned Revisions</b> | <ul style="list-style-type: none"> <li>• None</li> </ul>   | <ul style="list-style-type: none"> <li>• Reduce text input slightly</li> </ul>   | <ul style="list-style-type: none"> <li>• n/a</li> </ul>   |