Table of Contents

Introduction

ese guidelines are to help ensure that the best possible items are developed for PIRLS/prePIRLS 2016. It is important to follow some basic procedures so that the PIRLS assessments are uniform in approach and format. During the item-writing sessions, please ask sta or consult these guidelines if you have any questions.

Item Formats

e two item formats used most in PIRLS are multiple-choice and

Number of Items and Score Points per Passage

In order to have a reliable measure of reading comprehension, each passage should have questions worth a total of at least 15 score points. Considering that items sometimes are deleted during the eld testing and review process, please write items totaling 18–20 score points per passage. On average, this will be 12 to 14 items per passage.

- Multiple-choice questions are worth

Exhibit 1: Generalized Scoring Guide for One-Point Questions

PIRLS	L preP

Converting the percentages of the PIRLS assessment devoted to each reading purpose and comprehension process means developing questions for each passage that will yield the following:

- **F **: 3–4 score points (probably also 3 to 4 items, because the questions primarily will be multiple-choice or one-point constructed-response);

- ** I **: 5–6 score points (probably 4 or 5 items, because the questions primarily will be multiple-choice or one- and two-point constructed-response questions);

- **I I **: 5–6 score points (probably 2 or 3 items, because they primarily will be longer constructed-response questions worth two or three points); and

- **E C **: 3–4 score points (typically a range of item types, so they might range from 4 multiple-choice questions to 1 long three-point constructed-response question).

Converting these percentages for the prePIRLS assessment means developing questions for each passage that will yield the following:

- **F **: 8–10 score points (probably also 8 to 10 items, because the questions primarily will be multiple-choice or one-point constructed-response);

- ** I **: 4–5 score points (probably 3 or 4 items, because the questions primarily will be multiple-choice and one- or two-point constructed-response questions); and

asking more difficult questions. These questions should be of relatively low difficulty and invite the student into the passage. The question might ask about the main topic of the passage or about some reasonably important or basic

the last question, because this makes it quite tempting for students to skip it and be nished.

5. **M**

Make sure that the information in the item does not provide clues to answers to other items in the set. Also, answering an item correctly must NOT depend on answering a previous item correctly. Pay particular attention to information in both the stem and options for multiple-choice items. Also, check to see that the constructed-response questions elicit di erent answers and require more than repeating the theme or major idea from question to question.

General Issues in Writing Items for PIRLS

Item writing is a task that requires imagination and creativity, but at the same time demands considerable discipline in order to meet all of the criteria discussed in this manual. e previous sections of this manual have presented guidelines speci c to the PIRLS 2016 passages. e guidelines in this section pertain to good item and test development practices in general, and have been collected from a number of sources. ese issues also must be considered in judging the quality and suitability of an item for PIRLS 2016.

Testing Time

In developing items, it is important to consider the time required for students to complete the required task. In PIRLS, students have 40 minutes to read and answer the questions about each passage. As a general rule, a typical fourth grade student is expected to complete a multiple-choice item in approximately one minute. Constructed-response items are allocated more testing time in the PIRLS design, with 1–2 minutes for short-answer items and 3–5 minutes for

extended-response items. When writing a set of items, please keep the total time it will take students to respond into consideration.

Grade Appropriateness

The language, style, and reading level used in items must be appropriate to the age and experiences of the students in the target grade. The items should be written at a reading level such that students should be able to understand the demands of an item without di culty.

Item Difficulty

It is desirable that there be some relatively easy items and some challenging items. However, items that almost all students or almost no students are able to answer correctly reduce the e ectiveness of the test to discriminate between groups with high achievement and groups with low achievement.

Avoiding Bias

In preparing test items, be sensitive to the possibility of unintentionally placing groups of students at an unfair disadvantage. In an international study, extra care is required to consider the diversity of environments, backgrounds, beliefs, and cultures among students in the participating countries.

Considering National Contexts

Be particularly aware of issues related to nationality, culture, ethnicity, and geographic location. Items requiring background knowledge con ned to a subset of participating countries are unlikely to be suitable.

Geographic location has an e ect on the learning experiences students are exposed to, because aspects of the local environment have an impact on schooling.

Even though television and the Internet can provide students with some knowledge of remote places, rsthand experience of some phenomena enhances understanding and can give some students an advantage over others.

Gender

A gender-related context included in an item may distract some students from the purpose of the item. Situations in which stereotypical roles or attitudes are unnecessarily attributed to males or females, or in which there is implicit

The next sections present guidelines specific to multiple-choice items, including writing the stem, structuring the response options, and developing plausible distracters.

The Stem

For PIRLS, because the students are relatively young and clarity is of vital importance, please phrase all stems as a question.

Example of a stem formulated as a question:

Where did Labon put the mousetraps?

- Ⓐ In a basket
- Ⓑ Near the mouse holes
- Ⓒ Under the chairs
- Ⓓ On the ceiling

1. Provide sufficient information in the stem to make the task clear and unambiguous to students. Students should be able to answer the question before reading the options.

2. The stem should not include extraneous information. Extraneous information is liable to confuse students who otherwise would have determined the correct answer.

3. Do NOT use negative stems—those containing words such

Structure of the Response Options (or Alternatives)

1. As shown in the "Labon" example above, multiple-choice items for PIRLS have four response options, labeled A–D.

2. Make sure that one of the four response options or alternatives is the   ,   . Make sure there is only one correct or best answer. For example, response options cannot represent subsets of other options. Also, do not use subsets of response options that together account for all possibilities (e.g., day and night), because one of these must be the key.

3. Make sure that the grammatical structure of all response options " t" the stem. Inconsistent grammar can provide clues to the key or eliminate incorrect response options.

4. Make sure all (or sets) of the response options are parallel in length, level of complexity, and grammatical structure. Avoid the tendency to include more details or quali cations in the correct response, thus making it stand out.

5. If a word or phrase is repeated in each of the response options, try to reduce the reading burden by moving the word(s) to the stem. However, do not sacri ce clarity to save a few words.

6. Do not use words or phrases in the stem that are repeated in one of the response options and, therefore, act as a clue to the correct response.

7. Do **NO** use "none of these" and "all of these" as response options.

Plausibility of Distracters|

Developing a constructed-response item accurately targeted on the ability to be assessed, along with the accompanying scoring guide, is not a straightforward task. If the task is not well specified,

Writing a Full-Credit Response to the Question

As you are writing a constructed-response item, please write a full-credit answer to the question in terms of the language, knowledge, and skills that a fourth grade student could be expected to possess.

is is an essential rst step in producing a scoring guide and testing the viability of the question. If you cannot answer the question or disagree among yourselves, then the question should be reconsidered. In any case, thinking of the answer simultaneously with developing

Exhibit 2: Generalized Scoring Guide for One-Point Questions

Acceptable Response (Score = 1)
These responses demonstrate comprehension of the aspect of text addressed in the question. They include all elements required by the question. The responses are determined to be accurate based on ideas or information in the text.

Unacceptable Response (Score = 0)
These responses do not demonstrate comprehension of the aspect of text addressed in the question. They may attempt to provide some or all of the elements required by the question. The responses, however, are determined to be inaccurate based on information or ideas in the text. Or, they include only ideas or information that are too vague or unrelated to the question to be considered accurate.
Also give a score of "0" to uninterpretable responses. This includes crossed-out and erased attempts, illegible and off-task responses, and drawings and doodles.

Exhibit 3: Generalized Scoring Guide for Two-Point Questions

Complete Comprehension (Score = 2)

These responses demonstrate complete comprehension of the aspect of text addressed in the question. They include all elements required by the question. When required, they demonstrate a level of comprehension that goes beyond a literal understanding, and provide appropriate interpretations, inferences, or evaluations that are consistent with the text. Or, they include complete and adequate ideas or information from the text to support an interpretation, inference, or evaluation based on the text.

Partial Comprehension (Score = 1)

These responses demonstrate only partial comprehension of the aspect of text addressed in the question. They may include some, but not all, of the elements required by the question. Or, they may address all elements required by the question, but demonstrate only a literal understanding when the question asks for an interpretation, inference, or understanding of a more abstract concept. When required by the question to provide an explanation for an interpretation, inference, or evaluation, the responses may lack adequate textual support, or provide only unrelated or vague information.

No Comprehension (Score = 0)

These responses demonstrate no comprehension of the aspect of text addressed in the question. They may attempt to provide some or all of the elements required by the question, however, the response is determined to be inaccurate based on ideas or information in the text. Or, they may fail to address any element required by the question. Or, the responses include only information or ideas that are too vague or unrelated to the question to be considered evidence of comprehension.

Also give a score of "0" to uninterpretable responses. This includes crossed-out and erased attempts, illegible and off-task responses, and drawings and doodles.
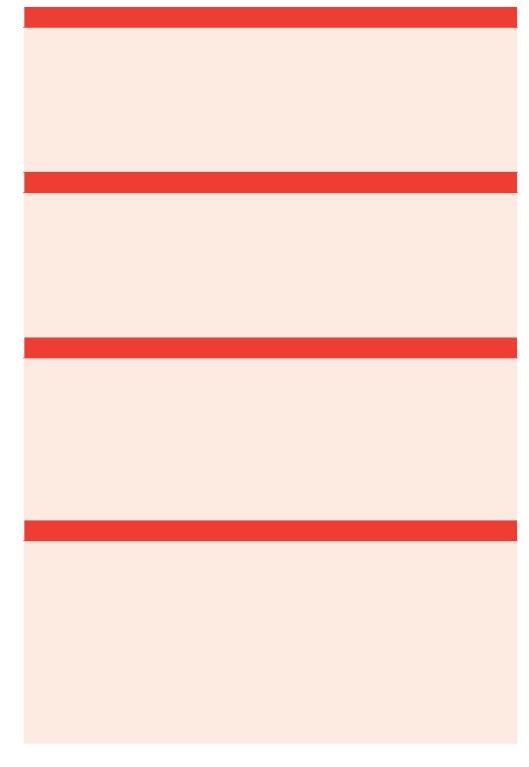
Exhibit 4: Generalized Scoring Guide for Three-Point Questions

Tailoring the PIRLS Generalized Scoring Guides for Each Unique Constructed-Response Item

Using the generalized guides or templates presented in the previous section, develop a unique scoring guide tailored to each constructed-response question. Two critical goals must be addressed in developing the scoring guides for each constructed-response item:

- Make the criteria as specic as possible in order to standardize scoring decisions across countries; and

- Provide for a range of responses within each score level.

ese somewhat conicting goals of specicity and exibility are addressed by providing both specic and general descriptions of comprehension at each score level. To provide examples of how this is accomplished, Appendix A contains the "Fly Eagle Fly" passage from PIRLS 2011 together with its items and scoring guides. For example, the scoring guide for Item 10 is for a one-point question developed to assess students' ability to evaluate language while reading for the purpose of literary experience. Please identify the "purpose" and "process" assessed by each question at the top of the rst page of every guide. Each scoring guide is divided into sections corresponding to the number of score levels, including a score of zero. Note that this one-point guide has two sections: the rst section provides criteria for a score of one, and the second section provides criteria for a score of zero.

For each score level in the guide, a general statement regarding the nature of comprehension that is characteristic of responses at that level is shown rst. e statement providing general guidance is followed by specific examples of expected student responses, although the examples are not an exhaustive list of all possibilities.

7. Explain what the farmer's friend meant when he told the eagle, "You belong not to the earth but to the sky."

*P   :   *

2 – Complete Comprehension

e response interprets the meaning of both parts of the quote – "belong not to the earth" and "belong to the sky" in terms of the story.

E a   :

It is supposed to be free in the sky and not stuck on the ground.

at it was not a chicken who walked on the earth. It was an eagle and meant to y.

It was meant to be ying with other birds of its kind, not among chickens.

It is meant to y, not walk.

e sky is his home, not the ground.

1 – Partial Comprehension

e response interprets only the rst or the second part of the quote.

E a   :

at it was not a chicken. /It was an eagle.

It was the king of the ying birds.

It was not a ground animal.

It is meant to y.

Or, the response describes the literal contrast only.

E a   :

It was not a chicken but an eagle.

0 – No Comprehension

e response may provide an explanation of the quote that is vague or inaccurate, or it may provide a simple rephrasing of the quote itself.

E a   :

It is supposed to be not of the earth but of the sky.

It belongs to the sky not on the ground.

Appendix B: Multiple-Choice and Constructed-Response Checklists

Exhibit B.1: Multiple-Choice Item Review Checklist

Item Characteristic	Yes	No