<table>
<thead>
<tr>
<th><strong>Title</strong></th>
<th>Scoring student tests in reading comprehension (OER 09-10 RS &amp; OER 40-12 RS)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Author(s)</strong></td>
<td>Foong Poh Yi and Rita Elaine Silver</td>
</tr>
<tr>
<td><strong>Published by</strong></td>
<td>National Institute of Education (Singapore)</td>
</tr>
</tbody>
</table>

This document may be used for private study or research purpose only. This document or any part of it may not be duplicated and/or distributed without permission of the copyright owner.

The Singapore Copyright Act applies to the use of this document.

Citation: Foong, P. Y., & Silver, R. E. (2016). *Scoring student tests in reading comprehension (OER 09-10 RS & OER 40-12 RS)*. Singapore: National Institute of Education.
SCORING STUDENT TESTs IN READING COMPREHENSION
(OER09-10RS & OER 40-12RS)

FOONG Poh Yi

with

Rita Elaine SILVER

Updated 27 June 2016

These procedures were initially trialed and used for scoring student tests for OER 09/10 RS. Subsequently the procedures were adapted slightly, using Cohen’s Kappa rather than percentage agreement, for OER 40-12 RS. For more information on OER 09-10RS see Silver, Png, Rasidir, Foong, Huynh & Kogut (2009) and Silver, Png, Kogut & Huynh (2010). For more information on OER 40-12 RS, see Silver and Png (2015), Silver and Png (in press).

Introduction

This document explains the procedure used to score student test responses from the project ‘Comprehending Reading Comprehension’ (OER09-10RS). Baseline data in the form of student test responses from a reading comprehension test were collected thrice in an academic year, January to December 2011, from three levels, Primary 3, 4, and 5.

Student test
For each reading comprehension test, there were 3 passages. Each passage had two sections, one which comprised of four traditional reading comprehension (TRC) questions and the other of four ‘Questioning-the-author’ (QtA) queries. Each TRC question or QtA query carried a maximum of two marks. In total, each student had to answer 24 questions and could score up
to a maximum of 48 marks potentially. Table 1 below shows the titles of the different passages for each of the three levels.

Table 1

*Passages for Primary 3, 4 and 5*

<table>
<thead>
<tr>
<th>Level</th>
<th>Primary 3</th>
<th>Primary 4</th>
<th>Primary 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>Passage 1</td>
<td>Hibernation</td>
<td>Society</td>
<td>The Development of Clothes</td>
</tr>
<tr>
<td>Passage 2</td>
<td>Skunks</td>
<td>Telling Time Without Clocks</td>
<td>Anaesthetics</td>
</tr>
<tr>
<td>Passage 3</td>
<td>Honey Makers</td>
<td>Brain and Body</td>
<td>Houdini</td>
</tr>
</tbody>
</table>

Passages were taken from children’s literature and reading assessments books. Each passage was checked for appropriateness for that grade level based on the expertise of three experienced primary school teachers in Singapore (all members of the research team) and by examination of ‘readability’ scores comparing results from the Fry Readability Formula (Fry, 1968), Gunning Fog, Flesch Reading Ease & Grade Level, and Power Sumner Kearl readability measures (Wyatt & Schnellbach, 2012).

**Training**

In preparation for the scoring of the student test responses, the research team underwent training in order to achieve a standardized method in coding all student responses. Two researchers who had coded the most number of student test responses in an earlier project (OER29-08RS) led the training by showing the team how to code the responses. The objective of the training was to achieve high inter-scorer agreement between the coders.

**Stage 1**

At the first stage, initial group training and inter scorer agreement checks were carried out. Copies of scored test scripts from the earlier project, (OER29-08RS), were presented to
everyone on the team. These scored test scripts were coded by consensus by the two researchers in charge of training. Next, the procedures and logic for scoring were explained. To further illustrate, all answers were marked on a 3-point scale. Hence, a correct answer would be awarded 2 marks, a partially correct answer, 1 mark, and an incorrect answer, 0 marks. In addition, students were not penalized for grammatical or spelling errors as these do not reflect incomprehension. A set of the marking scheme and another copy of a scored test script (marks already awarded to the answers do not appear on these copies) were then given to all six researchers. Individually, each researcher had to score this test script and then checked their scores with the master scores. This practice was repeated until everyone achieved high agreement with the master scores.

Stage 2
At the second stage, pair training and inter-scorer agreement checks were carried out. The team of six was divided into three pairs. Each pair was to score all the student responses from one of the three levels. During this stage of the training, two sets of marking scheme and two copies of student test scripts (previously scored by one of the two researchers in charge of the training) were given to each pair. Individually, each pair was to score these test scripts and check their scores with the master scores. This training repeated itself until both scorers achieved high agreement rate with the master scores. A total of 10 samples of student test scripts (previously scored) for each level were used at this stage.

Stage 3
At the final stage, each pair proceeded to do independent coding. Copies of unscored test scripts were given out and each coder was to score these independently. Later the scores for each student are checked for inter scorer agreement and disagreement in scores are resolved and marked as ‘agreement by consensus’. The final scores agreed by both scores were to be the final scores for that student. A total of 12 student test scripts for each level were used at this training stage.
Independent Coding

When the two coders of each pair have reached a high inter scorer agreement with each other, the remaining student test scripts are distributed between the two evenly. Each coder then scores her share of the student test responses independently. At this stage, a sample of 10 scripts was to be coded by both scorers independently. These ten scripts are then used for inter scorer agreement checks. Ideally, a inter scorer agreement rate of 90% should be achieved at this stage.

Updated Agreement Checking

As noted above, for OER 40-12 RS, agreement was confirmed using Cohen’s Kappa with .70 as ‘good’ and .80 as ‘excellent’ agreement. Coding was checked for agreement overall and for agreement on individual test items.

Acknowledgements

Our thanks to the Office of Educational Research, National Institute of Education, Singapore for funding on both projects. The procedures described here and the views expressed are those of the authors.

References


