

## Chapter 3

# **ANALYZING STUDENT RESPONSES IN MATHEMATICS USING TWO-DIGIT RUBRICS**

John A. Dossey, Chancey O. Jones, and Tami S. Martin

The Third International Mathematics and Science Study (TIMSS) was the first international assessment of student achievement in mathematics to make use of a large number of constructed response items in measuring and reporting student performance. The use of such items provides a great deal of data to teachers, curriculum developers, and researchers. Seeing student work allows individuals in each of these groups to examine more directly, rather than surmise about, evidence of students' misconceptions, error patterns, and problem-solving strategies. In addition to using a large number of constructed response items, TIMSS also forged new ground in employing two-digit rubrics for scoring student responses. Such scoring rubrics provide information on the correctness and methods students use to produce their work and on misconceptions noted in that work. In this chapter, we examine the patterns and findings associated with the use of such two-digit rubrics to score student responses in TIMSS.

### **TWO-DIGIT RUBRICS**

To maximize the amount of information that might be obtained from the use of constructed response items in the TIMSS mathematics assessment, two-digit rubrics were developed to score student responses. They were devised to retain as much information as possible about student work in the codes themselves. In particular, these rubrics were developed to permit scoring student responses for correctness and capture the analytical information embedded in student responses and to:

- Be clear, distinct, and readily interpretable, and based on empirical data... so as to account for the most common correct responses, typical errors, and misconceptions.
- Be capable of encoding the adequacy of an explanation, justification, or strategy as well as the frequency with which it is used.

- Be simple, in order to get high reliability and not to impose unreasonable time or resource burden.
- As far as possible, allow for nuances of language and idiosyncratic features of various countries, but avoid being so complex that coders are overwhelmed and tend to limit themselves to a few stereotypic codes.
- Have a number of codes that is not excessive, but sufficient to reduce coding ambiguity to a minimum (Lie, Taylor, & Harmon, 1996).

To achieve these goals, TIMSS used the same general two-digit rubric structure across all parts of the study. The ten's digit of the two-digit code represents the number of score points assigned to the response, based on its level of correctness. The unit's digit, when combined with the ten's digit, provides a code number to identify specific approaches or strategies by credit level, or, in the case of student difficulty with the item, common errors or misconceptions.

Figure 1 displays item V2 from the Population 2 assessment. This item, with the same rubric also was used as part of the Population 3 Mathematical Literacy Assessment as item A12. The rubric was developed through analyzing student field trial responses and in the actual administration of the TIMSS assessment. Student performance on this item for students in Population 2 and the Population 3 Literacy study will be reviewed later in the chapter.

In general, the structure of the two-digit rubric might be conceptualized as shown in Table 1. Variations of this general scheme were developed for problems worth 3, 2, and 1 points (Lie, Taylor, & Harmon, 1996). The overall aim was to develop a means of giving students credit for partially correct solution; and, at the same time, develop documentation of the kinds of approaches used by students from different countries. The latter kind of information could be invaluable to curriculum developers and teachers wishing to develop instructional materials that take into account the kinds of errors made by students on such items.

Because of the immense scope of the coding effort required to score the constructed response items across the countries participating in TIMSS, training sessions were designed and delivered at an international level. A series of 4-day training sessions were developed to train representatives of national centers who, in turn, had the responsibility for training personnel in their respective countries to reliably apply the two-digit codes. A total of 10 of these training sessions were held around the globe to prepare country-level teams to train their national coders (Mullis, Jones, & Garden, 1996). While there were concerns about whether this process would work, the final reliability results concerning the coding indicated that it did. An analysis of random samples of scored work indicated that the actual achieved average reliability coefficients ranging from 0.95 to 0.97 for the first-digit correctness codes and from 0.87 to 0.93 for the second-digit diagnostic codes (Mullis et al., 1997; Beaton et al., 1996; Mullis et al., 1998).

V2. The following two advertisements appeared in a newspaper in a country where the units of currency are *zeds*.

**BUILDING A**  
Office space available  
  
85 - 95 square meters  
475 *zeds* per month  
  
100 - 120 square meters  
800 *zeds* per month

**BUILDING B**  
Office space available  
  
35 - 260 square meters  
90 *zeds* per square meter  
per year

If a company is interested in renting an office of 110 square meters in that country for a year, at which office building, A or B, should they rent the office in order to get the lower price? Show your work.

Note: There is no distinction made between responses with and without units.

Code	Response
Correct Response	
30	Building A. Correct calculation of rents for both buildings. 9600 yearly/800 monthly and 9900 yearly/825 monthly, OR 825 to compare with 800 given.
39	Other correct.
Partial Response	
20	Building A. Correct calculation of rent for Building A OR B but not both.
21	Building B OR building is not named. Correct calculation of rents for both buildings.
Minimal Response	
10	Building A. Calculations or explanation are incorrect or inadequate.
11	Building A. No work shown.
12	Building B, OR building is not named. Correct calculation of rent for Building A OR B but not both.
16	Building A. Explanation is given only in the form of extracts from the advertisements.
19	Other minimal.
Incorrect Response	
70	Building B. Incorrect or inadequate calculations.
71	Building B. No work shown.
79	Other incorrect.
Nonresponse	
90	Crossed out/erased, illegible, or impossible to interpret.
99	BLANK

Figure 1. Population 2 mathematics item V2 and rubric.

Secondary Analysis of the TIMSS Data

Robitaille, D.F.; Beaton, A.E. (Eds.)

2002, XII, 429 p., Softcover

ISBN: 978-1-4020-0860-3