



TIMSS went to great lengths to ensure that comparisons of student achievement across countries would be as fair and equitable as possible. The *TIMSS 2011 Assessment Frameworks* were designed to specify the important aspects of science that participating countries agreed should be the focus of an international assessment of science achievement, and the assessment items were developed through a collaborative process with national representatives to faithfully represent the specifications in the frameworks and field tested extensively in participating countries. Finalizing the TIMSS 2011 assessments involved a series of reviews by representatives of the participating countries, experts in science, and testing specialists. At the end of this process, the National Research Coordinators (NRCs) from each country formally approved the TIMSS 2011 assessments, thus accepting them as being sufficiently fair to compare their students' science achievement with that of students from other countries.

Although the assessments were developed to represent an agreed-upon framework and were intended to have as much in common across countries as possible, it was unavoidable that the match between the TIMSS 2011 assessment (or test) and the science curriculum would not be the same in all countries. To restrict test items to just those topics included in the curricula of all participating countries and covered in the same sequence would severely limit test coverage and restrict the research questions that the study is designed to address. The tests, therefore, inevitably have some items measuring topics unfamiliar to some students in some countries.

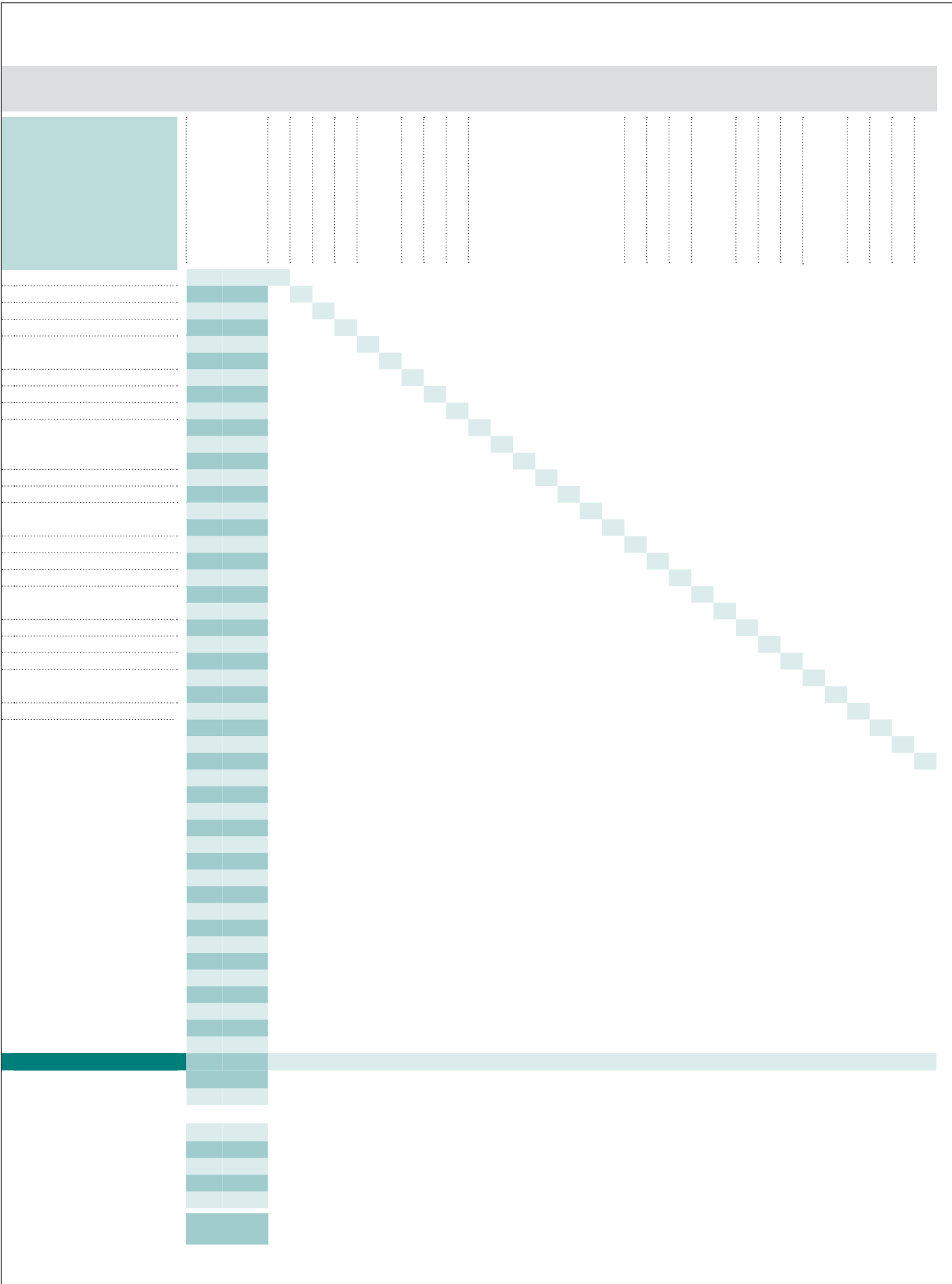
The Test-Curriculum Matching Analysis (TCMA) was conducted to investigate the extent to which the TIMSS 2011 science assessment was relevant to each country's curriculum. The TCMA also investigates the impact on a country's performance of including only achievement items that were judged to be relevant to its own curriculum.¹

To gather data about the extent to which the TIMSS 2011 tests were relevant to the curricula of the TIMSS countries and benchmarking participants, NRCs were asked to examine each achievement item and indicate whether the item was in their country's intended curriculum at the grade tested (fourth or eighth grade). The NRCs were asked to choose persons very familiar with the curriculum at these grades to make this determination. In some countries, the curriculum was prescribed for a range of grades and was not explicit about what was to be covered by the end of the fourth or eighth grades. For example, in Sweden the curriculum specifies the curricular goals to be achieved by the end of the eighth and ninth grades, but does not provide a grade-by-grade specification.

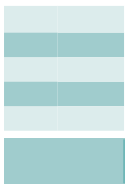
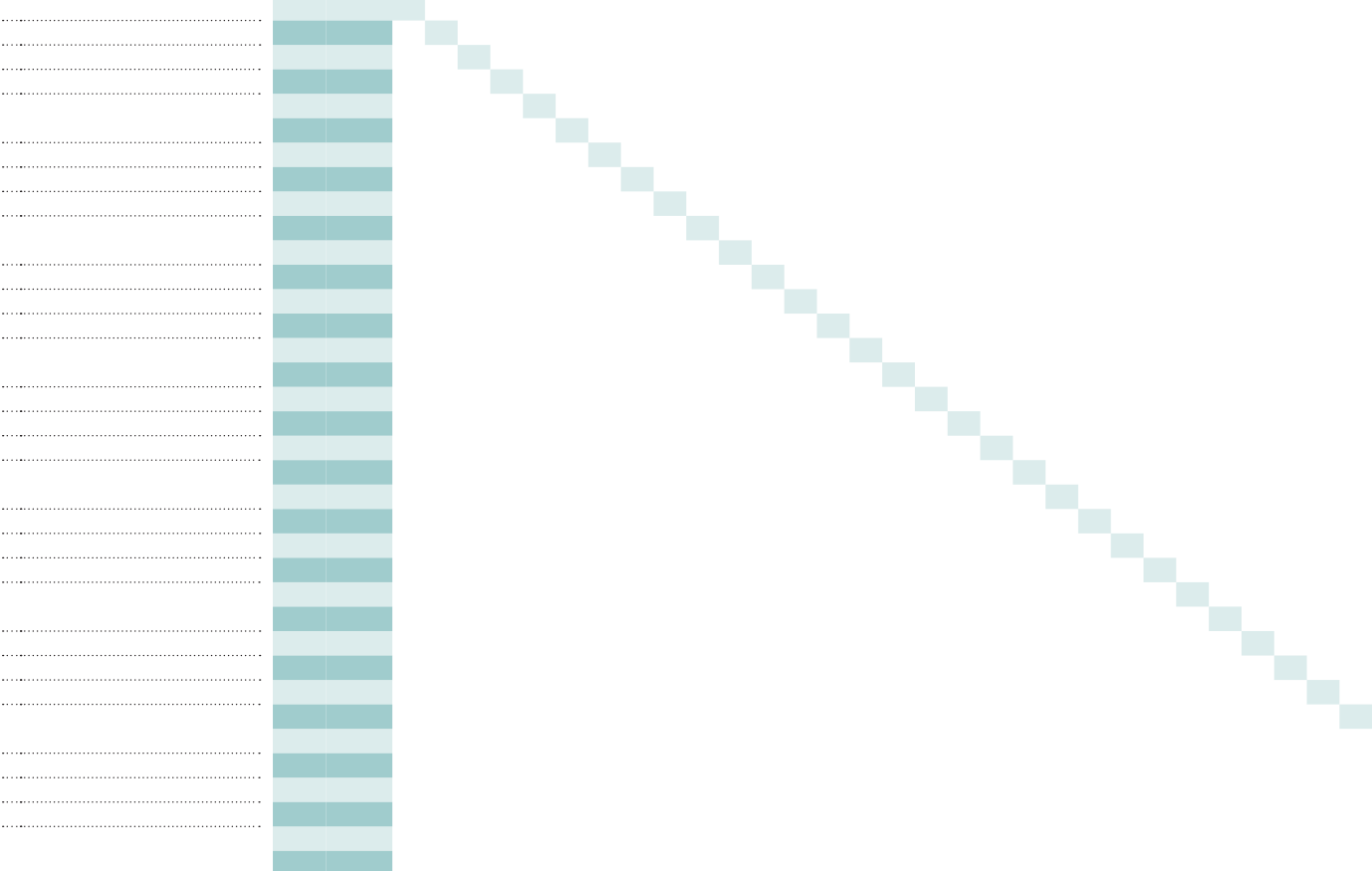
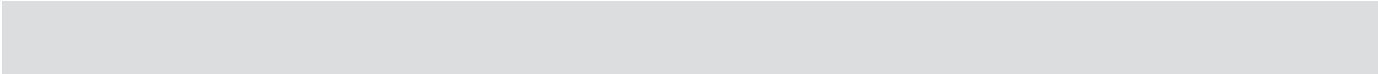
In such situations, coordinators were asked to make the best judgment possible.² Since an item might be in the curriculum for some but not all students in a country, coordinators were asked to consider an item included if it was in the intended curriculum for more than 50 percent of the students. All TIMSS 2011 participants took part in the TCMA analysis except Bahrain, Georgia, Saudi Arabia, Honduras (sixth grade participant), and the US benchmarking states at the fourth grade, and Bahrain, Georgia, Ghana, Indonesia, Saudi Arabia, Syrian Arab Republic, Honduras (ninth grade participant), and the US benchmarking states at the eighth grade.

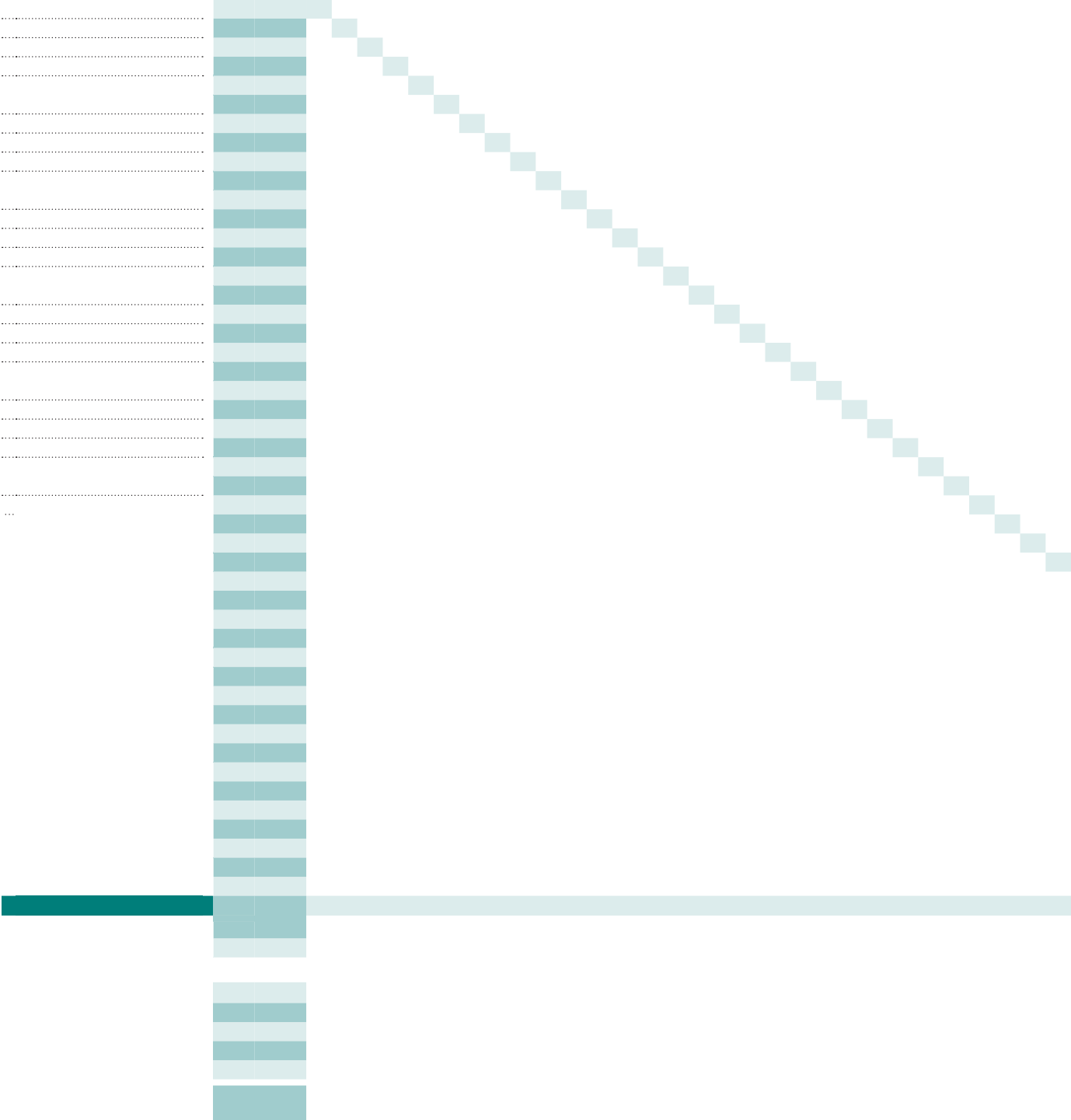
Exhibits F.1 through F.4 present the TCMA results for the TIMSS 2011 science test at the fourth and eighth grades. Exhibits F.1 and F.2 show the average percent correct on the science items judged appropriate by each country at the fourth and eighth grades, respectively. Exhibits F.3 and F.4 show the standard errors corresponding to the percentages presented in Exhibits F.1 and F.2.

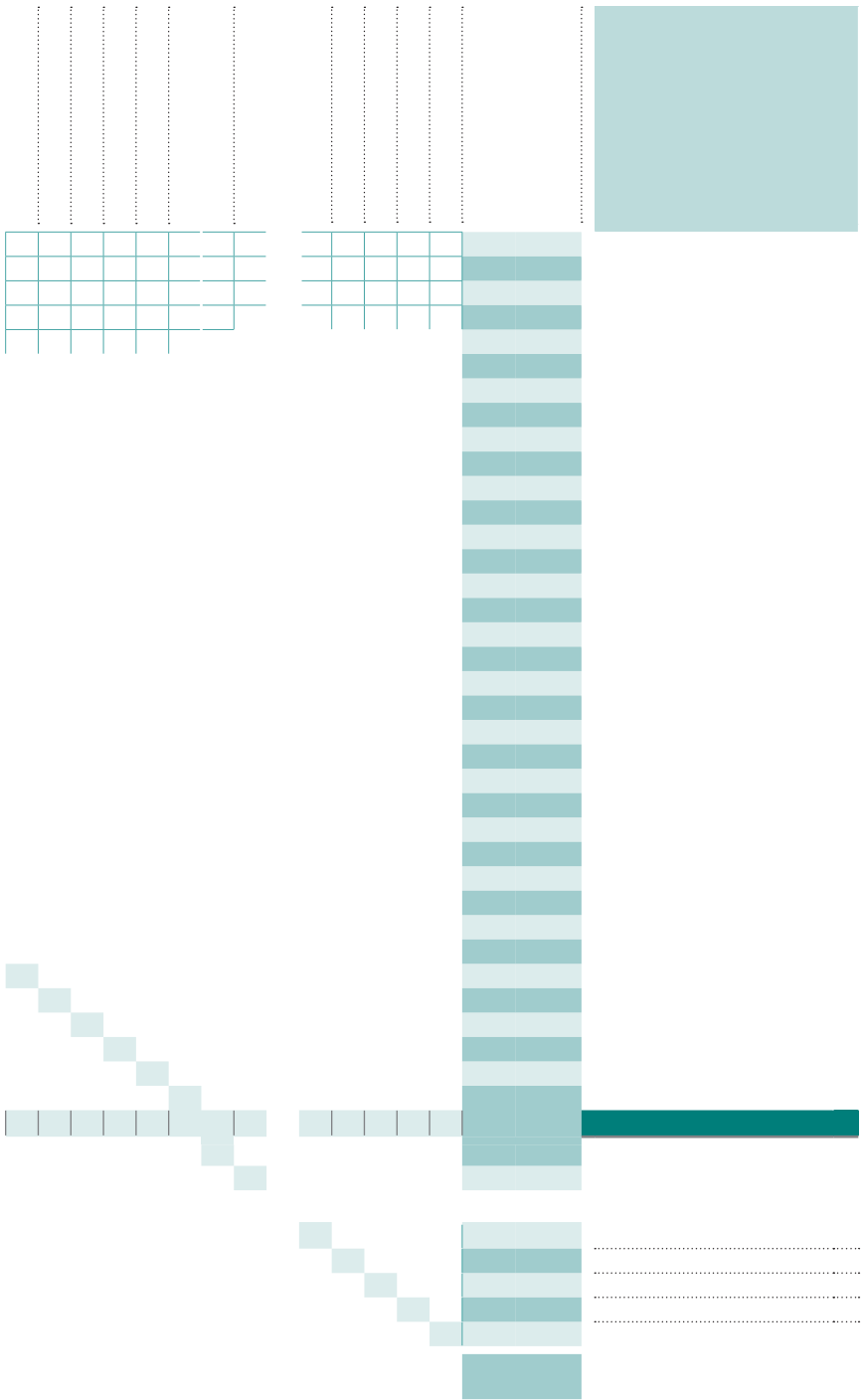
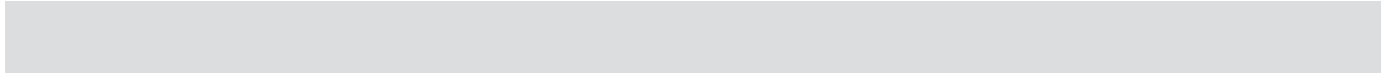
In Exhibit F.1, the bottom row of the exhibit shows the number of items, in terms of score points, identified as appropriate in each country. At the fourth grade, the maximum number of score points in the assessment was 181 points.³ Reading along the bottom row, it can be seen that only eight participants—Singapore, Korea, Japan, Chinese Taipei, the Russian Federation, Chile, Tunisia, and Yemen—judged less than half of the synd











development. The fact that the majority of countries indicated that most items

