

DTAMS Development Process Summary

Identifying Content Breadth & Depth

Teams of researchers analyzed a number of standards documents and research literature to synthesize the science content (see “*Content Subscore Topics*” link) middle school teachers should know. See “*Middle Science Content Chart*” for the summary of content identified in various sources. Four types of knowledge (see “*Types of Knowledge Subscore Details*” link) were also identified. This provided a 2-dimensional chart within which questions were generated to ensure both *breadth of coverage* (content) and *depth of coverage* (knowledge type).

Generating Initial Items

Teams of practicing science teachers, science teacher educators, and scientists (between 6-8 on a team) generated initial test items intended to simultaneously target a particular content subtopic and a particular knowledge type. Assessment-wide, items were targeted to be balanced across both dimensions. See “*Item Specification Grids*” link for the map of specific items onto the 2-dimensional chart of content breadth and type of knowledge depth.

Establishing Initial Construct Validity

Evidence of validity of the items for measuring teacher content knowledge in the various categories was established by asking external reviewers to review the items. Items were edited and sorted into randomized sets. They were sent to reviewers along with a review form that solicited: 1) the correct answer to the multiple choice items; 2) categorization of each item into a content category and subcategory; 3) categorization of each item into a knowledge type category; 4) a rating of the item as STS or not; and 5) a rating of the appropriateness of the item for middle school teachers.

Reviewers for each content assessment included scientists, science educators, and science teachers. Each item was reviewed by 27-31 reviewers in life science, 29-33 reviewers in physical science, and 20-22 reviewers in earth science. Each person reviewed about 75 items.

Data from the reviewers were analyzed to identify items that met criteria the DTAMS staff established for measuring the assigned constructs. The criteria for an item receiving verification as fitting a content category was that at least 75% of the reviewers identified the item as assessing a given category. To guarantee a balanced distribution within each category, subcategories within the categories had to be agreed upon by more than 50% of the reviewers; both of these criteria were required for an item to be accepted on the content category criterion. The knowledge type criterion was considered acceptable if more than 50% of reviewers rated the item as belonging to one type. For appropriateness, items that received an average ranking over 2.4 (on a scale of 1=low, 2=medium, 3=high) were considered appropriate. If an item met all three of those criteria, it was accepted to be included in the field tests. If it met two of the criteria, it was reviewed to determine if the wording could be clarified or improved.

Revised items were or will be sent out for a second review. The items that met review criteria were selected to be the prototype for items in the field tests.

Field Testing

Based on reviewer feedback, questions were selected, revised, and assembled into field tests. Parallel questions were generated to produce 6 versions of each content-area field test. Tests are designed to be completed by test-takers within an hour. Field test results were used to screen for items needing further revision or replacement.