

SMSO

Survey of Mathematics & Science Opportunities

CURRICULUM ANALYSIS
TECHNICAL REPORT SERIES
No. 2

DOCUMENT ANALYSIS: DATA COLLECTION AND PROCESSING.

June 14, 1995

Project Director:
William Schmidt

MICHIGAN STATE
University

DOCUMENT ANALYSIS: DATA COLLECTION AND PROCESSING PROCEDURES

Summary

- The Document Analysis is one of five procedures that make up the TIMSS Curriculum Analysis.
- Data were collected from textbooks and curriculum guides for the upper grades in TIMSS Populations 1, 2 and 3 irregardless of the populations for which the achievement tests will be administered in each country.
- The textbook and curriculum guide sample is intended to represent material that no less than 1/2 to 2/3 of students in each TIMSS population are likely to have in their classes.
- The procedures for Document Analysis - Data Collection were:
 - Selection of textbooks and curriculum guides
 - Delimitation and identification of units within the documents
 - Delimitation and identification of blocks within the units
 - Coding of identified blocks.
- Paper forms from the data collection were entered and processed in electronic form.
- Data processing for the first international reports included:
 - Cleaning
 - Routines for counting occurrence of codes and generation of percentages within documents.

- Routines to define a representative country document when more than one is available
- Routines to take document percentages and generate analysis tables.

Overview

The Document Analysis (DA) of textbooks and curriculum guides is one of five data collections that make up the TIMSS curriculum analysis.¹ This data collection is intended to provide detailed information on topics (contents), student performance expectations, and disciplinary perspectives intended at each one of the TIMSS focal populations.

The empirical data generated from the DA will yield comprehensive inventories of these features of textbooks and curriculum guides:

TEXTBOOKS	CURRICULUM GUIDES
<ul style="list-style-type: none"> • Intended contents (topics) <ul style="list-style-type: none"> • expectations for student performance • disciplinary perspectives • instructional formats • pedagogical approaches • use of practical activities/experiments 	<ul style="list-style-type: none"> • Intended contents (topics) <ul style="list-style-type: none"> • expectations for student performance • assessment requirements/suggestions • content and performance policies • pedagogical suggestions • disciplinary perspectives

The Document Sample

¹ Each data collection is the subject of a separate report in this series. An overview of each procedure and its function in the composite portrayal of curricula is in the first report of this series.

Before embarking on DA, countries had to determine which textbooks and curriculum guides would be analyzed. Representatives from each country collaborated on this decision during curriculum analysis training.

Countries were required to fulfill the following selection criteria for each of TIMSS populations 1, 2 and 3 (physics and advanced mathematics only):

- identify the upper of the two grades in their designed sample.
- select the national textbooks and curriculum guides for that grade, if these exist.
- if they do not select regional curriculum guides or syllabi if these exist.
- select the most widely used textbooks.
- collectively, the document sample must represent material used by 1/2 to 1/3 of students in identified grades.

Written agreements on document samples were drawn up at regional training sessions. Deviations from original samples due to changes in curricular policy, etc.--have been documented through two iterations of data review.

The Process of Data Collection

In this section we provide a very brief overview of the process of DA data collection. Coders were provided with 2 1/2 days of training and over 200 pages of documentation to master these techniques.²

1. Identifying units.

The first step in the data collection procedure is to divide a document into units for analysis (henceforth referred to as “Units”).

- For Textbooks: The unit is a lesson. A lesson is understood to be a segment of text material that typically covers a single main subtopic. A lesson is understood to be material covered in one to three periods of instruction.

- For Curriculum Guides: Defined as the smallest segments of the document identified by examination of the table of contents and/or major subheadings in the text.

2. Assigning Unit Types.

The second step is to identify the type of unit. This forms a part of the data on major features of the document.

- For Textbooks: Unit types are
 1. Introduction
 2. Lesson
 3. Multiple Lesson Pages
 4. Instructional Appendix
 5. Other

- For Curriculum Guides: Unit types are

² Training materials relevant to these techniques are in SMSO Report Series 42, the regional training

- | | |
|-------------------|------------------|
| 0. Introduction | 3. Content Unit |
| 1. Policy Unit | 4. Pedagogy Unit |
| 2. Objective Unit | 5. Other |

Unit information is registered on Form DA-2. See Figure 1 for example of this form for a sample unit of Japanese textbook material.

meeting training manual, and the Training Manual Supplement.

Unit Documentation Form DA-2

Use this form as the cover sheet for one unit. Attach all Form DA-3's for the unit to it.

Country: JAPAN Your name: Eizo NAGASAKI/Hanako SENUMA

Current Date: 30th August, 1993 Your title: Chief/Senior Researcher

Document ID Code: M181C (Copy the Document ID Code from form DA-1 or DA-4)

Curriculum Guide Textbook Science Mathematics

Unit ID Number 5 (Number the units sequentially within the document)

1. Title of unit as given in document: 単項式と多項式 / Monomials and polynomials

Indicate the type of unit:

2. For guideline, or	3. For textbooks
<input type="checkbox"/> 0. Introduction to guide	<input type="checkbox"/> 1. Introduction
<input type="checkbox"/> 1. Policy units	<input checked="" type="checkbox"/> 2. Lesson
<input type="checkbox"/> 2. Objective units	<input type="checkbox"/> 3. Multiple-lesson pages
<input type="checkbox"/> 3. Content units	<input type="checkbox"/> 4. Instructional appendix
<input type="checkbox"/> 4. Pedagogy units	<input type="checkbox"/> 5. Other:
<input type="checkbox"/> 5. Other units:	

4. Page range in unit is from page 8 to 10.

5. Is this unit primarily

<input type="checkbox"/> Exposition and narrative	6. Is this unit primarily presented
<input type="checkbox"/> Activity and exercise	<input type="checkbox"/> Concretely and pictorially
<input checked="" type="checkbox"/> Other (Please describe)	<input checked="" type="checkbox"/> By text and symbols
<u>Exposition and exercise</u>	<input type="checkbox"/> Other (Please describe)

7. Is this the first time that the content of this unit is covered in your country's curriculum?

<input checked="" type="checkbox"/> Yes	8. Will the content of this unit be covered again in your country's curriculum?
<input type="checkbox"/> No	<input type="checkbox"/> Yes
<input type="checkbox"/> I do not know	<input checked="" type="checkbox"/> No
	<input type="checkbox"/> I do not know

9. Have there been any recent changes in the official curriculum for the topic of this unit?

<input type="checkbox"/> Yes	10. Are any changes planned in the near future in the official curriculum for the topic of this unit?
<input checked="" type="checkbox"/> No	<input type="checkbox"/> Yes
<input type="checkbox"/> I do not know	<input checked="" type="checkbox"/> No
	<input type="checkbox"/> I do not know

If yes, please describe:

Figure 1

3. Dividing Units into Blocks and assigning block types.

- Units are further subdivided into segments called Blocks, these blocks are assigned types. Data from block types provide additional information on important features of the documents.
 - For Textbooks, block types are:
 1. Narrative
 2. Related Narrative
 3. Unrelated instructional narrative
 4. Related Graphic
 5. Unrelated Graphic
 6. Exercise/Question set
 7. Unrelated exercise/question set
 8. Activity
 9. Worked Example
 10. Other
 - For Curriculum Guides, block types are:
 1. Official policies
 2. Objectives
 3. Content elements
 4. Pedagogical suggestions
 5. Examples
 6. Assessment suggestions
 7. Other
- Country coders mark units and their blocks directly on document pages, which form part of the data submitted to the Curriculum Analysis Project. See Figures 2a and 2b for a specimen corresponding to the same Japanese unit in Figure 1. Figures 2a and 2b are for only the first two pages of this unit. These pages represent the first seven blocks of a total of fifteen for this unit.

1 式の計算

①



1 単項式と多項式

②

Q 次のような数の遊びがある。

- ① 適当な数を考えてください。
- ② 考えた数を2倍してください。
- ③ ②の結果に10をたしてください。
- ④ ③の結果から、最初に考えた数より5だけ大きい数をひいてください。
- ⑤ 最後に、④の結果から、最初に考えた数より2だけ小さい数をひいてください。

③



15

- ① 適当な数を考え、上の順序で計算し、最後の結果を右の の中に書き入れてみよう。
- ② 9ページの右下の ? の部分を折り返してみよう。
- ③ このように、結果を先まわりしてあてることができるわけを考えてみよう。

Figure 2a

④ 単項式と多項式

式のなかには、 $2x$ のように乗法だけでつくられているものと、 $2x+10$ のように加法と乗法でつくられているものがある。

$2x$, $\frac{1}{3}a^2$ などのように、数や文字についての乗法だけでつくられた式を**単項式**という。1つの文字や1つの数、たとえば、 x や -5 なども単項式と考える。

また、 $2x+10$, $3a^2+4ab+1$ などのように、単項式の和の形で表された式を**多項式**といい、そのひとつひとつの単項式を多項式の**項**という。たとえば、 $3a^2+4ab+1$ の項は

$$3a^2, \quad 4ab, \quad 1$$

である。

⑤ **例1** $3x^2-2x-5$ は

$$3x^2+(-2x)+(-5)$$

と書けるから、多項式であり、その項は

$$3x^2, \quad -2x, \quad -5$$

⑥

$$3x^2 + (-2x) + (-5)$$

└─── 項 ───┘

⑦ **問1** 次の多項式の項をいえ。

① $4a+3b$

② $2x+y-3$

③ $\frac{1}{2}x-y^2-\frac{1}{3}$

?

Figure 2b

4. Coding instructional material in each block.
- The instructional material in each block is described using categories from the three aspects of the TIMSS curriculum Frameworks. That is, country coders identify each block's topic(s), expectations(s) for student performance, and disciplinary perspective(s).
 - Once the coders have identified the Framework categories that best describe the block, they record the Framework numbers used to denote these categories on a paper form (Form DA-3). This process is referred to as Coding.
 - Figures 3a and 3b shows a specimen Form DA-3 corresponding to all fifteen blocks of the Japanese unit 5 in Figure 1 and Figures 2a and 2b.

Data Processing

All data forms are sent to the TIMSS Curriculum Analysis Project, along with the original documents with Units and Blocks clearly marked on them.

Original documents were archived and forms underwent initial evaluation to ascertain whether proper procedures were followed. After these checks, data are forwarded for data entry.

- All data, including full-text and framework coding data are entered in cross-referenced electronic data files.
- After entry, data undergo computer verifications to identify any irregularities. When identified, such irregularities are cleared by referring to original forms, archived textbooks or curriculum guides, or by querying country NRCs.
- Full-text, DA-2 and DA-3 coding data files are processed separately. They are linked in the Curriculum Data Bank through the unique document, unit and block identifications. These identifications also link the electronic data files with the document (textbook and curriculum guide) archive and with the archived paper forms.
- The major portion of the data is comprised by the DA-3 data files. These are also the most central to the first international report of the Curriculum Analysis.

- The DA-3 coding files contain one data line per block, with the following structure:
 - a. Country ID number (ISO standard)
 - b. Document ID
 - c. Unit ID
 - d. Block ID
 - e. Page number(s)
 - f. Related block information
 - g. Block type
 - h. Content code(s)
 - I. Performance Expectation code(s)
 - j. Perspectives code(s)
- A specimen of DA-3 electronic data can be observed in Figure 4. This is a portion of the data file for Japanese document M181C. The portion enclosed in a box corresponds to the blocks belonging to unit 5, which is the same unit used in Figure 1 through 3.

Generating Analysis Files

Data undergo a series of transformations in order to generate analysis files. Detailed treatment of these are found in the third report in this series. Following, a list of major steps in this procedure.

- Depending upon the analysis, appropriate document type (curriculum guide or textbook) and unit type(s) (e.g. Lesson Units) are selected from the master data file.
- Within each selected unit, if appropriate, blocks of a particular desired type are selected, e.g. exercise question set.
- The framework code (for topic, performance expectation, perspective or combinations of these) for each block is noted.
- For each framework code (or combination of codes) a count or binary observations (present or absent) is kept which indicates the number of blocks in the document that contain that content code or whether the content code is present in the document.
- These counts can be done for all block types or for particular block types such as narrative.
- These code counts are converted into percentages either by dividing by all selected block types or, where appropriate, by dividing against specific block types.
- The resulting percentages represent the proportion of blocks of a certain type (within certain unit types) that are coded with a given framework code.

- For the data represented in Figure 4 the following are the resulting percentages treating the 34 blocks as a complete document for illustration purposes

-- topic codes

$$1.6.2 \quad 34/34 = 100\%$$

-- performance expectation code

$$2.3.1 \quad 2/34 = 5.88\%$$

$$2.1.3 \quad 21/34 = 61.76\%$$

$$2.2.2 \quad 11/34 = 32.35\%$$

-- perspective codes

$$3.1 \quad 2/34 = 5.88\%$$

$$3.4 \quad 2/34 = 5.88\%$$

-- combination codes

$$1.6.2 \times 2.3.1 = 2/34 = 5.88\%$$

$$1.6.2 \times 2.1.3 = 21/34 = 61.76\%$$

$$1.6.2 \times 2.2.2 = 11/34 = 32.35\%$$

- These percentages constitute the sufficient statistics that make up the analysis files for the first international report of the TIMSS curriculum analysis.
- The process is straightforward when blocks with only single topic, performance expectations or perspective codes are involved. The more complicated process resulting from complex signatures is described in Report No. 3 in this series.
- The sufficient statistics described are applicable to a single document -- how they are combined to form a country level “representative document” is described in Report No. 4.