A TEI Model for TIMSS and PISA Assessments

Eva Forsbom
Department of
Linguistics and Philology
Uppsala University
evafo@stp.lingfil.uu.se

Ebba Gustavii
Department of
Linguistics and Philology
Uppsala University
ebbag@stp.lingfil.uu.se

Markus Saers
Department of
Linguistics and Philology
Uppsala University
markuss@stp.lingfil.uu.se

1 Introduction

The current research is part of a larger project[1], where we aim to study what kind of knowledge, linguistic or subject matter, is really measured in international assessments of students' knowledge of mathematics and natural science, e.g. TIMSS[2] and PISA[3]. The main purpose of the study is to measure readability of questions in the Swedish part of TIMMS 2003 and PISA 2003/2006, in connection with the actual test results, to see what effect language has for the results.

As a first step, we have defined a genre model for assessments, i.e. a model of the logical structure of assessments, in order to extract the information that is relevant for each question. Linguistic annotations to use for readability measures are under way.

2 Data

TIMSS and PISA are two international assessment frameworks, both including tests of mathematics and science. In addition to the actual tests, students, their teachers, and their school principals complete questionnaires about the contexts for learning mathematics and science.

TIMSS is a project of the IEA[4]. IEA has been conducting studies of cross-national achievement since 1959. TIMSS 2003 is the second most recent in a regular cycle of studies to measure trends in 4th- and 8th-grade students' mathematics and science achievement. The most recent study is TIMSS 2007.

PISA is an OECD[5] project, testing the reading, mathematical and scientific literacy of 15-year-old students and the performance of education systems. The first survey took place in 2000, the second in 2003 and the third in 2006.

3 Method

Assessment data for both TIMSS and PISA questions were already structured in databases, while the actual questions were only available in PDF and Word files. Text parts and graphics were automatically extracted, and semi-automatically annotated in TEI[6] format.

We adapted the generic TEI P5 annotation model to suit our needs. In particular, we defined a set of divisions representing the logical structure of the assessments and questions, and a number of relationships between various question components. The logical components can be visualised as boxes through a web application (cf. Figure 1).

One major objective was that the logical structure should as far as possible reflect our need to extract and classify the information to analyse for every question. But, it should also be straightforward to annotate the logical structure, and logical components should therefore optimally be supported by graphical signals.

A preliminary model was defined for TIMSS, slightly redefined for PISA, and finally applied to both TIMSS and PISA. The major difference con-

[1] "Text and Language in Assessment of Mathematics and Science", funded by The Swedish Research Council, in cooperation with the Dep. of Educational Measurement, Umeå University, and the Dep. of Curriculum Studies, Uppsala University. http://stp.lingfil.uu.se/timss/.

[2] Trends in International Mathematics and Science Study. http://timss.bc.edu/timss2003.html.

[3] Programme for International Student Assessment. http://www.pisa.oecd.org/.

[4] International Association for the Evaluation of Educational Achievement.

[5] Organisation for Economic Co-operation and Development.

[6] Text Encoding Initiative. http://www.tei-c.org/.

cerns grouped questions and the way sub questions share information.

4 The assessment genre model

At a top level, both the TIMSS and PISA material can be classified into three types of questions:

Simple questions - where there is only one item.
Multipart questions - where there are two or more sub items (a, b and c questions).
Grouped questions - where there are several simple or multipart questions grouped around a common topic.

In our model, all types have the same basic overall structure, the **question set**. The question set contains one or more **item**s. The item is the statistical unit in all assessments. Each item has a unique ID, which is connected to an item-specific assessment guide, and to all statistical variables collected for that item.

A question set can also have **meta information**, i.e. information on the testing context or the test itself, such as *End of Metal Crown section.* In addition, it can contain **figure**s, **table**s, and **leading text**, i.e. sections that put the actual question into context or contain information needed to solve the question, but not directly states what the student should answer or do. A question set can furthermore be recursive. Multipart and grouped questions, for example, are represented in this way.

The item, in turn, contains exactly one **core question**, and can have zero or more **prompt** types. The core question is the central question, stating what the student should answer or do. The prompt extends the core question, and prompts the student to react in a certain way, such as formulating an answer or filling in a table. We have specified the following prompt types:

Answer line - normally a solid line where students should write their answer, and often preceded by a word (*Answer:*).
Multiple choice - a list of options to choose from.
Order - a list of statements to be ordered.
Figure - a figure to draw something in.
Table - a table to be completed.
Specification - a more detailed instruction on how to respond, e.g. *Motivate your answer.*.
True or false - a set of statements to be answered with true or false (or yes/no).

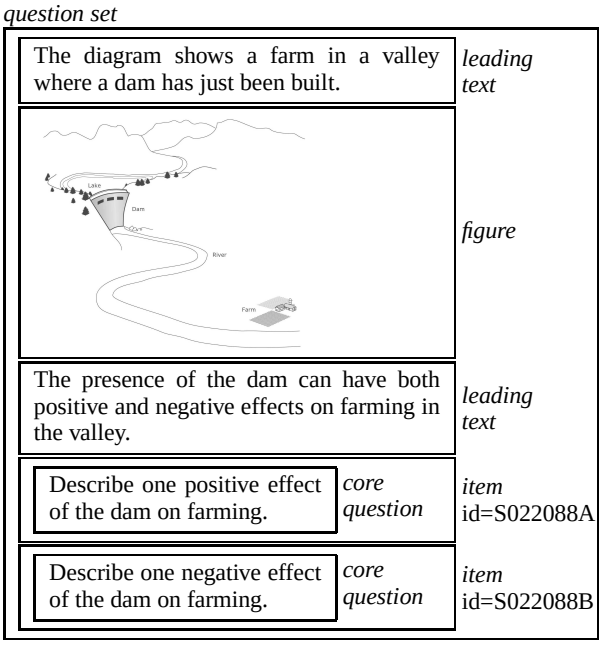

Figure 1: Boxed version of multipart question.

The item can also include any leading text, figures, and tables that are relevant for that item only.

Simple questions, therefore, have all their parts within the item, while some multipart and grouped questions have shared information outside of the items, not all of which is relevant for each single item. In some cases, particularly in TIMSS, the answer to one item is required to understand or answer a following item. In those cases, we use a **pointer** within the second item to link that information to the previous item.

Although our focus is on textual information, graphical elements, such as figures and tables, are often essential for the understanding of a question, and sometimes contain a fair amount of textual information. As a first step towards integrating textual and graphical information, we have also classified the type of relations between graphical elements and items.

5 Concluding remarks

The presented genre model of assessments will facilitate our readability research, by making it possible to extract for each single (sub)question the exact amount of information (linguistic or other elements) that have to be processed in order to answer the question. It also makes it possible to study individual types of elements closer.

The readability research in itself will focus on vocabulary usage and the amount of information packaging.