

Identification and analysis of text-structure and wording in TIMSS-items

Peter Nyström, Umeå University, peter.nystrom@edmeas.umu.se

Abstract

The extent to which students are able to understand the text in an item is important in order for them to show what they know and can do. In this study a word-recognition test is used to identify items where verbal skills have a significant contribution to the probability of getting the answer right. The items identified were analyzed in order to find critical features in the item texts. The analysis resulted in several categories pertaining to text structure and word complexity.

Keywords: *secondary analysis, item analysis, text, TIMSS*

Introduction

Construct-irrelevant variance is a threat to interpretation and use of results from all evaluations of educational achievement (Messick, 1989). One area identified as a potential (or even probable) source of this kind of validity threat, is readability of items designed and used for assessment. The dominant mode of presenting tasks for students in assessment situations is through the written language, and this is certainly true for international comparative studies. This makes assessments dependent on the written word and on reading comprehension. However, tests which are intended to measure achievement in mathematics and science should not measure reading. Differences in students reading ability should not affect the outcome of the test and contribute to construct irrelevant variance.

Even though it seems obvious that an assessment of achievement in mathematics and science should not assess reading comprehension, the relation between reading and knowing mathematics and science is complex. A modern interpretation of what it means to know mathematics and science includes language and communication as competencies within these subjects. It is a matter of judgment to evaluate the existence of improper or irrelevant language influence on achievement in mathematics and science.

One example of the complex relationship between language and subject matter is that words used in mathematics belong to three categories (Shorrocks-Taylor & Hargreaves, 1999): words that have the same meaning in mathematics as they have in everyday language, words that are used and spelled the same way in mathematics as in everyday language, but the

meaning of the words are different (e.g. *product*), and specific words which are not found outside mathematics. In order to read texts in mathematics (and science) it is necessary to be able to recognise which category words belong to in order to be able to interpret them correctly. Words with multiple meanings seem to cause difficulties for English first language and second language learners, who confuse the meanings used in different contexts (Wellington & Osborne, 2001, in Dempster & Reddy, 2007).

In her synthesis of research by applied linguists and mathematics educators Schleppegrell (2007) highlights the linguistic challenges of mathematics, including the multi-semiotic formation of mathematics, its dense noun phrases that participate in relational processes, and the precise meanings of conjunctions and implicit logical relationships that link elements in mathematics discourse. Schleppegrell talks about the grammar of mathematics, and argues that mathematics is highly technical, with characteristic patterns of vocabulary and grammar. She exemplifies the technical vocabulary with the mathematics words *sum* and *fraction*, but she also gives examples of words that have particular meaning in mathematics, such as *place*, *borrow* and *product*.

Just knowing mathematical words such as more, less, and as many as, for example, is not enough; students also need to learn the language patterns associated with these words and how they construct concepts in mathematics (Schleppegrell, 2007, p. 143).

We can conclude that in addition to potential problems with individual words, text structure and complexity has been found to be important for the readability of tasks in mathematics and science. In their study of TIMSS 2003 focusing South African second language learners, Dempster & Reddy (2007) found that sentence complexity was negatively correlated with percent correct in the nine items where more than 40 % of the learners chose an incorrect answer. Furthermore, they identified 5 features characteristic of the items that are most difficult for African learners to answer: passive voice, using when as a logical connective rather than to start a question, containing many qualifiers including a variety of prepositions and embedded clauses, nominalisation, words with dual meanings or specific terminology.

Many of the linguistic features identified as specific to mathematics and science (see e.g. Lemke, 1990; and O'Halloran, 2000), have been shown to be reading obstacles in other studies. Some of these features are subject specific vocabulary, terminology, grammar, and text structure. In research on texts and reading comprehension there are primarily four main characteristics that are relevant to consider (Liberg, Folkeryd, af Geijerstam, & Edling, 2002). The first characteristic is the content, what the text is about. The second characteristic is the complexity of the text. The third characteristic concerns how the text is built using different forms of connective markers and genre patterns, to make it possible for the reader to build a

cohesive, coherent understanding of the text. The fourth characteristic is the degree of involvement and active reader participation that the text invites to. A fifth characteristic might be the use of other forms of communication, pictures, tables and diagrams (Kress, 2003).

Our research builds on established theories and results concerning readability. However, even though readability is an important aspect to consider in the construction of problems, and in analysis of how students solve the problems, readability in these kinds of texts (tasks and items) does not seem to be possible to capture using the quantitative measures building on standard-procedures, particularly in mathematics (Homan, Hewitt, & Linder, 1994; Wiest, 2003). Systematic and research-based knowledge about text-features influencing readability in mathematics and science tasks is rare in the research literature. Despite the accepted fact that language skills, and in particular reading comprehension, is an important aspect of all assessment situations, surprisingly few studies have dealt with the significance of subject-specific discourse in written tests.

This lack of research might be a reason for the modest attention given to language in assessment tasks in some of the most influential handbooks for the development of assessments. In "Standards for educational and psychological testing" (American Educational Research Association, American Psychological Association, & National Council on Measurement in Education, 1999) the issue of language in tasks is only mentioned explicitly in connection to bilingualism, and briefly with respect to fairness: "In testing applications where the level of linguistic or reading ability is not part of the construct of interest, the linguistic or reading demands of the test should be kept to the minimum necessary for the valid assessment of the intended construct." (p. 82). The American "Code of fair testing practices in education" (Working Group of the Joint Committee on Testing Practices, 2005), a description of what it takes to guarantee fairness in assessment, which among other things is to guarantee that all test-takers are given equal opportunities to show what they know within the domain tested, has no explicit focus on the language aspects. However, as Gipps & Murphy (1994) have pointed out, language plays a critical role in the communication of meaning, and does so in many subtle ways. Language is therefore a factor that needs to be considered in the evaluation of fairness in assessment.

The study presented here is part of a project aiming at deepening the knowledge about aspects of language in assessment of what students know and can do in mathematics and science. The project is based on discourse- and text-analytic theory for the study of critical linguistic features in texts. Critical linguistic features refers to features which research has shown to have significance for readability, but also features which can emerge as potential problems from an analysis of student performance on assessment tasks. This study focuses the latter.

The purpose of this study is to identify TIMSS-items where student performance is influenced

by verbal ability. Furthermore, potential critical features concerning readability of these items are identified in an explorative manner. The study is financed by a grant from the Swedish Research Council.

Methodology

The object of analysis in this study is all of the mathematics and science items from TIMSS 1995, population 2 (13-year olds), i.e. 174 mathematics items and 142 science items. Each item was analyzed using an ordinal regression model with item score as the dependent variable, and TIMSS achievement measures (plausible values in mathematics or science) and scores on a word-recognition test as covariates. Items with a verbal covariate giving a significant contribution to explaining the item difficulty, while subject-specific achievement level is taken into account, were identified and subjected to further analysis. These items, in the following often denoted as *verbal items*, were subjected to an explorative analysis attempting at identifying critical features that could be plausible, initial explanations to the verbal influence found. Based on the emerging categories of verbal items, the other, non-verbal, items were also analyzed, aiming at investigating whether potential critical features were exclusively found in the verbal items, thereby validating the connection between verbal component and linguistic features.

The verbal component used in this study is based on results from a word-recognition instrument. This instrument was developed and added to the TIMSS 1995 cognitive instruments as a national option. The study reported here is the first study utilizing this data. The word-recognition instrument was developed from other existing Swedish instruments and consisted of 25 multiple-choice items. The instrument had a fairly high internal consistency, with a Cronbachs alpha of 0.87.

The study is primarily focusing a qualitative, explorative analysis of the verbal items found in the instruments of TIMSS 1995 in Sweden. In addition, a simple word count was performed, indicating number of words in the item stems as well as number of words in the longest alternative answer (for multiple choice items).

Results

In total, 45 mathematics items and 54 science items were identified as verbal items according to the definition presented earlier. For these items the verbal component measured by the word-recognition was significant in the multiple regression model ($p = 0.05$). The correlation between the result from the word-recognition test and TIMSS first plausible value was 0.632 for science and 0.607 for mathematics.

The results of the word count is presented in table 1.

[Table 1 in here]

Research on readability states that longer texts are not necessarily more difficult to read, and attempts to increase readability have often resulted in longer texts. The structure of the text is considered to be much more important for readability than the text length. From the results in table 1 we can conclude that in this study verbal items have more words on the average, both in the item stem and in the longest alternative answer of multiple choice questions.

The explorative evaluation of the verbal items resulted in two major categories, each with several sub-categories, and a few categories that are more tentative and only contain a very limited number of items. The majority of items identified as having a verbal component were possible to identify as belonging to these categories. Only 10-15 items were not categorized, and a plausible explanation for the verbal component of these items remains to be developed. In the presentation of each category, released items from TIMSS 1995 are used as examples. All of the examples shown have been downloaded from <http://timss.bc.edu/timss1995.html>.

In the first category, different kinds of complex text structures are identified as potential critical linguistic features. In 24 verbal items (9 in mathematics and 15 in science) a complex text structure was identified, caused by embedded clauses, nominalizations, use of the negative, long and complex alternative answers in multiple choice questions, etc. In particular four subcategories emerged; (1) general text complexity, (2) long and complex answers to choose from, (3) reading tasks in mathematics, and (4) use of the negative in questions.

Several items are characterized by a general text complexity, often due to embedded clauses and long texts. Figure 1 shows an item with a significant verbal component.

[Figure 1 in here]

In addition to general text complexity, a potential problem with the item in example 1 is that it poses two questions. Reading comprehension can influence the likelihood of actually doing both of the things required. Furthermore, the item contains words expressing relations (more than, less than) which will be discussed below.

Figure 2 shows a verbal item with a complex structure in the Swedish translation.

[Figure 2 in here]

In the translation from English to Swedish the first two sentences were almost literary translated. However, the third sentence was translated in a way that in English would be something like “How many laps does he run each day?” There is no mention of *laps* earlier in

the item, and this might have caused more problems for students with less developed reading skills. The item in Figure 2 can be compared to another item which was not identified as having a verbal component (Figure 3).

[Figure 3 in here]

The Swedish translation of the item in Figure 3 was consistent with the English original, using “laps” in both sentences.

Another verbal item where text complexity seems to have been introduced through translation is presented in Figure 4.

[Figure 4 in here]

If the Swedish translation of this item is translated back to English it would be something like:

One weighed a dolphin and then said that the weight, rounded to the nearest 10 kg, was 170 kg. Write down a weight that might have been the actual weight of the dolphin.

The passive form is kept, but vital information is presented using an imbedded clause, making the text more complex.

The second subcategory of complex texts is characterized by particularly long alternative answers in multiple choice items. Figure 5 presents an item where a verbal component seems influential.

[Figure 5 in here]

A third subcategory consists of what is here denoted as reading tasks in mathematics. Items belonging to this category seem to have a text complexity that is intentional, and an important part of the solution is to decode a dense and complex text. Figure 6 presents one of the totally four items found in this subcategory.

[Figure 6 in here]

Even though these reading tasks are problematic from a reading point of view, they can be defended if they capture a relevant aspect of the skills and competencies that are intended to be measured.

The final subcategory pertaining to text complexity captures the use of negatives, and two items in this subcategory are presented in Figures 7 and 8.

[Figure 7 in here]

[Figure 8 in here]

The items in Figures 7 and 8 are both influenced by a verbal component, and the use of negatives (not) is an emergent feature of these items. In addition the word *situated* (or rather its Swedish counterpart) might have caused some problems. There are several of these negatively formulated tasks among the ones identified as problematic from a reading point of view. There are however several examples of similar tasks that are not connected to the results on the word-recognition test (see Figure 9).

[Figure 9 in here]

Compared to this item, the items in Figures 7 and 8 resemble word-recognition items, adding a plausible cause of the verbal influence.

The second large category of items exhibiting a verbal component is characterized by words or phrases that might be problematic. The subcategories found are words or phrases from common language and words that are subject-specific.

Several of the verbal items contain words from common language that can be expected to cause some problems. One verbal item is about *live stock*, and at least the Swedish translation of that term is not a very frequent word in common language.

The item in Figure 10 belongs to the problematic tasks, where the verbal component is a significant factor. This can be understood by problems with words such as *release* and *absorb*.

[Figure 10 in here]

Figure 11 shows a similar item with a fairly strong verbal influence, even though the item is noted categorized as a verbal item because the significance level of the verbal factor ($p = 0.08$) is slightly higher than the criterion used in this study.

[Figure 11 in here]

We can note that there is a slight difference in the stems of these items, which might make a difference. The item in Figure 10 states “If you *are burning* wood”, while the item in Figure 11 talks about “When oil *is burning*”. This difference is accentuated in the Swedish translation through the use of a phrase closer to “If you *burn* wood”. To burn something might be interpreted as a process where energy is used, in particular for students with lower verbal abilities.

The words *release* and *absorb* in Figures 10 and 11 can be characterized as common words with a particular meaning and use in the scientific context, in particular in combination with

the word *energy*.

For several of the items where individual words or phrases are potential explanations to the relatively strong influence of a verbal component, the problematic words are subject-specific rather than taken from general language. This is particularly prominent in science. Some science items are very close to word recognition items in their appearance, which might explain the influence of verbal components (see Figure 12).

[Figure 12 in here]

A mathematics item, with subject-specific vocabulary, is presented in Figure 13.

[Figure 13 in here]

These items can be viewed as subject-specific word-recognition items, and the resemblance with the kind of items used in the verbal test (word-recognition test) is a plausible explanation to the verbal component found. The formulation of the item in Figure 13 can be compared to an item (Figure 14) where the use of subject-specific terminology has been avoided.

[Figure 14 in here]

In the Swedish version of the item in Figure 14, the word *congruent* is not used. Instead the first sentence says “These triangles have the same shape and size.” The avoidance of the mathematical term *congruent* might have saved this item for being identified as having a verbal component the method used in my study.

In the analysis, two verbal items containing the word *trait* were found (see Figures 15 and 16). The understanding of this particular word can be the critical feature making these items particularly connected to verbal ability.

[Figure 15 in here]

[Figure 16 in here]

Several of the items that seem to have a fairly large verbal component contain words expressing relations (see Figure 17). Words like more, less and as many are defined by (Schleppegrell, 2007) as words with specific meanings in mathematics.

[Figure 17 in here]

A few additional, more tentative categories were identified in the analysis of the verbal items. These categories contain very few items, but they might convey potential critical features that need to be considered in further studies.

It is far from obvious that decoding mathematical symbols or expressions should have something to do with the result of a word-recognition test, when mathematical ability is controlled for. However, for four items in mathematics the only feature found that could explain the verbal component is the mathematical notation or symbols. The item in Figure 18 is one of them.

[Figure 18 in here]

This item can be compared to the item in Figure 19 which is not indicated as having a verbal component in this study.

[Figure 19 in here]

In the item found in Figure 20, the order of the different answers might have had an impact on students' choice.

[Figure 20 in here]

Alternative E is the correct answer, but alternative D was chosen by 25 % of the Swedish students in grade 8. The description in alternative D works for the first two ordered pairs, and we can hypothesize that students choosing this alternative have tried on two ordered pairs of numbers but not the third. This hypothesis is supported by the fact that the third most common answer is alternative C, which works for the first of the ordered pairs. This behavior can be explained by an urge to reach closure as soon as possible (Biggs & Collis, 1982). The tendency towards closure is not a linguistic feature, but a possible connection to the word-recognition test is that the performance of students with a strong tendency toward closure is equally influenced in the word-recognition test and in TIMSS.

The analysis has primarily focused text, and to a lesser extent the relationship between written language and visual displays such as graphs, diagrams, and illustrations (see e.g. O'Halloran, 2000). However, one example of a potentially problematic interplay between text and illustration is found in the verbal item shown in Figure 21.

[Figure 21 in here]

There is an inconsistency between text and illustration, since the text is talking about a process but the picture does not strongly support the idea of a sequence of events.

Discussion

The explorative investigation of items identified as having a relatively strong verbal

component resulted in categories which to a large extent could be expected from the reading literature. This outcome was however not self-evident since very little research has been focused on the kind of texts found in the mathematics and science items of international comparative studies. Furthermore, the sub-categories described were not entirely possible to predict.

An overall observation from the analysis of the items is that there is a frequent use of passive form, which is contradictory to recommendations by reading experts. The use of passive form is however common in texts in mathematics and science and could be viewed as a typical feature for the genre. In addition, the requirement of international comparative studies that the item text must be possible to translate to a large number of languages might increase the use of the passive form. More active forms require an active subject which is assumed to be difficult to reproduce in different language and cultural settings.

Items identified as being particularly sensitive to students verbal capacity seem to be characterized by longer texts, higher text complexity, and difficult words and phrases. It is plausible that text structure and wording is inducing construct irrelevant variance into the cognitive instruments of TIMSS. However, this study is not possible to use to estimate the size of this problem. The study is designed to identify plausible linguistic features, not to evaluate the consequences of their existence. In addition, it is vitally important to consider whether some of the categories of items found in this study, and the potentially critical linguistic features characterizing them, actually describe irrelevant verbal components. We must conclude that science-words can not be excluded from science-task just because students can have an extra benefit from a verbal competence. (For example the word “traits”). Furthermore, if socializing students into the discourse of science is seen as a goal for the teaching and learning of mathematics and science, managing complex text structure typical for the subject can be seen as highly relevant for students at the end of compulsory school.

Translation is an additional complexity in international studies and Grisay (2003) claims that shortcomings in translation causes poor item functioning in international studies. Research has pointed out the particular difficulties inherent in the translation of mathematical texts (Wiest, 2003). Different languages are different when it comes to expressing mathematical concepts, which means that the mathematical register is difficult to translate. This study also identifies translation as a problem, and poor translation can induce irrelevant reading problems in items. Similarly, improvements of readability might be induced in the translation of the English original. The occurrence of such improvements is however not studied here.

A limitation in this study is that the corpus is fairly small in comparison to what is often considered necessary. The patterns of item features which this study aimed at identifying are rarely evident from a single item or even a small group of items; they emerge from the study of

many. Similarly, it is not certain that a single item showing a critical linguistic feature empirically can be found to require an irrelevant verbal component. This is the reason why some of the categories identified in this study must be seen as highly tentative. More research is needed to validate the existence of features only found in a couple of verbal items in this study.

A particularly problematic group of items are the ones that are identified as having a significant verbal component, even though almost all students can produce a correct item response. There are a few items in this category and they are excluded from the analysis. Even though the verbal component is statistically significant, the incorrect answers are too few to make a valid interpretation of the influence of critical linguistic features.

Conclusion and Implications

While controlling for subject specific ability in mathematics and science, a significant verbal factor explaining the performance was found in 26 % of the mathematics items, and 38 % of the science items in the Swedish version of TIMSS 1995. The word-count performed revealed that, on average, the number of words was higher in these items compared to items not identified as having a significant verbal factor in this study.

From the qualitative analysis of items identified as having a significant verbal component a general observation was the frequent use of the passive form. Furthermore, several features were identified in these items, that might explain the verbal influence, and categories of items were created and described based on these features. One major category consisted of items with complex text structure. This category was further divided into four subcategories; general text complexity, long and complex answers to choose from, reading tasks in mathematics, and use of the negative in questions. Another major category consisted of items with single words or expressions that can be problematic. These are either belonging to common language or to a subject-specific vocabulary. Another group of potentially problematic words can be found in common language but have a specific meaning or use in mathematics or science. In addition to the two major categories, three tentative critical features relevant for explaining verbal components have been identified.

Further research is needed to develop the analysis and validate the identification of critical linguistic features, by using other sets of items and researchers with other backgrounds. A better understanding of the complex relation between linguistic features and achievement in mathematics and science could be important for the interpretation of results from international comparative studies, but also for the construction of items which can be used for a valid assessment of what students know and can do in science and mathematics. Gipps and Murphy

(1994) referred to studies done in the UK where the results convincingly showed that students failed on tasks in science not because they made errors or didn't know, but because they tried to answer a completely different question. In the analysis of results from TIMSS, and other studies of achievement in mathematics and science, we must not only ask ourselves "why do students answer the way they do?" but also "which question was the student actually trying to answer?".

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing*. Washington, DC: American Psychological Association.
- Biggs, J. B., & Collis, K. F. (1982). *Evaluating the quality of learning. The SOLO taxonomy (Structure of the Observed Learning Outcome)*. New York: Academic Press.
- Dempster, E. R., & Reddy, V. (2007). Item readability and science achievement in TIMSS 2003 in South Africa. *Science Education, 91*(6), 906-925.
- Gipps, C., & Murphy, P. (1994). *A fair test? Assessment, achievement and equity*. Buckingham: Open University Press.
- Grisay, A. (2003). Translation procedures in OECD/PISA 2000 international assessment. *Language Testing, 20*(2), 225-240.
- Homan, S., Hewitt, M., & Linder, J. (1994). The development and validation of a formula for measuring single-sentence test item readability. *Journal of Educational Measurement, 31*(4), 349-358.
- Kress, G. (2003). *Literacy in the new media age*. London: Routledge.
- Lemke, J. L. (1990). *Talking science: Language, learning, and values*. Norwood, NJ: Ablex Publishing Corporation.
- Liberg, C., Folkeryd, J. W., af Geijerstam, Å., & Edling, A. (2002). Students' encounter with different texts in school. In K. Nauclér (Ed.), *Papers from the Third Conference on Reading and Writing. Working Papers no 50* (pp. 46-61). Lund: Lund university, Department of Linguistics.
- Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Educational Measurement* (Vol. 3, pp. 13-103). New York: American Council on Education.

- O'Halloran, K. L. (2000). Classroom discourse in mathematics: A multisemiotic analysis. *Linguistics and Education, 10*(3), 359-388.
- Schleppegrell, M. J. (2007). The linguistic challenges of mathematics teaching and learning: a research review. *Reading & Writing Quarterly, 23*, 139-159.
- Shorrocks-Taylor, D., & Hargreaves, M. (1999). Making it clear: a review of language issues in testing with special reference to the National Curriculum mathematics tests at key stage 2. *Educational Research, 41*(2), 123-136.
- Wiest, L. R. (2003). Comprehension of mathematical text. *Philosophy of mathematics education journal*(17).
- Working Group of the Joint Committee on Testing Practices. (2005). Code of fair testing practices in education (revised). *Educational Measurement: Issues and Practice, 24*(1), 23-27.

Electrical energy is used to power a lamp.

Is the amount of light energy produced more than, less than, or the same as the amount of electrical energy used?

The amount of light energy produced is

more than

less than (check one)

the same as
the amount of electrical energy used.

Give a reason to support your answer.

Figure 1 (item s012132)

Luis exercises by running 5 km each day. The course he runs is $\frac{1}{4}$ km long.
How many times through the course does he run each day?

Figure 2 (item m012112)

Alice can run 4 laps around a track in the same time that Carol can run 3 laps. When Carol has run 12 laps, how many laps has Alice run?

(A) 9

(B) 11

(C) 13

(D) 16

Figure 3 (item m012004)

Example 4 (m012148)

Rounded to the nearest 10 kg the weight of a dolphin was reported as 170 kg.
Write down a weight that might have been the actual weight of the dolphin.

Figure 4 (item m012148)

On cold days, snakes usually lie very still and eat little or nothing, while birds usually move around and eat a lot of food. Which statement best explains this?

A. Both animals are cold-blooded, but without feathers to keep warm, snakes get too cold to move.

B. Unlike birds, snakes are warm-blooded; they must hibernate during cold weather.

C. Unlike snakes, birds are cold-blooded; they are less affected by the cold than snakes.

D. Unlike snakes, birds are warm-blooded; they must eat food to maintain a constant temperature.

Figure 5 (item s012083)

Juan has 5 fewer hats than Maria, and Clarissa has 3 times as many hats as Juan. If Maria has n hats, which of these represents the number of hats that Clarissa has?

A. $5 - 3n$

B. $3n$

C. $n - 5$

D. $3n - 5$

E. $3(n - 5)$

Figure 6 (item m012123)

Most of the chemical energy released when gasoline burns in a car engine is not used to move the car, but is changed into

A. electricity

B. heat

C. magnetism

D. sound

Figure 7 (item s012008)

Which of the following organs is NOT situated in the abdomen?

(A) liver
 (B) kidney
 (C) stomach
 (D) bladder
 (E) heart

Figure 8 (item s012001)

In square $EFGH$, which of these is FALSE?

(A) $\triangle EIF$ and $\triangle EIH$ are congruent.
 (B) $\triangle GHI$ and $\triangle GHF$ are congruent.
 (C) $\triangle EFH$ and $\triangle EGH$ are congruent.
 (D) $\triangle EIF$ and $\triangle GIH$ are congruent.

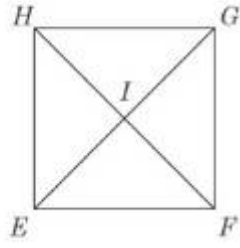


Figure 9 (item m012005)

If you are burning wood, the reaction will

A. release energy
 B. absorb energy
 C. neither absorb nor release energy
 D. sometimes release and sometimes absorb energy, depending on the kind of wood

Figure 10 (item s012048)

When oil is burning, the reaction will

A. only release energy
 B. only absorb energy
 C. neither absorb nor release energy
 D. sometimes release and sometimes absorb energy depending on the oil

Figure 11 (item s012088)

Which is the most basic unit of living things?

A. Cells

B. Bones

C. Tissues

D. Organs

Figure 12 (s012095)

A quadrilateral MUST be a parallelogram if it has

A. one pair of adjacent sides equal

B. one pair of parallel sides

C. a diagonal as axis of symmetry

D. two adjacent angles equal

E. two pairs of parallel sides

Figure 13 (item m012059)

These triangles are congruent. The measures of some of the sides and angles of the triangles are shown.

What is the value of x ?

A. 52

B. 55

C. 65

D. 73

E. 75

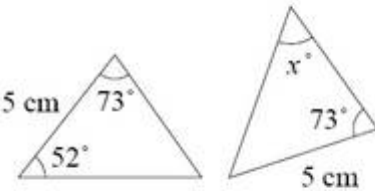


Figure 14 (item m012074)

Traits are transferred from generation to generation through the

- (A) sperm only
- (B) egg only
- (C) sperm and the egg
- (D) testes

Figure 15 (item s012039)

A son can inherit traits

- (A) only from his father
- (B) only from his mother
- (C) from both his father and his mother
- (D) from either his father or his mother, but not from both

Figure 16 (item s012026)

In the figure, how many MORE small squares need to be shaded so that $\frac{4}{5}$ of the small squares are shaded?

- (A) 5
- (B) 4
- (C) 3
- (D) 2
- (E) 1

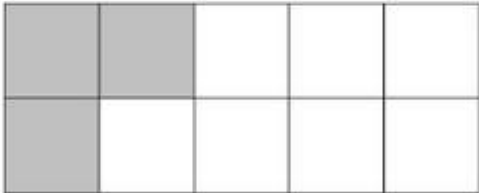


Figure 17 (item m012001)

Which of these expressions is equivalent to y^3 ?

- A. $y + y + y$
- B. $y \times y \times y$
- C. $3y$
- D. $y^2 + y$

Figure 18 (item m012120)

If m represents a positive number, which of these is equivalent to $m + m + m + m$?

A. $m + 4$

B. $4m$

C. m^4

D. $4(m + 1)$

Figure 19 (item m012115)

(3, 6) , (6, 15) , (8, 21)

Which of these describes how to get the second number from the first number in every ordered pair above?

(A) Add 3

(B) Subtract 3

(C) Multiply by 2

(D) Multiply by 2 and then add 3

(E) Multiply by 3 and then subtract 3

Figure 20 (item m012029)

The drawing shows an apple falling to the ground. In which of the three positions does gravity act on the apple?

A. 2 only

B. 1 and 2 only

C. 1 and 3 only

D. 1, 2, and 3

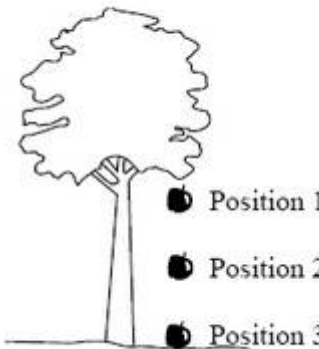


Figure 21 (item s012075)

Table 1 Average number of words in items from the Swedish version of TIMSS 1995

	Average number of words in TIMSS Science items		Average number of words in TIMSS Mathematics items	
	Stem	Longest alternative answer	Stem	Longest alternative answer
Items with verbal factor	19.3	6.3	26.0	2.5
Items without verbal factor	16.8	5.3	18.0	1.8