

- Swan, D. & U. Goswami. 1997. 'Phonological Awareness Deficits in Developmental Dyslexia and the Phonological Representations Hypothesis.' *Journal of Experimental Child Psychology* 66, 18–41.
- Søegård, A. & S. P. B. Petersen. 1974. *Ordstillelæsningsprøve OS400* ('Silent word reading test OS 400'). Copenhagen: Dansk psykologisk forlag.
- Torgesen, J. K., R. K. Wagner & C. A. Rashotte. 1994. 'Longitudinal studies of Phonological Processing and Reading.' *Journal of Learning Disabilities* 27, 276–286.
- Torgesen, J. K., R. K. Wagner, C. A. Rashotte, S. Burgess & S. Hecht. 1997. 'Contributions of Phonological Awareness and Rapid Automatic Naming Ability to the Growth of Word-Reading Skills in Second to Fifth Grade Children.' *Scientific Studies of Reading* 1, 161–185.
- Wagner, R. K. & J. K. Torgesen. 1987. 'The nature of phonological processing and its causal role in the acquisition of reading skills.' *Psychological Bulletin* 101, 192–212.
- Wagner, R. K., J. K. Torgesen & C. A. Rachotte. 1994. 'Development of reading-related phonological processing abilities: New evidence of bidirectional causality from a latent variable longitudinal study.' *Developmental Psychology* 30, 73–87.
- Wechsler, D. 1974. *Wechsler intelligence scale for children, revised*. New York: Psychological Corporation.

dkp@cphling.dk

## Who killed the young man? 15-year-olds' responses to a fable

Astrid Roe

Department of Teacher Education and School Development, University of Oslo, Norway

In 2000, more than 4000 Norwegian 15-year-olds participated in an international reading literacy survey, OECD PISA (Programme for International Student Assessment). OECD PISA is a new regular survey that assesses 15-year-olds' competence in three domains of literacy: reading literacy, mathematical literacy and scientific literacy. As reading was the main subject in 2000, it comprised 2/3 of the test material. In 2003, Mathematics will be the main subject and in 2006, it will be Science. The assessment was carried out in 32 countries during 2000, and around 100 000 15-year-olds participated.

In PISA *reading literacy* is to be understood in a broad sense rather than a technical sense. Technical skills like reading speed, and decoding of words and sentences are not tested, as 15-year-olds are considered to be able to read in a technical sense. Reading literacy here implies that readers should be able to construct, extend, and reflect on the meaning of what they have read across a wide range of texts associated with a variety of situations. The theory behind this reading framework is based on cognitive views of reading literacy, emphasising the interactive nature of reading and the constructive nature of comprehension (OECD PISA 1999). Reading is regarded as a process in which readers generate meaning in response to text by using their prior knowledge and understanding. It is implied that understanding a written text is more than just understanding the meaning of the words. Reading is the result of cognitive and verbal processes that are influenced by the reader, the context and the text itself. Readers will have different prior knowledge and experiences, and texts can affect different readers in different ways. Thus the process of reading and understanding becomes different from one reader to another (e.g. Fish 1987, Beach and Hynds 1991).

## Open constructed response items

The PISA 2000 survey is the first international reading literacy assessment where more than 50% of the items are open constructed response items. In the IEA-study of 1991, four open response items were tested out, but the results were not analysed (Mejding 1997). In multiple-choice items there is one correct answer given among four or five alternatives. In open response items, however, the reader has to produce a personal interpretation of the text expressed in his or her own words, which implies an assessment of writing skills as well.

Some items elicit a large variety of responses, particularly fictional texts and items that demand reflection or interpretation. There is considerable data in psychological and educational studies that focus on young children's responses to reading. There is much less evidence in print to suggest how older students and adults read and understand texts (Appleyard 1994). With so many open response items, PISA 2000 offers a unique opportunity to study adolescent readers' responses to texts qualitatively on a large scale. This adds a qualitative aspect to this quantitative study.

Marking open response items is quite challenging, although a detailed marking guide has been developed for this purpose. According to the marking guide, credited answers are either coded 2 or 1 hierarchically, meaning that simple or minimal answers get code 1 whereas answers that reflect a more advanced understanding can be given code 2. The two codes will to some extent categorise the credited responses roughly, but this is not the aim of the PISA study. The codes are transmitted into score values, and thus the qualitative aspect is lost.

## Two-digit coding

To preserve the qualitative aspect, the Norwegian PISA research group has developed a two-digit code system to categorise the responses into groups, defined by content. During the marking session after the Field Test in 1999, markers discovered that some texts were more open to various interpretations than others, and that one interpretation may be just as relevant and "correct" as another. Many of the items revealed a large variety of responses within one credited answer category. Based on these findings, a two-digit code-system was developed by adding a second digit to each of the codes. This made it possible to accumulate a larger variety of answers and, to some extent, preserve the qualitative aspect and thus give a more refined picture of how 15-year-olds read. The aim was to give insight into how different groups

of students use their metacognition and background knowledge during reading, and how a given text can affect readers differently.

The idea of developing two-digit codes for this purpose was highly influenced by researchers from the Norwegian PISA group in mathematics and science, who had already developed such codes for diagnostic purposes in the two domains (Kjærnsli, Lie & Turmo 1999). Two-digit codes were the official codes in mathematics and science in PISA 2000 – based on experience from the TIMSS<sup>1</sup> study, where such codes were used for the first time in an international comparative study (Lie, Taylor & Harmon 1996). Two-digit codes for reading were developed for ten PISA items, and were only used in Norway (Roe 2000).

Below, I use one of the ten items as an example of how the response categories may be used diagnostically. One question that will be raised is whether a certain way of responding to a text is typical for students who have certain background variables in common such as gender, reading habits, or mean over-all reading ability on the test.

## Response categories

The item chosen for analysis is a question about a fable. The fable tells the story of a man who dreams that a lion will kill his son. Even though he does everything to prevent this from happening, the son dies from an infection after having smashed his fist into a wall with a painted lion on it. The last sentence goes: "The lion, even though it was only a painted one, had indeed killed the young man, just as his father had foreseen."

The item is introduced with a conversation between two people. The first person argues that it is ridiculous to say that the painted lion killed the young man. The second person replies: 'I don't agree, the ending makes perfectly good sense to me'. The question is: *What do you think: Did the painted lion kill the son or not? Explain your answer, demonstrating your understanding of the fable's meaning.*

The intent of the question is to see whether the student can "reflect on the content of the fable, and to draw on values and beliefs to make a judgement about the moral of the story" (PISA Marking Guide 2000, p. 21).

Answers that go beyond a literal interpretation of the fable are coded 2 ("Full credit"), and answers at a literal level are coded 1 ("Partial credit"). By using a second digit, code 2 was divided into three categories: 21, 22 and 23, and code 1 was divided into two categories: 11 and 12.

<sup>1</sup> Third Mathematics and Science Study

**“Full credit”:**

**Code 21:** Linking **fate, destiny** or the father’s **dream** with the son’s death

**Code 22:** Explaining the cause of death in terms of **motives, attitudes** or **temperament** of the characters

**Code 23:** Referring to the idea that this is **not a realistic story** (may use terms such as “**metaphor**”, “**allegory**” or “**symbol**”).

**Note:** If one part of the answer reflected 21, while another part was type 22, we coded 21. If genre features were mentioned or implied we coded 23.

**“Partial credit”:**

**Code 11:** Commenting on the fact that **this had something to do with the lion**, for instance the fact that he hit a lion on the picture, not a zebra. May also mention the immediate physical causes of death or to the fact that a painting cannot kill.

**Code 12:** Responds at a **literal** level only. The answer must either refer to one of the immediate physical causes of death or to the fact that a painting cannot kill.

**Note:** If one part of the answer reflected 11 – 12 while another part reflected 21 – 23, we coded 21 – 23 as above.

**No Credit:**

**Code 01:** Insufficient, implausible, irrelevant or vague answers – or showing inaccurate comprehension of the story.

**Missing:**

**Code 99:** No answer

**Some examples of student responses:**

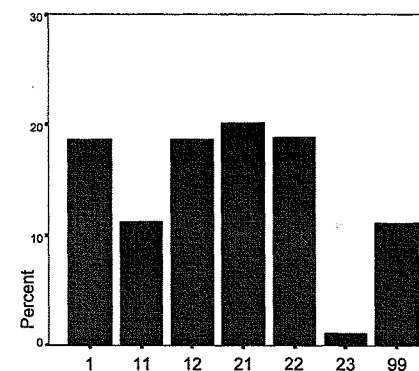
- The son caused his own death. He was angry because he was prevented from going hunting. (22)
- It was the father’s fault. He should not have tried to fight against what his dream told him (21)
- To prove that you cannot change Fate, the story has to end with a lion – real or painted or whatever – killing the son. (21)
- The son would have died anyway but the way he died was both the father’s and the son’s fault – the father for being over-protective and the son for getting angry. (21 and 22, coded 21)
- The son’s death has a symbolic meaning, as this is a fable. (23)
- In a way it was the lion, even if he died from infection. (11)
- He died from infection, but it was a lion on that picture....(11)
- He died from the fever after he had got an infected hand. (12)
- He died because he hurt himself when he smashed his hand into the picture. (*Identifies the son’s action, but not his motive*). (12)
- A picture can’t kill somebody. (12)
- It was the father. (01)
- The son killed himself. (01)
- Yes, the painted lion indeed killed the young man. (01)

**Reliability**

Open response items are sometimes quite challenging for markers. Neither a detailed marking guide nor several example responses can cover the variety of answers that students may come up with. One must accept some disagreement among markers, but if different markers code a large number of responses differently the reliability is invalidated. To check the reliability we organised a multiple marking session where 120 responses from the field test in 1999 were coded independently by six markers. The agreement was around 90% for the first digit and 85% for both digits. This proves that two-digit codes can add valuable diagnostic information without losing too much reliability.

**Distribution of responses**

Almost 90 % of the students made an attempt to answer to this question, which is not surprising, as the fable is short with a relatively simple plot. 19 % did not give credited answers, but altogether 40 % obtained score point 2 and 30 % got score point 1. With the exception of code 23, the responses are well distributed between the categories. It is interesting to note that only 1% mention genre features (code 23). On the other hand, meta-linguistic terms like allegory, metaphor and symbolic meaning may be rather advanced for 15-year-olds. Only one of 120 students used such terms in this item in the field test. A study of responses from nine items defined as “Reflection on form” from the field test showed that meta-linguistic terminology was hardly ever used. 15-year-olds seem to describe textual features and qualities in everyday language (Hertzberg & Roe 1999).



N = 1394

**Figure 1.** Distribution of responses

### Over-all reading ability in different response categories

In Figure 2, reading scores for the test as a whole (nationally standardised scores) are standardised to mean = 0 and standard deviation = 1. The bars indicate standard error from the mean. The figure shows that students who give code 12-answers generally are average readers. Students who give answers in the categories 11, 21 and 22, however, achieve clearly above the average. For code 23-answers it is not possible to generalise, as the number of cases is only 16.

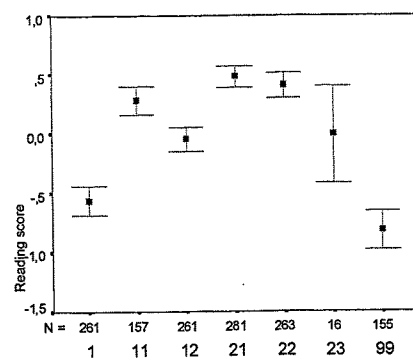


Figure 2. Mean over-all reading ability

Codes 11 and 12 will be collapsed into score point 1 in the official marking. It is, therefore, remarkable that students who give code 11-responses seem to outperform students in the code 12-category. They actually achieve as well as students who are given a full credit. This could imply a qualitative difference between code 11- and code 12-responses. One obvious difference is that code 12-responses are either quotes or paraphrases, whereas code 11-responses represent inferences. Code 12-responses repeat or paraphrase the fact that the son died from an infection, a fact that is explicitly described in the fable and that reflects a totally literal interpretation of the fable. Code 11-responses, however, also mention the connection with the lion. By suggesting that the lion also had something to do with the death, a causal link between the son's death and the actual cause of death is implied. This is what responses in the codes 21-, 22- and 23-categories represent as well. They explain the cause of death by means of inference based on information in the fable, without quoting or paraphrasing what is explicitly stated in the text.

Walter Kintsch discusses inferences in his book *Comprehension. A Paradigm for Cognition* (1998). He claims that understanding a narrative text sometimes demands a search in long-term memory for causal links. The reader may use some piece of pre-existing knowledge to bridge a gap in the text. This retrieval process accesses information available in long-term memory and is sometimes a resource-demanding search for bridging knowledge. In the search for bridging knowledge, the reader may also draw on *new information* extracted from the text and relevant information stored in long-term memory by some inference procedure. According to Kintsch, inferences may involve deductive reasoning that extends far beyond text comprehension, although text comprehension plays an important role here too. Reasoning comes into play when comprehension proper breaks down and the gaps in the text cannot be bridged in any other way, he claims (p. 192).

Do students who explain the son's death as resulting from fate disguised as a lion, frustration, protection, destiny, destiny or genre features (codes 11, 21, 22 and 23) produce some kind of new information based on the text, or are these responses a result of the students' search in long-term memory for a causal link? Thorough qualitative studies with interviews or think-aloud-studies could elaborate further on these questions. What these responses definitely have in common is that they add information that is not explicitly expressed in the text. They all reflect an understanding of the double meaning of the fable, more commonly spoken of as "reading between the lines".

We note that students who are not credited for their answers (code 01) achieve significantly below the average, and that students who have not made an attempt to answer this item (code 99) are the poorest achievers of all. One question is whether students who do not attempt to answer simply lack the ability to read and understand the fable and/or the question. In a very few cases this could be the reason, but the majority probably suffer from so-called "aliteracy", rather than illiteracy. Thomas and Moreman held that "the student who can read, but chooses not to is probably the most crucial concern confronting our education system today. It is not illiteracy we are combating, it is aliteracy" (in Granberg 1996, p. 34). Recent studies have indicated that this is an increasing problem in our country (Vaage 2000).

### Reading habits

The IEA reading literacy study showed a connection between voluntary reading and reading skills (Elley 1992). The PISA student questionnaire

contained several questions about reading habits and reading interests. Students were asked to indicate on a Likert scale from 1 ("Never") to 5 ("Several times per week") how often they read for pleasure: comics, textbooks, magazines, e-mail and web, newspapers and fiction. The two figures above show the average frequency reported by students for reading comics and fiction, distributed between code-response categories. Students from all response categories seem to read about the same amount of comics (Figure 3). The results are about the same for magazines, non-fiction, e-mail and web, and newspapers (not shown here).

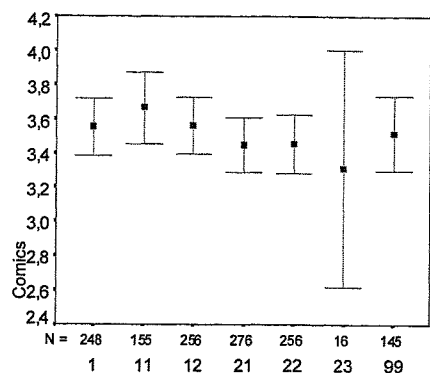


Figure 3. "How often do you read comics because you want to?"

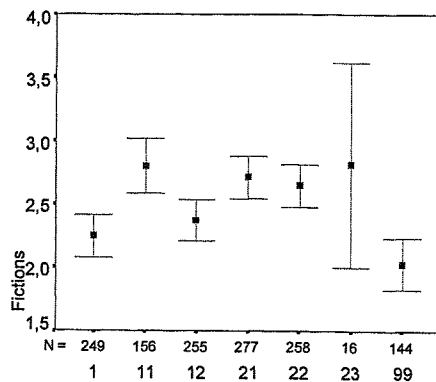


Figure 4. "How often do you read fiction because you want to?"

For fiction, however, the picture is different (Figure 4). Students in the code 01-, 12- and 99-categories report a lower frequency of fiction reading than students responding with codes 11, 21 and 22. Bearing in mind that the students in the latter categories represent the best achievers on the test as a whole, one may be tempted to suggest causality between fiction reading and reading literacy. Such an explanation is far too simple, although several Reading Literacy studies show a positive correlation between reading for pleasure and reading skills (Elley 1992, Solheim & Tønnessen 1999).

It is however not surprising that people who read a lot of fiction are better able to interpret a fable than people who are not familiar with fictional texts at all. Reading provides the reader with knowledge of content and text structures. The term *schema* is often used to describe long-term memory structures. The more extended schema knowledge a reader possesses the more able he or she is to make relevant conclusions and the more profound will the understanding of the text be (i.e. Iran-Nejad 1987).

### Gender differences

The distribution of answers between boys and girls shows that girls clearly dominate the code 11-, 21- and 22-categories whereas boys are in majority in the code 12-, 01- and 99-categories (Figure 5). It is a somewhat surprising that girls seem to be better able to understand the double meaning of this fable than boys, but I find it more striking that so many more boys than girls (23% and 14 %) give answers that are not accepted (code 01). As there is only one category for no-credit responses, we do not know what kinds of

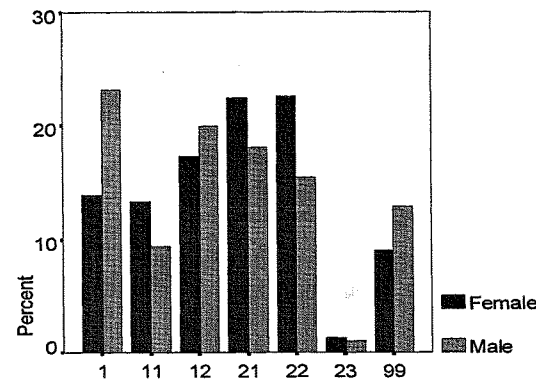


Figure 5. Gender differences in response categories

opinions students in this category have tried to express. Their answers may be inadequate or show misinterpretation, but they may just as well be a result of poor writing skills or they may be too vague or inaccurate.

Writing skills and verbal proficiencies are definitely taken into account in the assessment of open constructed responses. One hypothesis based on these findings is that boys generally do not express themselves as thoroughly and accurately as girls do or that boys have poorer verbal proficiencies than girls. Another hypothesis is that girls tend to give longer and more detailed answers, which may influence the markers positively. It is actually possible to carry out a qualitative examination of all the code 01-responses to verify or falsify these hypotheses.

Gender differences tend to be controversial, and at this point it is difficult to give a simple explanation of the differences shown in these results. Apart from writing skills and verbal proficiencies there are factors like maturity, interests, role expectancy, experience, physiological and psychological differences that may be involved. (See for instance Kimura 1999). If girls read more and have a broader field of interests, this will lead to better domain and schema knowledge and may also explain why they achieve better on this item (Taube 1996).

The general results from the IEA<sup>2</sup> reading literacy study in 1991 show only small gender differences among 14-year-olds, with girls performing slightly better. If we look at the IEA-results item by item, however, there is a tendency for boys to score better in tasks involving maps, charts and tables, and for girls to score better in narrative items. (Wagemaker 1996). A national study among 14-year-olds in Norway shows only minor general gender differences. (Solheim & Tønnessen 1999).

An issue that was brought up after the IEA study is the extent to which girls' and boys' opportunities to acquire socially required literacy are constrained or enhanced by social contexts. One conclusion was that cultural, social and educational contexts are related to reading literacy in the sense that they either create or do not create the opportunity for developing different literacy abilities (Elley 1992, Lundberg & Linnakyla 1993). If the gender differences shown in the item analysed above are representative of the results from the narrative items in the PISA survey in general, we may have to take a closer look at more of the background variables to search for an explanation. As only a few of the texts in the PISA study are defined as fiction, opposite gender differences might be revealed when it comes to maps, charts and

tables. If gender differences are generally greater in Norway than in other countries, there is reason to believe that the differences may be related to cultural, social or educational contexts in this country.

### Final reflections

So far I have only analysed one item and a few background variables from the PISA material. The results have shown that there are several ways of reading a fictional text and that there are connections between certain response categories and background variables such as gender, reading habits and reading achievement. The results from this item give reasons to believe that students whose responses reflect a deeper understanding than the literal description given in the fable seem to be better achievers in the test as a whole rather than students who just repeat what is actually described. This supports the theories about reading being more than just decoding and constructing meaning at a superficial level. Reading literacy also includes being able to understand what underlying messages a text is trying to impart. The reader must be able to recognise devices used by writers to convey messages and influence readers and to understand and interpret meaning from the structures and features of texts. Understanding literature implies being able to understand symbolic meaning, metaphors, and indirect messages.

Future analyses of more reading items, especially those connected to literary texts, where the responses are categorised by means of two-digit codes, will probably add interesting information to the patterns that have so far emerged from the analysis of this item; for instance, whether there are certain ways of responding to fictional texts that are typical of boys, or of students who read a lot, or of students who achieve well in the test. The frequency of responses in each category which will give information about the way young people respond to fiction. If any pattern emerges, this kind of qualitative analysis of quantitative data could contribute to better teaching of reading among secondary school teachers. If teachers are more aware of how different students get different meaning out of texts, and that reading is more than decoding and retrieval of surface meaning from text, they will probably be more aware of how to be good reading teachers for teenagers. Studies have shown that secondary school teachers pay little attention to teaching reading unless the student suffers from dyslexia or has other special needs. For "normal" students reading is regarded a skill that improves with practice (Durkin in Høien, Lundberg & Tønnessen 1994). The results from the Aesop item have shown that there are many different ways of reading and

<sup>2</sup> International Association for the Evaluation of Educational Achievement

interpreting a text and that some responses indicate better reading ability than others. This does not mean that one interpretation is right and another is wrong, rather that some interpretations may reflect better insight and the ability to detect nuances and thus be more advanced than others. Most secondary school teachers are probably aware of this, but they may not be taking it into account in their teaching. Teachers discuss students' different interpretations or just present students with the "right answer". However, they do not necessarily give them instruction about reading strategies to help them to go beyond the surface meaning to analyse, evaluate, and extend the ideas that are presented in the text. 15-year-olds need reading instruction from teachers who have knowledge and understanding of how young people respond to various texts and who are conscious of the fact that surface understanding is not enough.

## References

- Appleyard, S.J. 1994. *Becoming a reader. The experience of fiction from childhood to adulthood*. Cambridge University Press. UK.
- Beach, R. & S. Hynds 1991. Research on response to literature. In *Handbook of Reading Research Vol II*. N.Y.: Longman
- Elley, W. 1992. *How in the world do students read?* The Hague: International Association for the Evaluation of Educational Achievement
- Fish, S. 1980. *Is there a text in this class?* Cambridge: Harvard University Press.
- Granberg, M. 1996. *Lesing – en ferdighet i utvikling*. Oslo: Tano Aschehoug.
- Hertzberg, F. & A. Roe 1999. *Reading and reflection on form*. Paper presented at the 11<sup>th</sup> European Conference in Reading, Stavanger, Norway.
- Høien, T., I. Lundberg & F. E. Tønnessen 1994. *Norsk leseundervisning i internasjonalt lys*. Stavanger: Senter for leseforskning.
- Iran-Nejad, A. 1987. The Schema: A long-term memory structure or a transient structural phenomena. In R. J. Tierney, P. L. Anders & J. N. Mitchell (eds.). *Understanding readers' understanding. Theory and practice*. Hillsdale: N.J. Erlbaum.
- Kimura, D. 1999. *Sex and cognition*. Cambridge Mass.: Massachusetts Institute of Technology. The MIT Press
- Kintsch, W. 1998. *Comprehension. A paradigm for understanding*. Cambridge University Press. UK.

- Kjærnsli, M., S. Lie & A. Turmo 1999. *Two-digit codes for science and mathematics. Results from a Norwegian workshop*. University of Oslo. Faculty of Education.
- Lundberg, I. & P. Linnakylä 1993. *Teaching Reading Around the World*. The Hague: International Association for the Evaluation of Educational Achievement.
- Lie S., A. Taylor & M. Harmon 1996. Scoring techniques and criteria. In M. O. Martin & D. Kelly L. *TIMSS Technical report volume 1: Design and development*. Boston College. USA: Centre for the study and testing.
- Mejdning, J. 1994. *Den grimme ælling og svanerne? – om danske elevers læseferdigheder*. København: Danmarks Pædagogiske Institutt.
- OECD PISA 1999. *Measuring students' knowledge and skills. A new framework for assessment*. OECD Publications Paris. France.
- OECD PISA 2000. *Reading Marking guide 2000*.
- Roe, A. 2000. *Two-digit codes for reading*. Paper presented at the PISA Reading Functional Expert Group Meeting in Arnhem, Netherlands. PISA Report No. 8, ILS University of Oslo.
- Solheim, R. & F. E. Tønnessen 1999. *Kartlegging leseferdighet og lesevaner på 9. Klassestrinn*. Stavanger, Norway: Senter for leseforskning.
- Taube, K. 1996. Gender differences at the item level. In Wagemaker (ed.). *Are girls better readers? Gender differences in Reading Literacy in 32 Countries*. International Association for the Evaluation of Educational Achievement, Amsterdam
- Vaage, O. F. 2000. Barn og unge leser mindre. *Språknytt* 3, 15–17

astrid.roe@ils.uio.no