Sari Sulkunen

Text Authenticity in International Reading Literacy Assessment

Focusing on PISA 2000

JYVÄSKYLÄN YLIOPISTO

Sari Sulkunen

Text Authenticity in International Reading Literacy Assessment

Focusing on PISA 2000

Esitetään Jyväskylän yliopiston humanistisen tiedekunnan suostumuksella
julkisesti tarkastettavaksi yliopiston Athenaeum-rakennuksen salissa A103
marraskuun 17. päivänä 2007 kello 12.

Academic dissertation to be publicly discussed, by permission of
the Faculty of Humanities of the University of Jyväskylä,
in the Building Athenaeum, Hall A103, on November 17, 2007 at 12 o'clock noon.

UNIVERSITY OF JYVÄSKYLÄ

JYVÄSKYLÄ 2007

Text Authenticity in International Reading Literacy Assessment

Focusing on PISA 2000

Sari Sulkunen

Text Authenticity in International Reading Literacy Assessment

Focusing on PISA 2000

UNIVERSITY OF JYVÄSKYLÄ

ABSTRACT

The purpose of this study was to investigate how authentic in terms of text typicality and interest the texts used in PISA 2000 reading literacy assessment were to Finnish 15-year-old students. The study consisted of two phases. In the pilot phase, an 8-member panel reviewed the texts from the point of view of text typicality. The results showed that the most essential textual features in evaluating text authenticity were the topic, the genre, the layout and the language used. Based on these results a questionnaire for the main study was designed. Questionnaire was presented for 240 15-year-olds who evaluated and rated the typicality and interest of PISA texts' topic, genre, layout and language on a scale from 1 to 4. The results showed that the PISA texts were overall only moderately authentic in terms of typicality and interest to 15-year-old Finnish students. Furthermore, texts tended to be more typical than interesting. There were clear gender differences on which texts were found authentic but also texts that were equally typical and interesting for both genders. Overall, girls rated the texts more positively than boys. Based on the results it was concluded that the aim of choosing authentic texts for reading literacy test is important from the point of view of test validity. PISA texts, however, succeed only partially in achieving this aim as they form rather a pale reflection of texts authentic to 15-year-olds. In future, special concern should be in selecting texts more interesting to the target-group, particularly to boys.

Keywords: authentic texts, reading tests, test validity and reliability, literacy, discourse/text genres

Author's address   Sari Sulkunen
                            Institute for Educational Research
                            University of Jyväskylä
                            P.O. Box 35
                            FIN-40014 University of Jyväskylä
                            Finland
                            sari.sulkunen@ktl.jyu.fi

Supervisors   Professor emeritus
                            Matti Leiwo
                            Department of Languages
                            University of Jyväskylä, Finland

                            Professor
                            Pirjo Linnakylä
                            Institute for Educational Research
                            University of Jyväskylä, Finland

Reviewers   Senior researcher, PhD
                            Mirja Tarnanen
                            Centre for Applied Language Studies
                            University of Jyväskylä, Finland

                            Professor
                            Karin Taube
                            Department of Teacher Education in Swedish
                            and Social Sciences
                            Umeå University, Sweden

Opponent   Professor
                            Karin Taube
                            Department of Teacher Education in Swedish
                            and Social Sciences
                            Umeå University, Sweden

PREFACE

Although writing a doctoral thesis can be quite a lonely effort, I have been fortunate enough to be a member of a research community in which I have had the opportunity to work with a group of experts in interesting projects. While working in international reading literacy assessments has taught me a lot, most of what I have learnt comes from the talented people I have been working with. Thus, I wish to thank the members of the research community – colleagues, friends and supervisors – for all their help and support.

My deepest gratitude is addressed to my supervisors Professor Pirjo Linnakylä and Professor Matti Leiwo who in the course of guiding my work have shown dedication to this instructive and time-consuming task as well as profound expertise in the fields of reading literacy assessment and mother tongue instruction. Particularly I wish to thank my supervisors for their encouraging attitude and the kind supportive words they have used even in their most critical comments during the whole process. I am also grateful to the reviewers of my thesis, PhD Mirja Tarnanen and Professor Karin Taube. Their valuable and detailed comments and suggestions helped me to improve the final version of my thesis and made me feel confident about my work.

Institute for Educational Research has been my home base during the years of learning to become a researcher. I wish to thank the Director of the Institute, Professor Jouni Välijärvi, for giving me the opportunity to work and learn in such a unique community of experts and for the opportunity to write my thesis whenever there was time available from the projects. Likewise I wish to thank the Director of the Department of Languages, Dr. Bertold Fuchs, and Professor Aila Mielikäinen, responsible for the Finnish language studies, for giving me the opportunity to finish the thesis during the year I was working at the department with another inspiring group of experts.

I would like to thank all the members of the PISA research team at the Institute for Educational Research for offering a warm and challenging work environment. Likewise I am indebted to the Research group of Assessing Learning Outcomes, and especially the head of the team PhD Pekka Kupari who has not always had easy times with a group of young hot-tempered researchers but who has succeeded in being supportive to us all and in creating a forum for open and honest discussions.

There are many individual colleagues who have helped me with various ways during the process of writing my thesis. Antero Malin has taught me everything I know about statistical methods. The instruction has certainly been excellent, and I trust that there is still time to develop even further in this respect. He has also been of great help with the practical phases of analyzing the data. Eija Puhakka has helped me with the data entry phase and provided me with practical advice in analyzing the data. Kari Törmäkangas and Sakari Valkonen have also given me support, advice and recommended reading. Marja-Liisa Siikanen has helped me with converting the students' comments

into electronic form. With Kaisa Leino and Inga Arffman we have partly the same research interests, and they have patiently listened to my pondering of different solutions regarding to the study at hand, as has Tiina Nevanpää. With all these people I have shared many moments of discussing, not only of my thesis, but of all possible topics in the world. These discussions have included heated monologues and humorous dialogue alike, and have helped me to bond with the collegial community and to survive the most desperate moments of research work. Thank you all for your help, support and friendship.

During my year in Fennicum, I shared an office with Anna Pajunen, Minna Suni and Mikko Kajander. Together we shared also similar duties as teachers and doctoral students. Although everyone was busy with their own responsibilities, there was always time for humorous discussions and useful advice. In case one needed support in difficult times, it was also at hand. Thank you. And above all, thank you for being so patient with my frustrated monologues and thinking-aloud.

I am also grateful to David Underwood for proofreading my English manuscript. His comments taught me a lot and showed that he really aimed for understanding what I wanted to say and for improving the language used from that perspective.

I also want to thank my parents, Ritva and Pentti, for the growing environment where education was always valued and supported. It has been a solid stepping stone to my path of learning – path that still goes on.

Finally, I express my deepest gratitude in life to and for my own family, Joni, Juuso and Kaapo who make me happy when I just think about them. They make me complete, they bring sense of meaning to my life.

I dedicate this thesis to my sister Susanna who has shared with me the experience of writing a doctoral thesis. There have been numerous phone calls between us over the years during which frustration and despair about research but also joy has been expressed and practical problems solved. Her analytical skills are beyond compare – a constant source of inspiration for me.

Jyväskylä, November 2007

Sari Sulkunen

FIGURES

TABLES

CONTENTS

1 INTRODUCTION

1.1 The background of the study: authentic texts in international reading literacy assessments

It is widely recognised that the economic as well as the cultural and societal success of individuals and nations derives from their human capital. People continuously acquire new skills and knowledge throughout their lives making learning a lifelong process in which people should also have opportunities to continue learning. Consequently the Organisation for Economic Co-operation and Development (OECD) has initiated a large-scale programme of assessing student achievement in its attempts to improve the quality of education of its member countries as well as to provide information about the educational outcomes of the educational systems of these countries. (OECD 2001; 2002.)

This was the basis of the Programme for International Student Assessment (PISA), which aims to assess how well 15-year-old students are prepared for the challenges their life in the information society has to offer them, now and especially in their future studies and work. Rather than concentrating on the school curricula of key subject areas the focus is on young people's ability to apply and reflect on their knowledge and skills in real-life situations and their ability to continue developing these skills in order to meet the changing challenges of future societies. (OECD 2002.) PISA is planned to proceed in three-year cycles. The first cycle of assessment in the year 2000 covered the areas of reading literacy, mathematical literacy and scientific literacy. The main assessment area in year 2000 was reading literacy but, with the main assessment area varying, the focus was on mathematical literacy in 2003 and on scientific literacy in 2006. In 2009 reading literacy will be the major assessment area again, as it will be in the year 2018. (OECD 2001.)

In the year 2000 altogether 32 countries participated in PISA, with nearly a quarter of a million students assessed. A pen-and-paper test that lasted two hours was used. Additionally students answered a 30-minute background questionnaire. The test included text passages and related items including both multiple-choice and open-ended questions. (OECD 2001.) In the reading literacy

test of PISA 2000 almost half the items were open-ended tasks in which students constructed their own responses by relating the texts read to their own experience and knowledge. The test was designed based on an extensive assessment framework developed by an international group of reading experts who also guided the selection of texts to be used, the development of items and the marking guides for the constructed-response tasks. (OECD 2002.)

The results of PISA 2000 showed that Finnish students performed outstandingly in reading literacy. Finnish students' performance was significantly higher than that of students in any other country participating in PISA. Finland's good results were characterized by a high percentage of excellent readers and a low percentage of poor readers. The performance was especially high on the subscales for retrieving information and interpreting texts. Finnish students also performed well on the subscale for reflecting on and evaluating the text and Finland was ranked third among the participating countries. The performance on this subscale, however, was not as outstanding as on other subscales, which suggests that on this particular subscale there is room for improvement. Overall, the Finnish results were of high quality but also showed high consistency in that the standard deviation of overall mean scores was among the smallest in OECD countries. (OECD 2001; Linnakylä & Sulkunen 2002.)

The results of PISA 2000 also showed that in reading literacy there was a clear gender difference in favour of girls in all countries. In Finland, the gender difference was the greatest among all OECD countries. Since Finnish boys were the best readers in comparison with their male counterparts in PISA 2000, the clear gender difference reflected the outstanding performance of Finnish girls rather than poor performance on the part of boys. (OECD 2001; Linnakylä et al. 2002.) Furthermore, Finnish students proved to be relatively active and committed readers. Among the participating countries, they showed the highest level of interest in reading and the third highest level of engagement in reading. Additionally, three out of four students stated that they read for enjoyment every day. Finnish students actively read highly diverse materials, such as newspapers, magazines, comic books, e-mails and web pages, which all were read more actively by Finnish students than those of other OECD countries on average. Fiction and non-fiction books, on the other hand, were read less than in other OECD countries. There were some gender differences also in students' reading habits and materials. Finnish boys favoured e-mails and web pages slightly more than girls did. Moreover, comics and non-fiction books were read more by boys than by girls. Girls, on the other hand, clearly read fiction and magazines more than boys. In Finland, newspapers were frequently read by both genders. (OECD 2001; Linnakylä 2002.)

There has been lively discussion in seeking to interpret and explain the results of PISA 2000, and the PISA reading literacy test has been put under scrutiny. Due to Finland's excellent results, the Finnish education system has gained international attention and interest. Amongst the reasons suggested for these results are Finland's high level of teacher education, teacher and school

autonomy, the high grapheme-phoneme correspondence of our language, the low number of immigrants, subtitles on TV-programs, and even our long tradition of studying foreign languages (e.g. Välijärvi et al. 2002). In the countries that performed worse than they expected, the results have been explained in terms of the poor curricular match of the test or the test's consisting of culturally unfamiliar texts and tasks (e.g. Bonnet 2002, Prais 2003). Criticism has especially been directed at the Anglo-Saxon domination of test construction and the consequent predominant use of English materials, which reflects a neglect of the cultural and linguistic specificity of the participating countries. It has been stressed that the test should be organized allowing to normal national practices in test and task formats that are familiar to the participants. It has also been suggested that the translation process, even when as carefully planned and thorough as it is in PISA, necessarily alters the difficulty level of different national versions of the test due to differences between national languages. As a result of this, countries similar in culture have shown similar patterns in succeeding or failing on certain kinds of items. (Bonnet 2002, Lie & Roe 2003.)

The criticism presented has been associated with challenges to international comparative assessments in general, and especially concerns about cultural bias. Questions have been asked about whether it is even possible to construct culturally unbiased test with so many different countries, cultures and languages involved (e.g. Bonnet 2002). This is an important question related to the validity of the test, and is thus taken very seriously in international assessments. In the event of any group of participants, in terms of culture, gender, socio-economic background etc., being advantaged over another, the test is made biased and the validity of the test and interpretations based on the results are obviously put in question by construct irrelevant variance (see e.g. Messick 1994; 1995). In that case the test would not be equally fair to all participants: in fact, it would not even be the same test but a different test for different groups, thus invalidating the comparison of results (Standards 1999).

In PISA, the aim was to reduce cultural bias by inviting all participating countries to submit test material to the consortium organizing the test. Other measures to ensure cultural appropriateness included, for example, an extensive review process during which national experts rated the test materials in terms of, for example, students' familiarity with the content, concerns about cultural and other forms of bias, and anticipated problems with translation. The materials were further reviewed by an international cultural review panel. Field trial data was analyzed using various methods of item analysis in order to identify unacceptable interactions between item difficulty and country. Translation of the materials was subject to an independent verification process. (OECD 2002.)

Apart from the question of cultural bias, PISA has been criticized on the grounds that despite all efforts the tasks are not authentic real-life tasks such as PISA aims to present students with. Although the objective of constructing an

authentic test reflecting real-life reading as such has not been attacked, it has been questioned whether it is possible to create real-life tasks that are part of all the participants real life, since students have very different backgrounds in terms of social conditions and school systems, not to mention cultural and linguistic backgrounds. PISA has also been blamed for introducing gender and cultural bias in that the tasks set may be those of real life only for some participants, e.g. boys or girls. The PISA tasks have also been deemed inauthentic due to unrealistic simplicity in, for instance, some mathematical tasks. (Prais 2003.) Yet, PISA specifically aims at introducing students to authentic real-life tasks in which they can apply their skills and knowledge (OECD 2003). The aim of designing and creating an authentic reading literacy test, with authentic, real-life texts and tasks, reflects a wider interest in authenticity in the field of instruction and assessment, which originally derived from students' problems in applying the knowledge taught (e.g. Brown et al. 1989). The new instructional approach consequently stresses complex and meaningful problem-solving tasks in authentic real-life contexts (see e.g. Collins 1994; Collins et al. 1994; Edelson 1998). In terms of reading literacy, the demand for text and task authenticity has been further emphasized within a socio-cultural view of reading literacy (e.g. Barton 1994; Pitkänen-Huhta 1999; Street 1996).

Within the fields of assessment, interest in authenticity has arisen from the inconsistency between the new instructional approach and assessment methods which especially in large scale assessments included standardized multiple choice tasks and carefully constructed stimulus material (see e.g. Gottlieb 1997; Kohonen 1997). Authenticity in assessment is closely tied to the validity of testing (e.g. Bachman & Palmer 1996; Shohamy & Reves 1985; Wood 1993). Authentic tests reflecting real-life contexts and tasks are valid especially in terms of coverage, since they can capture the whole range of students' competencies (e.g. Huerta-Marcias 1995; Shohamy & Reves 1985; Wood 1993). Thus, authenticity is related to construct validity, since the test results can only be generalized beyond the testing situation if the test reflects language use outside that situation (Bachman & Palmer 1996). This is most important in assessments such as PISA that do not assess students' knowledge and abilities as covered and taught within school curricula but rather how students can apply and reflect on the knowledge and abilities they have acquired and how well they are prepared for future studies and working life. Overall, the concept of authenticity is related to many issues important in international studies of educational achievement, above all that of designing valid tests reflecting the construct chosen and capturing the real-life competencies of the students. The concept of authenticity also relates to concerns about bias.

Since issues of authenticity and cultural bias were central to the process of designing and constructing the PISA test, national teams were invited to submit authentic texts and tasks suitable in terms of the PISA framework to the consortium organizing the testing. Moreover, an extensive review process was implemented in which national subject matter experts reviewed the proposed

test materials, texts and tasks in terms of the target group's exposure to the material, whether in school or out of school contexts, as well as cultural concerns and other bias concerns, the familiarity of topic and text type and the degree of interest to 15-year-olds. The difficulty of the material was also estimated and anticipated translation problems were highlighted. The review process, however, was not without problems. Researchers and subject matter experts found the review process difficult, since it was often impossible to say how 15-year-olds (the target group) would feel about the texts in terms of the factors listed above. As for the level of the texts to be reviewed, there were also problems since text exposure and the interest of materials were to be assessed in terms of the whole text. However, genre, text type and topic might not be equally interesting or familiar and thus reviewing the texts would have been easier if topic and text type at least had been reviewed separately, as was the case in reviewing the familiarity of topic and text type.

The present study derives from the context described above. The general interest in authenticity within international reading literacy assessments as well as the practical problems in designing and creating an authentic test have raised several questions made more pressing by the critique of PISA from the point of view of cultural bias and authenticity. What is authenticity, especially in terms of texts? What are the attributes of authentic text in the context of international reading literacy assessments? How should the authenticity of texts be reviewed in the process of designing an authentic test? Which textual features should be assessed in determining text authenticity? Should the text be reviewed and assessed as a whole, or topic and text type separately, or topic and genre separately, etc.? How authentic were the texts used in the PISA 2000 reading literacy test in terms of the culture of Finnish 15-year old students', boys and girls? These are the issues the present study deals with and the aim is to bring some insight into both to concept of authentic text in international reading literacy tests and the cultural authenticity of the PISA texts in the lives of Finnish 15-year-olds.

The context in which issues described above are studied is that of the PISA 2000 reading literacy assessment, although other international reading literacy assessments are taken into account when appropriate. The PISA context is reflected at many levels in studying authenticity. At the level of international assessment, authenticity is related to issues of cultural bias. At the level of large-scale system-level assessment, authenticity can be studied within a national culture. At the level of international testing, authenticity is related to test validity (e.g. Bachman & Palmer 1996; Wu & Stansfield 2001), comparability of results and, to some extent, the motivation of the test-takers (e.g. Widdowson 1990; Bachman & Palmer 1996).

1.2   Research task

The aim of this study is two-fold. Firstly, the study aims to define and analyze the concept of authenticity as it applies to texts used in international reading literacy studies. The second aim is to consider the authenticity of the texts used in the PISA 2000 reading literacy test in relation to the culture of Finnish 15-year-old students.

The specific research questions of the empirical part of the study are:
1. Which textual features do readers find relevant to a text's authenticity?
2. How authentic were the texts used in the PISA 2000 reading literacy test to Finnish 15-year-olds?
3. Were the texts used in the PISA 2000 reading literacy test equally authentic to Finnish 15-year-old boys and girls?

The ultimate purpose of the study is to find criteria for selecting authentic texts for international reading literacy assessments. Creating a conceptual framework for establishing the authenticity of texts is thus a significant part of the study. This part of the study will not only help the practical work of researchers in the field but also and, most importantly, seek to enhance the validity of reading literacy tests.

Moreover, reviewing the PISA test in order to assess how authentic the texts chosen are in the culture of Finnish 15-year-olds will help to determine how fair and equal the PISA reading literacy test in 2000 was in terms of our national culture. Since the cultural and curricular authenticity of the test material has been paid attention to in interpreting the results of the PISA assessment (Bonnet 2002, Prais 2003; see also Cosgrove 2003, Sulkunen 2004), this part of the study will provide a perspective in evaluating the validity of the PISA 2000 reading literacy results.

In this part of the study special attention is given to gender differences in students' assessments of text authenticity. Since there was a wide difference in PISA reading literacy scores according to gender, especially in Finland, and since girls and boys showed clear differences in their reading habits (Linnakylä 2002), it seems probable that the texts are not equally authentic for girls and boys.

Where international assessments are in question the issues of cultural bias and validity are certainly relevant to discussion of the authenticity of texts. However, assessment and instruction on the national level also stand to benefit from the conceptual framework of the study, since it has bearing on e.g. recent discussion of the problem of reducing the gender bias found in the reading literacy skills of Finnish students (e.g. Linnakylä et al. 2002). The conceptual framework for establishing the authenticity of texts will benefit also the curriculum work being done both on national and local levels focusing on the enculturation of students to the world of authentic texts in preparing them for

the demands of the future in their studies, at work and in their lives as private individuals and citizens.

The empirical data of the study has been gathered through student questionnaires in which Finnish 15-year-old students assess the authenticity of the PISA texts. In the pilot study a panel of expert readers was also used to assess the authenticity of the texts. Data from the panel reviews were used as pilot data in developing the questionnaire for the 15-year-old students. Students' input is focused upon as an important source of information in ensuring the cultural suitability of international tests designed to assess their performance.

1.3   Organisation of the study

In order to describe the context in which the notion of authenticity in the present study appears, the conceptions of reading literacy applied in recent international reading literacy assessments are reviewed in Chapter 2. The chapter begins with a brief review of theories of reading in general. Thereafter four recent reading literacy assessments are briefly introduced and the definitions of reading literacy employed in these assessments are reviewed in the light of theories of reading literacy.

In Chapter 3 the focus is on the text. Since the concept of text is currently used in diverse ways, a brief definition is provided of what is meant by text in this study. Chapter 3 then proceeds to describe and define the concepts of text type and genre most commonly used in classifying texts for reading assessment. This is of importance for the study since, in designing and constructing a reading literacy test, a classification of texts is needed in order to ensure that the selection of texts covers the world of texts relevant to the target group being tested and thus reflects the definition of reading literacy used in the study. Chapter 3 aims to explain firstly what a text type is, and secondly, the concept of genre and its use. The relationship between the two concepts is also examined.

Chapter 4 concentrates on describing and defining the core concept of the study, i.e. an authentic text. In this chapter the aim is to explore what kinds of texts authentic texts are and what features, if any, they share. After describing the attributes of an authentic text, the theoretical background of the concept is briefly described in the contexts of both instruction and assessment in order to fully understand what is meant by the attributes of an authentic text. The chapter concludes with a discussion of the challenges and problems related to authenticity of texts in the context of international reading literacy assessments. Chapter 5 describes in detail how text was defined in PISA and how the texts were selected for the PISA 2000 reading literacy test.

The empirical part of the study begins with Chapter 6, in which the data, methods and materials of the study are described. Firstly the research questions

and phases of the study are introduced to the reader. Secondly, operationalizing the concept of text authenticity in the context of the empirical study in question is described and explained. Thirdly, the materials and methods of the study are described, concentrating first on the pilot study, in which a panel review process was used, and then on the main study, in which 15-year-old students responded to a questionnaire. The results of the panel review are also described in this chapter as a part of the design of the student questionnaire.

In Chapter 7 the results of the student questionnaire are described. In describing the results the focus is on the overall authenticity of the text as a whole as well as the authenticity of certain textual features (topic, genre, layout and language of the texts) reviewed separately by the students participating in the study. Special attention is paid to gender differences.

Chapter 8 focuses on the association between the different aspects of text authenticity treated in the present study, namely typicality and interest. The association between the authenticity of different textual features (topic, genre, layout, language) is also examined.

The thesis concludes by drawing conclusions based on the results of the study and discussing the results in Chapter 9. In the discussion section special attention is paid to the implications of the results for the large-scale international assessments of reading literacy but the instructional and pedagogical implications are also discussed. Chapter 9 also includes an assessment of the quality of the present study and suggestions for future research.

2 READING LITERACY IN INTERNATIONAL ASSESSMENTS

In order to give a fuller account of the context in which the notion of authenticity in the present study appears, the conceptions of reading literacy applied in recent international reading literacy assessments are reviewed in this chapter. The chapter begins with a short review of the different conceptions of reading literacy. Thereafter four recent reading literacy assessments are briefly introduced and the definitions of reading literacy used in these assessments are reviewed in the light of theories of reading literacy.

2.1 Defining the concept of reading and reading literacy

The change in paradigms of reading has shifted the focus of attention from purely cognitive aspects of reading onto contextual factors as well. Along with this change, the concept of reading has developed into that of reading literacy. This can be seen in the definitions used in recent international reading literacy assessments, none of which concentrates only on basic reading skills such as word-level decoding or literal comprehension. They reflect rather the more social views of reading now current. Cognitive theories of reading (Table 1) focus either on the steps necessary in decoding textual input into meaning (e.g. Gough 1985; LaBerge & Samuels 1985) or on text interpretation, taking into account also readers' knowledge and experience of the world (e.g. Rumelhart 1980; McNamara et al. 1991). These theories of reading, however, assume that the meaning which resides in the text is determinate and stable across contexts. A good reader is one that has the appropriate schemata to fit the text. A step towards current, more social views of reading was the transactional view represented foremost by Louise Rosenblatt (1988; 1994).

The transactional view of reading can be characterized as sociocognitive, since the reader's knowledge and experience gained socially have an important role in that meaning is constructed by the reader in transaction with the text. The

transactional view of reading was one of the first to acknowledge that meaning is indeterminate rather than determinate, and is constructed by the reader rather than being dictated by the writer or being fixed in the text. The role of context in the transactional view of reading is further underlined by the notion of stance. The reader's stance towards the text derives from the purpose of reading and this varies as, consequently, does the reading of the text (Rosenblatt 1994.) A somewhat similar but stronger emphasis on context has been presented in the sociocultural view of reading which relies on the work done by Scribner & Cole (1981) and Heath (1982), although Vygotsky had already presented ideas on how reading and writing are done for some reason and thus have purposes (1978). This view, however, stresses that instead of regarding reading and writing as separate skills of these should be treated as one unity of literacy, since they are connected through their uses in society. Literacy, just as well as reading literacy, is always used for specific purposes in specific contexts (Scribner & Cole 1981) and always has functions that go beyond acts of reading and writing (e.g. Barton 1994; Barton & Hamilton 2000; Gee 1990, 2000a, 2003).

The sociocultural view of reading further stresses that reading and writing are social practices rather than being just situated in a static context of the reader's (writer's) community. These social practices, like all social practices and all language use, are shaped by and shape culture at the same time (see Bruner 1986; Fairclough 1989, 1992). Thus, within the sociocultural view of literacy the notion of literacy practices is introduced and defined as "the general cultural ways of utilizing literacy which people draw upon in a literacy event" (Barton 1994, 37). A literacy event is something that can be observed but literacy practices are not observable since they involve people's values and attitudes, as well as their social relationships. Literacy practices are simply about what people do with literacy, how people talk about it and make sense of it. This does not mean, however, that literacy practices are only individual in nature. At the same time as individuals make use of their literacy practices, practices connect people, exist in the connections between people and involve shared cognitions that are represented in ideologies and social identities. (Barton & Hamilton 2000.) The sociocultural view of literacy emphasizes that people have multiple literacies that are associated with different domains of people's lives. Each literacy is connected to specific social, cultural, historical and institutional contexts. Literacy activities emerge from people's own interests and needs embedded in social goals and cultural practices. In different kinds of contexts with different texts people engage in different kinds of practices. (Barton & Hamilton 2000; Gee 2000b.)

As theories of reading have evolved into theories of reading literacy the focus has shifted to the functions of literacy and several categorizations have been developed. These categorizations have been based not only on the kinds of activities related to reading (Kadar-Fulöp 1985) but also on the perspective of society and the different contexts of reading (Goodman 1985). In a more recent

TABLE 1   Different approaches to reading.

	Cognitive approach				Sociocognitive approach	Sociocultural approach
view of reading	reading as information processing		reading as interactive meaning making		reading as transactional meaning construction	literacy as one unity of reading and writing in the context of use: reading and writing as social practices
proponents	bottom-up: e.g. Gough 1985 LaBerge & Samuels 1985	interactive: e.g. Ruddell & Speaker 1985	schema theories: e.g. Rumelhart 1980	mental model theories: e.g. McNamara et al. 1991	e.g. Beach & Hynds 1991 Rosenblatt 1994 Langer 1995	e.g. Vygotsky 1978 Scribner & Cole 1981 Heath 1980 Bruner 1986 Barton 1994 Barton & Hamilton 2000 Gee 1990, 2000a, 2000b, 2003
focus	steps in decoding textual input into meaning		reader's prior knowledge and experiences in meaning making	constructing mental models analogous to textual events in reading	transaction of reader and texts as a complex, nonlinear, recursive and self-correcting process; reader's knowledge and social experiences; reader's stance towards text guides reading and reflects the purpose for reading	reading and writing are used for specific purposes in specific context and have functions that go beyond mere act of reading and writing; power, ideology
meaning	reader translates text into meaning; reading is stable across contexts		meaning is determinate; interactive meaning making; text + interpretation		meaning is indeterminate and constructed by reader in transaction with text	meaning is social, constructed within/through social practices by the reader in the context of text, reader's prior knowledge and experiences and social context
role of text	meaning resides in text; text has central role		meaning resides in text; text has central role		starting point in transaction process; potential for meaning	mediator in social practices; reflects and is shaped by culture and context

categorization building on the work of previous studies, functional literacy is presented as consisting of the following aspects (Linnakylä 1995):

1. Reading connected with everyday tasks and the environment, including reading relevant to moving around and running errands. It can be reading street signs, instructions, bus timetables and advertisements.
2. Reading that supports social relations and communal intercourse, for instance writing e-mails to friends and relatives or following the news in order to be able to discuss it with neighbors or colleagues.
3. Occupational reading, i.e. reading at work or in applying for a job.
4. Reading to learn, which consists mainly of seeking information in various contexts and situations for various purposes.
5. Recreational and entertaining reading, which relates to people's use of their free time and is quite different in nature depending on the reader's interests. This category includes also reading for filling a void of communication.
6. Ritualistic reading, in which readers are satisfied by the reading event itself (e.g. the reading of religious texts).
7. Reflective reading, which involves discussing and arguing with the ideas presented in the text. This kind of reading characteristically expands thinking and requires involving reader's prior experience and knowledge.

The functional view of literacy emphasizes the richness of reading situations and the importance of reading in various everyday tasks. Reading is acknowledged to be a part of activities other than reading itself. In these activities, however, reading literacy has a central role as well as texts. Texts often correlate with functions of reading in the sense that for instance novels are more likely read for entertainment than in order to seek information, and an encyclopedia is more likely to be read in order to seek information than for pleasure or ritualistic purposes. Nevertheless, the functions of texts are not predetermined, since in addition to imposed uses of literacy (writer intended purpose) people also find self-generated uses of literacy (Barton 1994; Barton & Hamilton 2000). As Rosenblatt (1994) has already emphasized, the reader's purpose in reading guides his stance toward the text and thus no text itself nor the author can prescribe how any text should be read. Thus, texts have no characteristics that oblige readers to adopt a certain stance or approach towards them. The sociocultural view of reading stresses rather that texts both reflect and construct the surrounding culture at the same time (e.g. Fairclough 1989, 1992). It has therefore been suggested that in instruction and in assessment as well, there should be real reading materials from real reading situations (see Pitkänen-Huhta 1999; Street 1996). That is, instruction would benefit from the use of authentic reading materials reflecting the cultural context.

2.2 Views of reading reflected in reading literacy assessments

Reading literacy is considered to be one of the most important skills in lifelong and life wide learning and in recent years it has therefore been of great interest in international assessments. In 1990–1991 the IEA (The International Association for the Evaluation of Educational Achievement) conducted the Reading Literacy

Study (RLS) in which 32 countries around the world joined to investigate the reading literacy level of students at the ages of 9 and 14 (Elley 1992). In 2001 IEA started a new reading literacy study called Progress in International Reading Literacy Study (PIRLS), which focuses on young children at the ages of 9 and 10 (Framework and Specifications for PIRLS Assessment 2001).

Since literacy is now strongly associated not only with lifelong and life wide learning but also with the economic success of nations and individuals the OECD (Organization for Economic Co-operation and Development) has also been active on this front. The OECD initiated the International Adult Literacy Survey, which studied the level of reading literacy skills of adults aged 16 to 65. This survey was conducted in more than 20 countries around the world over the period 1994–2000. (OECD 2000.) PISA (the Programme for International Student Assessment) was also launched by the OECD. In the first cycle of PISA in the year 2000 the main assessment area was reading literacy, as it will be in 2009 and 2018. (OECD 1999.)

In reviewing the views of reading applied in recent international reading literacy assessments, it is noteworthy that these studies use the concept of 'reading literacy' rather than that of 'reading'. Reading literacy is seen as a functional ability that varies depending on different contexts and purposes. Overall, none of the recent studies concentrates on basic reading skills, such as decoding the text into meaning or merely literal comprehension, that were the focus of cognitive views of reading. Rather, the notion of functional literacy places these studies within the sociocultural view of reading literacy (Table 1). This has been reinforced in PISA by carefully avoiding the use of the concept of 'ability' that has cognitive connotations in other definitions of reading literacy and by including the purposes and contexts of reading in the definition (Table 2). Overall, the complexity of reading literacy as a social phenomenon and of the elements which make up meaning is reflected – necessarily in a reduced form – in the definitions of reading literacy used in the studies as well as in the actual tests.

All of the surveys mentioned above have the expression *understand and use* (employ) in the definition of reading literacy (see Table 2). It is thus acknowledged that reading is part of activities other than the mere act of reading. Only the PISA definition includes the idea that readers may need to take a critical stance towards the materials they read as *reflecting* includes the idea of relating the text – both content and form–to one's own experience and knowledge and evaluating the text from that perspective. All four definitions of reading literacy on the other hand describe what kinds of general purposes there are in reading. The IEA's Reading Literacy Study's definition is the most general and vague in this respect, since reading literacy involves reading texts that are required by society or valued by the individual. Progress in Reading Literacy Study (PIRLS), which relies on RLS, complements this by giving a few specific reading purposes, namely reading to learn, to participate in the community of readers and for enjoyment. These certainly are purposes in reading common to most readers but selecting the three over all other possible

purposes reflects the age (9 or 10) of the target population of the study (Framework and Specifications for PIRLS Assessment 2001). In that age group participating in the community of *readers* rather than the wider community, i.e. society in the PISA definition, is the choice of a significant community of peers.

TABLE 2   Organisation of the assessment framework in international reading literacy assessments.

	Definition of reading literacy	**Contexts and purposes of reading**	**Texts**
IEA/RLS 1990-91	The ability to understand and use those written language forms required by society and/or valued by the individual	society requires, individual values	written language forms: narrative prose, expository prose, documents
IEA/PIRLS 2001	The ability to understand and use those written language forms required by society and/or valued by the individual. Young readers can construct meaning from a variety of texts. They read to learn, to participate in communities of readers and for enjoyment.	society requires, individual values; *read to learn, to participate in communities of readers, for enjoyment*	written language forms: literary texts and information texts divided into chronological & non-chronological texts
OECD/IALS 1994–2000	The ability to understand and employ printed information in daily activities, at home, at work and in the community– to achieve one's goals, and to develop one's knowledge and potential.	at home, at work and in the community; to achieve one's goals, to develop one's knowledge and potential	printed information: prose, documents
OECD/PISA 2000-	Understanding, using and reflecting on written texts in order to achieve one's goals, to develop one's knowledge and potential, and to participate effectively in society.	*for private use, for public use, for work, for education;* in order to achieve one's goals, to develop one's knowledge and potential and to participate in society	written texts: continuous and non-continuous

The OECD studies, namely IALS and PISA, both emphasize the individual perspective of achieving one's goals and developing one's knowledge and potential in their definitions of reading literacy. PISA, however, stresses also the perspective of society in acknowledging that reading enables one's participation in society. This perspective can also be seen in the IALS definition, which stresses using printed materials in daily activities in the community as well as at home and at work. Furthermore, although this is not visible in the definition given of reading literacy, PISA framework is in part organised in terms of the different purposes of reading, referred to also as reading situations. The four purposes of reading are reading for private use, for public use, for work and for education. (OECD 1999.)

In the IEA's Reading Literacy Study the assessment area is defined in terms of text types, namely narrative texts, expository texts and documents (Linnakylä 1995). In this study it is taken that the reading process differs among these text types in that narrative texts are usually processed by understanding and drawing conclusions whereas, for instance, documents are used as a source of information in doing something. The PISA framework also includes a classification of texts by text type in order to ensure that the reading literacy test covers both continuous and non-continuous text types (OECD 1999). Likewise in IALS text type is central to the organisation of the assessment although the quality of tasks is also relevant. In this particular study there are three different literacy domains defined in terms of the reading materials and the aspects of information processing required in the tasks (OECD 1997). *Prose literacy* tasks are linked to various kinds of prose texts such as newspaper articles and texts from brochures, while literary prose is excluded. *Document literacy* tasks are linked to texts such as schedules, graphs and tables. *Quantitative literacy* tasks are linked to various prose and document texts but these tasks require arithmetical operations. (OECD 1997.) Quantitative literacy is thus defined in terms of tasks since the material the tasks are linked to consists of both non-fictional prose and documents.

These definitions and operationalizations of reading literacy stress the role of the different purposes of reading in different real-life situations. The purposes of reading mentioned in the four studies reflect the richness of real-life reading situations although only the two most recent studies (PISA and PIRLS) explicitly mention the specific purposes of reading that have been taken into account in devising the reading literacy test (Table 2). The IALS definition states only at a general level the contexts and purposes for reading assessed. The references to real-life situations and the consequent demand for real-life texts and tasks for the test make the notion of authenticity relevant to the reading literacy assessments. The two most recent international reading literacy assessments, namely PISA and PIRLS, both explicitly state in their frameworks that the aim is to simulate authentic texts and reading situations in the test (Framework and Specifications for PIRLS Assessment 2001; OECD 1999). In RLS and IALS authenticity is not explicitly mentioned although expressions such as 'familiar' and 'suitable' were used in describing the test material (e.g. Elley 1992; Linnakylä 1995). Such attributions relate to authenticity in terms of real-life texts and tasks.

The definitions mentioned also emphasize the importance of reading literacy from the point of view of both society and individual. The ability to read functionally gives the individual the potential to learn new things in non-formal contexts, to reach new worlds and cultures, to develop her/his own culture and grow emotionally and intellectually, as well as the ability to function actively in a society based largely on written materials, both printed and electronic (e.g. Linnakylä 1995). For society, the citizens' ability to read means the possibility of a thriving culture and economy created by knowledgeable and capable individuals willing to participate in the lifelong

learning required by the information society (see Linnakylä 1995; OECD 2000). The four studies described, especially those initiated by the OECD, stress the advantages for society. One could argue that this is natural since these studies specifically aim to investigate what kinds of abilities people must have to meet the needs of education and work-life as well as everyday life at home. The studies described here are not designed to explore individual skills or the meanings individuals give to literacy but to study the quality and equity of educational outcomes in the school systems of the participating countries.

Overall, these studies reflect the sociocultural view of reading literacy since literacy is seen as deeply contextual. Moreover, the fact that at least in the two most recent studies (PISA and PIRLS) the aim is that the texts and tasks chosen for the literacy test should be as authentic as possible for all the participants acknowledges that texts reflect the culture they emerge from and that different texts are used and read for different purposes (Framework and Specifications for PIRLS Assessment 2001; OECD 1999). In PISA different interpretations of the same texts are allowed when coding students' responses to the tasks. It is thus admitted that students with different background experience and knowledge may construct different meanings in interpreting the text.

International reading literacy studies have been criticised on the grounds that the tasks are not in fact authentic real-life tasks since they are decontextualized and any cultural variance has been eliminated in avoiding cultural bias in testing (Hamilton & Barton 1999; Bonnet 2002; Prais 2003). It should be acknowledged, however, that due to the nature of the survey and the large population assessed the studies cannot come near to incorporating the view of literacy as social practices important to individuals in their own community, as in the "new" literacy studies represented by e.g. Hamilton and Barton (1999). New literacy studies employ ethnographic methodology which suits their goals better since it allows studying individuals and their uses of literacy. Hamilton and Barton (1999) amongst others question the validity of international assessments since pursuing equality and comparability of test materials results in a test that only assesses the common transnational core of literacy practices and neglects the situated practices of individuals and their subcultures. On the other hand, assessing the transnational core of literacy is a necessary and a demanding task in globalizing world of texts and comes within the aims and scope of the large-scale studies. It is readily admitted in international large-scale reading literacy studies that individuals' use of literacy and the meanings individual people give to literacy are beyond the scope of the studies and their methodology. The test material can nevertheless reflect and simulate real-life reading situations and materials. Since the issue of test equality and comparability is crucial, the test construction process involves striking a balance between the demands of comparability and authenticity. This results in aiming for a test equally authentic or equally inauthentic to all participants.

In conclusion, the present study sees reading literacy as a sociocultural phenomenon in which the reader, the text, the context, and cultural values,

expectations and practices are intertwined. Readers' own individual interests and needs are embedded in social goals and cultural practices. In different kinds of contexts with different texts people engage in different kinds of literacy practices. (See Barton & Hamilton 2000; Gee 2000a.) However, it has to be admitted that although recent large-scale reading literacy assessments reflect the sociocultural view of literacy and thus aim for tests that are authentic, it is evident that they cannot incorporate the understanding of literacy as social practices to the extent described above. In large-scale international reading literacy assessments the sociocultural aspects of reading literacy are necessarily taken into account in a reduced form. Nevertheless, it is emphasized that reading literacy is used for different purposes in different contexts and that reading literacy activities as well as texts reflect the surrounding culture with its beliefs, values and practices. Within this context, the present study turns its focus on the text.

3   CONCEPTS USED IN TEXT CLASSIFICATION

A text has traditionally been defined as 'a continuous piece of spoken or written language, especially one with a recognizable beginning and ending' (Trask 1997) and as 'a piece of spoken or written language' (Richards et al. 1985). However, the emergence of new technologies and media has resulted in texts exploiting modes of meaning that go beyond writing and speech. Texts have become essentially multimodal in the sense that they integrate linguistic, visual, spatial, auditive and gestural meaning making systems (Cope & Kalantzis 2000). In the PISA reading literacy test in 2000 only written texts were used, mainly due to the practical limitations of testing, although the assessment framework included references to hypertexts and electronic texts as well (OECD 1999). Like real-life written texts, the PISA texts exploit both linguistic and visual meaning making systems, i.e. they contain not only written text but also pictures, photographs, and many non-continuous elements such as diagrams, charts and maps.

In designing and constructing a reading literacy test, a classification of texts is needed in order to ensure that the selection of texts covers the world of texts relevant to the target group being tested and thus reflects the definition of reading literacy used in the study. In recent international reading literacy studies texts have been classified for selection purposes mainly according to *text type* but exemplified often also as different *genres*. In the Reading Literacy Study (RLS) the test included three types of reading materials, namely *narrative prose* (e.g. fables, stories), *expository prose* (e.g. family letters, descriptions), and *documents* (e.g. charts, tables, maps, instructions) (Elley 1992). The primary basis for text classification was thus text type, but each type was exemplified by everyday real-life texts referred to as genres. In the Programme for International Reading Literacy Study (PIRLS) a decade later, the texts were classified according to text type as *literary texts* (e.g. narrative fiction) and *information texts* including both chronological texts (e.g. narrative and instructive texts) and non-chronological texts (e.g. persuasive and expository texts, documents) (Framework and Specifications for PIRLS Assessment 2001). The primary text types in the classification used in PIRLS were quite general

but they were further divided into more specific subcategories of text types exemplified by even more specific types of texts.

In the reading literacy studies initiated by the OECD similar criteria were used for classifying texts in the reading literacy test. In IALS, the reading materials were described as a selection of non-fictional *prose* including texts from newspapers, magazines and brochures and as *documents* such as schedules, charts, graphs, tables, maps and forms (OECD 1997). Two general types of text, prose and documents, were thus illustrated by referring to the context of appearance in the case of prose and to different formats of documents. In PISA also the texts were classified according to text type. *Continuous texts* were divided into description, narration, exposition, argumentation, instruction, and hypertext. *Non-continuous texts* were divided by format into ten different groups as follows: forms, information sheets, vouchers, certificates, calls and advertisements, charts and graphs, diagrams, tables and matrices, lists and maps. Each text type was further illustrated by giving more specific categories of text types, such as narrative, report or news story as examples of narration, or timetables as an example of tables and matrices. (OECD 1999.)

As a rule, reading literacy studies have applied classifications of texts in terms of text types. In different reading literacy studies text types have to some extent been labelled differently, and the level of specificity varies. For instance RLS mentions three types of materials and gives examples while PISA has two main categories, both of which are further divided into more specific text types with even more specific examples. In the most recent studies, namely PISA and PIRLS, text types have been described in detail, thus revealing the underlying criteria. These criteria mainly concern linguistic organization and typical formats of content although there are some references to the purpose of texts as well. (See Framework and Specifications for PIRLS Assessment 2001; OECD 1999.)

The diversity of text categorizations used in international reading literacy assessments employing text type suggests that the criteria behind the concept of text type can vary. In order to clarify the criteria behind different text classifications, this chapter begins by introducing two central concepts used in classifying and describing texts, namely *text type* and *genre*. Both concepts have traditionally been used in classifying texts according to the features that distinguish different kinds of texts from each other. The criteria according to which text types and genres are described and distinguished from each other vary as well. Sometimes the two concepts have been used vaguely and even interchangeably (e.g. Taavitsainen 2001).

This chapter aims first to explain what a text type is, and then the concept of genre and its use is explained. The relation between the two concepts is also explored.

3.1 Text type

Although the concepts of text type and genre have sometimes been used in a confusing manner, currently there seems to be consensus that the two concepts are distinct. Classification by genre relies on the context of texts, i.e. text external features and format, while text types are classified according to the linguistic features of texts alone. (See e.g. Biber 1989; Moessner 2001; Taavitsainen 2001.)

The most prominent classification of text types is that presented by Werlich (1983), whose text typology aims to describe universal text types having structural differences which derive from the forms and range of human cognition. Each text type is described in terms of typical surface structure, i.e. sentence types and the relations between sentences in the text. *Descriptive* text type deals with phenomena in space and typically consists of phenomena-registering sentences. *Narrative* text type deals with phenomena in time and typically consists of action-recording sentences. *Expository* text type focuses on explaining the relations of concepts and elements in a whole, typically exploiting phenomenon-identifying and phenomenon-linking sentences. *Argumentative* text type typically proposes relations between concepts and is related to the cognitive process of judging. The typical surface structure of this text type consists of the quality-attributing sentences. *Instructive* text type instructs the writer or others what to do. This text type relates to the cognitive process of planning and typically consists of action-demanding sentences. (Werlich 1983.) This classification of text types was used in describing continuous texts in the PISA reading literacy framework (see OECD 1999). For this purpose, however, Werlich's classification is less than ideal. Firstly, text type classification rarely works with whole texts since there usually are different parts in authentic texts that have different functions realized with different text types. Secondly, and perhaps most importantly, Werlich's text type classification in its strictest form fails to acknowledge that texts are pieces of language use (e.g. Karvonen 1995) as text types in text are not connected with the context of use. One can ask what relevance the text types as defined by Werlich have in relation to how texts are used in real-life contexts. This issue is returned to in chapter 9. (See also section 3.4 about multitype texts.)

Although Werlich's ideas have been widely used as a basis for text classification in international reading literacy assessments more recent studies concerned with linguistic similarities and differences in texts have often been based on large corpus studies. Within these studies, the linguistic features found in different types of texts have been explained with references to the texts' purposes and contexts of use. For instance, in Biber's multidimensional corpus study, the aim was to provide a typology of texts based on their linguistic features alone. The study was based on a corpus of 481 complete texts covering 23 different genres and subgenres. The corpus was analyzed using factor analysis of the co-occurrence of linguistic features in five different

dimensions, and this resulted in eight different text types of modern English. The text types identified were: intimate interpersonal interaction, informational interaction, scientific exposition, learned exposition, general narrative exposition, imaginative narrative, situated reportage, and involved persuasion. These text types are based on the core texts of each cluster found in the study, since there were no sharp distinctions between the text clusters but rather a continuous range of variation. The core texts are described as prototypes, while the rest were seen as peripheral texts resulting from speakers' and writers' individual exploitation of linguistic resources. (Biber 1989.)

For the Finnish language a somewhat similar study has been done by Saukkonen (1984) although he operates in a different conceptual framework. The aim of his study was to describe the style of certain rhetorical genres in terms of their linguistic characteristics. The rhetorical genres chosen for the study were artistic, scientific-theoretical and informative styles. The texts of the corpus were studied qualitatively in terms of their structure using Werlich's (1983) categories and in terms of the information structure of the texts. An extensive quantitative analysis of lexico-grammatical features was also conducted. (Saukkonen 1984.)

The linguistic characteristics used in analyzing and classifying text types are quite specific in nature. They consist of sentence types and the relations between sentences in Werlich's work, and several grammatical features in both Biber's and Saukkonen's studies, not to mention other studies of the English language adopting a similar approach (e.g. Taavitsainen 2001). Although Biber and Saukkonen deal with languages of different types, English and Finnish, some grammatical characteristics they take into account are the same. The interpretation of these characteristics may, however, be different, as well as their realization in each language. For instance, in Biber's study the use of past-tense verbs, 3rd person pronouns, public verbs and synthetic negation are interpreted as reflecting a narrative dimension. On the other hand, the use of time and place adverbials and adverbs, for instance, reflect situation dependent reference. (Biber 1989.) Moreover, Saukkonen (1984) pays attention to e.g. time and place adverbials and temporal conjunctions and concludes that they are typical of the rhetorical genre exploiting artistic style characterized by narration and description. Causative subordination and e.g. causative verbs, on the other hand, are typical of rhetorical genres exploiting scientific style and to some extent informative style. In addition to grammatical features, Saukkonen studied lexical differences between the styles exploited by rhetorical genres focusing on the semantic aspect of e.g. verbs and nouns used. According to his study, for instance, artistic texts typically exploit dynamic verbs referring to active movement, sensation or communication. Nouns typical to these texts refer to people, places and concrete objects while the most typical conjunctions and adverbials in artistic style express time, place, quantity and manner. (Saukkonen 1984.) Overall, the linguistic characteristics, both grammatical and lexical, identified in text type studies are quite specific.

To conclude, when texts are analyzed and classified in terms of text type the focus is mainly on linguistic criteria, as described above. However, although studies of text types concentrate on the linguistic characteristics of texts, contextual factors are referred to in explaining the occurrence of linguistic features. Contextual factors used more specifically are, for instance, the function or purpose of the text and perspective in the text (see Biber 1989; Saukkonen 1984; Taavitsanen 2001). To what extent contextual features actually are referred to in explaining linguistic occurrences in texts and to what extent explanations are successful in emphasizing the use and functions of linguistic occurrences is arguable.

3.2 Genre

While text types are described mainly in terms of the linguistic characteristics of texts, the concept of genre shifts the focus to the external features and format of texts (See e.g. Biber 1989; Moessner 2001; Taavitsainen 2001). In describing and analyzing genres, the primary interest lies in the context and the function of the text.

Genres have been studied widely within applied linguistics, and at least three different approaches can be distinguished (see Table 3), each having its own emphasis in defining and exploiting the concept of genre. One theory, based on Halliday's systemic functional grammar, aims to theorize and explain how and why texts are like and unlike each other (e.g. Kress 1993; Martin 1993; Martin & Rothery 1993; Eggins & Martin 1997). Another, the English for specific purposes (ESP) approach, aims to describe and explain different disciplinary genres in order to develop instruction in writing (e.g. Swales 1990; Bhatia 1993; Dudley-Evans 1994). A third is the sociocognitive approach which aims to describe the genre knowledge possessed and used by discourse community members and to study change in genres (Berkenkotter & Huckin 1995).

All three approaches agree that genre is fundamentally a social phenomenon which varies according to the purpose of communication. Texts that are used for similar communicative purposes are similar also in terms of structure, style, and content. The tasks genres are used for are culturally established, as are genres themselves. Thus there are genre conventions that constrain the use of linguistic and discursive devices in texts belonging to a genre. (See e.g. Swales 1990; Bhatia 1993; Eggins & Martin 1997.) Furthermore, all three approaches explain the linguistic organization of texts by connecting it to the social context and communicative purpose of the text. In other respects there may be substantial differences in emphasis between the different approaches, sometimes even within approaches.

Within the systemic functional approach, which is much influenced by M.A.K. Halliday's lexicogrammar (1978, 1985), the linguistic patterns of texts are studied and explained by referring to both register and genre as contextual

TABLE 3   Main approaches in studying genre within applied linguistics.

Approach	SYSTEMIC FUNCTIONAL		ENGLISH FOR SPECIFIC PURPOSES (ESP)			SOCIOCOGNITIVE
Advocates (e.g.)	Martin (1993) Eggins & Martin (1997)	Kress (1993)	Swales (1990)	Bhatia (1993)	Dudley-Evans (1994)	Berkenkotter & Huckin (1995)
Agenda	Pedagogical, access; classification of texts.	Pedagogical and political; access, empowerment.	Enhancing (writing) instruction in ESP.			Present genre knowledge as situated cognition.
Genre definition	The different types of text used in our culture to get things done (1993). Different genres are different ways of using language to achieve different culturally established tasks (1997).	Genre is a social process; a device to analyze the conventionalized nature of linguistic interactions and the way in which language both reflects and constructs certain relations of power and authority; genre is one of several categories needed to provide an account of what a text is, or of what social factors go into the linguistic formation of the text.	A genre consists of a class of communicative events, the members of which share some set of communicative purposes. These purposes are recognized by the members of the discourse community, and thereby constitute the rationale for the genre. This rationale shapes the schematic structure of the discourse and influences and constrains choice of content and style.–– In addition to purpose, exemplars of a genre exhibit various patterns of similarity in terms of structure, style, content and intended audience.	Relies on Swales (1990) and concludes: each genre is an instance of a successful achievement of a specific communicative purpose using conventionalized knowledge of linguistic and discoursal resources.	A genre is a means of achieving a communicative goal that has evolved in response to particular rhetorical needs and a genre will change and evolve in response to changes in those needs.	Genres are inherently dynamic rhetorical structures that can be manipulated according to the conditions of use.
Focus	Register analysis	Social factors that reveal themselves in textual structuring.	Textual organization in terms of moves (and steps).			Genre knowledge of the writer, academic genres. Change of genres.
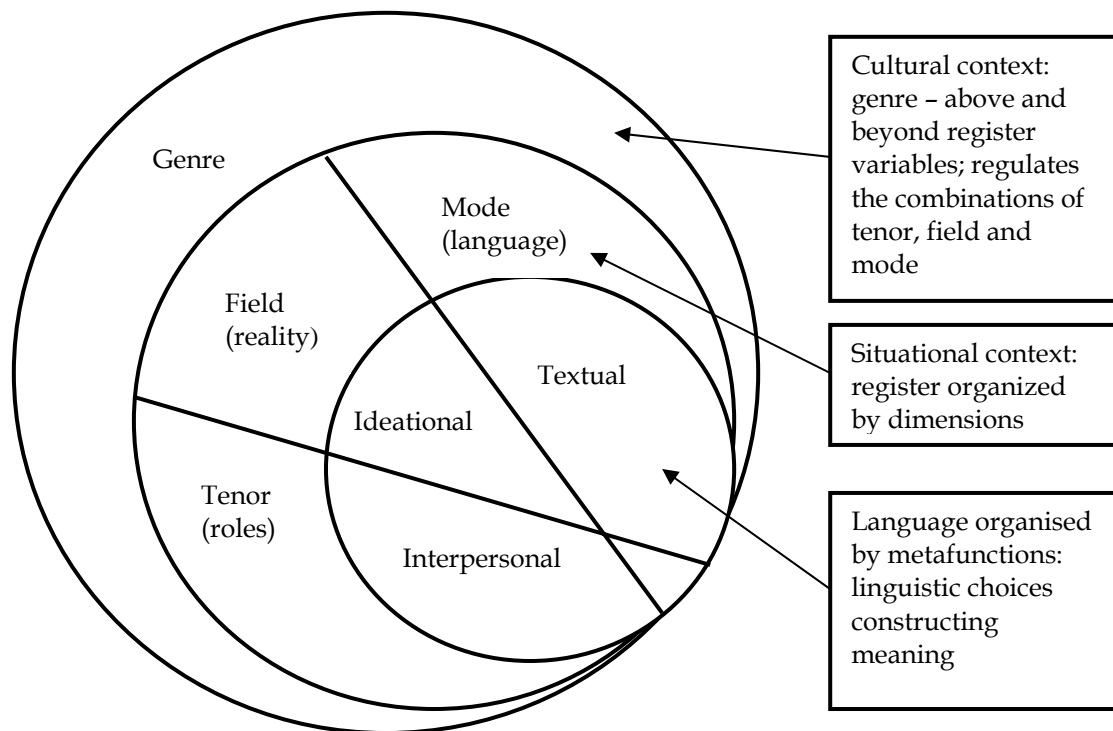
FIGURE 1   Genre in relation to register and language (adapted from Eggins & Martin 1997).

factors (see Figure 1). The concept of register is used to describe the situational context and its relationship with the text. The concept of genre, then, accounts for the cultural context of the text which is beyond and above register. From the perspective of the systemic functional approach analysis begins with the text, and its language is studied in terms of linguistic choices constructing interpersonal, ideational and textual meanings that correlate with the dimensions of the situational context, namely tenor, field and mode of discourse (Eggins & Martin 1997; see also Martin 1993; Martin & Rothery 1993.)

While mode, tenor and field are used in (analyzing the) organization of the context, they are realized in the text as linguistic choices. Tenor, field and mode condition interpersonal, ideational and textual meanings in the text, but on the other hand these meanings – and the linguistic choices made to express these meanings – at the same time construct different kinds of tenor, field and mode. It is thus presumed that the language used in a text can be predicted from the situational context, and on the other hand context can be deducted from the language used. The concept of genre, which sometimes seems to be neglected in favour of register within these studies, is used in describing which combinations of tenor, field and mode are possible in a culture and how the possible combinations work as goal-oriented social processes. (Eggins & Martin 1997; see also Martin 1993; Martin & Rothery 1993.)

Within the ESP approach, a great deal of what has been done is based on the pioneering work of Swales (e.g. 1990). Swales, like others relying on his work, acknowledges that it is the communicative purpose that defines which genre certain texts belong to (see Table 3). Because of their shared communicative purpose, texts also share other similar features in terms of e.g. structure, style and content. (Swales 1990). Since studies within the ESP approach aim to improve and develop writing instruction in various distinct areas of expertise, the concept of discourse community is central. Discourse community refers to a group of people whose members share a set of communicative goals. A discourse community possesses and uses one or more genres in furthering its aims, genres which create discoursal expectations and constraints. (Swales 1990.) The practical work done within this approach has focused on certain distinct genres used by certain discourse communities, such as research articles in an academic context (Swales 1990), sales promotion letters and legislative provisions (Bhatia 1993), and the discussion section of the academic dissertation (Dudley-Evans 1994) just to mention a few central studies in this field.

The analysis of genres within the ESP approach begins with texts and analyzes what kind of *moves* they consist of and how each move is realized linguistically. Moves are described in terms of communicative purpose. For instance, in the introduction of an academic research article, the first move, according to Swales (1990), is that of *establishing a territory*. The writer may have several strategies, i.e. *steps*, to choose from in realizing the move. In this case the writer may establish his territory within the research area by claiming centrality, or making topic generalizations, or reviewing previous research, or any combination of the three steps. The analysis then goes on to look at sentence level in more detail.

Within the sociocognitive approach the focus is also on academic genres and on the writers' point of view. However, the approach differs somewhat from those presented above in its theoretical emphasis. The concept of genre knowledge that writers possess and use is introduced and conceptualized in terms of situated cognition (based on Brown et al. 1989). The sociocognitive theoretical framework of genre includes five principles describing genre, namely dynamism, situatedness, form and content, duality of structure and community ownership. Dynamism refers to the fact that genres are dynamic forms that give coherence and meaning to experiences but also change according to the sociocognitive needs of their users. Situatedness means that genre knowledge is the result of an enculturation through participating in a community's practices as the role of an apprentice. Genre knowledge includes knowledge about both form and content, i.e. which topics in general are appropriate in a certain situation and in which form the topic is appropriate. Duality of structure refers to the fact that in using genres we both form and reproduce social structures. Community ownership means that genre reflects a discourse community's norms and ideology. (Berkenkotter & Huckin 1995.) Despite the different theoretical framework from that used by other approaches

in genre studies, it is evident that the five principles described above have many similarities with central concerns of other genre studies. It is generally acknowledged, for instance, that genres change according to changes in communicative purpose, and that genres are employed by the discourse communities whose values and norms are reflected and reproduced in those genres. However, the sociocognitive approach offers interesting insight into the process of learning the discourse rules of the expert community in its description of the process in terms of situated cognition and apprenticeship.

Since genres are used by specific discourse communities they are related to the professional expertise of such communities and seen as professional practice within an academic discipline or professional domain (Berkenkotter & Huckin 1995; Bhatia 2004). Genres thus tend to be discipline-specific or domain-specific. The relation between genres and disciplines (or domains), however, is not clear-cut. While disciplines and registers of language use may, and usually do, correlate with each other, genres often cut across disciplines, especially in the academic world. There are, for instance, genres such as textbooks and research articles that are common to many academic disciplines since they all have common communicative needs which form the basis for the genres. On the other hand, these genres also show disciplinary variation. For instance, the rhetorical strategies and modes of meaning making exploited in textbooks within business studies and law are quite different. (Bhatia 2004.) The same phenomenon can be seen in everyday genres as well. For instance, newspaper articles represent a genre that is spread widely in across different cultures. There are, however, cultural differences and variation within the genre. This phenomenon is further explored in the next chapter.

It is noteworthy, that especially Swales's model but also the one presented within the sociocognitive framework work well with relatively stable and standardized genres, exemplified above as different academic genres. Everyday genres, however, are more slippery in terms of description. For these written and spoken genres models presented above are too rigid and detailed since everyday genres show more variation according to the writer's or speaker's choices that do not necessarily obey the genre conventions very tightly. In this respect everyday genres show a range of variation. For instance in the news articles or stories in the newspaper readers expect a certain structure and choice of words and react rather easily if the text does not follow their expectations. Therefore the writers of the texts are likely to obey the genre conventions. With texts intended to express readers' opinions in the newspaper, however, there is much more room for individual choices in terms of the structure and the linguistic devices used. In order to describe genres by covering also the variation deriving from the individual writers' and speakers' choices, genres have in some models been presented as a dynamic potential with different choices that are possible to the writers and speakers and form the basis for the next choices as an opposite for linear and more rigid models (Ventola 2006).

The concept of genre colony, on the other hand, has been suggested as a tool for describing a group of genres that have similarities across disciplines

and domains. For instance, promotional genres form a genre colony the members of which share a common communicative purpose on general level. Members of genre colonies may have other communicative purposes different from each other, and they may otherwise have very little in common. Genres representing the promotional genre colony include, for instance, advertisements and job applications, which have distinct communicative purposes at a specific level. Advertisements can further be viewed as subgenres, for instance as TV and radio advertisements, which can further be viewed as e.g. car advertisements, sports club advertisements etc. (Bhatia 2004.) Genres distinguished according to the communicative purpose can thus be studied on many levels of specificity.

To conclude, in analyzing and classifying texts in terms of genre, the main focus of attention is the communicative purpose and function of the text. The linguistic patterns of the text are also studied but only from the point of view of the context of use. The language is studied in terms of how it reflects and constructs both the situational and the cultural context at the same time (e.g. Eggins & Martin 1997) and how cultural context constrains the linguistic realization of genres. There is also a critical side to genre analysis in that it studies the power relations revealed in texts (Kress 1993) and this is implicated in the pedagogical agenda common to each approach in genre studies represented in Table 3. As the genre studies emphasise the functional and extralinguistic criteria in texts, it seems more relevant to the reader than the criteria used in text type classification. After all, readers encounter texts in using them for various purposes and have many expectations of the text based on the context of use and the appearance of the text alone. Furthermore, this approach works with whole texts and offers possibilities to study the cultural aspect in genre variation found important in the assessment context. This aspect will be further explored in the next section.

3.3 Variation in genres

As noted above, all three approaches agree that genre is a social phenomenon in the sense that it defines itself in terms of social situation and communicative purpose. Texts belonging to the same genre are texts that are used for accomplishing a similar task in the culture, they have a similar communicative purpose and they thrive in similar social situations. As a result, texts belonging to the same genre are alike also in terms of structure, style, linguistic organization and to some degree even in terms of content. It is further acknowledged within all three approaches that although genres are relatively stable there is still a fair amount of variation between texts even within the same genre (e.g. Bhatia 2004; Berkenkotter & Huckin 1995; Swales 1990). Stability in genre naturally arises from stability in the situation in which the

genre is used (e.g. Cope & Kalantzis 1993; Kress 1993; Eggins & Martin 1997) or in "the conditions of use" as Berkenkotter & Huckin (1995) express it.

Variation within the same genre can be both diachronic and synchronic. Diachronic variation comes about through historical change in genres. As the world changes, cultures change and thus also the needs of those using different genres for different purposes. The communicative purposes and needs of discourse community members may change radically and this leads to certain genres fading away and new ones emerging. Alternatively, communicative needs may stay the same but genres used to fulfil them change altogether, as has been the case with personal letters being often replaced by e-mail messages that have at least partly different linguistic realization, not to mention the medium. However, genres often change gradually too. The communicative purpose and the task genre is used to accomplish may remain part of the culture while the linguistic realization of genre is adjusted to meet the changing needs of its users over the course of time. (See e.g. Berkenkotter & Huckin 1995; Taavitsainen 2001.) For instance, the scientific journal article has been and continues to be a genre of concern to any member of the academic discourse community. According to a study by Berkenkotter & Huckin (1995) the changes in the context of the journals brought about by the information explosion, have led to a new way of reading scientific articles similar to the way a newspaper is read. The bulk of journals and articles is skimmed through by readers looking for what is new and for most important findings. Therefore readers give most attention to such things as titles and graphic presentations which most summarise the results. Berkenkotter & Huckin (1995) found that over a 45-year period, scientific journal articles have changed gradually to better meet their readers' needs in this respect since among other things titles have become more informative and graphic presentations are more frequently used.

Although genres are relatively stable synchronically they are not without variation. Even though genres create discoursal expectations and constraints, writers are not necessarily bound to obey these conventions. Expert members of the discourse community in particular are likely to exploit and bend the rules to create an effect according to their own individual needs (e.g. Berkenkotter & Huckin 1995; Bhatia 2004; see also e.g. Biber 1989). Synchronic variation within a genre has been described in terms of prototype theory by Swales (1990), who states that "instances of genres vary in prototypicality" (p. 49). The most typical genre members are prototypes. The category of each genre, however, is broad and borderlines are fuzzy. Thus, there may be instances of the same genre that have no common textual features at all, but they are still seen to belong to the same genre so long as they share at least some features with other exemplars of the genre, possibly even with the prototypical text. The privileged feature defining genre members is the communicative purpose, while other features (e.g. form and structure) identify the extent to which a text is prototypical within a genre. (Swales 1990.) The exemplars of a genre that are far from the prototype near the borderlines of the category are extreme examples of the

individual writer's choices in bending the linguistic constraints typical of the genre.

In manipulating and exploiting the genre conventions, or "appropriating generic resources" as Bhatia (2004) expresses it, individual writers create hybrid genres, either mixed or embedded. Although most writers usually respect the generic integrity of genres, expert writers especially may take advantage of its dynamic and blurred nature for individual purposes. This results in mixed genres, such as infomercial or advertorial that combines communicative purposes and linguistic resources of several genres in an innovative way. On the other hand, the conventions of a genre may sometimes be under dispute and the generic integrity of the genre thus contested. For instance, the traditional form of legal discourse in the British context seems to be contested with a reform movement aiming for more straightforward legal writing. (Bhatia 2004.) In the Finnish context an example closer to the average people could be the language used by e.g. government officials. There are constant pleas in opinion forums in different media for simpler and clearer expression in the officials' communication to the people.

Genres also exhibit variation across cultures. For one thing, there are bound to be different genres in cultures that are dissimilar. Japanese culture, for instance, is often contrasted with western culture in this respect (Bhatia 1993). On the other hand, due to globalization there are many socially defined tasks and communicative purposes that are common in many, if not in the majority of, cultures. Thus the genres represented by the texts used to accomplish these tasks and purposes are also common. Some of these genres are very uniform. Academic genres, for example, are well-established and standardized and thus the linguistic realizations of academic genres across cultures are quite similar. However, there is also cross-cultural variation within the same genre common to cultures where, say, western influence prevails. The rhetorical structures of languages differ, and thus the linguistic realization of the same moves can also differ, not to mention the linguistic strategies applied. (Bhatia 1993; Swales 1990.) This is especially true with the various genres encountered in everyday reading events.

For the purposes of international reading literacy assessments, the aim has been to choose texts representing genres that are common in all participating countries and cultures in order to ensure that the test is equally fair and authentic for all. This poses a challenge in itself. However, a greater problem arises from the fact that there is cross-cultural variation also within the genres common to participating countries and cultures. The texts – and the test – are hardly equivalent if some read e.g. a news report with typical linguistic realization while others read the same familiar genre with untypical linguistic realization.

In all, the framework of genre gives possibilities not only to describe and classify texts based on extralinguistic criteria but also to study the variation in texts both in the synchronic, i.e. national, level and within the cross-cultural context. In international reading literacy assessments it is to be defined where

goes the limit between the cross-cultural variation acceptable for the readers and the variation deriving from cultural differences not acceptable for the readers in individual national cultures participating in the assessment.

3.4   Multitype texts

Despite the tension between stability and variability in real-life genres, it usually seems to be quite clear to readers which genre a text belongs to. In identifying the genre, readers evidently rely on textual features external to the language of the text (Bhatia 2004) since they often identify the genre before actually reading it. The external format of the text and not just the context of its appearance seems to be an important criterion in determining which genre certain texts belongs to (e.g. Biber 1989). External criteria on the basis of which readers identify genres include for instance visual clues that characterize the format typical to a genre, such as the use of titles and subtitles, font style and size, the use of pictures and graphs, and the use of columns. Readers easily identify texts in terms of everyday labels specifying genres, for instance news report, bedtime story, fable, advertisement, letter, and e-mail. It is noteworthy that in labelling or classifying texts in terms of genres based on text external criteria the process involves whole texts.

The classification of texts in terms of linguistic characteristics and text types does not necessarily involve the whole text, and it is thus noteworthy that whole texts rarely represent purely one text type. Authentic texts are often heterogeneous and consist of sections that may have different functions. Thus different text types are used in different sections of the text. Texts of this kind are referred to as multitype texts by Virtanen (1992) and multigeneric texts by Callaghan et al. (1993).

The focus of Werlich's original work is on parts of texts and the relations between sentences rather than classifying whole texts. There are indeed texts whose main function is to describe, tell a story (fictional or not), explain phenomena, persuade or instruct but in this case whole texts are described in terms of their purpose or function. Functional criteria as such, however, reflect the level of genre (as defined above) and do not belong to Werlich's theory of text types and so must be kept distinct from it. This is not to say that there are not texts whose function coincides with their text type. For instance, the majority of sections in a text describing a phenomenon are likely to represent the descriptive text type. These two need not coincide, however. (See e.g. Virtanen 1992.)

The need to create a text typology that connects both the functional and linguistic characteristics of texts has led to a two-level model of text typology presented by Virtanen (1992). In this model the notion of discourse type is introduced in referring to the level above text types. Discourse type is based on the function or purpose of discourse and it dictates the overall strategy of the

text. Text type then illustrates the level of actual texts, and is characterized as "the aggregate of prototypical surface features" (p.298). Overall, the description of text types relies on Werlich's text types – namely, narrative, descriptive, argumentative, instructive and expository (1983) texts. Discourse types are named using the same labels but on this level the criteria are text-external and refer to the function of discourse instead of its structure and surface level characteristics.

Discourse types and text types do not necessarily coincide in Virtanen's model (1992). Text types can be used secondarily or indirectly to realize other functions than that of the corresponding discourse type. For instance, instructive discourse type can be realized as descriptive text type, i.e. the purpose of the text is to instruct but the text describes the process rather than gives direct instructions. According to Virtanen (1992) narrative text type, for example, can realize any type of discourse and thus fulfil many functions or purposes. Conversely, argumentative discourse type can be realized as a range of different text types.

Within the systemic functional approach in genre studies, the two levels of text typologies have been combined by Callaghan and colleagues (1993). Genres are described as social processes that describe, explain, instruct, argue or narrate. The processes are commonly used in text types that are described as products of the processes involved. In contrast to the view presented by Martin and his associates (Martin 1993; Eggins & Martin 1997), who concentrate on texts and label texts as genres, this view stresses the social origin of genres and labels the products of the process as text types. Text types are then something concrete that students, for example, can work with. (Callaghan et al. 1993.)

Although these two attempts to create a text typology that combines functional and linguistic criteria have different theoretical frameworks – that of text type studies and genre studies – they otherwise have a common aim and solution. Functional criteria form a level above actual texts (discourse type; genre) and text types form the concrete level with linguistic insight. However, discourse type is defined in terms of purely functional criteria and genre within a wider social context as a process. Nevertheless, there seems to be a need to combine the two levels of text analysis, reflecting the difficulty of keeping the form and function of text apart. In international reading literacy assessments, both levels of text classification need to be taken into account, as was done in PISA 2000. Rather than applying the notion of genre, the functional criteria of text classification were covered by the classification of texts according to the situation for reading. Covering also the functional criteria of text classification is consistent with the notion of functional literacy applied in PISA as well as other reading literacy studies (Chapter 2). It is, however, more important to include both the linguistic and the functional criteria in the text classification for reading literacy assessments than to hang on to certain labels.

4   AUTHENTIC TEXT

4.1   The attributes of authentic text

In recent international reading literacy assessments the aim has been to choose authentic texts (and tasks) for the test. Traditional carefully written stimulus texts have evolved into large variety of real-life texts as the authenticity of test material has become an essential issue in test design. In this section the aim is to explore what kinds of texts authentic texts are and what kinds of features, if any, they share.

The Oxford Advanced Learner's Dictionary defines *authentic* primarily as something "real and genuine", the opposite of a copy. Moreover, conceptions of authentic text within the fields of instruction, learning and assessment reflect the idea that authentic text is always genuine in the sense that it has been written for a genuine communicative purpose rather than being a carefully designed and written stimulus for instructional or assessment purposes. Genuine text is thus real-life text that has not been produced for pedagogical or testing purposes only (Lewkowicz 2000). Lately there has been lively discussion about other features of authenticity. In fact, definitions of authentic text have evolved to the extent that one can ask whether genuineness is an absolutely necessary condition for text authenticity. (E.g. Widdowson 1978, 1979; Bachman 1990; Bachman & Palmer 1996.) This question will be returned to at the end of the section.

In the Literacy Dictionary authentic text is defined as "text that has not been altered in form or content" and as "representative of the real world" (Harris and Hodges 1995). Both definitions relate to essential features of authentic texts. The first definition emphasizes that authentic text is written and used for a genuine communicative purpose, and when used in instruction or assessment it is not altered or edited in any way for instructional or assessment purposes. This feature of authentic text is clearly associated with the genuineness of the text. The latter definition in the Literacy Dictionary relates to

the diversity of texts used in instruction or assessment. It is emphasized that the selection of texts should reflect the real world in that the whole range of texts that readers may face in their lives can be used in the literacy assessment or brought into the classroom.

The interpretations of authentic text applied in the most recent international reading literacy studies, PISA and PIRLS, reflect the above definition of authentic text by referring to typicality and the relevance of test material to *everyday life* and *the real world*. More specifically, PISA aims to simulate authentic reading situations and therefore the texts selected for the test represented reading situations and text types relevant to 15-year-old students' everyday life in and out of school (OECD 1999; 2003; Linnakylä & Sulkunen 2002). For PIRLS also, authenticity was central in choosing text passages for the test, as is indicated in the following quotation: "*To reach the goal of approximating an authentic reading experience in the assessment, the reading passages presented to students must be typical of those read by students in their everyday experiences.*" (Framework and Specifications for PIRLS Assessment 2001, p. 36.)

In addition to the real world and everyday connotations of authentic text, the PISA and PIRLS definitions of authentic text described above also reflect other features found essential to text authenticity–relevance and typicality. Authentic text is found to be meaningful and relevant in some way in the reader's world of texts, and as such typical of a reader's culture (Valencia et al. 1994; see also Brown et al. 1989; Newmann et al. 1996 on the nature of authenticity). Basically this means that the reader has a need or a desire (purpose) to read the text, preferably other than that a teacher has prescribed it. The reader has an interest in the text (Weiss 1994). It is noteworthy that these definitions of an authentic text emphasize that texts should be typical and relevant to the students, not just any genuine real-life texts. It is important to choose texts that are authentic in the life of the readers in question. After all, it is always the individual reader who defines whether or not a text is authentic to her/him. (Bachman 1990; Widdowson 1979; Bachman & Palmer 1996; Garcia & Verville 1994; Weiss 1994; see also Brown et al. 1989; Honebein et al. 1993; Newmann et al. 1996 on the nature of authentic activity.)

In choosing texts for instruction or assessment purposes it has been further stressed that texts should be not only genuine and meaningful but also *appropriate* as well as *interesting*, *engaging* and *rich* for readers (Garcia & Verville 1994; Kapinus et al. 1994; Weiss 1994). It has also been stated that the concepts and *language* used in texts should be *familiar* to students since it is imperative that texts and other test materials are accessible to most, if not all, students. As for texts and their features, it has further been argued that in addition to familiar language the text selection should cover a wide range of *topics* and *genres* (Garcia & Verville 1994; see also Kapinus et al. 1994).

The various attributes of an authentic text can be listed in summary form as follows:

- Naturally occurring, genuine, and unedited
- Typical and familiar in the reader's culture
- Meaningful and relevant to the reader
- Representative of the real world
- Appropriate for the reader
- Interesting, engaging, and rich
- True need or purpose for reading

The attributes of an authentic text seem to be partially overlapping but also quite different from each other. Attributes that are more or less the same appear on the same line in the list; typical and familiar, for instance, refer to basically the same feature of a text, as do interesting and engaging. Typical and interesting, on the other hand, are quite distinct features. In fact, the two features may be in tension with one another, since what is typical and familiar can hardly be very interesting since it holds no novelty value. However, typical texts are necessarily interesting to some degree; otherwise they would not be read frequently enough to become typical. To continue this line of thought somewhat further, one could argue that rather than being interesting typical texts can be relevant and meaningful. Relevant and meaningful as attributes of authenticity derive from a true need to read the text, and it has been argued that as a consequence of this need, readers become interested in it (Weiss 1994). In order to illustrate better the relations between the various attributes of authentic text, they are presented in the Figure 2 below.

As Figure 2 shows, the primary attribute of an authentic text is that the reader has a true need or purpose for reading it. The nature of the purpose for reading the text is unimportant as long as it is genuinely there. It is the reason why the reader reads the text which is the necessary condition for text authenticity. Because of this need or purpose, a text becomes meaningful to the reader and the reader takes an interest in it and wants to know more about it. Typicality as an attribute of text is clearly associated with this line of thought, but also with the frequency of encountering texts of the kind. Texts of a kind that the reader frequently encounters are typical in the reader's culture and familiar to the reader. Meaningfulness, interest and typicality are then related to each other as attributes of authentic text. The attributes of being a real world text and genuineness are also related to each other and derive from the reader's true need for reading the text. Since the need for reading the text is a genuine communicative purpose, the text too is a real-world text written for a genuine communicative purpose. Real-world texts are not only genuine in terms of communicative purpose but also in terms of other qualities: they reflect and construct the world surrounding them (e.g. Fairclough 1992) by means of the discoursal devices typical to different kinds of text (e.g. Swales 1990). Texts do this despite the fact that they become decontextualized (Widdowson 1990) when brought into the classroom for instructional or assessment purposes.
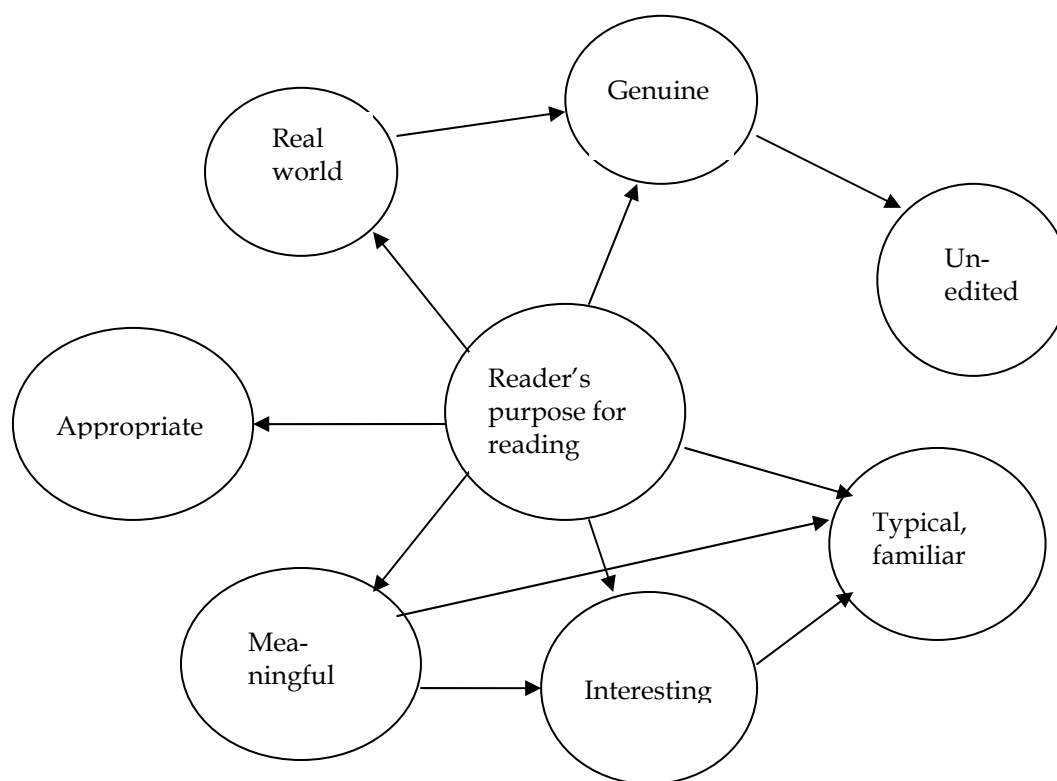
FIGURE 2   Attributes of authentic text in relation to each other.

The requirement that authentic texts should not be edited for instructional or assessment purposes is associated with the genuineness of real-world texts. The idea is that real-world texts are brought into the classroom as they are and were in their original context of appearance. Appropriateness as an attribute of authenticity is a somewhat isolated requirement in relation to the other attributes of an authentic text and echoes the concerns that students should not be presented with material unsuitable to them. In a way this is contradictory to the ideal of authentic text since it gives the power of choosing the texts to someone other than the reader. This is very common in instructional contexts, since even if authentic texts are brought into the classroom, they are usually chosen by teachers, not students themselves. However, texts deemed appropriate by a teacher are not necessarily meaningful and interesting to the students, who thus may not authenticate (Widdowson 1990) them, i.e. engage with the texts in a meaning making process. If appropriate texts are chosen bearing in mind the other requirements for authenticity, they can be or become genuinely authentic for the students.

As has become evident, the concept of authenticity involves many attributes other than genuineness. Recent discussion of authenticity has concentrated on these attributes but also stressed that the authenticity of a text should be defined in terms of interaction between the text and the reader (Widdowson 1978; Bachman 1990) rather than as a quality of the text alone. In

line with this reconceptualisation (Lewkowicz 2000), authenticity has been distinguished from genuineness, and it has been stressed that genuine text is not necessarily authentic. Consequently the notions of situational authenticity and interactive authenticity have been presented by Bachman (1990), who in the context of language testing defined *situational authenticity* as the correspondence between target language use (in this case real-life texts) and test material and *interactive authenticity* as interaction between the text and the reader, as applied to texts. It has been stressed that authenticity is a relative concept (Bachman & Palmer 1996), and thus we can not classify texts as authentic or inauthentic but rather as less or more authentic, thus placing them on a continuum of authenticity.

These definitions of authenticity and authentic text suggest that the essential feature of an authentic text is that as a meaningful and relevant real-life text it engages the reader in an active process of meaning construction. Thus, it is necessary to return to the question whether genuineness is a necessary condition for authenticity of text. For instance, to take a real-life example, a primary school teacher herself wrote a text for instructional purposes. The text referred to places familiar to her second grade students and the students were to put capital letters in the text where appropriate. The teacher reported that the students were enthusiastic about working with the text that was meaningful and relevant to them. Thus the text in question was not a genuine real-world text but still it seemed authentic to the students. Of course, instructional texts such as the text above are authentic texts too, and such texts can genuinely be written by the teacher. However, it has become evident by now that the idea in choosing authentic texts for instruction is to reflect the real-life world of texts. From this perspective, genuineness is still an essential attribute of text authenticity. In the above example the text written by the teacher was genuine – and authentic–only as an instructional text. Instructional texts form a quite distinct group of texts relevant to students' lives (see e.g. Breen 1985) but since the idea is to widen students' textual horizon to bring in real-life contexts such texts are not authentic in the sense meant in discussion of the use of authentic texts in instruction. Genuine text is not necessarily authentic but authentic text is always genuine. Therefore there is no need to separate genuineness from authenticity altogether as long as it is noted that authenticity involves much more than just genuineness.

Many attributes of authentic text are better understood in their context of appearance. Therefore the role of authentic text in instruction will be further described in the next section. The above discussion of genuineness as an attribute of authentic text will be returned to when the rationale behind using authentic texts in instruction and assessment becomes clearer.

4.2 Authentic text in instruction and learning

In the previous section, the attributes of an authentic text were described. However, in order to fully understand their implications the appearance of the notion of authenticity in learning theories and instruction will be described briefly. The main focus of this section will be on explaining the rationale behind using authentic texts in instruction, particularly literacy instruction.

Within theories and views of learning and instruction, the notion of authenticity has derived from the emphasis on *contextualized use* of knowledge instead of merely acquiring knowledge in isolation from its context of use. For instance, in terms of the socio-constructivist view of learning, students' inability to apply the knowledge acquired at school lay behind the model of learning through cognitive apprenticeship (Brown et al 1989). The main idea was that learning is a result of acting in situations. Thus learning new concepts is a process of enculturation during which students, like apprentices, should be exposed to the culture of experts and to their *authentic activities i.e. the ordinary practices of the community*. Teachers in the role of experts should guide students and familiarize them with the relevant concepts and activities. It was argued that in this way students learn both the new concepts and their authentic use in the relevant situations and culture. (Brown et al. 1989.) Situated cognition thus grounds education in the practical world and ties learning to real world experience. This gives students a sense of agency and also helps them in the early stages of learning to see how the knowledge they are building up can be applied. (Palincsar 1989). The practical applications of the notion of cognitive apprenticeship have created learning methods and environments which simulate expert processes of knowledge construction and capture the essential authentic characteristics of the expert cultures of each discipline (Collins et al. 1994, Edelson 1998, Honebein et al. 1993).

Within the field of reading literacy and texts, the idea of cognitive apprenticeship has been applied by the sociocognitive approach of genre studies, which offers interesting insight into the process of learning the discourse rules of the expert community, referred to as genre knowledge (see Table 3, p.25). It is emphasized that genre knowledge is the result of enculturation through participating in a community's practices in the role of an apprentice. (Berkenkotter & Huckin 1995.) Current socio-cultural literacy studies take the idea of enculturation further by stating that reading and writing are social practices used for specific purposes in specific contexts and thus have functions that go beyond the mere acts of reading and writing (see Table 1, p. 11). Part of literacy instruction is thus enculturating students to the culture of reading and writing and to the practices involved. This, however, seems to shift the focus of attention away from the text in literacy activities by emphasizing the social aspects of reading and cultural literacy practices (Barton 1994) that define the ways people use literacy for different purposes. On the other hand, social views on literacy have stressed the different uses and

functions different texts have and resulted in demands for the use of authentic texts in literacy instruction (e.g. Pitkänen-Huhta 1999; Street 1996; Luukka 2004a; Luukka & Leiwo 2004). Authentic texts in reading literacy instruction familiarize students with the large variety of literacy activities and the large variety of different purposes and texts. During such a process of enculturation students learn to know different kinds of texts and at the same time the contextualized uses of the texts with discourse rules and literacy practices intertwined and reflected in the texts. The essence of such instruction is preparing students for the challenges of real-life outside the school context (see also Luukka 2004b).

Similar issues with regard to authentic text have emerged in foreign language (FL) instruction. The communicative approach to language instruction shifted the focus of attention from views of language merely as isolated skills of reading, writing, speaking and listening (Canale & Swain 1980; see also Huhta 1993, Bachman 2000) onto the *use* and *functions* of language. It has been emphasized that if students are to learn to use a language within real-life situations (and thus to gain communicative competence), the language teaching methodology should engage students in natural language use in the classroom (Canale & Swain 1980; see also Widdowson 1978, 1990). Overlooking language use in instruction would place language too far from the learner's own experiences and needs (Widdowson 1978).

Simulating natural language use in the classroom resulted in using authentic material in instruction. During the early years the only criterion for authentic material, e.g. texts, was its genuineness. This view was, however, strongly criticized and it was emphasized that authenticity is not a feature of the text but of the interpretation that the reader makes of it (Widdowson 1979). That is, authenticity "is a function of the interaction between the reader/hearer and the text which incorporates the intentions of the writer/speaker" (p. 165). Shared conventions of the discourse community, for instance genre knowledge (Berkenkotter & Huckin 1995), help the reader understand the writer's intentions. This response to a text, however, is not automatic. From the individual reader's point of view, text reflecting the shared conventions of the community and thus being genuine may not be authentic should the reader in question not realize the writer's intentions. Therefore, it was argued that a genuine text, e.g. a novel, may still be inauthentic in the classroom if it is treated only e.g. as providing grammatical examples. (Widdowson 1979.)

Later on, the idea was developed further and the authenticity of *material* used in the classroom and the authenticity of *language use* simulated by tasks linked to the material were distinguished from each other. It has been pointed out that while it is possible to use authentic material in the classroom it is not, however, always possible to use it in an authentic way. Once the material, e.g. texts, is brought into the classroom it has already been decontextualized and thus its authenticity is reduced. A greater problem in terms of authentic instruction arises from the fact that it is sometimes necessary to use language material in the classroom inauthentically in order to learn to use it

authentically. It is still accepted that authentic material should be used since it motivates students. (Widdowson 1990.)

Despite the theoretical problems described above, the social views on literacy have given new insights into the use of authentic texts in language instruction, similar to those in mother tongue instruction. The aim of using authentic texts in literacy instruction is to enculturate students into the large variety of real-world texts and practices in order to provide thems with the knowledge and skills to cope with different texts in the world outside school. (See e.g. Leiwo 2005; van Lier 1996.) It has been stressed that due to the variety and diversity of real-world reading literacy practices, enculturation into literacy practices should be developed through a variety of different experiences in different contexts and with different kinds of texts, paying attention to the relations between text and the purposes as well as ways of reading (Kern 1995).

Since it is impossible to present students with all possible types of texts they may or may not encounter later in their lives, an essential part of authentic literacy instruction is to enhance students' awareness of the processes, practices and above all motives and purposes that are behind textual forms of different kind. (Leiwo 2005; van Lier 1996.) Most importantly, students should become aware of the extra linguistic issues related to texts: who is the writer, what is the purpose of the text, what kind of linguistic and stylistic devices have been used and why, and what kind of values and ideologies the text reflects and whose values they are (Pitkänen-Huhta 1999; see also Luukka 2004b, Leiwo 2005). Critical awareness of these issues prepares students to meet the demands of new literacy practices irrespective of the language taught. All this necessarily means that authentic texts have to be dealt with inauthentically in instruction at least to some degree. After all, authenticity is something students should achieve, not something they necessarily need to start with. (Widdowson 1990.)

In instructional contexts using authentic texts in the classroom aims ultimately at empowering students in terms of language use in the various areas in society. Empowering students by instruction involves autonomy, awareness and authenticity: instruction aims to produce autonomous language users who are aware of the extra linguistic factors constraining the style and form of the text and who produce authentic texts in the sense that they take *responsibility* for the texts, both content and form. (van Lier 1996; Leiwo 2005; see also Taylor 1992 on self-responsibility related to authenticity). From this perspective, the future orientation often associated with authentic texts in instructional but also in assessment contexts (Bopry 1999; Leiwo 2005; OECD 1999, 2003; Luukka 2004a, Luukka & Leiwo 2004) is well argued for since the educational goal relevant in this context is to prepare and empower students for their future lives as competent language users and full members of society.

Against this background some attributes of authenticity described in the previous section acquire a new perspective. Genuineness is one of these. In order to enculturate students to the numerous real-world texts and practices involved it seems clear that they need to be presented with various real-world texts that are necessarily genuine. Although empowering students to school

texts is also important (Leiwo 2005) it has become evident that school texts are often very distinct in nature and unlike any real-life texts in e.g. information density or cohesion devices (Karvonen 1995; also Leiwo 2005). Therefore the emphasis on empowering students relates to the out-of-school context, since the idea is to provide students with awareness, knowledge and skills to meet the challenges they will face in their lives after and out of school. The demand that real-life texts should not be edited for instructional purposes is in accordance with genuineness: editing texts would reduce their genuineness.

Demands for meaningful, interesting and typical text in terms of authenticity on the other hand seem to concern text qualities that are necessary in order to get readers to engage in a meaning making process with the text, i.e. authenticate it (Widdowson 1979). They derive from the true need for reading a text, a true communicative purpose. In an instructional context, where texts selected for the class are often chosen by the teacher, this clearly creates a challenge. The teacher can provide students with texts and then try to create the need for reading them or he can simply create a need for reading a text, i.e. a task, and then give the students latitude in choosing the text themselves either from a selection of texts or independently. However, in literacy instruction the main objective is to enculturate students to the large variety of real-world texts and literacy practices around them. Thus, it is important to present students with authentic real-life texts. In making visible the literacy practices involved in the texts in question and the whys and hows behind them, it is often necessary to approach texts inauthentically. Therefore, compromises with meaningfulness, interest and typicality as attributes of authentic text may have to be made. On the other hand, if these texts represent topics relevant and interesting to the students they may engage students in a true meaning making process. Furthermore, the relevance of texts that the students at first find inauthentic may increase if students find the aim of instruction as explicated by the teacher to be meaningful.

4.3 Authentic text in assessment

In the field of assessment, particularly in testing, the issues related to authentic text are to some extent similar to the considerations in instruction. The development of traditional tests with multiple choice tasks and carefully written stimulus texts into more authentic forms of assessment derives from the developments in instruction and the consequent inconsistency between instruction and assessment.

In the context of testing, the demand for authentic assessment first arose in the field of language testing (see e.g. Shohamy & Reves 1985, Spolsky 1985) after the discussion of communicative competence and FL instruction described above. At first the discussion of authentic assessment was concerned mostly with the genuineness of testing material, that is the texts (oral as well as written

ones) and tasks used in tests. It was considered important that the language produced in language tests should reflect real-life language use and it was believed that genuine input material was necessary in achieving this goal. (Shohamy & Reves 1985, Spolsky 1985, see also Lewkowicz 2000.) Although quite early on there were ideas about genuineness being an inadequate criterion for the authenticity of a text (Widdowson 1978) and views emphasizing that authenticity is essentially characteristic of the reader's response rather than the text itself, these considerations were not readily accepted and teachers' as well as researchers' interest was focused mainly on using genuine material for quite some time. Genuine text was understood as real-life text that was not produced for pedagogical or testing purposes only (Lewkowicz 2000).

The theoretical problems relating to authentic tests were much the same as those in FL teaching. Attempts to design authentic tests that duplicate real-life language use situations as closely as possible were criticised by pointing out that language in a testing situation can never really capture real-life language use, no matter how authentic the tasks and texts are, since recontextualising texts, etc. will reduce their authenticity. (e.g. Shohamy & Reves 1985; Widdowson 1990). Attention was also drawn to the test taker's role in participating in tests that simulate real-life language use (Spolsky 1985; also Widdowson 1990). It was stressed that such simulations can only succeed if test takers "play along" and treat the testing situation as authentic language use. It was also questioned whether it is always meaningful to use authentic test material, since such texts may be inappropriate for the language level of the test takers (Davies 1988). This concern echoed views that it may be necessary to use simple, adapted texts in inauthentic ways while learning to use a foreign language in an authentic situation (Widdowson's 1979). It was pointed out that authenticity is what language learners should achieve, not something they should necessarily start with. On the other hand, it was suggested that inauthentic language use in instruction and assessment might be considered authentic language *learning* behaviour since there are various areas of authenticity (Breen 1985).

Only in the 1990's did the discussion of authentic language testing take a step forward. This was partly fuelled by the demand for authentic assessment in general education emerging at the same time, especially in United States, which has been characterised as reconceptualising authenticity (Lewkowicz 2000). The problem of recontextualising genuine test material resulting in reduced authenticity (Bachman 1990) was solved by redefining authenticity. It was stressed that instead of copying the whole language use situation researchers should pay attention to the essential features of communicative language use. Thus, a more complex notion of authenticity evolved once situational authenticity and interactive authenticity had been distinguished from each other. Situational authenticity was defined as the correspondence between test tasks and target language use (TLU) tasks and interactive authenticity as the interaction between the test task and the test taker. (Bachman 1990.) Later on interactive authenticity was distinguished from the

notion of authenticity and renamed interactiveness (Bachman and Palmer 1996).This view focuses on the tasks used in tests instead of the texts. It is not necessary to copy the essential features of authentic texts in testing since it is possible to bring authentic texts as such to the testing situation. Thus, authentic text can be seen as one essential feature of language use in case of reading literacy. This view disposes of the problem of recontextualising texts, since as it can be argued that authentic text is an essential part of the language use situation simulated in language testing, the extent to which it is authentic in testing very much depends on the tasks associated with it. Since authenticity is characteristic of the interaction between the text and the reader rather than the text alone (e.g. Widdowson 1978), the focus should be on tasks that define the purpose of reading the text in a test.

During the last decade the discussion of authentic language assessment has widened from authentic language tests to national and international large-scale assessments, e.g. NAEP or PISA, and classroom assessment. With respect to international large-scale assessments the discussion has concentrated on using authentic real-life stimulus texts and tasks as described above in section 4.1 (see also OECD 2003) in order to create tests that are valid and capture the whole range of students' performance. Authentic classroom assessment on the other hand has been developed as an alternative to more traditional forms of assessment usually consisting of standardised multiple choice tests that are considered too narrow and unable to capture the full range of students' competence. Thus, authentic classroom assessment is also referred to in the U.S context as alternative assessment. (Gottlieb 1997, Kohonen 1997, Pollari 1997; see also Hiebert et al. 1994.) In the Finnish context the situation has been somewhat different since there has not been such a tradition of testing. It is noteworthy that both developing international large-scale assessments in the direction of greater authenticity and authentic classroom assessment have similar aims and starting point. However, international assessments such as PISA can only be developed within the limitations of their large-scale nature. Thus the form of assessment continues to be testing, which even in its authentic forms must still be reliable and psychometrically valid. Classroom assessment, on the other hand, does not have such restrictions and can employ any forms and devices of assessment needed or desired.

Authentic classroom assessment differs from traditional testing in many ways. It emphasizes the student's progress in achieving his goals and the learning process itself rather than just the learning outcomes, student's active role in learning and (self) assessment, contextualised and complex problem solving tasks and meaningfulness to the students' life in or out of school. (Valencia & Pearson 1987; Wiggins 1989; Linnakylä & Kupari 1996; Linnakylä & Välijärvi 2005.) Authentic classroom assessment values tasks that represent classroom curricula goals and instruction and emphasizes the communicative meaningfulness of evaluation in terms of feedback and feed forward to students. This type of assessment, moreover, is committed to measuring what is valued in education. Regardless of the form of assessment (e.g. teacher-made

test, learning logs, essays, diaries, portfolios), authentic assessment aims to improve individual performance and it is seen as an integral part of instruction. (Kohonen 1997, p. 13–14.) Students as well as teachers are usually involved in such assessment by giving them both a central role in developing, interpreting and using the assessment which has been reported to have positive effect on learning and instruction through feedback and feed forward (Hiebert et al. 1994; Linnakylä & Välijärvi 2005).

Forms of authentic assessment vary a great deal. In the context of language instruction, be it first, second or a foreign language, it can include e.g. performance assessment, such as teacher-made exams or essays and other writing tasks, learning logs, journal entries, diaries, projects of various kinds, language portfolios and self-assessment, or any combination of those (see e.g. Hiebert et al. 1994). In reading literacy assessment, the portfolio has been used a great deal as an authentic collection of student work, since it combines many features of authentic assessment, one of the most important being the on-going nature of collecting the evidence of both the student's learning process and its outcomes, and perhaps most importantly the student's reflection on the work included in the portfolio. (See e.g. Hancock et al. 1994; Hansen 1994; Valencia & Place 1994.) Portfolios are considered to be a multifaceted source of information about students' learning but also a valuable source of feedback both to teachers and especially to students who in traditional tests receive little if any feedback on their progress (Hiebert et al. 1994; see also Marzano 1994).

In addition to the use of classroom portfolios, large-scale portfolio assessment for reading has also been developed in some U.S. projects both at district and at state level (Garcia & Verville 1994; Valencia & Place 1994; Weiss 1994). These projects aim to combine authentic classroom assessment and reliable large-scale assessment. Portfolios of this type consist, for instance, of a selection of texts that students respond to independently, work in small collaborative groups to explore the text further and write an essay related to the text read (Weiss 1994) or high-quality literature with various tasks appended, such as sentence and chart completion, true/false statements, drawing and writing essays (Garcia & Verville 1994). Current reading literacy assessments similarly aim to develop reliable large-scale assessment into a more authentic form of assessment. The assessment device has continued to be a test, but the selection of texts and the nature of the tasks appended have become more authentic. The tasks in PISA, for instance, include multiple choice items but also open items requiring students to construct their own responses, some short some longer ones. Open constructed response items were a majority task type in PISA 2000. Multiple choice items were not considered to be as authentic as these, albeit more reliable to score cross-nationally. (See e.g. Framework and Specifications for PIRLS Assessment 2001; OECD 2003; Linnakylä & Sulkunen 2002.)

A challenge in using large-scale portfolios is selecting the texts to be read by all the students assessed, since texts should be about issues equally familiar to all students and written in a language equally familiar to all in order to avoid

bias (Weiss 1994). The challenge then is to find texts equally authentic to all participants, a challenge which is common to all forms of reading literacy assessment aiming to compare the results of the assessment at the individual or group level. It has been noted that although authentic assessment, e.g. the use of portfolios, consists of tasks that are more complex, realistic and appropriate for students and reflects contexts that are more natural, in many cases they include fewer tasks than traditional tests. Therefore, the limited selection of tasks may not reflect students' literacy very well, in which case the results should not be generalized beyond assessment context. (Valencia et al. 1994.) Overall, it is recommended to use multiple sources of information to ensure both reliability and authenticity of assessment, especially where a great deal is at stake for the participants.

To summarize, it seems that the role of authentic text in assessment is twofold. First, the demand for authentic texts (and tasks) is central in the context of testing in which it is stressed that testing material should include genuine real-life texts. Authentic tests would then be more valid since they cover the whole range of students' real-life competencies and reflect their natural language use outside the testing situation. Authentic tests are also a more direct form of assessment than traditional tests. Additionally, it has been emphasized that authentic test materials motivate students. International reading literacy assessments, e.g. PISA, belong to this category since the form of assessment has continued to be the test although this has been developed in the direction of greater authenticity in terms of texts and item formats. Secondly, there is the issue of authentic texts in the context of classroom assessment. Within classroom assessment the authentic text is not in focus, since authenticity in this context concerns the nature of the formative assessment supporting learning rather than developing assessment materials. The assessment materials in authentic classroom assessment are produced by the students based on the tasks and projects teachers ask them to do. In that sense all students' work at school is authentic within an instructional context. Of course, teachers can choose to use authentic texts as stimulus material in classroom tasks and projects. Tests can also be used in authentic classroom assessment, however, in which case the discussion of authentic texts in testing become relevant also to classroom assessment.

4.4   Text authenticity and the validity of authentic assessment

The expanding interest in authentic assessment has led to a discussion about the validity of such assessments, especially in the classroom context. As the discussion in the previous section implies, authenticity and reliability of assessment are often considered to be incompatible properties of assessment. It is then asked if there can be validity without reliability, i.e. can authentic assessment be valid if it cannot guarantee reliable results (e.g. Moss 1994). This

discussion is mainly concerned with authentic classroom assessment since in authentic tests in, for instance, international reading literacy assessments authenticity is subordinated to reliability and comparability rather than the other way around.

In the field of language testing, authenticity is taken to be one of the main criteria in evaluating the usefulness of a test, and it is argued that authenticity is closely connected with the notion of validity. Conclusions drawn from a test score are found valid only if they are appropriate and meaningful. One aspect of construct validity is the extent to which test results can be generalised. This depends on the authenticity of the test because test results cannot be generalised beyond the testing situation unless the test reflects language use outside that situation. (Bachman and Palmer 1996; also Spolsky 1985; Wu & Stansfield 2001.) This view of authenticity emphasizes that authentic assessment is valid since it can capture the whole range of students' competencies in actual real-life performances (see also Huerta-Marcias 1995; Shohamy & Reves 1985; Wood 1993). Authenticity is seen as an important aspect of content validity as well, since the authenticity of test tasks can be seen as their relevance to natural language use situations (Bachman and Palmer 1996; also Wu & Stansfield 2001). Additionally, authenticity can be regarded as relating to face validity, which refers to the appearance of the test from the point of view of the test-takers. The face validity of a test helps to obtain the co-operation of the test-takers since it means the test will appear to be relevant to them. (Linn 1993; Linn & Gronlund 1995.)

Traditionally validity was seen as a test's ability to measure what it is intended to measure (Nummenmaa et al. 1996; also Moss 1992). Thus, validity was a property of the test. In the 1970's the Standards for Educational and Psychological Testing defined validity as "the appropriateness of inferences from test scores or other forms of assessment" (APA 1974, p. 25), and was held to consist of three aspects–criterion, content and construct validity. The notions of concurrent, predictive and face validity also emerge in the literature concerning validity issues (e.g. Borg & Gall 1989). A more recent view of validity, based on the work of Messick (1989a), stresses the centrality of construct validity and the importance of the social consequences of test use. According to this view validity is to be seen as a unitary concept and defined as "an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (Messick 1989a, p. 13). This definition indicates that both empirical evidence and theoretical reasoning are to be used in evaluating the validity of test use, which is essentially a matter of degree and always context specific. (See e.g. Anastasi 1986, Angoff 1988, Luoma 2001 and Moss 1992 for descriptions of the evolution of the concept of validity.)

Instead of providing content, construct and criterion related evidence of validity a matrix is formed based on the two dimensions of validity (Messick 1989b). The dimension of *function or outcome of testing* can be either test

interpretation or test use, and the dimension of *the justification of testing* can be either appraisal of evidence or appraisal of consequence (see Table 4). It has been emphasized that tests are used in society, and the meanings of test scores are to be studied in the context of their (intended) use. Thus, it is stressed that values and social consequences are an essential part of score meaning. (Messick 1989b.)

TABLE 4   Facets of validity according to Messick (1989b, p. 10).

	Test interpretation	Test use
Evidential basis	construct validity	construct validity + relevance/utility
Consequential basis	construct validity + value implications	construct validity + relevance/utility + value implications + social consequences

The first cell in Messick's matrix calls for evidence for test interpretation that is, evidence for score meaning. Appropriate evidence here includes content relevance and representativeness, examining the internal structure of the test responses (their relationship to the tasks or items), and the external structure of the test (the relationship between scores or sub-scores and other measurements or background variables). The second cell calls for evidence of test use which in practice means evidence for the relevance and utility of the scores for the applied purpose. The third cell calls for consideration of the consequences of test interpretation, which means e.g. consideration of the values of score interpretation, including consideration of the values of the constructs, theories and ideology behind the test. The fourth cell calls for evidence of the consequential basis of test use, which requires evaluations of the social consequences of test use. (Messick 1989a, pp. 34 – 57, 1989b, p. 6; see also Luoma 2001; Moss 1992.)

The emphasis on social consequences in validity research has been welcomed by the proponents of authentic classroom assessment (see e.g. Moss 1992) since authentic assessment is said to be a step towards validity (as a more direct method of assessment) but a step away from reliability (Wood 1993, cited in Lewkowicz 1997, p. 165). Traditional standardized tests fail in educational relevance to teachers and students, whereas authentic assessment is more successful in improving performance and instruction (Newmann 1998; Teasdale & Leung 2000; Wiggins 1989; 1998). This has been interpreted as enhancing the validity of authentic classroom assessment based on appraisal of Messick's (1989a) consequential evidence for validity (see e.g. Moss 1992). On the other hand, the validity of authentic assessment has also been defended by stating that such assessments are more valid than standardized tests in assessing the whole range of students' competencies in actual performance of tasks

resembling those of real-life (e.g. Bachman 1990; Huerta-Marcias 1995; Shohamy & Reves 1985; Wood 1993).

The idea that the mere resemblance to real life of an assessment device is an adequate criterion for validity has, however, been criticised by pointing out that stressing only the consequential evidence in support of validity is inadequate, since it is only one facet of Messick's unitary concept of validity (Teusdale & Leung 2000). Furthermore, there has been strong criticism of the validity of authentic assessment as psychometrically weak because the complex performances elicited in such assessments are difficult to control (Shohamy & Reves 1985; Terwilliger 1997). Moreover, reliability – consistency both across tasks and among scorers – has proved to be difficult to achieve (Shohamy & Reves 1985; see also Moss 1994). It has been suggested that validity defined in terms of psychometrics should not be used to warrant interpretations made on the basis of authentic classroom assessment since the epistemological bases of the two types of assessment are different (Teusdale and Leung 2000).

The issue of validity in authentic classroom assessment has been approached by Frederiksen and Collins (1989), who present the concept of systemic validity and stress that a test is valid if it causes curricular and instructional changes in the educational system that result in the development of the cognitive skills the test aims to measure. The issue of validity has also been approached by focusing on the reliability of assessment (Moss 1994). It has been argued that reliability is an aspect of construct validity in the sense that there should be consonance among multiple sources of evidence supporting the score interpretation (that is, the construct to be validated). From this perspective it has been stated that authentic classroom assessment as a less standardized form of assessment presents a problem for reliability since students have latitude in interpreting and responding to tasks; they result in fewer but more complex responses that require using multiple skills and knowledge; and the responses require expert judgment for evaluation. Especially in using portfolios as an assessment device acceptable consistency between markers may be difficult to achieve since portfolios consist of an individual selection of multiple tasks that need to be scored simultaneously. (Moss 1994.) However, there are studies in which an acceptable level of inter-rater reliability has been achieved in assessing (reading) portfolios even for large-scale assessment purposes (see e.g. Garcia & Verville 1994; Valencia & Place 1994). Task reliability is reportedly much more difficult to achieve (Dunbar et al. 1991; Shavelson et al. 1993).

Moss (1994) has presented arguments for a hermeneutic approach as a solution to the problems of authentic forms of assessment with respect to reliability. The hermeneutic view complements the traditional view of reliability and offers an alternative path for validity research. The hermeneutic approach to assessment emphasizes holistic interpretations of students' selected performance made by expert members of the community, and grounding interpretations in textual and contextual evidence but also in a debate among the community. Assessment practice of this kind involves raters who know the context of the assessment (e.g. school, subject matter) in a transparent and

collaborative effort to evaluate students' abilities on the basis of multiple sources of evidence. (Moss 1994.)

It has been argued that both psychometric and hermeneutic processes of interpreting students' abilities involve "inferences from observable parts to an unobservable whole" and in both processes consistency supports the validity of the interpretations made (Moss 1994, p. 8). In the hermeneutic approach *inconsistency across tasks* does not invalidate the assessment but creates a challenge in the search for a comprehensive interpretation of a student's ability that explains the inconsistency. Unlike the psychometric approach, the hermeneutic approach additionally allows students to choose their own products for assessment, which enhances fairness. In the psychometric approach fairness is enhanced by giving all students the same tasks to be completed and by reducing bias. Moss claims, however, that "neither approach ensures bias" but "deserve discussion" (Moss 1994, p. 9). With respect to *consistency among readers*, Moss states that inconsistency would not invalidate the assessment but rather be a starting point for discussion in which multiple perspectives and interpretations are introduced, refined, and enriched into shared conclusions which are open to evaluation by any stakeholder. The validity and generalization of the resulting interpretations would thus be warranted through open and critical dialogue. (Moss 1994.)

These views of authenticity relate to construct under-representation and construct irrelevant variance as threats to validity (Messick 1995). Proponents of authentic classroom assessment often balance the technical requirements of reliability and validity with criteria such as authenticity (Frederiksen & Collins 1989; Linn, Baker & Dunbar 1991). This is found unnecessary in Messick's view of validity, since the criteria for validity cover also authenticity, which is guaranteed by minimal construct under-representation, while directness is guaranteed by minimal construct irrelevant variance (Messick 1994, p. 14). In any assessment, validity begins with construct definition (see Luoma 2001 on the importance of construct definition in the validation process). It seems only reasonable to aim to create an assessment device that represents the construct as well as possible. For instance in a reading literacy test it is desirable that the test should reflect different texts, tasks and reading situations as widely and as authentically as possible in order to achieve a score interpretation that reflects the construct, given that the construct is defined in a way that aims to cover reading literacy beyond the testing situation itself. Based on Messick's facet of evidence for test interpretation, the criteria for validity in this case include content relevance and content representativeness.

4.5   Challenges concerning authentic text in international reading literacy assessment

The concept of authenticity and authentic text in particular derive from theoretical considerations that may be incompatible with large-scale international studies. It has readily been acknowledged by the designers of such tests, that the cross-cultural nature of the studies as well as the need to translate the texts constrain the authenticity of the reading experiences, texts and tasks chosen for the studies (e.g. Framework and Specifications for PIRLS Assessment 2001). In this section the concept of authentic text is further explored and described in the context of international reading literacy assessment and the particular challenges in this case are considered.

Whole text as the target in reviewing text authenticity. The majority of the literature on authentic texts refers to the whole text in describing what authenticity in texts is. That is, the attributes of authentic texts are found to be qualities of a whole text. However, earlier in this chapter reference was made to other kinds of views and concerns. For instance, it has been stressed that the *language* used in texts should be familiar to the students (Weiss 1994) and that authentic texts should cover a wide range of real-life *topics* and *genres* (Kapinus et al. 1994). In the PISA 2000 reading literacy assessment, during the test design phase texts were reviewed as whole texts and in some cases topic and text type were reviewed separately. Texts were reviewed by researchers and subject matter experts in each country in terms of e.g. the target group's exposure to such material in and out of the school context, familiarity of *topic* and *text type* as well as interest to 15-year-olds. Reviewing a text's topic and text type separately was found easier than reviewing and assessing the text as a whole. Familiarity of topic and familiarity of text type may be quite different although the text itself is the same. The topic may be unfamiliar but the text type familiar, or vice versa. On the other hand, both topic and text type may be relatively familiar but the language used unfamiliar to the students, i.e. the language used may contain strange and difficult expressions. Text may also include expressions reflecting an unfamiliar context. Furthermore, the layout of the text, on the basis of which readers usually recognise its genre and which thus guides the approach readers take to the text might also be strange. It is to be noted that texts also have several characteristics that can be more or less authentic since authenticity is not an absolute quality. Usually even the most inauthentic texts have some authentic elements in them. All this suggests that the attributes of authentic text may have to be considered in terms of different textual features rather than just in terms of the whole text.

Whole text vs. extract. Since in the PISA 2000 reading literacy test the aim was to present students with texts that were as authentic as possible, the texts were usually whole texts, such as articles, advertisements, stories, information sheets, etc. However, many texts used in the PISA 2000 reading literacy test were extracts, i.e. they represent only part of the original text. In the case of

literary texts an extract was taken from Anouilh's play, for example. Part of the reason for this is restrictions of space and time since students have to read within the time limits of the test several texts of different types and respond to the tasks attached. If a text is otherwise suitable for the test it may be reasonable to select only an extract of it for the test. This does not necessarily reduce the authenticity of the texts since readers often encounter extracts in the real life contexts as well and are able to interact with the extract in the same way as they would interact with the whole text. On the other hand, some non-continuous texts in the PISA 2000 reading literacy test appear alone in the test although in real-life contexts non-continuous elements are usually part of larger texts also consisting of written elements. In the case of these texts, the degree of authenticity may be reduced because the rest of the text, i.e. the context of the non-continuous element, is lacking.

Translated texts. In international reading literacy assessment the texts presented to students are necessarily translated. In PISA 2000 even the texts that were submitted from students' own country, e.g. Finland, were first translated into English or French before submission to the consortium. Official versions of the texts selected for the PISA 2000 reading literacy test were then distributed to the participating countries in the official languages of the OECD (English and French) to be translated into the national languages. (See e.g. Arffman 2004.) From the point of view of text authenticity translations are not a problem. In the real world of texts, many texts readers encounter in their daily lives are translated: fiction, instructions, manuals, etc. As the world globalizes so does the world of texts. A greater problem may arise if the text is poorly translated, possibly even to the extent that the equivalence of the test is endangered (Arffman 2004), or the text clearly reflects an unfamiliar cultural context. The latter case is not, of course, a problem with fictional texts that often reflect and describe strange contexts and worlds. In the case of non-fictional texts, in PISA an attempt was made to solve the problems that arose due to unfamiliar cultural contexts reflected in the texts by national and cultural adaptations. These adaptations, however, meant only relatively superficial changes to the texts. For instance, a foreign currency mentioned in a text was changed to the national one. Thus, cultural differences in texts and in the contexts they reflect may result in reduced authenticity for the students participating in an international test of reading literacy.

Curricular authenticity. Materials used in international reading literacy assessments can be reviewed in terms of curricular authenticity. Curricular authenticity refers to the extent to which texts and tasks are typical and familiar to students in the school context. In the PISA reading literacy assessment no curricular analysis was made of the assessment domain or the texts and tasks used in the test was made since PISA aims to assess reading literacy proficiency and determine how well 15-year-old students are prepared for the challenges their life in the information society has to offer them, now and especially in their future studies and work. Rather than concentrating on the school curricula of key subject areas the focus is on young people's ability to apply and reflect

on their knowledge and skills in real-life situations and on their ability to continue developing these skills in order to meet the changing challenges of future societies. For this reason, an international group of experts defined reading literacy and created guidelines for the test without analyzing the school curricula of the participating countries. After all, reading literacy proficiency is developed contexts also other than that of school. (OECD 2002.) In explaining the results of the assessment curricular analysis can, however, be referred to (See e.g. Cosgrove 2003; Sulkunen 2004b).

Cultural aspects. In international reading literacy assessments in which the test takers have to read texts in their mother tongue and then respond to questions the issue of the authenticity of text has significant cultural aspects. In order to make the test fair and culturally unbiased, all respondents should have an equal opportunity to respond to a real-life text that is typical and meaningful within their national culture as well as their own subculture. This is the only way texts can be authenticated (Widdowson 1990), since authentication presupposes the respondents' engagement with the text in an authentic way by taking interest in it. The challenge is to find texts that can be equally authenticated by all the participants from different cultures in order to ensure the equivalence of the test.

Within the context of an international reading literacy assessment several dimensions of authenticity can be distinguished from the point of view of the readers and test takers (Figure 3). Firstly, all the texts chosen for a reading literacy test should be genuine texts written for real communication purposes somewhere in the world. Secondly, from the point of view of the national culture of each participating country it is important that texts should reflect the textual conventions typical within the culture of their citizens. Thirdly, in addition to these two dimensions of authenticity, texts should be similar to those in the test takers' culture. For instance, in the PISA assessment with its target group of 15-year-old students, the texts should be meaningful to them and typical in their subculture and their future life. The dimension of cultural authenticity at the level of the group (level 3 in Figure 3) covers also the subcultures of the two genders. In the world of reading literacy the test participants' gender has proven to be relevant since boys and girls show differences in both reading habits and reading literacy performance (e.g. Roe & Taube 2003; Linnakylä et al. 2002; OECD 2002).

The fourth dimension of text authenticity in this context is that of each individual test taker's culture. Even among 15-year-old Finnish students there are numerous individuals who might actually have a different perception of texts that are typical for them. The last two dimensions appearing in Figure 3. refer to the school context. Since the PISA reading literacy test was (and will be) administered at school, texts should be authentic in that context as well. If students encounter at school texts and tasks that are untypical in that particular context it may confuse their reading and their performance in the test (Hambleton 2002). For the same reason authentic text should also be that what

is used in testing. In the case of PISA, testing takes place at school, but this is not necessarily the case.

1.The global dimension: genuine text

2. The national dimension: typical real-life text within the national culture

3. The group dimension: typical real-life text within the subculture

4. The individual dimension: typical and meaningful text within individual's culture

5. School context: typical and meaning-ful school text

6. Testing situation: typical and meaningful text in a test

FIGURE 3   Dimensions of text authenticity from the reader's perspective.

Large-scale nature of the assessment. Since authenticity is essentially an individual phenomenon (Breen 1985; Bachman & Palmer 1996) the large-scale nature of international reading literacy assessments constrains the extent to which the authenticity of texts can be taken into account. It is always the individual reader who either authenticates (Widdowson 1990) the text or not, that is individual reader decides to engage with the text in a meaning making process. In an international large-scale assessment, however, the aim is to describe, explain and compare the educational outcomes of education systems rather than those of the individual respondents.

In selecting the texts for the test, only the three first dimensions of text authenticity (Figure 3) can be taken into account: texts chosen for the test should be globally genuine texts; they should be equally authentic (or inauthentic) in all the national cultures of the countries participating the test; and, lastly, they should be chosen taking into account important subgroups and subcultures of the participants. The aim is to ensure that the texts are authentic not only to participants in different countries but also, for instance, to different genders, to students with different home backgrounds and to different ethnic minorities in order to make the test as fair and equivalent as possible for all participants. In other words, the aim is to reduce bias in the test. The individual

dimension of text authenticity cannot be taken into account since, bearing in mind the aim of these large-scale studies, it is not necessary and it would be impossible to design a test that would cover all the individual preferences of thousands of participants as in PISA. Overlooking the individual level is, however, at odds with the ideal of authentic text and one of the many compromises that need to be made in constructing a large-scale international test. In such tests the degree to which authenticity can realistically be achieved is not overall very high since the assessment element alters the language use situation and thus compromises authenticity (Spence-Brown 2001). This is especially true when the form of assessment is a test.

Appropriateness as an attribute of authentic text. Appropriateness is a relatively isolated requirement for text authenticity, as pointed out in Section 4.1 (Figure 2). In international reading literacy studies, however, this particular attribute of authentic text is useful. In the context of language testing the concept of appropriateness derives from concerns that authentic text is not necessarily appropriate for the language skills of the students (Davies 1988). That is, authentic texts and especially the language in them may be too difficult for students learning a foreign language. In reading literacy studies in which students are tested in their own mother tongue this concern is relevant in terms of the content of the text. Texts need to be appropriate in the sense that they do not include concepts and vocabulary that are too strange and unfamiliar to the students. This again is an aspect of authentic text that is to be judged in relation to the readers or the target group of the assessment. From this point of view appropriateness is related to the age of the respondents and the life experience they have. In the case of the PISA reading literacy test, the respondents are 15-year-olds and thus the content of the texts used in needs to be on a quite general level in order to be understood by the students. Apart from the age of the respondents, appropriateness as an attribute of authentic text also has cultural aspects, particularly in international assessments. The demand for the appropriateness of the text derives from cultural sensitivity since, regardless of the cultural background of the respondent, the aim is to present the respondent with texts that are not offensive or otherwise inappropriate. Therefore there are no texts about, for instance, religious or political issues in the PISA 2000 reading literacy test. This is yet another issue related to reducing bias and ensuring the fairness and equivalence of the test.

Familiarity as an attribute of authentic text. Although familiarity has been presented as an attribute of text authenticity it is desirable only to the extent that makes the text accessible to students participating in the test (Garcia & Verville 1994; Kapinus 1994). As already discussed in Section 4.1, texts that are completely typical and familiar in a reader's culture can hardly be interesting because they lack novelty value. Texts should have something familiar in them in order to enable the reader to engage with them but if all the textual features are familiar the text as a whole will be simply too familiar and will not necessarily provide any incentive for reading. In an international reading literacy test familiarity also has cultural aspects since there is unavoidably some

cultural unfamiliarity in texts selected for an international reading literacy test. This is especially the case in PISA, which values cultural diversity and reflects cultural and linguistic plurality in the test. It has been argued that a proficient reader necessarily has to deal with unfamiliar materials. (McQueen & Mendelovits 2003.) In order to balance the demands for familiarity as an attribute of authenticity and unfamiliarity as a motivating element, texts should be reviewed and judged in terms of separate textual features such as topic, language, etc. rather than reviewing the text as a whole.

Unedited text. One of the attributes of authentic text related to genuineness is that is has not been edited for instructional or assessment purposes but is used as it appeared in its original context. This attribute of authentic text may have to be compromised in a testing context in order to avoid confusion among readers that might distract their performance in the test. In choosing authentic real-life texts from various countries and cultures, as was done e.g. in PISA, it is necessary to adapt some surface elements of texts to prevent bias occurring in other participating countries; for instance, reference to a foreign currency may be adapted to a local one. Furthermore, in large-scale assessment students read the text and respond to tasks only in the test situation. Therefore, it is preferable to avoid texts that include bad or inappropriate language or false information. In some cases, such a text might otherwise be suitable for the test, in which case these flaws should be corrected and the text edited. In instruction, such defects are not a problem but a topic for discussion that at best might be lively, interesting and educative. However, in large-scale testing there is no room for discussions of this kind, and unfamiliar and inappropriate features in a text can confuse students.

Authentic text today or in the future. In defining what is authentic to readers it is often the present world of texts that is paid attention to. That is, texts that are typical, interesting, relevant and familiar to readers at the present time are authentic to them. However, there is a future aspect to authenticity as well (Bopry 1999; Leiwo 2005), since in instruction and assessment the aim is to provide students with skills they will need in their future lives and to assess these skills. The PISA reading literacy test includes texts that students are likely to encounter in their future, e.g. work-related texts, for precisely this reason (OECD 2003). Texts that are not authentic to such readers now but possibly will be in the future are necessarily chosen by others, not the students themselves. However, this is also the case in instruction and assessment (testing) contexts where presently authentic texts are used, since it is usually teachers and test designers who determine which texts are used.

To summarise, in an international comparative reading literacy assessment the authenticity of the test contributes not only to the validity of the test but also to cultural equivalence between the participating countries. For the test to be valid, the text chosen and the tasks set should be authentic and reflect the texts encountered in real life by the test takers. Then the whole range of students' competencies will be covered in the test and the results of the test can be generalized beyond the test situation. (See e.g. Bachman 1990; Bachman &

Palmer 1996.) For the sake of cultural equivalence it is important that the intendedly authentic texts and tasks in a reading literacy test should be equally authentic or equally inauthentic to readers in all participating countries. This ensures test fairness and equivalence and reduces bias. In international reading literacy assessments, however, there are challenges to the use of authentic texts that often result in compromises with the ideal of an authentic text. This is the case with large-scale studies in which what is basically an individual level phenomenon must be treated on a group level only. Furthermore, the requirement that unedited texts should be used for assessment purposes may have to be compromised in order to avoid confusion among the test takers if confronted by strange expressions, false information etc. Text authenticity has significant cultural aspects with implications for the validity of assessments in international contexts, and therefore the process of selecting the texts for the test is of crucial importance in ensuring test fairness and equivalence.

5   THE SELECTION OF TEXTS IN PISA 2000

The reading literacy test in PISA 2000 included 37 different text units chosen to represent as wide a range of text types and situations for reading as possible. A twofold classification was used, distinguishing between *continuous texts* further classified by text type adapted from Werlich (1983) and *non-continuous texts* further classified by structure and format (Kirsch & Mosenthal 1989-1991). The classification was central in ensuring that a large variety of texts was represented. The continuous texts in the test took the form of narrative texts, descriptive texts, expository texts, argumentative texts and instructions. The non-continuous texts were forms, information sheets, vouchers, certificates, calls and advertisements, charts and graphs, diagrams, tables and matrices, lists and maps. (OECD 1999; OECD 2002.)

As shown in Table 5, the majority were types of continuous text (27 out of 37 texts), of which expository texts formed the largest group. Although there were only 10 non-continuous texts they were quite varied including tables, forms, graphs, maps and a diagram. Although the distribution of texts by text type shows that nearly three quarters of them were continuous texts, the distribution of items attached to the texts was somewhat more even, two thirds of the items in the reading literacy test being attached to continuous texts and one third to non-continuous texts (OECD 2002, p. 29).

While the classification by text type covered the linguistic and structural aspect of the texts in PISA, contextual factors were covered by classifying the texts according to four situations for reading. Situation for reading refers to the use or purpose that the author has intended for the text. Furthermore, in dividing texts and tasks into different situations for reading, social aspects and general contents were taken into account since reading usually involves others, albeit often implicitly or indirectly, and the general content of a text is often oriented to its intended use. The four situations for reading according to which the tasks and texts were classified were reading for education, reading for occupational purposes, reading for private use and reading for public use. Text books, for instance, are typically read for educational purposes in which case the others involved would be teachers and fellow-students.

TABLE 5   Texts used in the PISA 2000 reading literacy test by text type

Text type	Number of texts
Continuous texts:	
Narrative	4
Expository	10
Descriptive	4
Argumentative/persuasive	5
Injunctive	4
Non-continuous texts:	
Charts and graphs	2
Tables	4
Schematics	1
Maps	1
Forms	2
Total	37

In the case of 15-year-olds, instructions and forms related to applying for a job are usually texts read in occupational contexts and the others involved would be potential employers. Narratives, personal letters and e-mail are mostly read for private and in this case the other people involved are e.g. relatives and friends. Official documents and timetables are usually read for public use and the other people involved are anonymous to the reader. (Linnakylä & Sulkunen 2002; OECD 2002.) As Table 6 shows, approximately 40 % of the texts in the PISA 2000 reading literacy test were texts intended for public use. The remainder were drawn in approximately equal proportions from the three remaining contexts. The distribution of items across the four situations overall reflects the distribution of texts, although occupational texts were given less weight than other contexts (OECD 2002, p. 27).

TABLE 6   Texts used in the PISA 2000 reading literacy test by situation.

Context	Number of texts
Educational	9
Occupational	6
Private	6
Public	16
Total	37

The challenge in an international reading literacy test is to represent the world of texts as well as possible in all the participating cultures and countries in order to ensure a linguistically and culturally equivalent test for all. Thus, during the process of devising the test, a diversity of cultural perspectives was ensured by inviting reading literacy experts in every participating country to submit to the consortium texts (and tasks) suitable for the test. Participants were encouraged to recommend authentic texts reflecting the definition of

reading literacy and text classification used in the PISA 2000 framework. Two thirds of the participating countries did contribute materials to the consortium. Other measures taken to ensure cultural relevance for all the participants included, firstly, an extensive review process during which national subject-matter experts rated the materials in terms of students' familiarity with the content, cultural concerns, other bias concerns and anticipated translation problems; secondly, a review of the materials by an international cultural review panel; thirdly, analysis of the field trial data in order to detect unacceptable interactions between item-difficulty and country; fourthly, translation of the materials under rigorous procedures including an independent verification process and extensive and uniform training of markers in all countries. (See OECD 2002.)

The texts used in the PISA 2000 reading literacy test are listed in Table 7. In the table, the texts are characterised in terms of text structure, text type and reading situation in the same way as they were classified for the PISA assessment. Information is also provided about the sources, based on the information included in the PISA test booklets. Context of appearance describes on a general level the original context of the text. Genre refers to the everyday label given to the text based on text external characteristics of the text. For some purely non-continuous texts it has been difficult to define the genre of the text and the cell has been left blank. Furthermore, the modal elements in the text are listed.

TABLE 7   Texts used in PISA 2000 reading literacy assessment.

PISA text	text structure (PISA)	text type (PISA)	situation (PISA)	source (if known)	context of appearance	genre	modalities in the text
1. Lake Chad	non-continuous	charts and graphs	Public	The Times Atlas of Archaeology	encyclopaedia		two graphs
2. Drugged spiders	continuous	expository	Public	The Independent, 12.3.1994.	newspaper	newspaper article	plain text
3. Macondo	continuous	narrative	private	Gabriel Garcia Marquez: One hundred years of solitude.	novel	story	plain text
4. Aesop	continuous	narrative	private	The complete fables. 1998.	fable collection	fable	plain text
5. Beach	continuous	expository	Public	The Guardian 28.5.1998.	newspaper	newspaper article	plain text
6. Iran Air	non-continuous	table	Public	Iran Air winter timetable 1997–1998	timetable	timetable	lists of abbreviations used + timetable
7. Flu	continuous	argumentative	occupational		information sheet for ACOL personnel	information sheet	text + drawings + slogan with distinctive font
8. Graffiti	continuous	argumentative	Public		web pages	expression of opinion	two texts by different writers
9. Household work	non-continuous	charts and graphs	educational	Family matters, no 46, fall 1997.	research journal	research report	text + graphs
10. If	continuous	argumentative	Public			commercial advertisement	poem + photographgraph + logo of the motorbike
11. Labour force	non-continuous	diagram	educational	D. Miller: Form 6 Economics, p.64.	textbook		a tree diagram
12. Library	non-continuous	map	Public		library information board	floor plan	map of library
13. News agencies	non-continuous	table	Public				table listing news agencies
14. Plan Int	non-continuous	table	Public	Quarterly report, first quarter 1997.	report	financial report	table

15. Police	continuous	expository	educational	Le Ligueur, 27.5.1998.	magazine	magazine article	text + drawing + information boxes
16. Rhinoceros	continuous	descriptive	Public	Swedish Daily News 16.1.1994.	newspaper	newspaper article	plain text
17. Shirts	continuous	expository	private	Daily Mail 5.2.1998.	newspaper	newspaper article	text + table comparing shirts
18. Telephone	non-continuous	table	Public			instructions	list of teleph. numbers + instructions
19. Running shoes	continuous	expository	educational	Revue ID (16), 1. – 15.6.1997.	magazine	article	text + drawing
20. Exchange	continuous	expository	educational	YFU Finland.	web pages	information sheet	plain text
21. Gift	continuous	narrative	private	Mississippi writers: reflections of childhood and youth, vol. 1. 1985.	short story collection	short story	plain text
22. Student opinions	continuous	argumentative	educational			expression of opinion	five independent texts
23. Art store	non-continuous	advertisement	educational		catalogue	sales catalogue	list of products + order form
24. Amanda	continuous	narrative	private	Jean Anouilh: Léocadia, end of scene II.	play	play	extract of a play + definitions of professionals in theatre
25. Job application	non-continuous	form	occupational		fast food restaurant's job application form	application form	plain form
26. South Pole	continuous	expository	educational	Bertrand Imbert: Le grand défi des poles.	textbook	textbook text	text + map + diagram
27. Nuclear waste	non-continuous	charts and graphs	public		newspaper	graph	a graph
28. Optician	continuous	descriptive	occupational			brochure	a bar graph + drawn pictures +

							information box
29. Etiquette	continuous	instructive	educational	YFU Finland.	web pages	instructions	list of instructions
30. Personnel	continuous	descriptive	occupational			information sheet	
31. New rules	continuous	expository	public		newspaper	editorial	plain text
32. Job interview	continuous	instructive	occupational			instructions	plain text
33. Bicycle	continuous	instructive	public		bicycle owner's manual	instructions	table, picture
34. Allergies	continuous	expository	public		newspaper	news article	two short news articles
35. Warranty	continuous	descriptive	public			warranty instructions	plain text with decorative edgings
36. Movie reviews	continuous	argumentative	private		newspaper	review	four short reviews
37. Contact employer	continuous	instructive	occupational			instructions	plain text

6 RESEARCH QUESTIONS AND METHODS

6.1 Research questions and phases of the study

The aim of the empirical part of this study is to examine which textual features readers pay attention to in reviewing text authenticity and to evaluate the authenticity of the texts used in the PISA 2000 reading literacy test in the culture of Finnish 15-year-old students.

The specific research questions of the study are:

1. Which textual features do readers find relevant to a text's authenticity?
2. How authentic are the texts used in the PISA 2000 reading literacy test for Finnish 15-year-old students?
3. Are the texts used in the PISA 2000 reading literacy test equally authentic to Finnish 15-year-old girls and boys?

In order to make the concept of authenticity more accessible to the informants participating in the empirical part of the study, the concept of authenticity needed to be operationalized in terms of the literature on authenticity and authentic texts with a review of the attributes of authentic text. Of all the numerous attributes of authentic text, 'typicality' was chosen for use in the study since it captures two essential aspects of text authenticity: that of a text's being typical and familiar in the reader's culture and that of its resemblance to real-life texts. (See e.g. Valencia et al. 1994; Harris & Hodges 1995.)

In the pilot phase of the study a panel of expert readers was asked to assess and rate the typicality of each PISA text among the texts that panellists read. Explanations for the ratings were also called for. In the main study questionnaire, the respondents of which were 15-year-olds, 'typicality' was also used to operationalize the concept of authenticity whose Finnish language equivalent would not belong to the vocabulary of students of that age. However, since it was found that another essential part of authenticity is overlooked if only 'typicality' is used, an additional attribute of authenticity

was brought in, namely 'interest'. As shown in the literature review presented in Section 4.1, authentic text is not only typical in the reader's culture but also engaging, rich and interesting for the individual reader (Garcia & Verville 1994; Kapinus et al. 1994; Weiss 1994). As expressed by Widdowson (1990), the reader authenticates a text by taking an interest in it. Since typicality and interest as attributes of a text may not and need not coincide, it was seen as appropriate to include them both in the questionnaire presented to the students.

The research process designed to answer the questions presented above is illustrated in Figure 4 below. As shown in the figure, the empirical part of the study begun with the pilot study, in which expert panellists reviewed, rated and commented on the typicality of the PISA texts. Based on the results of the pilot study, four textual features that were found to be most significant in the analysis were selected for use in the student questionnaire of the main study. The selection of the four textual features for the questionnaire was also influenced by the literature review described in Chapter 4.

```
┌─────────────────────────────┐
│ Theoretical review          │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ Pilot study:                │
│ Panel review by expert      │
│ readers to find out textual │
│ features relevant to readers│
│ in assessing text typicality│
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ Main study:                 │
│ Student questionnaire for   │
│ reviewing text typicality and│
│ interest in PISA texts      │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ Validity study:             │
│ Exploratory factor analysis to│
│ determine the relationships │
│ between typicality and interest│
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ Analysis of the data resulting│
│ in a ranking of the PISA texts│
│ in terms of typicality and  │
│ interest; special focus on  │
│ gender differences          │
└─────────────────────────────┘
```

FIGURE 4   Phases of the study.

In the main phase of the study, a questionnaire was first developed based on the pilot study and then presented to a selected group of 15-year-old students. In the questionnaire the students were asked to rate the typicality and interest of the textual features of the PISA texts selected as a result of the pilot study. In analyzing the data, typicality and interest indices were calculated for each text as a whole and also separately for the specific textual features. Special attention was focused on gender differences.

In the final phase of the study the results obtained through the questionnaire revealed the need to explore the extent to which students actually rate the typicality and interest of the separate textual features as opposed to rating each textual feature more or less the same. Therefore, the correlations between the typicality of the four textual features were calculated and the same was done for the perceived interest of the textual features. Based on the patterns appearing among the correlations, exploratory factor analysis was conducted for a closer examination of the relationships between the variables of the study. The results of the analysis were used to reach conclusions as to the validity of the concept of authenticity and the usefulness of the questionnaire, as well as the results of the study and the conclusions based on those results. The validity of the way the concept of authenticity was operationalized and of the textual features chosen were subjected to particular scrutiny.

The pilot study and the main study are described in detail in the next section. Special interest is given to the data and materials used in the studies as well as the methods of analyzing the data.

6.2   Materials and methods

The present study aims to investigate one possible aspect of test validity and ultimately improve the validity of tests like the PISA reading literacy test. Thus it can be characterized as a validation study. Current validation studies usually employ conceptions of validity defined as "an integrated evaluative judgement of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of inferences and actions based on test scores or other modes of assessment" (Messick 1989a, p. 13). This definition indicates that both empirical evidence and theoretical reasoning are to be used in evaluating test validity. It has been emphasized (Banerjee & Luoma 1997) that applying such a conception of validity requires thorough understanding of the test from the early phases of test design to the conclusions and the uses to which the results of the test are put. The present study concentrates on one aspect of the reading literacy test validity, namely text authenticity, and aims to achieve a thorough understanding of the phenomenon through both theoretical and empirical examination.

In order to achieve this goal the present study exploits a *mixed methodology*. More precisely, the present study can be characterized as a work of mixed

method research, in which qualitative and quantitative data collection and analysis are used in the same study (Teddlie & Tashakkori 2003). In mixed methodology the most suitable and purposeful methods for data collection and analysis are used to answer the research questions and to meet the demands of the context of study. Emphasizing the research question and the context in determining the methods to be used at the expense of the methodological and paradigmatic traditions derives from pragmatism as a practical research philosophy. (Teddlie & Tashakkori 2003; Maxcy 2003; Greene & Caracelli 2003; Dillon et al. 2000.) It has been argued that using mixed methods increases the validity of the study, since by using multiple methods on the same phenomenon (methodological triangulation) a more complete understanding of the phenomenon can be achieved. Furthermore, the weaknesses of any one method can be compensated for by another method, which also increases the validity of the study. (Frechtling & Sharp 1997; Teddlie & Tashakkori 2003.)

The mixed approach used in the present study is sequential (Tashakkori & Teddlie 1998; Morse 2003) in the sense that multiple approaches to data collection and analysis follow in sequence in different phases of the study. More specifically, the method used is sequential qualitative-quantitative analysis since qualitatively identified groups are studied by using quantitative data collected after qualitative analysis (Onwuegbuzie & Teddlie 2003). Usually quantitative analysis is employed on either extreme or negative cases detected in qualitative analysis (Onwuegbuzie & Teddlie 2003) but in the present study qualitatively identified textual features can be characterized as essential rather than extreme or negative. Since the aim of the study is to discover and describe rather that test a theory or hypothesis (Morse 2003) the theoretical thrust of the study is inductive making the qualitative approach dominant in the study despite the strong presence of quantitative analysis in the main study phase. The main study is characterized by a parallel use of quantitative and qualitative approaches.

In the pilot phase of the study, data was gathered from expert panellists reviewing, rating and commenting on the typicality of the PISA texts. Using expert judgments is the most common qualitative method in validation studies (Banerjee & Luoma 1997; Bachman 2000). Qualitative validation studies have concentrated on providing information on the content of the test and the properties of tasks. In the case of the present study this applies also to the texts. Usually the experts used are linguistics or teachers, at least in the case of test content. However, in the case of studies relating to face validity and acceptability the experts are usually stakeholders. (Banerjee & Luoma 1997.) In this study, the experts referred to are panellists who knew the texts exceptionally well after marking open ended responses to questions attached to the texts and can be characterized as key informants (Frechtling & Sharp 1997). Most of them had at least some linguistic training or experience. In analyzing the pilot study data, the ratings of text typicality were ignored and analysis focused on the written comments panellists produced to explain their ratings.

Thus the actual pilot study data consists of these openly constructed written comments on the familiar and unusual features in PISA texts.

In the main study, the approach used in gathering and analyzing the data was both quantitative and qualitative although the small number of respondents restricted the use of more sophisticated quantitative methods of analysis. In gathering the main study data a questionnaire was used and the data of the main study consists of the ratings provided by 15-year-old students in assessing the authenticity of the PISA texts. In the main study written explanations for the ratings were also provided by the students responding to the questionnaire and they were used as illustrative examples for the numerical data.

Without the frame of authenticity, typicality and interest can be reduced simply to qualities of texts. Within the frame of authenticity, however, they are rather qualities or characteristics of the relation between text and reader (e.g. Widdowson 1978, 1979; Bachman 1990). Thus, in this study, both in the pilot and the main phases, the data gathering methods were chosen bearing in mind the value and position the reader gives to a text. This is to emphasize that typicality and interest are in this study attributes of authenticity, not qualities related to texts only. As in other studies examining the authenticity of assessment material (see e.g. Wu & Stansfield 2001; Spence-Brown 2001) in this study also it is assumed that text authenticity can validly be determined only by the readers. In case of testing, text authenticity is to be assessed by the students belonging to the target group of the test.

In this study specific attention is paid on gender differences in students' assessments of text authenticity. Since the gender difference in the PISA reading literacy scores was wide, especially in Finland, and since girls and boys have clear differences in their reading habits and interests (Linnakylä 2002; Linnakylä et al. 2002; Roe & Taube 2003), it was expected that different texts would be authentic for girls and boys.

The aims and the different phases of the study as well as the methods of analysis are summarized in Table 8 and described in detail below.

TABLE 8 Phases of the study and methods of analysis.

Aim of each phase of the study	Phase of the study	Method of analysis
1. To discover textual features relevant to readers in terms of text typicality in order to construct a structured questionnaire for 15-year-olds.	Pilot study: *Panel review* of PISA texts in terms of their typicality with ratings and written explanations.	Content analysis: close reading and classification of textual features arising from the data.
2. To find out how authentic PISA texts were for Finnish 15-year-old students with special focus on gender.	Main study: Presenting the *questionnaire* to a selected group of 15-year-old students who rated the typicality and interest of the topic, genre, layout and language of the texts.	Calculating means of ratings as indices for typicality and interest for whole texts and the four textual features separately, based on which the texts were placed on rank order. A t-test was used to compare the means to see whether the differences were statistically significant. Cross tabulations of rating means by gender. Introducing descriptive written explanations for students' ratings.
3. To study the extent to which the elements of authenticity, typicality and interest, are related to each other and to distinct textual features.	Validity study: Studying the *relations* between typicality and interest of genre, topic, layout and language.	Explorative factor analysis.

6.3 Pilot study

6.3.1 Method of the study

To find out which features of a text contribute to its authenticity for the reader a small panel of Finnish expert readers were asked to review and evaluate the authenticity of the PISA texts both in the reading culture of 15-year-olds and in their own reading culture.

In the first pre-pilot phase, 6 panellists reviewed the authenticity of the PISA texts in terms of their familiarity to 15-year-old students. Given the PISA context of the present study the point of view of the target population of PISA (15-year-olds) was essential. Therefore the panellists were instructed to assess text familiarity from this perspective. However, it soon came apparent, both from practical experience and theory, that readers are most competent in judging text authenticity from their own perspective, since the authenticity of text is always ultimately defined by the individual reader himself (e.g.

Widdowson 1979). Therefore, in the second phase, which is considered to be the actual pilot study, 8 panellists were asked to review the authenticity of the PISA-texts from their own perspective in terms of the texts' typicality in their own lives. This approach to reviewing text authenticity was thus more valid and revealed the cultural background of Finnish young adult readers in general rather than that of 15-year-olds. Panellists were instructed to review the typicality of the PISA texts among the texts they read and particularly to point out what familiar or unusual features there were in the texts. (Appendix 1: Example of the questionnaire presented to the panellists in 2000.)

The pilot study panellists were all young female university students, aged from 21 to 27, and they knew the PISA texts well since they had worked as markers in evaluating the open-ended tasks in the PISA field trial in 1999 and in the PISA main study in 2000. They can thus be characterized as experts in regard to reading PISA texts. Most of the panellists had at least some linguistic training: four of them were Finnish language majors within a teacher training programme, one was a history major, one a major in sociology, one a major in adult education and one in political science. Overall, the panellists possessed at least to a certain degree the linguistic knowledge for analyzing texts.

The fact that all the panellists were female might have caused biased ratings of text typicality by gender. However, the ratings of the typicality of the PISA texts were not in focus in this phase of the study and thus they were ignored in the analysis. Ratings were asked for in order to give the panellists a starting point for approaching the texts in reviewing their typicality. The actual pilot study data consisted of the openly constructed written comments of unusual and/or familiar textual features provided by the panellists to explain their ratings. The analysis of the written comments did not aim to reveal who commented on which textual features and how many times, but simply to examine what kind of textual features panellists referred to in explaining their ratings. In this kind of approach to data analysis, possible gender differences would be irrelevant, which is why it did not seem necessary to recruit any male panellists. Besides, to give male panellists equal opportunities to become expert readers in terms of the PISA texts, they would have had to be already members of the Finnish PISA marking team.

The written comments produced by the panellists on unusual and/or familiar textual features of the PISA texts were analyzed using procedures typical of content analysis. Content analysis is typically used in analyzing various kinds of textual data systematically and objectively (Kyngäs & Vanhanen 1999; Ryan & Bernard 2000), especially free-flowing textual data such as free comments in response to open-ended questions as in the case of the present study. The principal idea in content analysis is to reduce texts to categories describing phenomena on a general level. Categories can be quantified and thus also analyzed quantitatively. (Kyngäs & Vanhanen 1999; Ryan & Bernard 2000.)

The pilot study data consisted of the written comments produced by panellists on the typicality of the PISA texts. These comments varied in length:

the shortest being an incomplete sentence, the longest several sentences. A typical item of data is as follows: "*[Movie] reviews could be from a TV-guide. In a very typical manner movies have been introduced in a sentence or two. It is probably most important to introduce film makers, actors and other facts. Short reviews usually include stars rating the film that are missing here. In a longer article-like review there is typically thematic description of topic and plot which is totally missing in this review, naturally.*" As the example illustrates, one comment may deal with one or more themes. The example above includes five sentences, the first of which refers to the context of appearance of the text and the other four to the structure of the text, since they deal with the elements used in text. (For a more detailed example, see Appendices 2 and 3.)

The unit of analysis in this study was the thematic unit (Krippendorf 1980; Ryan & Bernard 2000) i.e. a piece of text that reflects a single theme. In practice, thematic units in the pilot data consisted in some cases of only one sentence, sometimes of several sentences. One written comment might include one or more thematic units, as in the above example, which includes two themes (context of appearance and structure of text). In addition to this, the same sentences might include more than one theme. Each thematic unit was categorized according to the textual feature it comments on, i.e. its structure. An inductive approach (Kyngäs & Vanhanen 1999; Ryan & Bernard 2000) was used in the analysis in the sense that themes were induced from the data although the researcher's prior knowledge and experience necessarily had an effect on the categories the researcher paid attention to in analysis. Knowledge of the literature on textual features and text analysis as well as experience of text analysis and the PISA texts certainly had an influence, present in the analysis as a kind of background theory, which is characteristic of the deductive approach (Kyngäs & Vanhanen 1999) to content analysis. Naming the categories that resulted in analysis, represents a deductive approach in the sense that existing labels for textual features were used. The content analysis performed in this study was thus not purely inductive.

Content analysis of the data was conducted by the researcher herself following careful line-by-line reading. The object of the analysis was the manifest content (Kyngäs & Vanhanen 1999) of the comments, i.e. only the expressions used by the panellists were under analysis. Interpretations of the panellists' intentions were thus kept to a minimum. The data was read through several times. During the first two readings, the data became familiar in content to the researcher. During the third round, the data was reduced to expressions representing themes induced from the data during which the categories present in the data began to take shape. The focus was not on the comments made on the themes but on the themes themselves. For instance, if two comments both referred to the genre of the text they were both categorized under the same theme 'genre' although one may have included positive evaluation of the genre and the other negative evaluation. During the fourth reading, similarities and differences were looked for in the expressions representing the different themes, the categories in the data became clearer and the analysis became final. The validity of the analysis was enhanced by a second analysis done by the

same researcher one year later. During this second analysis some refinements to the categories was made. (See Appendix 2 for an example of the analysis performed and Appendix 3 for a list of codes and their explanations.)

6.3.2 Textual features essential to authenticity

The categories of textual features found in the data consisting of the written comments provided by the panellists are presented in Table 9 together with the number of thematic units representing each category. As the table shows, the thematic units found in the data were classified under 11 different categories of textual features.

The category including most thematic units is that of 'genre', which refers to the type of text in terms of its purpose. Thematic units representing genre often included references to the context of use (purpose) or context of appearance.

> "Text seems familiar, typical *newspaper article* that I read daily."
> "*As a genre* this might be quite familiar. If one is interested in something or searches for information, informative and persuasive texts are very typical. Like travel brochures or web pages."

TABLE 9   Textual features found in the content analysis.

Category of textual feature	Number of thematic units referring to the feature
Genre	136
Context	98
Content	96
Structure	89
Topic	61
Layout	60
Form	59
Language	43
Style	39
Target group	16
Text type	3

The category with the second most thematic units was that of 'context'. 'Context' covers thematic units commenting on the text's context of appearance and/or context of use (i.e. purpose) and/or lack of context. The context of appearance and context of use were often speculated upon when they were unclear. The lack of appropriate context was usually commented on in the case of non-linear documents such as tables, diagrams and charts.

> "In the *opinion section of the magazines* one could see texts like this. The texts seem too formal, however, compared to typical opinions."
> "This is difficult to answer since I have read *One hundred years of solitude* so I know the text. When *one knows the context* the text originally was in it changes the way one

approaches the text. If I didn't know this, the text would be untypical, a rather isolated piece."
"The chart must be familiar *from newspapers*. It has been however taken *out of context*. It certainly has not appeared as such."

The category of 'content' includes nearly as many thematic units as 'context'. Content refers to what is said in the text. In contrast with the category of 'topic', content refers to what is said in the text in detail.

"It is strange that the title is so vague in this editorial. The text is almost *solely about the technology of human fertilization*. Therefore the title could have been more specific."
"It was more familiar that little news tells *about accidents or minor news*."
"In Figure 1, *2000 B.C. – o – 1000 AD* is not logical although it can be understood. Maybe in the figure below *it would be good to mention that animals appeared in...*"

The category of 'structure' was also well represented in terms of thematic units. This category covers comments on the elements the text consists of.

"Very typical article: *the order of bringing up issues, short pieces of interviews, short, simple paragraphs…*"
"In a very typical manner *movies have been introduced in a sentence or two*. It is probably most important *to introduce film makers, actors and other facts*. Short reviews *usually include stars rating the film* that are missing here. In a longer article-like review there is typically *thematic description of topic and plot*, which is totally missing in this review, naturally."

The category of 'topic' refers to the theme of the text. It was also quite often mentioned in the panellists' comments.

"*The topic* is familiar, this could be in some science magazine."
"*The topic* is unfamiliar. I am not interested in *animals* and I can't imagine where I would run into text like this."

The category of 'layout' covers the appearance of text: choice of font, use of bold letters, poor quality of pictures, etc. The layout of texts was quite often commented on by panellists. The appearance of a text is quite important in determining which genre the text belongs to and is often related to the context of appearance. Knowing the context and genre of a text activates many expectations about the text and these expectations guide the way reader approaches the text. (e.g. Biber 1989.) Thus, the untypical appearance of a text can have an effect on how it is read and interpreted. In some cases panel members found it difficult to say which genre certain texts belonged to since the layout and the text itself did not give them enough clues.

"The *word WARRANTY is way too big*. And the *edging is too decorative* to fit Finnish taste."
"In this text there are at least *some kinds of blocks marked with dots*. It is easier to read when *important words have been written with capital letters or in bold*."

The category of 'form' covers the length of text or the type of text in the case of a non-linear text.

> "A job application as a *form-kind-of-text* is not typical since usually applications are freely written. Otherwise *similar forms* that are used in applying or registration are frequently bumped into."
> "The text feels strange because it is so *short*. The tales I usually read are *longer*."

The category of 'language' refers to thematic units commenting on the choice of words, familiar or unfamiliar expressions, and the quality of expressions. This category includes thematic units commenting the use of foreign language.

> "There are a *lot of names in foreign languages*, also ones that I don't understand, but it's ok in this text."
> "It is untypical that these *instructions are in Finnish* although the numbers are Turkish. The text has been *translated* and therefore seems untypical."
> "The writer's *choice of words is sometimes weird*: "in which director Patrick Leonte demonstrates both intelligence and rigor absent from his earlier films" , "Carmen Salina's acting is effectively melodramatic.""

The category of 'style' takes in thematic units commenting on the way things have been said in the text. Often these thematic units included comments on the general style or tone of the text rather than on details of expression or the choice of words, the latter being placed in the category of 'language'.

> "For a bulletin in a workplace this was *quite warmly written.*–– The *informative style* is familiar."
> "These opinions are not the most typical ones by their *style of writing*. Style in these is a bit more *aggressive, nasty and ironic* than in opinions usually."

The category of 'target group' covers thematic units with references to the target group the text might be aimed at. Panellists often contrasted themselves with 15-year-olds in these thematic units.

> "This is not typical since it is *targeted at people interested in exchange programs*."
> "The text is aimed *at people younger than me* and thus feels strange. A few years ago this would have been quite familiar."

The category of 'text type' is concerned with text type in terms of linguistic features only. There were only 3 thematic units in this category including mention of Werlich's text types (1983) such as description, narration etc.

> "Typical *narrative*: includes description, long sentences, yet understandable."

Based on the results of the content analysis as well as theoretical considerations, four textual features were selected for the main study questionnaire to be presented to 15-year-old students. These textual features were genre, topic, layout and language. Genre was the most typical feature in terms of which panellists commented on and classified texts. Furthermore, genre has been considered in the literature a textual feature to be taken into account in assessing the authenticity of a text (Kapinus et al. 1994). Topic is another such textual feature (Kapinus et al. 1994). The topic of the text is, furthermore, a textual feature related to students' interests and it was relatively frequently

mentioned in the panellists' comments. Content was referred to more often than topic but it was considered to be a textual feature easily commented on by expert readers but not by the readers encountering the text for the first time as would the students in the main study. For the same reasons the structure of text was also excluded from the questionnaire.

Layout, on the other hand, was selected for the questionnaire, since it can be related to the identification of genre through the context of appearance determining the way a text looks. Form as a textual feature was considered to be on the one hand too vague, and on the other hand to require too much defining to be presented to 15-year-olds in the main study questionnaire. Language, however, was mentioned in the literature (Weiss 1994) as one of the textual features that should be appropriate for students where instructional and assessment purposes are in question and it was therefore selected for the questionnaire. Furthermore, the categories of 'language' and 'style' were quite close to each other since both include thematic units referring to the way things are expressed in the texts (see Appendix 3 for a description of the categories). If these categories were combined, the new category would include 82 thematic units, giving it the fifth largest number of thematic units. Target group and text type were excluded from the questionnaire because of their assumed irrelevance for 15-year-olds and the limited number of thematic units found in the content analysis.

Genre, topic, layout and *language* were, then, considered to be the most essential textual features for this age-group in reviewing and evaluating text authenticity. They were therefore selected as the main constructs to be focused on in the student questionnaire. In addition to these immediate results, the panel review resulted in other interesting findings that will be discussed along with the conclusions and discussion presented in Chapter 9.

6.4   Main study

6.4.1 Questionnaire for 15-year-olds

In order to find out how authentic the texts in the PISA 2000 reading literacy test were in terms of the culture of Finnish 15-year-old students, a selected group of students (altogether 240) was presented with a questionnaire in which they assessed the authenticity of the 37 texts used in the test. The questionnaire asked students to rate the typicality and interest of each text in terms of its topic, layout, language used and genre (Appendix 4). The ratings were made on a Likert scale ranging from 1 (not typical/interesting at all) to 4 (very typical/interesting). Students were also asked to write down explanations for their ratings. In order to minimize the effort involved in rating the large number of texts they were allocated to nine different questionnaires using a rotated matrix, the design being an application of the PISA 2000 test booklet design

(PISA National Project Manager's manual for PISA 2000). Thus, each text appeared in two (out of nine) versions of the questionnaire and was rated by 43–54 students. The number of ratings per text varies since not all students were able to rate all the texts in the given time of one lesson. (See appendix 5.)

Altogether 240 15-year-old students answered the questionnaire during the spring of 2001. Both genders were represented nearly equally well among the respondents since there were 117 girls and 123 boys. Students answering the questionnaire were from 13 different schools that had participated in PISA 2000 the year before. The schools were located in Alavus, Helsinki, Joensuu, Juuka, Jyväskylä (2), Laukaa (2), Rovaniemi (2), Tampere (2) and Vantaa. In each school one group of students chosen by the teacher was selected. In most cases the teacher was the school's guidance counsellor who had been the test administrator in the PISA 2000 test. Among the selected students only one had taken the test. However, she had been given another test booklet on that occasion and therefore the texts rated were equally as unfamiliar to her as to her peers.

At the beginning of each lesson during which the students completed the questionnaire they were motivated by describing the general aim of the study and by emphasizing their input as an important source of information in ensuring the appropriateness and suitability of international tests in a Finnish context. Students then received instructions for completing the questionnaire emphasizing that there were no correct or incorrect responses but that each student was to give their own personal opinion about the typicality and interest of texts presented to them. Students were further instructed to assess the typicality of each text by comparing it to the kinds of reading materials they usually read and by judging the typicality of the given textual features. The interest of the text was to be assessed simply by judging how interesting and appealing the text and its topic, layout, language and genre were. Students were verbally given an example in order to illustrate and clarify the meaning of topic, layout, language and genre by describing a magazine article about horses and how topic of the article would be horses, the genre would be the type of text (article), the layout would refer to the appearance of the text (how it looks: font, titles and headings, the use of space, etc.) and the language would refer to the expressions, sentences and words used. Students were also instructed to freely write down their explanations for their ratings and the given textual features as well as any other comments they might have on the texts.

The students were relatively well motivated to complete the questionnaire although there were a few who filled in the boxes quite quickly without concentrating on giving any explanations for their ratings (which in these cases often seemed to be quite negative). The given time of one lesson (45 minutes) seemed to be enough for most students, who skimmed through each text, rated it and gave explanations for their ratings. Some students, however, were more thorough in reading the texts and would have needed more time to complete the questionnaire. The ratings these students gave were usually carefully explained. Some students complained about the scale from 1 to 4 which lacked

a value meaning 'can't really say' or 'not typical/interesting/nor untypical/uninteresting'. This, however, was intentional in order to avoid the extensive use of this potentially appealing choice in many cases. The explanations students gave for their ratings turned out to be quite general. The most specific characterisations of the texts were about the textual features mentioned in the questionnaire, mainly topic and genre.

Based on the student ratings each text received numerical indices of typicality and interest on a scale from 1 to 4. Each text received an average value (of typicality and interest) based on students' rating of the four textual features (topic, layout, language, and genre). Furthermore, a mean of all students' average ratings was calculated to form the indices of overall typicality and interest. Based on the indices, the texts were placed in rank order of typicality and interest. A paired samples t-test (2-tailed with significance level of 5 %) was used to compare the means to see whether the differences between typicality and interest indices were statistically significant. The mean values between the two genders were not tested due to the small number of respondents representing each gender per text. The results of the study are described in Chapter 7.

6.4.2 Relations between students' evaluation of typicality and interest

In the course of analyzing results of the main study, the need to study the extent to which the students were able to assess distinct textual features became apparent. Based on the indices of typicality and interest for each textual feature, texts were placed in rank order of typicality and interest as described above. The order of texts in terms of topic, layout, language and genre suggested, however, that students may have rated each textual feature similarly. Furthermore, it was found necessary to study the relation between typicality and interest as elements of authenticity from the point of view of the validity of the questionnaire, in which authenticity was operationalized as typicality and interest.

To study the relations between the variables, the correlations using Pearson's product moment correlations coefficient were calculated. Since the correlations between the eight variables showed patterns of high correlation among typicality variables and interest variables but lower correlations across typicality and interest variables, a further analysis was conducted to reveal and summarize the relations between variables. Thus, an exploratory factor analysis (SPSS Principal component analysis with Varimax rotation; Harman, 1967) was performed for each of the 37 texts in order to see whether the eight variables combined into fewer variables, i.e. factors. The factor analysis sought to discover simple patterns in the relationships of the variables and whether the observed variables could be explained by a smaller number of variables, called *factors*. (Darlington 1997.) It was expected based on theory behind the construct of authenticity and thus also behind the questionnaire designed for the main study that two factors would be extracted: one with loadings on item variables

measuring the typicality of the four textual features and the other with loadings on item variables measuring the interest of the textual features.

Although the study focused on examining gender differences in text authenticity, factor analysis was not carried out for each gender separately due to the small number of respondents representing each gender per text.

The results of the analysis are presented in Chapter 8.

7 THE AUTHENTICITY OF THE TEXTS OF THE PISA 2000 READING LITERACY TEST IN THE CULTURE OF FINNISH 15-YEAR-OLDS

7.1 The overall typicality and interest of the PISA texts

7.1.1 The PISA texts by typicality and interest

The texts were ranked by their typicality and interest to see which PISA texts were most typical and interesting to the students participating in the study. However, since the differences in average typicality and interest values between the texts are relatively small and only some of them are statistically significant (p-value < 0,05) the rankings of particular texts are not to be taken as an absolute indication of the interrelations between individual texts. Rather the rankings serve as tool for showing up what came at the top and bottom of the lists as well as for describing the general tendencies on which are typical and interesting PISA texts.

The overall indices of text typicality and interest (Table 10 below) show that the students participating in the study on average found the texts in the PISA reading literacy test only moderately authentic. Text typicality ratings vary between 1,81 and 2,94 and text interest ratings between 1,73 and 2,75. Thus, only the most typical and interesting texts achieve a value near 3, signifying 'typical' or 'interesting'. It is to be noted, however, that the overall indices are the mean of each student's mean of the four textual features and this obscures individual differences and preferences. The students may also have a tendency to rate texts on a scale from 1 to 4 cautiously and avoid the extreme values. Nevertheless it seems likely that the students participating in the study on average found the PISA texts to be only moderately authentic.

TABLE 10   Overall typicality and interest of the PISA 2000 texts for a group of Finnish 15-year-old students.

OVERALL TYPICALITY	mean	OVERALL INTEREST	mean
1. news texts about allergies and explorers*	2,94	1. article about NASA tests on drugged spiders	2,75
2. movie reviews*	2,84	2. motorbike advertisement	2,74
3. argumentative texts about space research	2,69	3. article about using DNA technology in police investigations	2,67
4. article about NASA tests on drugged spiders	2,66	4. news texts about allergies and explorers	2,66
5. information sheet for potential exchange students	2,64	5. argumentative texts on space research	2,64
5. job application form*	2,64	6. etiquette for exchange students	2,60
7. article about the quality of men's shirts	2,63	7. article about men's shirts	2,58
8. argumentative texts on graffiti	2,61	8. information sheet for potential exchange students	2,56
8. report on children's attitudes toward housework*	2,61	8. argumentative texts on graffiti	2,56
8. article about using DNA technology in police investigations	2,61	10. movie reviews	2,54
11. instructions on preparing for a job interview	2,59	11. report on children's attitudes toward housework	2,44
11. editorial on problems that arise with new fertilization technology	2,59	12. graphs about nuclear waste	2,43
13. fable on father's over-protectiveness	2,56	13. narrative about woman and panther trapped in flood	2,40
14. etiquette for exchange students	2,55	14. article about choosing running shoes	2,39
15. article about rhinoceros	2,54	15. editorial on problems that arise with new technology	2,38
16. narrative about woman and panther trapped in flood	2,53	15. fable on father's over protectiveness	2,38
17. article about protecting beaches*	2,48	17. instructions on preparing for a job interview	2,37
18. motorbike advertisement*	2,47	18. article about rhinoceros	2,36
19. brochure describing optician's work*	2,45	19. job application form	2,34
20. article about choosing running shoes	2,41	20. floor plan of a library	2,31
21. graphs about Lake Chad's altitude and fauna	2,37	21. graphs about Lake Chad's altitude and fauna	2,31
22. instructions for contacting employer	2,36	22. brochure describing optician's work	2,29
23. graphs about nuclear waste	2,35	23. article about protecting beaches	2,27
24. information sheet about flu vaccination	2,30	24. table from a report of an aid agency	2,26
25. narrative about villagers' first encounter with movies*	2,28	25. text on conquering the South Pole	2,25
26. warranty instructions for a refrigerator*	2,26	26. information sheet about flu vaccination	2,20
27. information sheet about a company's job centre*	2,25	27. instructions for contacting employer	2,17
28. text on conquering the South Pole	2,21	28. extract from a play	2,15
29. table from a report of an aid agency	2,17	29. narrative about villagers' first encounter with movies	2,05
29. floor plan of a library	2,17	30. information sheet about a company's job centre	2,04

31. bicycle assembly instructions for the buyer	2,14	31. bicycle assembly instructions for the buyer	2,00
32. a sheet from a catalogue and a mail-order form from an art store*	2,10	32. warranty instructions for a refrigerator	1,92
32. tree diagram of structure of labour force*	2,10	33. tree diagram of structure of labour force	1,91
34. telephone use instructions*	2,09	34. a sheet from a catalogue and a mail-order form from an art store	1,86
35. extract from a play	1,97		
36. flight timetable	1,90	35. telephone use instructions	1,85
37. table of news agencies	1,81	36. flight timetable	1,80
		37. table of news agencies	1,73

* Difference between means is statistically significant (p-value< 0,05).

In addition to the relatively modest degree of authenticity attributed in terms of typicality and interest, it is noteworthy that comparing each text's index of typicality and that of interest shows that nearly all the texts were deemed to be more typical than interesting by the students, though only for a third of the texts was the difference statistically significant. Thus the selection of texts in the PISA 2000 reading literacy test seems to have been to some degree authentic in the sense that it consisted of the kinds of texts students usually encounter in their everyday life, but texts did not offer many new and interesting reading experiences. There are, however, some exceptions to the rule, for instance the extract of a play and the motorbike advertisement including a poem and a photograph of an old man. The former text was placed 35th on the list by text typicality (typicality value 1,97) and 28th on the list by text interest (interest value 2,15). The latter text was placed 18th on the list of PISA texts in terms of typicality (typicality value 2,47) but second on the list in terms of interest (interest value 2,74). Other texts found to be more interesting than typical were the article about NASA tests on drugged spiders, the etiquette for exchange students, the two graphs on nuclear waste, the table from a report of an aid agency, the article about using DNA technology in police investigations, the text on conquering the South Pole and the floor plan of a library. In these cases, however, the interest value was only slightly higher than the typicality value, with a difference that was not statistically significant although the ranking may have clearly improved.

The students who participated in the study found that the most typical PISA text was the textual unit consisting of two short news texts, which had a typicality value of 2,94. One is about gene deficit linked to allergies and the other is about explorers at Antarctic. Students explained their ratings by stating that they read a lot of newspapers and news articles are thus familiar with them. The second most typical textual unit consisted of four short movie reviews receiving clearly lower typicality value than the texts ranked first. A female student commented: "*I read a lot of movie reviews and this is quite a typical one*". The text found third most typical consisted of five argumentative texts written by young students about space research and its resources (typicality value 2,69).

Students found that the fourth most typical text in the PISA reading literacy test was an article describing a NASA test in which spiders are given

different drugs to test their toxicity. The text was found only slightly less typical (2,66) than the argumentative texts deemed third most typical. The fifth most typical PISA text was an informative text describing the qualities that applicants to be exchange students should have. Many students explained their rating by referring to the familiar topic of the text. The same typicality value was attributed to a job application form that students found familiar as a text type.

In terms of the PISA classifications of text structure and situations for reading, the most typical PISA texts were continuous ones, with the exception of the job application form which belongs to the non-continuous category of texts (see Table 7). The four situations for reading that PISA texts were to cover, on the other hand, were well represented among the most typical PISA texts, since among the six most typical PISA texts for the students participating in the study there were two texts from a public reading situation, two from an educational situation, and one from private and occupational situations each.

The most interesting PISA text was the article about drugged spiders. The topic of the text was of particular interest to students. A male student commented:"*Drugs on spiders? The idea is kind of weird and therefore I had the energy to read it"*. However, the interest value of the article was quite low (2,75) although the text was one of the few deemed more interesting than typical. The second most interesting text among the PISA reading literacy texts was found to be a motorbike advertisement consisting of a photograph of an old man and a poem that tells how the narrator would live given a chance to live his life again. The text was given nearly the same interest value as the most interesting text (the article on NASA tests). On the list of most typical PISA texts, the text was placed 18th, so it was not deemed commonplace in students' everyday lives. Many students explained that the topic of the poem was appealing:

> "The language of the text was quite strange…but interesting. Although it looked boring the text's content was interesting. That is the most important part. It made you think that you only live once…" (Girl)

The third most interesting text was the article about exploiting DNA technology in police investigations. On the list ranking the PISA texts by the overall typicality, the article shared eighth place with two other texts receiving the same value (2,61). Fourth on the list of most interesting texts, with an interest value of 2,66 was the news item on allergies and explorers. A girl referred to the exhausted explorers in the news article by saying: "*Upsetting news is more interesting to read than good news"*.

The students' opinions on space research were deemed the fifth most interesting text by those participating in the study (interest value 2,64). The genre of students' opinions seems to have been found both typical and interesting but also the topic was of interest even to some boys:

> "I read a lot of texts like this in youth magazines… readers' opinions expressed clearly…It is interesting to read other people's opinions. You can compare them with your own." (Girl)
> "The topic has been discussed in the media for years. It is very interesting. They are trying to find a solution to the problems mentioned in the text(s)." (Boy)

In terms of text structure and PISA situations for reading, the most interesting PISA texts were all continuous ones. The reading situations these texts represent, however, were varied–public as well as educational ones.

The least typical PISA text was the table consisting of a list of news agencies (typicality value 1,81). The same text was also deemed to be the least interesting among the 37 PISA texts receiving an interest value clearly lower than other texts. The second least typical text was the flight timetable which was also the second least interesting PISA text, with an average typicality value of 1,90 and an average interest value of 1,80. The telephone use instructions of a hotel, the tree diagram of the structure of a country's labour force, the art store catalogue, the mail-order form and the bicycle assembly instructions for a buyer were deemed both untypical and uninteresting.

It is noteworthy that in terms of the text structure categories of PISA, nearly all the least typical and least interesting texts mentioned above are non-continuous. While the bicycle assembly instructions are continuous, the text does also contain non-continuous elements. The reading situations for the least typical and interesting PISA texts are either public or educational (see Table 10). The exception to the rule is the play extract rated the third least typical PISA text since it is classified as continuous and placed in a private context.

Overall, among the ten least typical PISA texts there are nine texts that are either non-continuous or include both non-continuous and continuous elements, the text on conquering the South Pole and the bicycle assembly instructions representing the latter type. The play extract deemed the third least typical PISA text (typicality value 1,97) on the other hand is classified as continuous narrative in the PISA framework. On the other hand, there are also non-continuous elements in some texts among the most typical PISA texts, namely the job application form, which is classified as non-continuous, and the article on the quality of men's shirts and the report on children's attitudes toward housework which include a table and bar graphs respectively. This suggests that the reason for texts being found untypical is somewhat more complicated than just their non-continuous structure.

Since the number of students participating in the study was quite low and each text was assessed by approximately 50 students the conclusions reached as to the authenticity of the PISA texts based on the results presented above reflect the opinions and experiences only of the students participating in the study. There are individual differences among students in evaluating text typicality and interest as there are differences between sub-groups of students.

Next, text typicality and interest are studied in terms of gender.

7.1.2 Gender differences in the typicality of the PISA texts

Finnish students' ratings of the PISA texts revealed some clear differences between the two genders with respect to both the typicality and the interest of the texts. Girls and boys had to some extent different experiences of what kinds of text are typical in their everyday lives (Table 11). For both genders the most

TABLE 11   Typicality of texts in the PISA 2000 reading literacy test by gender.

| Text | Typicality of text | | | |
| | Girls | | Boys | |
	mean	ranking	mean	ranking
1. argumentative texts about space research	3,06	1.	2,33	22.
2. report on children's attitudes toward housework	2,87	2.	2,35	20.
3. job application form	2,86	3.	2,38	18.
4. news texts about allergies and explorers	2,85	4.	3,01	1.
5. information sheet for potential exchange students	2,84	5.	2,41	16.
6. movie reviews	2,79	6.	2,89	2.
7. article about using DNA technology in police investigations	2,73	7.	2,48	11.
8. narrative about woman and panther trapped in flood	2,73	7.	2,34	21.
9. article about NASA tests on drugged spiders	2,72	9.	2,59	5.
10. etiquette for exchange students	2,72	9.	2,39	17.
11. article about the quality of men's shirts	2,71	11.	2,53	7.
12. article about protecting beaches	2,70	12.	2,24	27.
13. argumentative texts on graffiti	2,69	13.	2,53	7.
14. motorbike advertisement	2,64	14.	2,28	24.
15. article on rhinoceros	2,63	15.	2,43	13.
16. fable on father's over protectiveness	2,61	16.	2,51	10.
17. instructions on preparing for a job interview	2,55	17.	2,63	4.
18. editorial on problems that arise with new technology	2,50	18.	2,68	3.
19. graphs on nuclear waste	2,40	19.	2,30	23.
20. article on choosing running shoes	2,38	20.	2,43	13.
21. brochure describing optician's work	2,37	21.	2,54	6.
22. instructions for contacting employer	2,37	21.	2,36	19.
23. narrative about villagers' first encounter with movies	2,34	23.	2,23	28.
24. information sheet about flu vaccination	2,33	24.	2,28	24.
25. information sheet about a company's job centre	2,31	25.	2,19	30.
26. bicycle assembly instructions for the buyer	2,30	26.	1,98	35.
27. warranty instructions for a refrigerator	2,24	27.	2,28	24.
28. graphs about Lake Chad's altitude and fauna	2,20	28.	2,52	9.
29. floor plan of a library	2,18	29.	2,15	31.
30. a sheet from a catalogue and a mail-order form from an art store	2,16	30.	2,04	33.
31. extract from a play	2,14	31.	1,83	37.
32. telephone use instructions	2,13	32.	2,06	32.
33. table from a report of an aid agency	2,12	33.	2,22	29.
34. text on conquering the South Pole	1,97	34.	2,44	12.
35. flight timetable	1,88	35.	1,93	36.
36. tree diagram of structure of labour force	1,74	36.	2,42	15.
37. table of news agencies	1,63	37.	2,00	34.

typical text received a typicality value above 3,00, meaning 'typical', but the texts were not the same. Girls used a wider scale than boys in rating the PISA texts by typicality, girls' average ratings ranging between 1,63 and 3,06 and

boys' between 1,83 and 3,01. Either boys tended to rate the texts more cautiously, avoiding the extreme values (1 and 4), or they did find especially the texts at the end of the list of PISA texts by typicality generally more typical than girls did.

The most typical textual unit for girls was the group of argumentative texts in which five young students expressed their opinions on space research and the resources used in it (typicality value 3,06). The second most typical text for girls was the report on children's attitudes to doing household tasks, which was given a typicality value clearly lower than the most typical text unit. The text included a bar diagram describing percentages of students representing different attitudes in various countries. The typicality of the two texts reflects the popularity of texts dealing with people's opinions on various topics among girls, as is illustrated in the quotations about the text dealing with housework below:

> "Questionnaires like that are familiar, and the responses…it is interesting to know what other people think." (Girl).
> "Quite interesting because this involves me too." (Girl)
> "Questionnaires and studies in general are interesting! A big plus for this one!" (Girl)

The third most typical text for the girls was the job application form for a fast food restaurant. This text received a typicality value that was nearly the same as the second most typical text. Forms were said to be familiar text type, and for some students applying for a job was a familiar experience:

> "I haven't seen a job application form before but it reminds me of the form used to apply for upper secondary school or other forms. I have filled in a lot of forms lately." (Girl)
> "I have filled in bits of paper (where you must write your name, address etc.) before but I haven't filled in a job application form." (Girl)

The news about gene deficit linked to allergies and explorers in Antarctica was put fourth on the list of most typical PISA texts by girls, and the information sheet for potential exchange students describing the desirable qualities of applicants came fifth. In the latter, the topic was appealing to students: "*I think that the text is quite boring, just plain text with no variation. The topic was interesting but they could have illustrated it a bit. It would be nice to have exchange students' experiences in it.*" (Girl) The texts placed second, third, fourth and fifth on the list received typicality values very close to each other. Thus, the rank of each of these texts is not very informative since the texts were almost equally typical for the girls involved in the study.

In terms of text structure and format as described in PISA, the most typical PISA texts for girls proved to be continuous texts with the exception of the job application form. The reading situations that the most typical texts for girls represent, however, vary since they included three texts classified as to be read in educational reading situation, and one each in occupational and public reading situations.

From the point of view of typicality, boys rated PISA texts somewhat differently. Among the five most typical texts for boys, the short news item about allergies and explorers were placed first (typicality value 3,01), movie reviews were placed second (typicality value 2,89), and the editorial on ethical problems that emerge with new fertilization technology was placed third (typicality value 2,68). The information sheet giving instructions for job interviews ranked as fourth most typical (typicality value 2,63) and the article about NASA tests on spiders fifth. Although the text was ranked higher by boys than girls, it was still rated more typical by girls than by boys, receiving a typicality value of 2,72 for girls and 2,58 for boys. Girls rated especially the editorial placed third by boys, but also the instructions for a job interview as clearly less typical than boys did. The editorial was placed 18th on the list of most typical texts for girls and the job interview instructions were placed 17th. Moreover, the text found sixth most typical by boys – the brochure describing optician's work – was clearly less typical for girls, who rated it to be 21st with a typicality value of 2,37 for girls (2,54 for boys).

In terms of text structure, the five PISA texts most typical for boys are all continuous texts. Three of the texts, including that judged the most typical, are categorised as texts for reading a public situation. The movie reviews, on the other hand, belong to a private reading situation since they are usually read to satisfy personal interest, while the instructions for job interview were – for obvious reasons – assigned to an occupational situation for reading.

Looking at the content of the five most typical texts as judged by boys and girls shows that boys favoured texts that deal with issues in natural sciences and technology. Girls, however, seem to find texts including expressions of opinion most typical. At the age of 15 Finnish boys and girls are at the common stage of leaving school and applying for further education and possibly for summer jobs. Thus, it is no surprise that on the list of the five most typical PISA texts for both genders there is a text dealing with applying for a job. The instructions on preparing for a job interview are on the boys' list and the job application form on the girls' list. Otherwise the contents of the five most typical PISA texts reflect issues common in youth culture, such as movies and a potential year as an exchange student.

It is noteworthy that girls find the most untypical PISA texts to be non-continuous documents. Among the ten most untypical PISA texts for girls there are nine non-continuous documents, such as the table of news agencies (typicality value 1,63), the tree diagram of the structure of a country's labour force (1,74), the flight timetable (1,88), and the floor plan of a library (2,18). On the other hand, the second and third most typical PISA texts for girls (the report on young people's attitudes to doing housework and the job application form) include a substantial amount of non-continuous elements. Based on the girls own comments (examples given above) it seems that the topic and content of the two texts is closer to girls' everyday lives than the topics of the texts deemed untypical.

Boys also deemed many non-continuous texts untypical since they placed six non-continuous documents among the ten most untypical PISA texts. There were some clear differences as to which PISA texts were found untypical by boys and which by girls. The least typical text for boys was the play extract that received the typicality value of 1,83. Girls put the text in 31st place with typicality value of 2,14. The second least typical text for girls was the tree diagram of a country's labour force. Boys clearly found it to be more typical since it was placed 15th with a typicality value of 2,42. The text on conquering the South Pole was clearly more typical to boys than to girls as boys' ratings placed it 12th and girls' ratings 34th. The graphs about Lake Chad's altitude and fauna were deemed the ninth most typical PISA text by boys but 28th most typical by girls. On the other hand, there were also similarities, since the flight timetable, the table of news agencies, and the telephone use instructions were deemed to be among the least typical texts by both gender groups. Overall, boys assessed non-continuous texts more positively than girls, since on the boys' list there were non-continuous documents or texts containing significant non-continuous elements also in places 9, 12, 15, 18, and 20.

The four literary texts in PISA were all found more typical by girls than by boys. In PISA terms, these texts were all continuous narratives read in a private context. The most typical of the four texts for girls was the story about the woman and the panther trapped in a Mississippi flood, placed eighth on the girls' list and 21st on the boys' list with the typicality values of 2,73 and 2,34 respectively. The fable about the overprotective father, on the other hand, was the most typical of the literary texts for boys who rated it the tenth most typical whereas girls found it to be the 16th most typical. The story about villagers' first encounter with movies was 23rd on the girls' list and, with a slightly lower typicality value, 28th on the boys' list. As mentioned above the play extract was deemed the least typical of the PISA texts by boys. Since the content of the play was romantic, involving a prince and a girl looking like the prince's dead bride it was to be expected that the text would not be part of the boys' textual world.

7.1.3 Gender differences in the interest of the PISA texts

In evaluating the interest of the PISA texts, as with their typicality, girls' and boys' different experiences affected which texts were interesting to them (Table 12). Furthermore, girls again used a wider scale than boys, since girls' average ratings of text interest varied from 1,63 to 2,96 and boys' average ratings from 1,65 to 2,76. This time the difference lies at the top of the list by interest since the most interesting PISA text for girls was attributed a 0,2 point higher interest value than the most interesting text for boys. Overall, the PISA texts were not found as interesting as they were typical.

TABLE 12 Interest of texts in the PISA 2000 reading literacy test by gender.

Text	Interest of texts			
	Girls		Boys	
	mean	ranking	mean	ranking
1. argumentative texts about space research	2,96	1.	2,34	13.
2. motorbike advertisement	2,84	2.	2,63	3.
3. argumentative texts about graffiti	2,81	3.	2,30	16.
4. article about using DNA technology in police investigations	2,80	4.	2,53	6.
5. etiquette for exchange students	2,80	4.	2,40	10.
6. report on children's attitudes toward housework	2,78	6.	2,09	26.
7. information sheet for potential exchange students	2,73	7.	2,36	11.
8. article about NASA tests on drugged spiders	2,73	7.	2,76	1.
9. news texts about allergies and explorers	2,66	9.	2,66	2.
10. narrative about woman and panther trapped in flood	2,63	10.	2,19	21.
11. movie reviews	2,60	11.	2,48	7.
12. article about the quality of men's shirts	2,56	12.	2,61	4.
13. article on rhinoceros	2,55	13.	2,16	23.
14. job application form	2,53	14.	2,11	25.
15. article on choosing running shoes	2,49	15.	2,31	15.
16. fable on father's over protectiveness	2,48	16.	2,26	18.
17. article about protecting beaches	2,47	17.	2,03	28.
18. instructions on preparing for a job interview	2,47	17.	2,28	17.
19. extract from a play	2,43	19.	1,91	32.
20. table from a report of an aid agency	2,42	20.	2,09	26.
21. editorial on problems that arise with new technology	2,41	21.	2,35	12.
22. graphs on nuclear waste	2,41	21.	2,44	8.
23. brochure describing optician's work	2,32	23.	2,25	19.
24. a floor plan of a library	2,29	24.	2,34	13.
25. information sheet about flu vaccination	2,27	25.	2,15	24.
26. graphs about Lake Chad's altitude and fauna	2,20	26.	2,41	9.
27. instructions for contacting employer	2,11	27.	2,22	20.
28. narrative about villagers' first encounter with movies	2,11	27.	2,00	31.
29. information sheet about a company's job centre	2,06	29.	2,01	30.
30. warranty instructions for a refrigerator	2,02	30.	1,82	33.
31. a sheet from a catalogue and a mail-order form from an art store	2,02	30.	1,69	35.
32. telephone use instructions	2,01	32.	1,68	36.
33. bicycle assembly instructions for the buyer	1,97	33.	2,02	29.
34. flight timetable	1,93	34.	1,65	37.
35. text on conquering the South Pole	1,91	35.	2,57	5.
36. table of news agencies	1,71	36.	1,74	34.
37. tree diagram of structure of labour force	1,63	37.	2,17	22.

The most interesting PISA unit according to girls consisted of five argumentative texts on space research (interest value 2,96). Based on the comments written by students the genre of the texts was of particular interest to girls. Reading other people's opinions and expressing their own seemed

popular among girls. The second most interesting PISA text for girls was the motorbike advertisement with a photograph of an old man and a poem, which had an interest value that was clearly lower than the most interesting text, although overall the differences between the ratings were quite small. The advertisement was one of the few texts that both boys and girls found interesting. The topic of the poem especially was found interesting; however, judging by their comments not all the students realized that the poem was part of an advertisement. The third most interesting PISA text unit for girls was made up of argumentative texts about graffiti (2,81) and, as illustrated in the following quotes, the genre was the most interesting aspect of the texts:

> "Nearly all magazines have pages where people can express their opinions. I read a lot of magazines so these pages are familiar to me. The text is interesting because it is fun to see what people think about different issues. Others couldn't care less about graffiti but others are very upset about it." (Girl)
> "I write a lot about my opinions." (Girl)

The fourth most interesting PISA text for girls was the article about the police using DNA technology in their investigations, which was attributed nearly the same interest value as the third most interesting text. The article, like the motorbike advertisement, was a text appealing also to boys, who placed it sixth. Some students appraising the article found test designers' concern to use texts that also look authentic quite positive. As one boy wrote: "*The text wasn't just a big chunk of writing, which is positive (and a familiar feature in newspaper texts). There was even a picture! The layout was nice to look at.*" There was also another text rated at the same interest value (2,80), which thus shared the ranking of the fourth most interesting text to girls, namely that about etiquette for exchange students designed to guide them as to how to behave abroad. Overall, among the five most interesting PISA texts for girls, there were only continuous texts, representing either public or educational situations for reading. Texts expressing opinions on current issues were of particular interest to girls but absent from the boys' list. The report about young people's attitudes towards doing housework was also interesting for girls and placed sixth, since it describes opinions on an issue that is relevant in young peoples' lives.

Boys evaluated the interest of the PISA texts quite differently. The most interesting PISA texts for boys were continuous texts representing a public situation for reading. The most interesting text according to the boys participating in the study was the article about NASA tests on drugged spiders (interest value 2,76). Generally boys found that the topic and content of the text were of great interest to them. This article was only the seventh most interesting text for girls, but it was still one of the texts that both genders alike found interesting since the interest value was nearly the same for both genders. Boys found that the second most interesting textual unit in the PISA reading literacy test consisted of the two short news items about gene deficit linked to allergies and explorers in Antarctica. This text, too, was of the same interest value for both genders (2,66) although it was only ninth on the girls' list of most interesting texts.

The third most interesting PISA text according to boys was the motorbike advertisement consisting of a photograph of an old man and a poem in which the narrator told what he would do differently given the chance to live his life again. As stated above, the text was also found interesting by the girls participating in the study. Although the ranking of the text was nearly the same for the two genders, the average interest value for girls (2,84) was clearly higher than that for the boys (2,63). The fourth most interesting PISA text for boys was the article about men's shirts, describing the results of a study in which cheap department store shirts were found better in quality than expensive designer shirts (interest value 2,61). The text included a table of the qualities of the different shirts in the study, and belongs in a private reading situation in PISA terms. Girls found the article nearly as interesting (interest value 2,56) although the ranking was clearly lower. The textbook extract about conquering the South Pole was the fifth most interesting text for boys, with an interest value of 2,57. The text includes a map of Antarctica and a diagram. The topic of the text appealed to boys, and as one of them commented: "*It is very textbook like. The text is very interesting description of conquering the South Pole.*" This text clearly divided girls and boys into different textual worlds since girls found it to be 35th on their list by text interest (average interest value 1,91). Additionally, boys found that the article about the police using DNA techniques in their investigations was nearly as interesting as the South Pole text. Both texts belong in an educational reading situation in PISA terms. The article was attributed an interest value of 2,53. Girls deemed the same text to be clearly more interesting, although the ranking (fourth) was only two places higher than that of the boys.

The least interesting PISA texts for boys were the flight timetable, the table including the telephone use instructions of a hotel, the mail-order catalogue of an art store, the table of news agencies, and the warranty instructions for a refrigerator. These texts were also among the least interesting PISA texts for girls. Apart from the warranty instructions, all the texts listed above were non-continuous. Moreover, they all represent a public context, except the art store catalogue which is classified as educational text. A female student wrote about the table of news agencies as follows: "*It reminds me of stock exchange tables…listing things like that looks boring.*" Generally students found that the content of the texts at the end of the list was uninteresting because it was irrelevant to them. For instance the table including telephone use instructions was commented on as follows:

"Lists like that are familiar. There is no need to know the area codes in Turkey." (Boy)
"It is about area codes in Turkey! I would never read it voluntarily/for fun. It has information that I don't need!" (Girl)

Despite the general picture of the uninteresting texts seeming to be judged equally so by both genders, there were, however, some texts at the end of the list of texts by interest that boys clearly found more interesting than girls. One of them is the text on conquering the South Pole as described above. Another is the tree diagram of a country's labour force, which girls deemed to be the least

interesting PISA text but boys placed 22nd on the list by interest, with an interest value 0,54 higher than that found by the girls. The graphs on Lake Chad's altitude and fauna were also more interesting to boys than girls, since boys placed it ninth but girls only 26th on the list of PISA texts by interest. The latter two of these three texts, which appear to be part of boys' textual world rather than girls', consist solely of non-continuous elements.

The literary texts in PISA were generally found quite uninteresting, especially by boys. Girls gave higher interest values to all four literary texts in the study. The play extract was 32nd on the boys list having a clearly lower interest value for boys than girls, who ranked it 19th. The extract from a narrative about villagers' first encounter with movies did not interest boys either, since they placed it 31st on the list by interest. Girls also found this particular text to be the least interesting literary text in the PISA reading literacy test (27th). Two other narratives in PISA were also deemed quite uninteresting by boys, the fable ranking 18th and the story about a woman and a panther trapped in a flood 21st on the boys' list of the 37 PISA texts by interest. Girls, on the other hand, found the story about the woman and the panther to be the most interesting literary PISA text. The story came 10th on their list by interest with an interest value of 2,63.

To summarize, continuous texts from public and educational contexts were deemed interesting by both genders, although not necessarily the same texts. The most uninteresting texts, especially for girls, were non-continuous and also from public and educational reading situations. The two contexts were well represented in PISA, since altogether 25 out of the 37 PISA texts were from these situations for reading.

7.2   The typicality and interest of the PISA texts by topic

7.2.1 Overall typicality and interest of topics

When students had the opportunity to freely explain their ratings of the PISA 2000 reading literacy texts they usually referred to either the topic or the genre of the text. The topic of the text was usually referred to in the case of non-continuous texts deemed untypical or uninteresting by the students, as with texts clearly found more interesting than typical. As reported above e.g. the article about the NASA test on drugged spiders and the motorbike advertisement were such texts. Students found the topics of the two texts appealing. The genre of the text was referred to especially in the case of texts originally appearing in newspapers or magazines and with texts expressing the writer's opinions. The former were of special interest to boys and the latter to girls. It is interesting to see, however, which texts were typical and interesting solely in terms of topic and genre. What kind of topics and genres among PISA

texts were on average found authentic by the 15-year-old Finnish students participating in the study?

In assessing topic typicality and interest students again used the scale from 1 to 4 cautiously. The most typical topic received an average typicality value of 2,93, and the most interesting topic an average interest value of 2,90 (Table 13). The least typical topic received a typicality value as low as 1,52 and the lowest interest value was only slightly higher. Thus it seems that no topic was on average very interesting or even very typical. In fact, the topics did not quite reach the value signifying interesting or typical (value 3). On the other hand, it is striking that the least typical and likewise the least interesting topic was the same. Students did not find news agencies as a topic either typical or interesting.

TABLE 13   Overall typicality and interest of the topic of the PISA 2000 texts for a group of Finnish 15-year-old students.

TYPICALITY OF TOPIC	mean	INTEREST OF TOPIC	mean
1. news texts about allergies and explorers	2,93	1. article about NASA tests on drugged spiders	2,90
2. movie reviews*	2,89	2. news texts about allergies and explorers	2,86
3. argumentative texts about space research	2,63	2. information sheet for potential exchange students	2,86
4. report on children's attitudes toward housework	2,56	4. motorbike advertisement	2,80
5. information sheet for potential exchange students*	2,53	5. article about DNA technology in police investigations	2,73
6. argumentative texts on graffiti	2,49	5. argumentative texts on space research	2,73
7. etiquette for exchange students	2,47		
8. article about the quality of men's shirts*	2,43	5. article about the quality of men's shirts	2,73
9. article about DNA technology in police investigations*	2,40	8. argumentative texts on graffiti	2,66
10. job application form	2,39	9. etiquette for exchange students	2,61
		9. movie reviews	2,61
11. editorial on problems that arise with new technology	2,37	11. editorial on problems that arise with new technology	2,57
12. instructions on preparing for a job interview	2,35	12. graphs about nuclear waste	2,48
12. motorbike advertisement*	2,35	13. report on children's attitudes toward house hold work	2,45
14. article about rhinoceros	2,32	14. article about rhinoceros	2,44
15. fable on father's over protectiveness	2,31	15. fable on father's over protectiveness	2,43
16. graphs about Lake Chad's altitude and fauna	2,30	16. table from a report of an aid agency	2,42
17. instructions for contacting employer	2,28	17. instructions on preparing for a job interview	2,41
18. article about protecting beaches	2,27	18. job application form	2,36
19. narrative about woman and panther trapped in flood	2,25	19. article about choosing running shoes	2,35
20. article about choosing running shoes	2,22	20. narrative about woman and panther trapped in flood	2,33

21. brochure describing optician's work	2,15	21. article about protecting beaches	2,31
22. article about NASA tests on drugged spiders*	2,14	22. graphs about Lake Chad's altitude and fauna	2,30
23. graphs about nuclear waste*	2,13	22. text on conquering the South Pole	2,30
24. text on conquering the South Pole	2,11	24. brochure describing optician's work	2,28
25. information sheet about flu vaccination	2,04	25. instructions for contacting employer	2,24
26. warranty instructions for a refrigerator*	2,00	26. floor plan of a library	2,22
26. floor plan of a library	2,00	27. information sheet about flu vaccination	2,10
28. table from a report of an aid agency*	1,96	28. extract from a play	2,02
29. extract from a play	1,92	29. narrative about villagers' first encounter with movies	1,92
29. a sheet from a catalogue and a mail-order form from an art store	1,92	30. information sheet about a company's job centre	1,86
31. narrative about villagers' first encounter with movies	1,90	30. bicycle assembly instructions for the buyer	1,86
32. tree diagram of structure of labour force	1,89	30. a sheet from a catalogue and a mail-order form from an art store	1,86
33. flight timetable	1,85	33. tree diagram of structure of labour force	1,85
34. bicycle assembly instructions for the buyer	1,81	33. telephone use instructions	1,85
35. telephone use instructions	1,80	35. flight timetable	1,80
36. information sheet about a company's job centre	1,78	36. warranty instructions for a refrigerator	1,71
37. table of news agencies	1,52	37. table of news agencies	1,58

* Difference between means is statistically significant (p-value< 0,05).

Even though the PISA texts overall tended to be more typical than interesting, the situation in terms of topic only is exactly the opposite. Comparing the typicality and interest values of each text revealed that most of the topics (26 texts out of 37) were deemed more interesting than typical. Some of the differences between average values were quite small and not statistically significant, e.g. the topic typicality value for the article about protecting beaches was 2,27 and the topic interest value for the same text was 2,31. On the other hand, there were also texts the topic of which was clearly considered to be more interesting than typical, e.g. the article on NASA tests on drugged spiders, the motorbike advertisement, and the information sheet for potential exchange students. With these texts, the difference between topic typicality and topic interest values was statistically significant.

There were only 8 texts out of the 37 where the topic had higher typicality value than interest value. With some texts the differences in the average values were quite small, but there were a few texts the values for which had clear differences. For instance, the warranty instructions for a refrigerator was attributed a typicality value of 2,00 and an interest value of 1,71. Three of the texts had the same or nearly the same typicality and interest values.

Only five of the texts had a topic familiar enough to receive an average typicality value over 2,50. The news texts about allergies and explorers were deemed to be the text unit with the most typical topic. However, students did

not make many comments about the typicality of the topic (allergies, explorers). The interest of the topic was commented upon more. Some students, however, referred to the personal relevance of the allergy related text, like one girl who stated: "*Since I have asthma myself these things keep coming and I should read them and I do.*" Movie reviews came second with a typicality value nearly the same (2,89). Students found movies to be a familiar and relevant part of their lives. A student recognized one of the films reviewed:

> "The fact that it was a movie review was familiar [not e.g. a book review or play review]." (Girl)
> "These are familiar because I see a lot of films." (Boy)
> "Sister Act is a familiar movie." (Girl)

The texts about space research written by young students had the third most typical topic with a typicality value of 2,63. There were many critical comments about the topic of the texts, especially its interest. However, many found the topic in general or some details of the content familiar or typical:

> "It is about opinions, which is familiar. Space research is a stranger feature." (Girl)
> "One meets opinions like that every day. The topic students are writing about is also familiar." (Girl)
> "I don't usually talk about deceases-- they talk about space but they don't know about it too much." (Boy)
> "We are talking about energy production at school and we have talked about pollution.--" (Boy)

Young people's attitudes towards housework were found to be another typical topic by the students. The report was rated the fourth most typical topic among the PISA texts with a typicality value of 2,56. Students' comments on the topic revealed that while some students found that the typical feature of the text was the description of a survey many felt that the topic was close to their personal experience of doing housework and negotiating about it with their parents:

> "The questionnaires and answers are familiar." (Girl)
> "I don't read non-fictional texts. There is lot of discussion over the topic however." (Girl)
> "I don't have problems like that! It is natural that children should learn to take responsibility…at least sometimes." (Boy)
> "It was about housework, which is familiar." (Boy)
> "We have thought about this at home too." (Boy)
> "We discuss this topic a lot at home." (Girl)

The fifth most typical topic among the PISA texts was found to be the information about the qualifications of exchange students (typicality value 2,53). Another text relating to the same topic, etiquette for exchange students, was deemed to be close in typicality of topic (2,47). Students' comments about the information sheet on exchange students' qualifications reveal that the topic is familiar to students since at the age of fifteen they are the target group of student exchange organizations:

> "I know many exchange students." (Boy)
> "I have read texts about exchange students many times but usually they are advertisements." (Girl)
> "They advertise [the issue] for people of this age.." (Girl)
> "Typical and familiar information about the exchange student year. Everybody has heard of that." (Girl)
> "This is very familiar because my older sister is going and she has told me a lot about exchange student stuff." (Girl)

Overall, topics of PISA texts found most typical seemed to be those related to generally relevant scientific issues (allergies, space research) and to issues familiar to young people in their personal lives (movies, housework, and exchange students).

The least typical topics, on the other hand, seemed to be those that students could not see as relevant in any way to their lives. There were ten texts the topic of which was deemed untypical to the extent that they received an average typicality value under 2,00 (Table 13). The topic of the telephone use instructions including the area codes for Turkish cities, given a typicality value of 1,80 was especially criticized by students for being irrelevant and unnecessary:

> "There is no reason to know the area codes for Turkey" (Boy)
> "I don't need that information, at least not now." (Girl)

The table including a list of news agencies, attributing the lowest typicality value to its topic (1,52) was criticized for its unfamiliar topic but comments were also made about the content more specifically and complaints that it included strange names and abbreviations.

> "Some names were odd and I don't read tables anyway." (Girl)
> "–– strange names–– the topic is hopeless." (Boy)
> "Unfamiliar features…are abbreviations." (Girl)

The flight timetable was another of the texts with the least typical topics (typicality value 1,85). Students criticized the timetable for being a kind of text that they do not encounter in their everyday life. They acknowledged that they read timetables but not flight timetables. Students also complained about the strange abbreviations. However, no student made comments about the timetable being from a country they do not know and only a couple of students made comments about the text including unfamiliar city names. Thus it seems that they did not find the unfamiliar cultural context of the text the most relevant textual feature in rating its typicality from the point of view of the topic.

> "The abbreviations [are unfamiliar]" (Boy)
> "I don't get involved with texts like that." (Girl)
> I read a lot of timetables but they are train or bus timetables." (Girl)

Although most PISA texts proved to have a topic more interesting than typical the highest average interest value was no higher than the highest typicality

value, in fact slightly lower (2,90). However, there were 11 texts with an average topic interest value higher than 2,50 while only five texts had an average topic typicality value higher than that (see above). The most interesting topic was the description of the NASA tests on drugged spiders, although the article was only 22nd on the list of topic typicality (typicality value 2,14). Based on the students' comments below it seems that what attracted young readers in this text was the unusual combination of drugs and spiders.

> "It was very interesting because of the unusual topic." (Boy)
> "I think it is interesting to read about the effect different drugs have and about studies like this." (Girl)
> "Well it was exciting that they did the tests on spiders!" (Girl)
> "Unusual is interesting." (Boy)
> "The topic is interesting and draws the attention. The text involves space, that I love, and spiders that I hate." (Girl)

The topic interest of the news texts about allergies and explorers and the information sheet for potential exchange students was the same since both emerged with an average topic interest value of 2,86. This value was only slightly lower than that of the article about drugged spiders. Students' comments on the two short news texts revealed that they found the content of the both texts appealing:

> "The topic is quite interesting because in the first text [allergies] it is new and so little but still so great." (Boy)
> "The layout in the newspaper often looks like this, black and white, no pictures and looks boring. One doesn't notice it, but if one does one realizes that the text is very interesting. Illnesses and reading about them is interesting." (Girl)
> "Interesting=did the man [explorer] die or not." (Boy)

The text about qualifications for exchange students was deemed generally interesting:

> "The topic is interesting but the language does not appeal to me." (Boy)
> "This would be interesting if I was about to go abroad to be an exchange student. Otherwise it is boring again…" (Girl)
> "–– The topic is interesting but they could have illustrated it a bit. It would be nice if it included exchange students own experiences too." (Girl)

The fourth most interesting topic was that of the motorbike advertisement consisting of a poem and a photograph of an old man (interest value 2,80). In the poem the narrator thought back over his life and what he would do differently if he had the chance to relive it. It seems that many students found the idea of the poem interesting although not everyone realized that it was part of an advertisement, at least at first. They concentrated more on the poem and its message than the purpose of the whole textual unit.

> "Poems are nice and in my opinion this was too. I have read poems about similar topics before. Harley Davidson??? Was this some kind of advertisement???" (Girl)
> "Although the text looked boring, the content was interesting which is the most important thing. It made you think that you only live once…" (Girl)

> "Interesting, the text is short, clear and unpredictable. On the other hand it is boring since it doesn't relate to my life." (Boy)
> "I suppose many people think they could have done something differently. This was nice to read." (Girl)

Three texts came out with the same value for topic interest (2,73). Using DNA technology in police investigations, space research and the quality of men's shirts were all found equally interesting topics. Issues concerning modern police work were thought interesting in the article about DNA technology in police investigations:

> "The beginning was interesting; this kind of thing gets one's attention (about current issues)" (Girl)
> "I am interested in methods used in solving crimes." (Girl)
> "It is interesting to know what the police can do to identify people." (Boy)

Young students' opinions about space research were most often commented upon from the point of view of genre, since students found other people's opinions interesting. However, some students made comments about the topic itself:

> "It is interesting to read other young people's opinions. One can compare them with one's own opinions." (Girl)
> "As I already mentioned I like hearing other people's opinions." (Girl)
> "It is always nice to read about other people's opinions and topic was interesting, too." (Girl)

The topic of the quality of men's shirts was found interesting since it involves issues relevant to the young consumer:

> "I find it interesting to read information about different products." (Girl)
> "The relationship between price and quality is interesting." (Boy)
> "It would have been more interesting if it was about young people's clothes." (Girl)

Nine texts emerged with a topic interest value under 2,00. Thus, the topics of these texts were on average deemed worse than uninteresting or boring. It is noteworthy that eight of the nine least interesting texts were also among the nine least typical texts. The same texts were thus found both untypical and uninteresting by topic. The 9th least interesting topic was the description of how the people of Macondo reacted when they had their first encounter with the movies (interest value 1,92). From the comments students made it seems that many of them did not understand the text due to the lack of context and background information. The text was an extract from a novel and clearly quite difficult to comprehend for many students:

> "There is no sense in the text, you know what I mean. Who cares about unknown people of Macondo who get angry at films? Depth! I mean some background, even a bit…" (Boy)
> "The text doesn't even tell who the people of Macondo are." (Boy)
> "The topic does not interest me. In fact, I don't even understand the text." (Girl)
> "Boring because there was nothing sensible happening in the text." (Girl)

> "[The layout is quite boring.] But it is always nice to learn of other cultures and the reactions they have to new "modern" things." (Boy)

The 8th least interesting topic was a company's job centre (interest value 1,86). Students found the topic unfamiliar and of no current interest although there were also students who saw that the topic might be relevant in their future. Students' comments revealed that the lack of relevance to their current lives made the topic uninteresting. Thus, relevance and interest were intertwined:

> "For grown ups." (Boy)
> "I have never heard of the job centre (TPVK) so I couldn't care less." (Boy)
> "This doesn't really concern me because I haven't had a job, so the topic was boring." (Girl)
> "This is of no interest just now." (Girl)
> "Interesting because this has got to do with the future." (Boy)

The second least interesting topic was the warranty instructions for a refrigerator (1,71). Students found them remote from their everyday lives:

> "These things belong to adults." (Boy)
> "The text was boring to read for pleasure but if one needs information on the matter, it is necessary." (Girl)
> "It was boring because I am not much interested in buying a refrigerator." (Boy)
> "Someone should shoot the person coming up with these topics. Warranty card for a refrigerator! For God's sake! The only positive thing was that the frame of the card is pretty." (Girl)

The least interesting topic was news agencies. Students' comments expressed the view give that a simple list of news agencies with no relevance to their lives is something that provides no sensible reason for reading:

> "Boring: what could be interesting in a list of news agencies??" (Girl)
> "I doubt that young people of my age would be interested in news agencies." (Girl)
> "The text is quite boring and unpleasantly like a list. The topic is hopeless." (Boy)

Overall the least interesting topics were criticized for being irrelevant and for having no current role in the lives of 15-year-old students. Irrelevance as such was given as an explanation for the lack of interest. To summarize, the most typical topics in the PISA texts seem to be those related to issues of general interest, usually scientific or environmental issues or health, and issues relevant to young students' lives, such as movies, housework and the exchange year abroad. The most interesting topics in the PISA texts were also either issues of general interest or issues relevant to 15-year-olds, although the individual texts at the top of the lists by topic typicality and interest are at least partly different.

7.2.2 Gender differences in topic typicality

In assessing topic typicality on a scale from 1 to 4 girls used the scale more boldly than boys. Especially at the lower end of the scale girls clearly rated topic typicality more negatively than boys, since the lowest average typicality rating

given by girls was 1,26 while the lowest value given by boys was 1,69 (Table 14). At the higher end of the scale the difference was smaller, since the highest average typicality value given by girls was 3,00 and that given by boys was 2,96. Overall, girls found more typical topics among the PISA texts than boys,

TABLE 14   Topic typicality of texts in the PISA 2000 reading literacy test by gender.

| Text | Topic typicality | | | |
| | Girls | | Boys | |
	mean	ranking	mean	ranking
1. argumentative texts about space research	3,00	1.	2,27	16.
2. news texts about allergies and explorers	2,90	2.	2,96	1.
3. movie reviews	2,81	3.	2,96	1.
4. report on children's attitudes toward housework	2,80	4.	2,32	11.
5. information sheet for potential exchange students	2,73	5.	2,30	15.
6. etiquette for exchange students	2,64	6.	2,31	12.
7. job application form	2,63	7.	2,12	25.
8. argumentative texts about graffiti	2,56	8.	2,42	6.
9. motorbike advertisement	2,52	9.	2,17	22.
10. article about using DNA technology in police investigations	2,44	10.	2,35	10.
11. article about the quality of men's shirts	2,41	11.	2,46	3.
12. instructions on preparing for a job interview	2,39	12.	2,31	12.
13. fable on father's over-protectiveness	2,38	13.	2,23	18.
14. article about protecting beaches	2,38	13.	2,13	24.
15. instructions for contacting employer	2,33	15.	2,24	17.
16. narrative about woman and panther trapped in flood	2,32	16.	2,19	20.
17. editorial on problems that arise with new technology	2,27	17.	2,46	3.
18. article on rhinoceros	2,23	18.	2,42	6.
19. graphs about Lake Chad's altitude and fauna	2,23	18.	2,38	8.
20. extract from a play	2,18	20.	1,69	37.
21. graphs on nuclear waste	2,15	21.	2,12	25.
22. a floor plan of a library	2,15	21.	1,84	31.
23. information sheet about flu vaccination	2,14	23.	1,96	28.
24. article on choosing running shoes	2,13	24.	2,31	12.
25. brochure describing optician's work	2,11	25.	2,19	20.
26. a sheet from a catalogue and a mail-order form from an art store	2,04	26.	1,79	34.
27. article about NASA tests on drugged spiders	1,92	27.	2,38	8.
28. narrative about villagers' first encounter with movies	1,88	28.	1,92	29.
29. telephone use instructions	1,87	29.	1,73	36.
30. bicycle assembly instructions for the buyer	1,85	30.	1,77	35.
31. table from a report of an aid agency	1,84	31.	2,08	27.
32. warranty instructions for a refrigerator	1,83	32.	2,16	23.
33. flight timetable	1,81	33.	1,90	30.
34. text on conquering the South Pole	1,74	34.	2,46	3.
35. information sheet about a company's job centre	1,71	35.	1,84	31.
36. tree diagram of structure of labour force	1,55	36.	2,20	19.
37. table of news agencies	1,26	37.	1,80	33.

since there were nine texts on the girls' list of PISA texts by topic typicality that were attributed an average typicality value higher than 2,50, whereas there were only two such texts on the boys' list.

Looking at the typicality of topics by gender reveals common trends in evaluating the typicality of topics in the PISA texts but also some substantial differences between genders with certain texts (Table 14). The movies and allergies and explorers seemed to be typical topics to both boys and girls. Girls rated the news texts about allergies and explorers as having the second most typical topic (typicality value 2,90) and boys rated the text at the top of the list (typicality value 2,96). The movie reviews were third on the girls' list and first on the boys' list with the same typicality value as news texts (2,96). On the other hand, there are also differences at the very top of the list of PISA texts in terms of topic typicality. Girls found that space research was the most typical topic among the PISA texts (typicality value 3,00). Boys, however, found the same topic to be the 16th most typical with a typicality value much lower than that given by the girls (2,27). The fourth most typical topic for girls was children's attitudes towards housework (typicality value 2,80), whereas boys found the topic to be the 11th most typical topic (2,32).

There were other texts with a topic more familiar to boys than to girls. The third most typical topic according to boys was the quality of men's shirts (typicality value 2,46). The same average typicality value was also attributed to the editorial on the problems caused by the new fertilization technology and the text on conquering the South Pole. Girls however judged the topics of these texts to be less typical. Girls found the quality of men's shirts to be a slightly less typical topic than boys did since they gave a typicality value of 2,41 which placed the article 11th on the girls' list of PISA texts by topic typicality. The problems caused by the new fertilization technology on the other hand clearly seemed less typical topic to girls than to boys. Girls found the topic to be the 17th most typical with a typicality value of 2,27. The difference between the genders was clearest with the text on conquering the South Pole. While boys put the topic of the South Pole text in equal third place on their list of PISA texts by topic typicality, girls found it to have the fourth least typical topic (average typicality value 1,74). The topic was clearly not appealing to girls at all. In all, the topics of these three texts seem to belong to boys' areas of interest rather than that of girls, at least it was so among the students participating in the study.

The topic of the article about the NASA test on drugged spiders was another clearly more typical to boys than to girls. The article had the eighth most typical topic with an average typicality value of 2,38 on the boys' list, whereas girls clearly found the topic of the article untypical, since it was given an average typicality value of 1,92 placing it 27th on the girls' list. Still another text that can be cautiously characterized as a boys' text is the tree diagram of a country's labour force. The topic of the diagram was deemed the second least typical by girls (1,55) but on the boys' list it was placed 19th with a clearly higher average typicality value (2,20).

The job application form, the poem in the motorbike advertisement and the extract from a play describing the encounter of a prince and Amanda had topics clearly more familiar to girls than boys. The job application form had the seventh most typical topic (2,63) among the PISA texts according to the girls while the boys' ratings placed it only 25th (2,12). The motorbike advertisement was also more familiar in its topic to girls than to boys. This can probably be explained by the poem dealing with issues of living one's life to the fullest and regretting past experiences, or lack of them, in life. Girls found the poem to have the 9th most typical topic (2,52) while on the boys' list of PISA texts by topic typicality it was only 22nd (2,17). The Anouilh play about Amanda who reminds the prince of his deceased lover was expectedly not familiar in its topic to boys, who found it to have the least typical topic among the PISA texts (1,69). The girls' ratings on the other hand placed the text 20th on their list (2,18).

Otherwise, the lower end of the list of PISA texts by topic typicality seems to be quite similar for both genders. Among the least typical texts by topic were, for instance, the table about news agencies, the flight timetable, the information sheet about a company's job centre and the telephone use instructions of a hotel. Three of the four texts commonly judged untypical were non-continuous ones by structure and format. There were, however, slight differences in average typicality values between the two genders. Most of the texts at the end of the list were more typical in their topic to boys than to girls, although the telephone use instructions were an exception to the rule. The most significant difference in this respect at the end of the list was the extremely low typicality value given to the topic of the table about news agencies by girls (1,26). The boys' ratings resulted in an average typicality value of 1,80 although the table was also among the least typical topics on the boys' list (32nd).

Overall, there seemed to be no tendency to rate texts from certain reading situations more typical than others in either gender's ratings. The reading situations represented by the most typical topics were educational, public and private for girls as well as for boys. The work related topics of occupational texts were not considered typical by either gender, with the exception of the job application form placed seventh in the girls' ratings. The topics found untypical by both genders represented all four reading situations in PISA terms, although the only untypical topic from a private context was that of the play extract at the bottom of the boys' list of PISA texts by topic typicality.

7.2.3 Gender differences in topic interest

Looking at the interest of the topics in the PISA texts (Table 15) shows that most of the texts had topics that were more interesting to girls than to boys (24 texts out of 37). With some texts the differences in interest values are quite small (e.g. 0,05) while at the greatest the difference is as much as 0,88. Boys used the scale from 1 to 4 overall more cautiously than girls, since the interest values on their list vary from 1,58 to 2,92. Girls were slightly bolder and assessed the interest of the topics by using more the extreme ends of the scale. Their interest values vary from 1,48 to 3,04.

TABLE 15   Topic interest of texts in the PISA 2000 reading literacy test by gender.

| Text | Topic interest | | | |
| | Girls | | Boys | |
	mean	ranking	mean	ranking
1. argumentative texts about space research	3,04	1.	2,42	12.
2. information sheet for potential exchange students	3,04	1.	2,65	6.
3. news texts about allergies and explorers	2,95	3.	2,79	2.
4. motorbike advertisement	2,93	4.	2,67	4.
5. etiquette for exchange students	2,92	5.	2,31	14.
6. argumentative texts about graffiti	2,89	6.	2,42	13.
7. report on children's attitudes toward housework	2,88	8.	2,00	28.
8. article about NASA tests on drugged spiders	2,88	7.	2,92	1.
9. article about using DNA technology in police investigations	2,80	9.	2,65	5.
10. movie reviews	2,70	10.	2,52	9.
11. article about the quality of men's shirts	2,70	11.	2,75	3.
12. editorial on problems that arise with new technology	2,64	12.	2,50	10.
13. fable on father's over protectiveness	2,62	14.	2,19	22.
14. article on rhinoceros	2,62	13.	2,25	19.
15. table from a report of an aid agency	2,60	15.	2,24	20.
16. job application form	2,59	16.	2,09	25.
17. instructions on preparing for a job interview	2,57	17.	2,27	18.
18. article about protecting beaches	2,50	18.	2,09	24.
19. extract from a play	2,50	19.	1,62	36.
20. narrative about woman and panther trapped in flood	2,44	20.	2,22	21.
21. article on choosing running shoes	2,43	21.	2,27	17.
22. graphs on nuclear waste	2,42	22.	2,54	8.
23. brochure describing optician's work	2,37	23.	2,19	23.
24. instructions for contacting employer	2,20	24.	2,28	16.
25. information sheet about flu vaccination	2,18	25.	2,04	27.
26. a floor plan of a library	2,15	26.	2,28	15.
27. graphs about Lake Chad's altitude and fauna	2,14	27.	2,46	11.
28. a sheet from a catalogue and a mail-order form from an art store	2,12	28.	1,58	37.
29. text on conquering the South Pole	1,96	29.	2,63	7.
30. telephone use instructions	1,92	30.	1,77	33.
31. flight timetable	1,88	31.	1,70	34.
32. narrative about villagers' first encounter with movies	1,85	32.	2,00	29.
33. information sheet about a company's job centre	1,79	33.	1,92	31.
34. bicycle assembly instructions for the buyer	1,77	34.	1,96	30.
35. warranty instructions for a refrigerator	1,63	35.	1,80	32.
36. tree diagram of structure of labour force	1,59	36.	2,08	26.
37. table of news agencies	1,48	37.	1,68	35.

Girls and boys had quite a lot in common in terms of topic interest. Among the ten most interesting topics for girls there were 6 texts with a topic among the ten most interesting also for boys (Table 15). Five of the six texts in the top ten of both genders were more interesting to girls than boys. For instance allergies and explorers received the average interest value of 2,95 in girls' ratings and an

interest value of 2,79 in the boys' ratings. The interest values given placed the text third on the girls' list and second on the boys' list. The qualifications of potential exchange students was attributed an average interest value of 3,04 by the girls participating in the study, and 2,65 by the boys. On the girls' list of PISA texts by the topic interest the text was placed equal first but on the boys' list it was equal fifth. The motorbike advertisement which was fourth on both genders' list of PISA texts by topic interest received higher average interest value on the girls' ratings than on the boys'. The NASA tests on drugged spiders, on the other hand, was also among the ten most interesting topics on both genders' lists but was given slightly higher average interest value (2,92) in the boys' ratings than in the girls' (2,88). On the boys' list the article was placed first and on the girls' list seventh.

The four texts that were not on both genders' top ten list of PISA texts by topic interest had topics that clearly divided girls and boys. One example was opinions about space research, which was the most interesting topic according to the girls' ratings (interest value 3,04). Whereas the boys placed the text's topic 12th with an average interest value clearly lower than that given by girls (2,42). Etiquette for exchange students was another topic more appealing to girls than to boys. Girls found it to have the fifth most interesting topic, while the boys' ratings placed it 14th with clearly lower interest value. Children's attitude towards housework was another girls' topic. According to the girls' ratings it was the 7th most interesting topic (2,88) at the opposite end of the list from the boys' ratings, which placed it only 28th (2,00). The text on conquering the South Pole, on the other hand, was boys' text in the sense that its topic appealed to boys but not to girls. On the boys' list of PISA texts by interest the report was in seventh place (2,63) but on the girls' list only 29th (1,96). The low average interest value given by girls means that they found the topic to be worse than 'boring'.

Several texts appeared at the bottom of both the boys' and the girls' lists of PISA texts by topic interest. Among the nine least interesting topics were seven texts the topic of which both boys and girls found uninteresting or boring. For instance news agencies were a topic that was deemed distinctly worse than boring in both genders' ratings. Girls especially found it uninteresting. The warranty of a refrigerator, bicycle assembly, a company's job centre, and villagers' first encounter with movies were topics found uninteresting by both genders. However, the girls' average ratings of the interest of the topics mentioned were more negative than those of the boys'. For instance the story about the people of Macondo seeing films for the first time was given an average interest value of 1,85 by girls but 2,00 by boys. Nevertheless, the ranking of the text was nearly the same for both genders, the girls placing it 32nd and the boys 29th. It thus seems that girls were more critical than boys in assessing the interest of the topics of many texts at the bottom of the list. There were, however, exceptions to the rule, since the flight timetable as well as the telephone use instructions were both placed three positions lower on the boys' list of PISA texts by topic interest than on the girls' list. The interest values that

the topics of the texts were given were slightly lower on the boys' list as well. In all, the differences in interest values and rankings of the topics at the bottom of the list of PISA texts by topic interest were not substantial. In this sense it seems that there were more common than separating elements for both genders at the bottom of the lists.

However, there were two texts at the bottom of the list of PISA texts by topic interest for each gender that were not on the list of the other gender. As at the top of the list, these texts not judged in common by boys and girls held substantial differences in topic interest for the two genders. The text about conquering the South Pole was one of these texts. As mentioned above, it had the seventh most interesting topic according to boys. However, girls found that the topic of the text was unappealing since it was only 29th on their list with an interest value clearly lower than that for the boys. Another text with a topic more appealing to boys than girls was the tree diagram of a country's labour force. Girls clearly found its topic boring since their ratings resulted in average interest value of 1,59 which placed the text second to bottom. Boys, on the other hand, found the topic of the text somewhat more interesting, since their ratings placed the text 26th.

There were two texts with a topic extremely unappealing to boys. The play extract about Amanda, who reminded the prince of his past love, had the second least interesting topic according to boys (interest value 1,62). The romantic topic clearly appealed to girls more, since their ratings placed the text 19th with clearly higher interest value (2,50). The text with the least interesting topic according to boys was the sheet from an art store catalogue with a mail-order form attached. The average interest value of the topic was as low as 1,58. The topic of the text did not appeal very much to girls either although they did rate the topic more positively. The girls' ratings placed the text 28th on their list of PISA texts by topic interest.

Overall, there was much in common between boys and girls in what they found to be topics of interest, but even more common in what they found uninteresting. At the top of the two lists of PISA texts by topic interest were several texts boys and girls assessed differently. The texts at the top of the list were usually assessed more positively by girls than by boys, which show up in a comparison of the average interest values. At the bottom of the list, however, the boys' average ratings were more positive than those of the girls, who used the extreme ends of the scale more boldly. Topics clearly more interesting to girls were space research, etiquette for exchange students, children's attitudes towards housework, applying for a job, Amanda's affair with the prince, and the art store. The most dramatic case of difference in topic interest for boys was revealed by the ratings of the text on conquering the South Pole. The most interesting topic for boys was the NASA tests on drugged spiders.

In terms of reading situation as defined in the PISA framework it seems that the most interesting topics for girls were from educational and public reading situations. Both the argumentative texts about space research as well as the information sheet for exchange students were classified as educational texts,

as was the text with the fifth most typical topic for girls, also about exchange students. For boys, on the other hand, the most interesting topics in the PISA texts were from public reading situations, namely the NASA tests on spiders, allergies and explorers and the fourth most typical topic–the essentials in life in the motorbike advertisement. The quality of men's shirts, from a private context, and DNA in police investigations from an educational context were also among the most interesting topics for boys. There were topics from all four reading situations among the least interesting topics in the PISA texts for girls as well as for boys.

7.3 The typicality and interest of the PISA texts by genre

7.3.1 Overall typicality and interest of genres

In evaluating genre typicality and interest students used the four-point scale even more cautiously than in assessing topic within the same framework. The typicality and interest of the genres of PISA texts were judged to quite low. The average ratings for genre typicality and interest ranged only from 1,81 to 2,85 for genre typicality and from 1,73 to 2,76 for genre interest (Table 16).

TABLE 16   Overall typicality and interest of the genre of the PISA 2000 texts for a group of Finnish 15-year-old students.

TYPICALITY OF GENRE	mean	INTEREST OF GENRE	mean
1. movie reviews*	2,85	1. article about NASA tests on drugged spiders	2,76
2. article about DNA technology in police investigations	2,83	2. movie reviews	2,69
3. news texts about allergies and explorers	2,81	3. article about DNA technology in police investigations	2,68
4. article about the quality of men's shirts	2,75	4. motorbike advertisement	2,67
5. argumentative texts about space research	2,74	5. news texts about allergies and explorers	2,66
5. article about NASA tests on drugged spiders	2,74	6. argumentative texts on space research	2,62
7. narrative about woman and panther trapped in flood	2,67	7. argumentative texts on graffiti	2,60
8. report on children's attitudes toward housework*	2,65	8. article about men's shirts	2,55
8. argumentative texts on graffiti	2,65	9. information sheet for potential exchange students	2,54
8. instructions on preparing for a job interview*	2,65	9. etiquette for exchange students	2,54
8. brochure describing optician's work*	2,65	11. editorial on problems that arise with new technology	2,49
12. information sheet for potential exchange students	2,61	12. narrative about woman and panther trapped in flood	2,48
12. editorial on problems that arise with new fertilization technology	2,61	13. graphs about nuclear waste	2,39
14. job application form*	2,60	13. report on children's attitudes toward housework	2,39
15. etiquette for exchange students	2,55	15. article about choosing running shoes	2,35

16. article about protecting beaches	2,53	16. fable on father's over protectiveness	2,34
17. article about rhinoceros	2,52	17. article about rhinoceros	2,32
18. fable on father's over protectiveness	2,50	18. instructions on preparing for a job interview	2,31
19. motorbike advertisement	2,49	18. article about protecting beaches	2,31
20. graphs about nuclear waste	2,45	20. brochure describing optician's work	2,29
20. information sheet about flu vaccination*	2,45	21. job application form	2,26
22. article about choosing running shoes	2,41	21. text on conquering the South Pole	2,26
23. graphs about Lake Chad's altitude and fauna	2,40	23. graphs about Lake Chad's altitude and fauna	2,24
24. warranty instructions for a refrigerator*	2,39	24. table from a report of an aid agency	2,20
25. instructions for contacting employer*	2,38	24. floor plan of a library	2,20
25. narrative about villagers' first encounter with movies*	2,38	24. information sheet about flu vaccination	2,20
27. information sheet about a company's job centre*	2,35	27. extract from a play	2,19
28. text on conquering the South Pole	2,28	28. narrative about villagers' first encounter with movies	2,12
29. bicycle assembly instructions for the buyer	2,25	28. information sheet about a company's job centre	2,12
30. tree diagram of structure of labour force*	2,21	30. instructions for contacting employer	2,11
31. table from a report of an aid agency	2,20	31. bicycle assembly instructions for the buyer	2,04
32. a sheet from a catalogue and a mail-order form of art store*	2,12	32. tree diagram of structure of labour force	1,98
33. telephone use instructions*	2,09	33. warranty instructions for a refrigerator	1,92
34. floor plan of a library	2,08	34. a sheet from a catalogue and a mail-order form from an art store	1,86
35. extract from a play*	1,96		
36. flight timetable	1,93	35. telephone use instructions	1,78
37. table of news agencies	1,81	35. flight timetable	1,78
		37. table of news agencies	1,73

* Difference between means is statistically significant (p-value< 0,05).

Most PISA texts had a genre more typical than interesting since 30 texts out of 37 were attributed higher average genre typicality value than average genre interest value. For instance, the movie reviews were given a genre typicality value of 2,85 and a genre interest value of 2,69, and the text with the least typical and least interesting genre, the table of news agencies, was given an average typicality value of 1,81 and an average interest value of 1,73. Only three of the 37 texts were of a genre more interesting than typical and four texts were of a genre found equally as typical as interesting. It is to be noted, however, that not all the differences between typicality and interest values were statistically significant and in some cases they were quite small. The fact that the genres of the PISA texts were more typical than interesting was further underlined by the number of texts attributed average genre typicality and interest values higher than 2,50. There were 18 texts found sufficiently typical in genre enough to

come out with an average typicality value above that figure but only 10 texts in a genre considered interesting to the same extent. There were 3 texts where the genre was given an average genre typicality value under 2,00, meaning 'untypical' but 6 cases of an average genre interest value under 2,00, meaning 'uninteresting'.

The PISA texts with both most typical and most interesting genres were texts that within the PISA framework were classified as continuous. Furthermore, they represent all the PISA reading situations except the occupational one. Many were genres originating in newspapers and magazines. Among the six most typical genres were movie reviews, articles, news texts, and argumentative texts expressing opinions, such as are typical of newspapers and magazines. The texts at the top of the list of PISA texts by genre interest were much the same since among the six most interesting genres were five cases where the same texts were among the six most typical genres. The texts both typical and interesting in genre were the movie reviews, the article about using DNA in police investigations, the news about allergies and explorers, the article about NASA tests on drugged spiders, and the argumentative texts expressing students' opinions on space research. The order of the texts varied between the two lists by genre typicality and genre interest. For instance, the text with most typical genre, movie reviews, had the second most interesting genre. On the other hand, the text with the sixth most typical genre, the article about tests with drugged spiders, was in the most interesting genre according to the students participating in the study. Students' comments on the genre of texts were restricted to referring to the type of materials they usually read. For instance the argumentative text unit with young students' opinions was commented on as follows:

> "It is interesting to read other young people's opinions. One can compare them with one's own opinions." (Girl)
> "As I already mentioned I like hearing other people's opinions." (Girl)
> "It is about opinions which is familiar. Space research is a stranger feature." (Girl)
> "One meets opinions like that every day. The topic students are writing about is also familiar." (Girl)

There were two texts that were not both in the top six of the list of PISA texts by genre typicality and genre interest. The fourth most typical genre was found to be that of the article on the quality of men's shirts but it was only deemed to be the eighth most interesting (2,55). The fourth most interesting genre was that of the motorbike advertisement with the poem (2,67). The genre of the advertisement was not found very typical since it was only ranked 19th by genre typicality (2,49).

The students' association of the typicality and interest of certain genres in their assessment was fairly consistent. For instance, articles such as the article about DNA in police investigations, about the quality of men's shirts and about NASA tests on drugged spiders were all considered both typical and interesting. On the other hand, there were seemingly similar articles the genre of which was clearly deemed less typical and less interesting. The articles about

saving beaches, about the rhinoceros, and about choosing running shoes were placed 16th, 17th, and 22nd by genre typicality and 19th, 17th, and 15th by genre interest respectively. Thus there seems to be a slight contradiction in the assessment of the articles representing the same genre. Yet, it is to be noted that seemingly similar articles are not necessarily similar in terms of structure and style. Furthermore the content of the article may influence the assessment of the genre, since it is sometimes difficult to keep the two apart, especially for young students. It seems that the topic interest of the articles deemed less typical and interesting by genre was indeed lower than the topic interest of the articles found typical and interesting by genre (see Table 16 above and Table 13 on topic typicality and interest). With regard to the typicality of the newspaper or magazine articles in PISA, it is interesting that with all six texts of that genre, typicality of topic was deemed lower than typicality of genre. Thus, it is evident that in the PISA newspaper and magazine articles the genre was found more typical than the actual topics (the police using DNA, the quality of men's shirts, tests on drugged spiders, saving the beaches, the rhinoceros, choosing running shoes).

As mentioned above, occupational texts were not among the most typical or interesting PISA texts by genre. The text types represented by texts from occupational reading situations were primarily the informative and the instructive. These text types clearly appealed to 15-year-old students participating in the study less than newspaper and magazine articles, which in PISA are either expository or descriptive by text type.

The three texts the genres of which were more interesting than typical were the motorbike advertisement with the poem, the floor plan and the extract from a play. As mentioned above, the genre of the motorbike advertisement was found 19th most typical among the PISA texts. However, students' comments revealed that not everyone realized that the text was an advertisement but evaluated it only as a poem, as the following quotation illustrates: "*Poems are nice and in my opinion this was too. I have read poems about similar topics before. Harley Davidson??? Was this some kind of advertisement???*" (Girl) Presumably the position of the text on the list of PISA texts by genre typicality reflects the typicality of the poem and not necessarily of an advertisement. Students, however, found the genre of the text to be one of the most interesting ones among the PISA texts (fourth) with the average genre interest value clearly higher than the average genre typicality value. The genre of the floor plan was found to be the fourth least typical (2,08) genre. The genre was, however, found to be somewhat more interesting, ranking 24th (2,20). The play extract proved to have similar differences between genre typicality and interest. It had the third least typical genre (1,96) but was placed 27th in terms of genre interest (2,19).

The two least typical and least interesting genres coincided. The table about news agencies and the flight timetable came last on both lists of PISA texts by genre with average genre typicality values (1,81 and 1,93) slightly higher than the average genre interest values (1,73 and 1,78). The genre of the

telephone use instructions was also deemed untypical (2,09) and especially uninteresting (1,78). In the PISA text classification all three texts are non-continuous documents as are the floor plan, at the bottom of the list by genre typicality, and the page from an art store catalogue, at the bottom of the list of PISA texts by genre interest. Overall, the least typical and interesting genres mainly proved to be non-continuous, and represented usually public but also educational and private reading situations.

In all, the least typical and least interesting PISA texts seem to be the same ones despite the textual feature assessed. The most typical and the most interesting genres likewise seem to coincide and originate in newspapers and magazines.

7.3.2 Gender differences in genre typicality

In assessing genre typicality, girls once again used the scale from 1 to 4 more boldly than boys. In fact, the highest average value in the present study was reached by girls in assessing genre typicality since no other feature of any PISA texts assessed in the study was given as high a value as 3,28 (Table 17). The lowest average genre typicality value given by girls on the other hand was 1,59. Overall, girls found the genres of the PISA texts to be relatively typical. There were 21 texts that were attributed an average genre typicality value over 2,50. Boys, however, did not find the genres of the PISA texts to be as typical as girls did, since the average typical values given by boys varied from 1,85 to 2,92 and only 10 texts were considered sufficiently typical to be given an average typicality value higher than 2,50. On both genders' lists of PISA texts by genre typicality only two texts came out with an average typicality value under 2,00 meaning 'untypical'.

Comparing the average genre typicality values given by girls and boys to same texts underlines the impression that girls found the genre of many PISA texts more typical than boys did. For 24 texts the average genre typicality value given by the girls was higher than that given by the boys. For 11 texts the average genre typicality value given by the boys was higher, and for two texts the value was exactly the same.

The clearest differences between girls and boys in assessing genre typicality were at the top of the lists (Table 17). The first eight texts on the girls' list of PISA texts by genre typicality were all texts which girls found more typical in genre than boys, and in five of the eight texts the difference was substantial (more than 0,50). Only one text, the article about using DNA technology in police work was among the eight most typical PISA texts by genre on both genders' lists. The article was second on the girls' list (genre typicality value 3,00) and sixth on the boys' list (2,65).

TABLE 17 Genre typicality of texts in the PISA 2000 reading literacy test by gender.

Text	Genre typicality			
	Girls		Boys	
	mean	ranking	mean	ranking
1. argumentative texts about space research	3,28	1.	2,20	28.
2. article about using DNA technology in police investigations	3,00	2.	2,65	6.
3. narrative about woman and panther trapped in flood	3,00	2.	2,35	21.
4. job application form	2,93	4.	2,22	27.
5. article about NASA tests on drugged spiders	2,92	5.	2,54	10.
6. information sheet for potential exchange students	2,88	6.	2,30	24.
7. report on children's attitudes toward housework	2,88	6.	2,42	17.
8. article about protecting beaches	2,88	6.	2,13	30.
9. movie reviews	2,85	9.	2,85	2.
10. article about the quality of men's shirts	2,81	10.	2,67	5.
11. etiquette for exchange students	2,72	11.	2,38	19.
12. argumentative texts about graffiti	2,70	12.	2,60	7.
13. article on rhinoceros	2,69	13.	2,33	22.
14. news texts about allergies and explorers	2,68	14.	2,92	1.
15. editorial on problems that arise with new fertilization technology	2,65	15.	2,58	8.
16. motorbike advertisement	2,63	16.	2,33	22.
17. instructions on preparing for a job interview	2,59	17.	2,69	4.
18. fable on father's over-protectiveness	2,58	18.	2,41	18.
19. brochure describing optician's work	2,58	18.	2,72	3.
20. graphs on nuclear waste	2,52	20.	2,38	20.
21. narrative about villagers' first encounter with movies	2,50	21.	2,27	26.
22. information sheet about flu vaccination	2,43	22.	2,46	13.
23. information sheet about a company's job centre	2,42	23.	2,28	25.
24. bicycle assembly instructions for the buyer	2,38	24.	2,12	31.
25. article on choosing running shoes	2,35	25.	2,46	13.
26. graphs about Lake Chad's altitude and fauna	2,32	26.	2,48	11.
27. instructions for contacting employer	2,30	27.	2,44	16.
28. warranty instructions for a refrigerator	2,29	28.	2,48	11.
29. table from a report of an aid agency	2,24	29.	2,16	29.
30. a sheet from a catalogue and a mail-order form from an art store	2,15	30.	2,09	33.
31. telephone use instructions	2,13	31.	2,05	34.
32. extract from a play	2,09	32.	1,85	37.
33. text on conquering the South Pole	2,09	32.	2,46	13.
34. floor plan of a library	2,04	34.	2,12	31.
35. flight timetable	2,00	35.	1,86	36.
36. tree diagram of structure of labour force	1,82	36.	2,56	9.
37. table of news agencies	1,59	37.	2,04	35.

The most typical genre among the PISA texts for the girls was the argumentative text unit in which a group of young students expressed their opinions on space research. This came out with the top average typicality value of 3,28. On the boys' list, it was only 28th with substantially lower average genre

typicality value (2,20). The other text representing the same genre in PISA, the argumentative texts about graffiti, was attributed an average genre typicality value of 2,70 and placed 12th on the girls' list. Among the eight most typical PISA texts by genre for girls were three newspaper or magazine articles, the most typical of which was the article about using DNA technology in police investigations. The article was placed second with a genre typicality value of 3,00. The other two articles among the eight most typical PISA texts by genre were the article about NASA tests on drugged spiders, which ranked fifth and the article about protecting beaches which ranked sixth equal. On the boys' list the article about drugged spiders was tenth (2,54) but the article about saving beaches only 30th (2,13).

The short story about a woman and a panther trapped by a flood was attributed the same average genre typicality value (3,00) by girls as the article about using DNA in police work. Thus the two texts shared second place on the girls' list of PISA texts by genre typicality. According to the girls the story was the most typical literary text by genre in PISA. It seems, however, that although the genre was typical the topic could have been closer to young students' lives since in terms of topic typicality the story ranked only 16th (2,32). The fourth most typical genre in girls' opinion was that of the job application form (2,93). The application form is exceptional text among those at the top of the girls' list of PISA texts by genre typicality since it is non-continuous by text type and classified as an occupational text while the other typical genres according to the girls were of continuous type and from three other reading situations. On the boys' list the job application form ranked only 27th with once again clearly lower genre typicality value (2,22). As the students' comments quoted above showed, they were not necessarily familiar with applying for a job, but the act of filling in different kinds of forms, e.g. in applying for further studies, was familiar. The last three texts in the girls' top eight by genre typicality received the same average genre typicality value (2,88). These texts were the article about saving beaches, the report on children's attitudes towards housework and the information sheet about the qualifications of exchange students.

Of the eight genres found most typical by the boys six usually appear in a newspaper or magazine. The most typical genre in the boys' opinions was that of the short news texts, given an average genre typicality value of 2,92, which was higher than the girls' value (2,68) for the same text. The second most typical genre was also of newspaper origin, namely the genre of the movie review. Movie reviews were rated at exactly the same average genre typicality value by both genders (2,85) although the ranking on the boys' list was higher. The third and the fourth most typical genres were from immediate contexts other than newspaper or magazine. The third most typical genre was that of the brochure describing an optician's work (2,72) and the fourth most typical genre was that of the instructions on preparing for a job interview (2,69). Both were attributed lower average genre typicality value by girls than boys, and were placed only 17th and 18th on the girls' list. Interestingly enough, the two texts, like the job application form on the girls' list, represented the occupational situation for

reading in the PISA classification of reading situations. Texts like these are, however, likely to be encountered at school within guidance counselling.

The rest of the eight texts with the most typical genres for boys had genres more typical to girls than to boys in terms of average genre typicality values although the position of the texts was higher on boys' list. These four texts also had genres originally appearing in newspapers or magazines as did the two most typical genres for boys. The genre found fifth most typical by boys was that of the article on the quality of men's shirts and the sixth most typical was that of the article on using DNA in police investigations. The seventh most typical genre on the boys' list of PISA texts by genre typicality was that of the argumentative text unit consisting of two opinions on graffiti (2,60), which had the 12th most typical genre according to girls. The eighth most typical genre for boys was that of the editorial on problems that arise with new fertilization technology (2,58).

Other texts in genres clearly more typical to boys than to girls were the tree diagram of a country's labour force, the textbook extract about conquering the South Pole, two graphs about Lake Chad's altitude and fauna and surprisingly enough, the warranty instructions for a refrigerator. The tree diagram was the ninth most typical genre for boys but girls rated it as having the second least typical genre. In fact, apart from the job application form, the girls did not have any plain non-continuous texts among their 19 most typical PISA texts by genre although two of texts include essential non-continuous elements. On the boys' list non-continuous documents were placed 9th and 11th since the graphs about Lake Chad were deemed to have the 11th most typical genre. The girls, however, found the graphs to have only the 26th most typical genre (2,32). The warranty instructions for a refrigerator shared 11th position on the boys' list, since it received exactly the same average genre typicality value as graphs about Lake Chad. The warranty instructions were not equally typical for the girls, who ranked the text 28th. The textbook extract on conquering the South Pole was clearly more typical to boys than to girls, the boys ranking it 13th equal in genre typicality, while the girls placed it only 32nd with clearly lower typicality value.

Apart from the textbook extract and the diagram of a country's labour force the lower end of the boys' and the girls' lists of PISA texts by genre typicality seems to be quite similar. The positions of individual texts somewhat vary and there are small differences in average genre typicality values but still the texts in a genre which was found untypical were the same. Apart from those cases already referred to, the least typical genre for girls was that of the table of news agencies, followed by the flight timetable, the floor plan of a library, and the extract from a play. The telephone use instructions and the sheet from an art store catalogue were also among the least typical PISA texts by genre for girls. The least typical genre for boys was that of the play, followed by the flight timetable, the table of news agencies, the telephone use instructions, and the art store catalogue. Thus, all the PISA texts untypical in genre, except for the play extract and the textbook extract on the South Pole, were non-continuous by text

type in the PISA classification. The reading situations represented by the untypical texts by genre were usually public but sometimes also private and educational (see Table 7). With regard to the genre of the PISA texts it is noteworthy that several genres found typical were genres originally appearing in newspapers or magazines, such as articles, reviews, news texts and opinions.

7.3.3 Gender differences in genre interest

In evaluating the genre interest of the PISA texts, boys and girls seemed to have their own favourites (Table 18). The scale from 1 to 4 was again used in a different way by the two genders but the differences were not as clear as in evaluating genre typicality. Overall, the genres of the PISA texts did not give rise to great enthusiasm in terms of interest since the top average genre interest values were quite modest, especially so for the boys. The highest average genre interest value given by the girls was 2,96, while that given by the boys was 2,79. The lowest average genre interest value was 1,68 for the girls and 1,55 for the boys. In contrast with their evaluations of topic interest and typicality and of genre typicality the boys gave lower average values at the bottom of the list of PISA texts by genre interest. In rating the other textual features mentioned the girls were more critical than the boys. It seems that the boys did not find the genres of the PISA texts of any great interest.

The lack of interest of the PISA genres for boys was further underlined by the fact that in their assessment only seven PISA texts were in a genre that was interesting enough to be attributed an average interest value higher than 2,50. On the girls' list by genre interest there were 15 texts with average interest value above 2,50. At the same time, 6 texts were attributed an average interest value lower than 2,00 on the boys' list, while on the girls' list there were only 3 such texts. The greater interest for girls of the genres of the PISA texts is further emphasized by the fact that 26 texts out of 37 were given a higher average genre interest value by the girls than by the boys. Only 6 texts had a higher genre interest value for boys, while 5 texts had the same value for both gender groups.

For girls, the most interesting genre among the PISA texts was clearly the argumentative text expressing opinions. The texts in which young students expressed their opinions on space research were given an average interest value of 2,96 and were placed first on the girls' list of PISA texts by genre interest. The texts expressing students opinions on graffiti were placed second with a slightly lower average interest value (2,89). The rating of the two texts by genre interest was consistent and emphasized girls' interest in expressing and reading about opinions on current issues. The boys too were consistent in rating the two texts representing the same genre, although they did not find the genre to be as interesting as the girls did. On the boys' list of PISA texts by genre interest the two texts shared 13th position with the same average genre interest value of 2,28.

TABLE 18   Genre interest of texts in the PISA 2000 reading literacy test by gender.

| Text | Genre interest | | | |
| | Girls | | Boys | |
	mean	ranking	mean	ranking
1. argumentative texts about space research	2,96	1.	2,28	13.
2. argumentative texts about graffiti	2,89	2.	2,28	13.
3. narrative about woman and panther trapped in flood	2,88	3.	2,11	25.
4. movie reviews	2,85	4.	2,52	5.
5. article about using DNA technology in police investigations	2,84	5.	2,50	6.
6. motorbike advertisement	2,78	6.	2,54	4.
7. article about NASA tests on drugged spiders	2,77	7.	2,75	2.
8. report on children's attitudes toward housework	2,72	8.	2,04	29.
9. etiquette for exchange students	2,72	8.	2,36	9.
10. information sheet for potential exchange students	2,69	10.	2,36	9.
11. editorial on problems that arise with new fertilization technology	2,57	11.	2,42	8.
12. job application form	2,52	12.	1,96	32.
13. article about protecting beaches	2,50	13.	2,09	26.
14. article on rhinoceros	2,50	13.	2,13	23.
15. news texts about allergies and explorers	2,50	13.	2,79	1.
16. instructions on preparing for a job interview	2,48	16.	2,15	21.
17. article about the quality of men's shirts	2,44	17.	2,67	3.
18. fable on father's over-protectiveness	2,42	18.	2,24	16.
19. graphs on nuclear waste	2,42	18.	2,36	9.
20. brochure describing optician's work	2,36	20.	2,23	18.
21. article on choosing running shoes	2,36	20.	2,35	12.
22. table from a report of an aid agency	2,36	20.	2,04	29.
23. extract of a play	2,33	23.	2,08	27.
24. information sheet about flu vaccination	2,27	24.	2,15	21.
25. graphs about Lake Chad's altitude and fauna	2,23	25.	2,25	15.
26. narrative about villagers' first encounter with movies	2,19	26.	2,04	29.
27. floor plan of a library	2,19	26.	2,20	19.
28. information sheet about a company's job centre	2,13	28.	2,12	24.
29. instructions for contacting employer	2,05	29.	2,16	20.
30. bicycle assembly instructions for the buyer	2,00	30.	2,08	27.
31. warranty instructions for a refrigerator	2,00	30.	1,84	33.
32. a sheet from a catalogue and a mail-order form from an art store	2,00	30.	1,71	34.
33. telephone use instructions	2,00	30.	1,55	37.
34. text on conquering the South Pole	2,00	30.	2,50	6.
35. flight timetable	1,92	35.	1,60	36.
36. table of news agencies	1,78	36.	1,68	35.
37. tree diagram of structure of labour force	1,68	37.	2,24	16.

The third most interesting genre for girls was that of the story of a woman trapped in a flood with a panther. The story was attributed an average interest value of 2,88, and it was clearly more interesting than other literary texts in PISA. Other literary narratives were deemed less interesting, since the fable was 18th and the story about the people of Macondo only 26th by genre interest. Boys

found the genre of the story of the woman and panther quite uninteresting since it was only 25th on their list of PISA texts by genre interest. The most interesting literary genre for boys was that of the fable which was placed 16th with an average interest value of 2,24.

The next four texts on the girls' list are texts such as usually appear in newspapers or magazines. The fourth most typical genre for girls was that of the movie reviews. The same text unit was also near the top of the boys' list (fifth) but the average interest value given by boys was clearly lower. The fifth most typical genre was that of the article about using DNA in police investigations. The article was given an average interest value nearly the same as the movie reviews. The genre of the article was of interest also to boys, ranking sixth on their list, but again the average interest value for the genre was clearly lower than that given by the girls. The motorbike advertisement including a poem was in an interesting genre according to the girls, ranking sixth most interesting, but the students' comments quoted above (p. 106) reveal that they assessed the genre as a poem rather than an advertisement. It was of interest also to boys, being in the fourth most interesting genre for them, but with an average interest value once again lower than the value given by the girls.

The seventh most interesting genre for girls was that of the article about NASA tests on drugged spiders. The article was attributed nearly the same average genre interest value by the boys although it was second on their list of PISA texts by genre interest. Overall, there were several cases where the genre of the text was interesting to both genders. Among the seven most interesting genres for boys and girls were four texts on both the girls' and the boys' lists, namely the movie reviews, the motorbike advertisement with the poem, and the articles about tests on drugged spiders and using DNA in police work. The argumentative texts about space research and graffiti, as well as the story about a woman trapped in Mississippi flood were clearly in genres more appealing to girls than to boys. Naturally also boys had their favourite genres that did not appeal to girls. The most interesting genre in their assessment was that of the news texts about allergies and explorers with an average interest value of 2,79. On the girls' list the news texts were 13th. The third most interesting genre after the article on tests on drugged spiders was the article about the quality of men's shirts, which was only 17th on girls' list. The genre ranking sixth equal on the boys' list of PISA texts by genre was clearly more appealing to boys than girls. This textbook extract about conquering the South Pole was attributed an average interest value of 2,50 while the corresponding rating by the girls was only 2,00 and the text shared 30th position on their list of PISA texts by genre interest.

Below the top seven there still were some texts that girls and boys rated differently enough to make them clearly more appealing to one gender than the another. The text reporting young people's attitudes towards housework was expectedly of greater interest to the girls. The report shared 8th place on the girls' list with an average genre interest value of 2,72 which was substantially

higher than the value given by boys (2,04). On the boys' list the report came only 29th. Even more clearly a girls' text was the job application form, 12th on the girls' list (2,52) but only 32nd on the boys' list with an average genre interest value under 2,00. To some extent also the article about saving beaches was a girls' text although boys found other articles quite interesting. The article shared 13th place on the girls' list (2,50) but came only 26th on the boys' list (2,09).

Apart from the news texts, which were in the most interesting genre for the boys and the article on men's shirts and the text on conquering the South Pole there was still one text the genre of which was far more interesting to boys than to girls. This was the tree diagram of a country's labour force, placed 16th in the boys' ratings of the PISA texts by genre interest (2,24) while the girls rated it to be in the least interesting genre of all the PISA texts with an average genre interest value as low as 1,68. Otherwise the least interesting texts by genre were the same as the texts at the bottom of the lists by other textual features studied. Texts in genres of which appealed neither to girls nor boys were the table of news agencies, the flight timetable, the telephone use instructions, and the catalogue sheet of an art store. All of the texts deemed uninteresting by genre were non-continuous in the PISA text classification and represented public and educational contexts. Interestingly enough, the warranty instructions for a refrigerator represented one of the least interesting genres for the boys, placed 33rd with an average genre interest value of 1,84, although it was deemed to represent quite a typical genre in the boys' experience, ranking 11th with an average genre typicality value of 2,48 (see Table 17 for gender differences on genre typicality).

Both on the girls' and the boys' lists the texts deemed interesting by genre were continuous. The most interesting genres for both genders were genres originally appearing in newspapers and magazines, as was the case with the most typical genres as well. The genres most favoured by girls were those of texts expressing opinions, also found interesting by girls as were articles and a narrative. Boys found the most interesting texts by genres to be news texts, articles, movie reviews and the motorbike advertisement.

7.4   The typicality and interest of the PISA texts by layout

7.4.1 Overall typicality and interest of layout

The layout of the PISA texts was relatively authentic in terms of typicality. Students participating in the study found the layout of the texts on average more typical than interesting. The average typicality value for the layout of texts ranged from 1,89 to 3,00, and the average interest value from 1,87 to 2,78 (Table 19). The higher typicality of layout than interest was further underlined by the fact that 25 texts were attributed higher average typicality value than

interest value for the layout. For instance, in the short news texts about allergies and explorers the average typicality value for the layout was 3,00, while the average interest value was only 2,50. Moreover, 17 texts were given average typicality value higher than 2,50 but only 10 texts were given an average interest value above the same value.

TABLE 19 Typicality and interest of the layout of the PISA 2000 texts for a group of Finnish 15-year-old students.

TYPICALITY OF LAYOUT	mean	INTEREST OF LAYOUT	Mean
1. news texts about allergies and explorers*	3,00	1. motorbike advertisement	2,78
2. article about NASA tests on drugged spiders	2,82	2. article about DNA technology in police investigations	2,71
3. movie reviews*	2,80	3. floor plan of a library	2,63
4. fable on father's over-protectiveness*	2,73	4. argumentative texts on space research	2,59
5. editorial on problems that arise with new technology*	2,72	5. article about NASA tests on drugged spiders	2,58
6. job application form*	2,69	5. etiquette for exchange students	2,58
6. narrative about woman and panther trapped in flood*	2,69	7. graphs about nuclear waste	2,56
8. argumentative texts about space research	2,67	8. report on children's attitudes toward housework	2,53
8. article about the quality of men's shirts	2,67	9. article about men's shirts	2,51
10. information sheet for potential exchange students*	2,63	10. news texts about allergies and explorers	2,50
11. report on children's attitudes toward housework	2,61	11. argumentative texts on graffiti	2,45
12. argumentative texts on graffiti	2,60	12. article about choosing running shoes	2,43
13. instructions on preparing for a job interview	2,59	12. graphs about Lake Chad's altitude and fauna	2,43
14. article about DNA technology in police investigations	2,58	14. job application form	2,42
15. article about protecting beaches*	2,57	15. movie reviews	2,39
16. article about rhinoceros*	2,55	16. instructions on preparing for a job interview	2,37
17. motorbike advertisement*	2,51	17. fable on father's over-protectiveness	2,36
18. narrative about villagers' first encounter with movies*	2,46	17. brochure describing optician's work	2,36
19. etiquette for exchange students	2,43	19. information sheet for potential exchange students	2,33
20. brochure describing optician's work	2,41	20. narrative about woman and panther trapped in flood	2,31
20. information sheet about a company's job centre*	2,41	20. information sheet about flu vaccination	2,31
22. graphs about nuclear waste	2,38	22. article about rhinoceros	2,24
23. article about choosing running shoes	2,35	22. table from a report of an aid agency	2,24
23. graphs about Lake Chad's altitude and fauna	2,35	24. text on conquering the South Pole	2,21
25. instructions for contacting employer	2,33	25. editorial on problems that arise with new technology	2,20

26. telephone use instructions*	2,22	26. warranty instructions for a refrigerator	2,18
27. warranty instructions for a refrigerator	2,20	27. extract of a play	2,15
27. table from a report of an aid agency	2,20	28. information sheet about a company's job centre	2,14
29. a sheet from a catalogue and a mail-order form from an art store	2,16	29. bicycle assembly instructions for the buyer	2,10
30. information sheet about flu vaccination	2,15	30. instructions for contacting employer	2,09
30. text on conquering the South Pole	2,15	31. article about protecting beaches	2,08
30. tree diagram of structure of labour force*	2,15	32. narrative about villagers' first encounter with movies	2,02
33. floor plan of a library*	2,10	33. a sheet from a catalogue and a mail-order form from an art store	1,96
34. bicycle assembly instructions for the buyer	2,06	34. tree diagram of structure of labour force	1,94
35. extract from a play	1,98	35. table of news agencies	1,88
36. table of news agencies	1,94	36. telephone use instructions	1,87
37. flight timetable	1,89	36. flight timetable	1,87

* Difference between means is statistically significant (p-value< 0,05).

The most typical layout was that of the news texts about allergies and explorers on Antarctica with average typicality value of 3,00. The second most typical was that of the article on NASA tests on drugged spiders (2,82), and the third most typical, with nearly the same average typicality value, was that of the movie reviews. The fourth most typical layout was that of the fable (2,73) and the fifth most typical, with nearly the same average typicality value was the editorial on ethical problems arising from the new fertilization technology. All these texts consisted of written text alone. The sixth most typical layout was that of the job application form, which was the only non-continuous text attributed an average typicality value above 2,50 for its layout.

The least typical layouts among the PISA texts proved to be those of the flight timetable and the table of news agencies, both non-continuous texts. The third least typical layout, also with an average typicality value below 2,00, meaning 'untypical', was that of the play extract. The bicycle assembly instructions and the floor plan of a library were also deemed to have quite untypical layouts.

Among the five most interesting layouts only one was among the five most typical ones, namely the article about NASA tests on drugged spiders. This was placed second by typicality of layout but fifth by the interest of the layout. The average interest value (2,58) attributed to the layout was, however, lower than that for typicality of layout (2,82). The most interesting layout was considered to be that of the motorbike advertisement consisting of a poem and a full-page photograph of an old man. The second most interesting was that of the article about the use of DNA technology in criminal investigations by the police. The article was in multiple columns with a box giving background information and a drawing. Surprisingly enough, the floor plan of a library, deemed to have the fifth least typical layout, was rated as having the third most interesting layout. The fourth most interesting layout was that of the argumentative texts about space research. Another text with the same average

interest value for the layout was that concerning etiquette for exchange students. Among the six most interesting layouts there was only one non-continuous text, the floor plan of a library, all the others being continuous texts by text type as defined in the PISA framework. However, the seventh most interesting layout was found to be that of the graphs about nuclear waste, i.e. a non-continuous document.

The least interesting layouts among the PISA texts were those of several non-continuous texts. The layouts given the lowest average interest value were that of the flight timetable and the telephone use instructions organized as a table. The table of news agencies was deemed to have a layout nearly equally uninteresting. The tree diagram of the structure of a country's labour force was placed fourth last on the list of PISA texts by interest of layout.

In addition to the floor plan of a library with the third most interesting but fifth least typical layout, there were 11 texts with a more interesting than typical layout. With most texts the difference between average typicality value and interest value was quite small. For some texts the difference was clearer. Apart from the floor plan of the library, such texts were e.g. the article about using DNA technology in police investigations and the motorbike advertisement. Four of the 12 texts with a layout more interesting than typical were non-continuous. The floor plan of the library and the table from the report of an aid agency were among these texts as well as the graphs about nuclear waste and about Lake Chad's altitude and fauna.

To summarize, the texts rated the least typical and the least interesting seemed to be the same whether in terms of layout, topic or genre. Most of these texts deemed inauthentic in terms of typicality and interest were non-continuous. However, some non-continuous documents were considered on average relatively interesting for their layout although not very typical. Such documents were the floor plan of the library, the table from the report of an aid agency and the graphs of nuclear waste and Lake Chad's altitude and fauna. However, the differences between average typicality and interest values were quite small.

It should be noted that in terms of layout the PISA texts were only moderately authentic since the top average value given for layout typicality was 3,00, meaning 'typical' and the top average value for layout interest only 2,78 on the scale used in the study.

7.4.2 Gender differences in layout typicality

Girls and boys participating in the study rated the typicality of layout of individual PISA texts quite differently although the highest and lowest average typicality values were nearly the same. The lowest average typicality value given by the girls was 1,77 and by the boys 1,81. The highest value for both genders was the same (3,00). Girls, however, on average rated the typicality of the PISA texts' layout more positively than boys since the layout of 24 texts out of the 37 was attributed a higher average typicality value given by the girls than by the boys. The layout of only 12 texts was rated more positively by the boys

than by the girls. The girls' average ratings resulted in 20 texts emerging with an average typicality value for layout above 2,50, while the boys' ratings resulted in only 11 texts emerging with an average typicality value above that figure.

TABLE 20   Typicality of layout of texts in the PISA 2000 reading literacy test by gender.

| Text | Typicality of layout | | | |
| | Girls | | Boys | |
	mean	ranking	mean	ranking
1. argumentative texts about space research	3,00	1.	2,35	18.
2. news texts about allergies and explorers	3,00	1.	3,00	1.
3. narrative about woman and panther trapped in flood	3,00	1.	2,38	16.
4. article about NASA tests on drugged spiders	2,96	4.	2,67	5.
5. job application form	2,85	5.	2,50	10.
6. report on children's attitudes toward housework	2,84	6.	2,38	16.
7. information sheet for potential exchange students	2,81	7.	2,43	14.
8. article about the quality of men's shirts	2,81	7.	2,50	10.
9. article about protecting beaches	2,81	7.	2,30	25.
10. fable on father's over-protectiveness	2,77	10.	2,68	4.
11. motorbike advertisement	2,70	11.	2,29	26.
12. article about using DNA technology in police investigations	2,68	12.	2,48	12.
13. article on rhinoceros	2,68	12.	2,42	15.
14. argumentative texts about graffiti	2,67	14.	2,54	9.
15. movie reviews	2,63	15.	2,96	2.
16. editorial on problems that arise with new fertilization technology	2,59	16.	2,83	3.
17. narrative about villagers' first encounter with movies	2,58	17.	2,35	18.
18. instructions on preparing for a job interview	2,57	18.	2,62	7.
19. etiquette for exchange students	2,56	19.	2,31	22.
20. information sheet about a company's job centre	2,54	20.	2,28	27.
21. graphs on nuclear waste	2,44	21.	2,31	22.
22. article on choosing running shoes	2,39	22.	2,31	22.
23. instructions for contacting employer	2,38	23.	2,28	27.
24. a sheet from a catalogue and a mail-order form from an art store	2,27	24.	2,04	33.
25. bicycle assembly instructions for the buyer	2,27	24.	1,85	36.
26. information sheet about flu vaccination	2,23	26.	2,08	32.
27. telephone use instructions	2,21	27.	2,23	30.
28. warranty instructions for a refrigerator	2,21	27.	2,20	31.
29. extract from a play	2,18	29.	1,81	37.
30. brochure describing optician's work	2,18	29.	2,65	6.
31. graphs about Lake Chad's altitude and fauna	2,09	31.	2,58	8.
32. table from a report of an aid agency	2,08	32.	2,32	21.
33. a floor plan of a library	1,96	33.	2,24	29.
34. text on conquering the South Pole	1,96	33.	2,33	20.
35. table of news agencies	1,89	35.	2,00	34.
36. flight timetable	1,85	36.	1,95	35.
37. tree diagram of structure of labour force	1,77	37.	2,48	12.

The girls participating in the study gave three texts the same top average typicality value of 3,00 for the layout. These were the argumentative texts about space research, the short news items about allergies and explorers, and the story about the woman and the panther trapped in a Mississippi flood. The fourth most typical layout in the girls' ranking was that of the article on NASA tests on drugged spiders, given an average typicality value of 2,96. The four most typical layouts were examples of continuous texts with nothing but written text, but the fifth most typical layout for girls was that of the job application form, which was the only non-continuous text among the 20 texts with an average typicality value above 2,50 in the girls' ratings.

The least typical layouts in the girls' ratings were mostly those of non-continuous documents. In fact, only two of the eight least typical layouts were in the for of continuous texts, namely those of the brochure describing the work of an optician placed 30th and the text on conquering the South Pole, placed 34th. The least typical layouts were those of the tree diagram of the structure of a labour force, the flight timetable, the table of news agencies, and the floor plan of the library, all given an average typicality value under 2,00, meaning 'untypical'.

The top average typicality value given by the boys was the same as that given by the girls (3,00), and it was for the same text, namely the news article about allergies and explorers. The second most typical layout in the boys' ratings, however, was that of the movie reviews, which were placed only 15th on the girls' list of PISA texts by layout typicality. The third most typical layout in the boys' opinion was that of the editorial on the ethical problems arising from new fertilization technology. The fourth most typical layout was that of the fable about a father's over-protectiveness. The fifth most typical layout for boys was that of the article about NASA tests on drugged spiders, which was placed fourth by the girls. Overall, the layouts boys found most typical were in the form of continuous texts containing written text alone. The ratings given by boys and girls on the typicality of layout were somewhat different since among the five most typical PISA texts by layout only two texts were the same, one of which was attributed a much higher typicality value by the girls. The second and third most typical layouts for the boys, the movie reviews and the editorial, on the other hand, were deemed clearly less typical by girls than boys.

There were some differences between the two genders in the layouts found least typical. The least typical for boys was that of the play extract, followed by the bicycle assembly instructions for the buyer with a picture of a bicycle and a table. This text was deemed more typical by girls. The third and fourth least typical layouts for the boys, however, were also found among the texts deemed untypical in terms of layout by the girls, namely the flight timetable and the table of news agencies. The fifth least typical layout for the boys was that of the art store catalogue and the accompanying mail order form.

The text deemed to have the least typical layout by the girls, namely the tree diagram of a country's labour force, was attributed a markedly higher

average typicality value by the boys (2,48) than by the girls (1,77). The diagram was placed 12th on the boys' list of PISA texts by layout typicality.

To summarize, there were many differences between the two genders in rating the typicality of layout of the PISA texts in the case of both the texts deemed most typical and least typical. The clearest differences emerged over non-continuous texts placed at the bottom of the list of PISA texts by layout typicality. On the other hand, there were also texts in a layout rated in a similar way by boys and girls. The news texts about allergies and explorers as well as the article on NASA tests on drugged spiders were found typical in their layout by both genders. Another example was the article on using DNA in police investigations, ranked 12th on both genders' lists of PISA texts by layout typicality. Among the least typical PISA texts by layout were at least two texts similarly ranked by both boys and girls, namely the telephone use instructions and the flight timetable. It is to be noted that the average typicality values given for layout were overall quite modest in the case of both gender groups, and gender differences in the rating of individual texts were usually quite small.

7.4.3 Gender differences in layout interest

The interest of the layout of the PISA texts was in general found to be lower than the typicality of layout. Both the highest and lowest average interest values were lower than the corresponding typicality values for both gender groups. This was also the case with the highest typicality and interest values when students' ratings were studied together instead of by gender (see Table 19). Girls on average were more positive than boys in rating the interest of layout (Table 21) at least in their use of the scale from 1 to 4. The highest average interest value attributed by girls was 2,96, which was quite near their highest typicality value, whereas the highest average interest value given by boys was only 2,75. Thus boys especially seemed to be more critical of the interest than the typicality of layout. At the bottom of the list of PISA texts by layout interest, however, boys and girls seemed to be almost equally critical, since the difference between their lowest average interest values was only 0,05.

The more positive assessment by girls of layout interest was further underlined by the fact that 14 texts were attributed an average interest value above 2,50 in the girls' ratings while in the boys' ratings this was the case for only 6 texts (see the Table 21 below). Moreover, the layout of 24 individual texts was rated more positively by girls than by boys and only 12 cases in which the layout was more positively rated by boys than by girls, which also applies to their ratings of typicality of layout.

TABLE 21   Interest of layout of texts in the PISA 2000 reading literacy test by gender.

| Text | Interest of layout | | | |
| | Girls | | Boys | |
	mean	ranking	mean	ranking
1. argumentative texts about space research	2,96	1.	2,23	19.
2. article about using DNA technology in police investigations	2,88	2.	2,52	5.
3. report on children's attitudes toward housework	2,84	3.	2,21	20.
4. motorbike advertisement	2,81	4.	2,75	1.
5. etiquette for exchange students	2,72	5.	2,44	8.
6. argumentative texts about graffiti	2,70	6.	2,19	21.
7. job application form	2,63	7.	2,17	22.
8. news texts about allergies and explorers	2,60	8.	2,42	10.
9. article about NASA tests on drugged spiders	2,58	9.	2,58	3.
10. information sheet for potential exchange students	2,58	9.	2,04	28.
11. article on choosing running shoes	2,57	11.	2,31	13.
12. article about the quality of men's shirts	2,56	12.	2,46	7.
13. narrative about woman and panther trapped in flood	2,52	13.	2,11	25.
14. graphs on nuclear waste	2,50	14.	2,62	2.
15. fable on father's over-protectiveness	2,46	15.	2,24	16.
16. warranty instructions for a refrigerator	2,42	16.	1,96	30.
17. table from a report of an aid agency	2,40	17.	2,08	26.
18. article on rhinoceros	2,38	18.	2,08	26.
19. extract from a play	2,36	19.	1,96	30.
20. graphs about Lake Chad's altitude and fauna	2,36	19.	2,50	6.
21. instructions on preparing for a job interview	2,35	21.	2,38	12.
22. movie reviews	2,33	22.	2,44	8.
23. information sheet about flu vaccination	2,32	23.	2,30	14.
24. article about protecting beaches	2,31	24.	1,83	34.
25. brochure describing optician's work	2,30	25.	2,42	10.
26. information sheet about a company's job centre	2,29	26.	2,00	29.
27. narrative about villagers' first encounter with movies	2,15	27.	1,88	33.
28. editorial on problems that arise with new technology	2,14	28.	2,25	15.
29. a sheet from a catalogue and a mail-order form from an art store	2,12	29.	1,79	35.
30. telephone use instructions	2,08	30.	1,64	37.
31. instructions for contacting employer	2,05	31.	2,12	24.
32. bicycle assembly instructions for the buyer	2,04	32.	2,16	23.
33. flight timetable	2,00	33.	1,70	36.
34. a floor plan of a library	1,96	34.	2,24	16.
35. table of news agencies	1,85	35.	1,92	32.
36. text on conquering the South Pole	1,83	36.	2,58	3.
37. tree diagram of structure of labour force	1,59	37.	2,24	16.

The layout found the most interesting among the PISA texts by the girls participating in the study was that of the argumentative texts on space research, which they also found the most typical layout. The second most interesting was that of the article about using DNA technology in police investigations, which consisted of multiple columns, an information box and a drawing, while the third most interesting was that of the report on children's attitudes towards

housework, which included graphs. The fourth most interesting layout for girls was that of the motorbike advertisement with a poem and an old black-and-white photograph of an old man. The fifth most interesting layout for girls was that of the guidelines on etiquette for exchange students.

The least interesting layout among the PISA texts for the girls participating in the study was that of the tree diagram of the structure of a country's labour force, given an average interest value of 1,59. The diagram also had the least typical layout in the girls' assessment. The second least interesting layout was that of the text on conquering the South Pole, consisting of a map and a diagram as well as written text, and the third least interesting layout with nearly the same average interest value was that of the table of news agencies. The fourth least interesting layout for girls was that of the floor plan of a library and the fifth least interesting was that of the flight timetable. All five texts with the least interesting layouts for girls were also the PISA texts they deemed to be the least typical in layout, although the average values and thus the ranking of the individual texts may differ. Among the five texts only the flight timetable was given an average interest value for its layout (2,00) higher than the average typicality value (1,85). Four of the five texts were non-continuous and even the text on conquering the South Pole, though categorised as continuous, included non-continuous elements.

The most interesting layout for boys was that of the motorbike advertisement, which was also among the most interesting layouts for girls. Boys found the layout of the advertisement to be markedly more interesting than typical. The second most interesting layout for boys was that of the graph on nuclear waste, a text the girls placed only 14th on their list of PISA texts by layout interest. In equal third position came the article on NASA tests on drugged spiders and the text on conquering the South Pole. The layout of the latter text was clearly more authentic to boys than to girls since the girls rated the layout of the text to be second least interesting and fourth least typical (see Table 20). The fifth most interesting layout for boys was that of the article about using DNA technology in police investigations.

In the boys' rankings, but not the girls', one non-continuous text appeared among the five most interesting texts by layout, namely the graph on nuclear waste. The graph is printed in black and white with symbols such as the flags of different countries and waste barrels, and there is very little written text in it. This was drawn from a newspaper as were also the motorbike advertisement, the article about NASA tests on drugged spiders and the article about the use of DNA technology by the police. Nearly all of these texts included elements other than just plain written text. The article about NASA tests that included only written text looked like a genuine newspaper article with two columns an and appropriate font type.

The least interesting PISA texts by layout for boys were quite different than those for girls. The least interesting layout for boys was that of the telephone use instructions. The second least interesting layout was that of the flight timetable. The third least interesting layout for boys was that of the sheet

from an art store catalogue, and the fourth least interesting layout was that of the article about protecting beaches. The girls' ratings placed the same text 24th by layout interest. The fifth least interesting layout was that of the story about the people of Macondo and their first encounter with the movies, which consisted of plain written text. Girls again rated the text as having a more interesting layout.

Girls and boys had quite different views on which layouts were and were not interesting, since only two texts ranked among those with the most interesting layouts (the article about using DNA technology in police work and the motorbike advertisement) and one (the flight timetable) among those with the least interesting layouts as assessed by the two genders. Individual texts were often rated more positively by girls than by boys. However, both the text on conquering the South Pole and the tree diagram of a country's labour force were among the least interesting PISA texts for girls but rated clearly more positively by boys.

Overall, there were many differences in how the two genders rated the layout interest of the PISA texts. In most cases the differences were, however, quite small and insignificant and must therefore be dealt with cautiously. It is also noteworthy that the authenticity of the PISA texts in terms of layout interest was rather low.

7.5   The typicality and interest of language of the PISA texts

7.5.1 Overall typicality and interest of language

The average typicality and interest values obtained revealed that the language used in the PISA texts was only moderately authentic, since the highest average typicality value was 3,00 (Table 22) and the highest average interest value was 2,74. That the highest value for language typicality was greater than that for language interest suggests that the language used in the PISA texts was in general more typical than interesting to the students participating in the study, as was the case for both the genre and layout of texts. Comparing the average ratings of individual texts shows that the language of 34 texts was found to be more typical than interesting. Moreover, in the three cases where the language was found more interesting than typical there were very small differences in average typicality and interest values. Overall, there were 22 cases where the differences between the typicality and interest values were statistically significant. With other textual features there were distinctly fewer differences with statistical significance. The fact that the language used in the PISA texts was found more typical than interesting was further underlined by the number of texts emerging with an average value above 2,50. In terms of language typicality there were 21 such texts but only 10 in terms of language interest.

TABLE 22   Typicality and interest of the language in the PISA 2000 texts for a group of Finnish 15-year-old students.

TYPICALITY OF LANGUAGE	mean	INTEREST OF LANGUAGE	mean
1. news texts about allergies and explorers*	3,00	1. article about NASA tests on drugged spiders	2,74
2. article about NASA tests on drugged spiders*	2,94	2. motorbike advertisement	2,71
3. job application form*	2,86	3. etiquette for exchange students	2,70
4. movie reviews*	2,81	4. argumentative texts on space research	2,65
5. instructions on preparing for a job interview*	2,80	5. news texts about allergies and explorers	2,61
6. information sheet for potential exchange students*	2,78	6. article about DNA technology in police investigations	2,56
7. etiquette for exchange students	2,76	7. article about men's shirts	2,55
8. article about rhinoceros*	2,74	8. argumentative texts on graffiti	2,52
9. fable on father's over protectiveness*	2,71	9. information sheet for potential exchange students	2,51
9. argumentative texts about space research	2,71	10. movie reviews	2,50
9. report on children's attitudes toward housework*	2,71	11. narrative about woman and panther trapped in flood	2,48
12. argumentative texts on graffiti	2,70	12. article about rhinoceros	2,44
13. editorial on problems that arise with new technology*	2,67	13. article about choosing running shoes	2,42
13. article about the quality of men's shirts	2,67	14. report on children's attitudes toward housework	2,41
15. article about choosing running shoes	2,65	15. fable on father's over protectiveness	2,40
16. brochure describing optician's work*	2,64	16. instructions on preparing for a job interview	2,39
17. article about DNA technology in police investigations	2,62	17. article about protecting beaches	2,37
18. information sheet about flu vaccination*	2,60	18. job application form	2,30
19. article about protecting beaches	2,57	19. editorial on problems that arise with new technology	2,28
20. narrative about woman and panther trapped in flood	2,55	20. graphs about nuclear waste	2,27
21. motorbike advertisement*	2,53	21. graphs about Lake Chad's altitude and fauna	2,26
22. floor plan of a library*	2,49	22. instructions for contacting employer	2,24
23. information sheet about a company's job centre*	2,48	23. extract from a play	2,23
24. instructions for contacting employer*	2,46	23. brochure describing optician's work	2,23
25. warranty instructions for a refrigerator*	2,45	25. floor plan of a library	2,22
26. bicycle assembly instructions for the buyer*	2,44	26. information sheet about flu vaccination	2,20
27. graphs about nuclear waste	2,43	26. text on conquering the South Pole	2,20
27. graphs about Lake Chad's altitude and fauna	2,43	28. table from a report of an aid agency	2,16
29. narrative about villagers' first encounter with movies*	2,38	28. narrative about villagers' first encounter with movies	2,16
30. table from a report of an aid agency	2,32	30. information sheet about a company's job centre	2,02

31. text on conquering the South Pole	2,30	31. bicycle assembly instructions for the buyer	1,98
32. telephone use instructions*	2,26	32. telephone use instructions	1,91
33. a sheet from a catalogue and a mail-order form from an art store*	2,20	33. tree diagram of structure of labour force	1,89
34. tree diagram of structure of labour force*	2,15	34. warranty instructions for a refrigerator	1,86
35. extract from a play	2,02	35. a sheet from a catalogue and a mail-order form from an art store	1,78
36. table of news agencies*	1,96		
37. flight timetable	1,91	36. flight timetable	1,77
		37. table of news agencies	1,71

* Difference between means is statistically significant (p-value< 0,05).

The PISA texts attributed the most typical language were the news texts about allergies and explorers, which were also found quite interesting in terms of language, coming fifth on the list of PISA texts by language interest. The text found second most typical in its language was the article about a NASA test on drugged spiders, which was also the text found to have the most interesting language. In third, fourth and fifth places for most typical language were the job application form, the movie reviews, and the instructions on preparing for a job interview.

The PISA text assessed as having the least typical language was the flight timetable, and immediately above it came the table of news agencies. The play extract, which came just above that, was attributed an average interest value of 2,02 for its language and was one of the few texts to be assessed as having language more interesting than typical. The average language interest value for the text was 2,23, which placed it 23rd on the list of PISA texts by language interest. The fourth least typical language was that of the tree diagram of the structure of a country's labour force. The fifth least typical language was that of the art store catalogue. Apart from the play extract, the texts found least typical in their language were also found to be the least interesting in terms of language. The flight timetable, the table of news agencies, the tree diagram and the art store catalogue all came out with average language interest values below 2,00 and were deemed less interesting than typical in terms of language. These four texts were also all non-continuous and either from educational or public reading situations.

The PISA texts with the most interesting language were in many cases the same as the ones with most typical language. The most interesting language proved to be that of the article on NASA tests on drugged spiders. Second came the motorbike advertisement, followed by the guidelines on etiquette for exchange students placed third on the list of PISA texts by language interest. Fourth came the argumentative texts about space research and fifth the news texts on allergies and explorers (2,61), which was found to be the most typical in terms of language of all the PISA texts by the students participating in the study.

In terms of text format and situation for reading, as defined in the PISA framework, the texts with most interesting language were continuous and represented either public or educational reading situations. Amongst the texts

with the most typical language, however, was one non-continuous text, namely the job application form. The application form was one example of an occupational reading situation being represented among the texts found most typical for their language, while another was the instructions on preparing for a job interview, placed fifth. The other texts among those found to have the most typical language were continuous and represented public or private reading situations (the movie reviews).

To summarize, the authenticity of the language used in the PISA texts appears to be low both in terms of typicality and interest. The typicality of the language used, however, exceeded its interest in nearly every cases, often by a margin that was statistically significant. Two texts–the news texts and article about NASA tests – were found to have both typical and interesting language. Otherwise, the texts with most typical language and those with the most interesting language were different. The texts rated as having the least typical and the least interesting language, however, were the same non-continuous texts as were similarly rated in terms of genre, topic and layout.

7.5.2 Gender differences in language typicality

Although the overall typicality and interest of the language of the PISA texts was quite low, girls and boys rated the typicality of language differently to the extent that their top average typicality values exceeded those of the overall values for typicality and interest. As in terms of topic, genre, and layout typicality, there were again clear differences in which PISA texts were deemed most typical in terms of language by the two genders (Table 23). However, on this occasion the boys' rating of the text they found to have the most typical language was more positive than that of the girls, since the top average language typicality value given by the boys was 3,17 while that of the girls was 3,08. In general, girls were more positive in their ratings, since for 22 individual texts their average ratings were higher than those given by the boys. Moreover 25 texts were attributed an average typicality value of 2,50 or above by the girls while on the boys' list of PISA texts by language typicality there were 17 such texts. At the bottom of the list, boys were again more positive than girls since the lowest average language typicality value given by the boys was 1,96 and by the girls 1,78. Moreover, the boys only attributed one text an average language typicality value under 2,00 while the girls did so in three cases. Overall, in terms of the average language typicality values given to individual texts, girls were more positive in their ratings than boys, but at the extreme ends of the lists of PISA texts by language typicality the boys' ratings resulted in higher average values.

TABLE 23   Typicality of language of texts in the PISA 2000 reading literacy test by gender.

Text	Typicality of language			
	Girls		Boys	
	mean	ranking	mean	ranking
1. article about NASA tests on drugged spiders	3,08	1.	2,79	4.
2. job application form	3,04	2.	2,67	7.
3. etiquette for exchange students	2,96	3.	2,58	13.
4. argumentative texts about space research	2,96	3.	2,46	20.
5. report on children's attitudes toward house hold work	2,96	3.	2,46	20.
6. article on rhinoceros	2,92	6.	2,54	14.
7. information sheet for potential exchange students	2,92	6.	2,61	12.
8. movie reviews	2,85	8.	2,78	5.
9. argumentative texts about graffiti	2,85	8.	2,54	14.
10. article about the quality of men's shirts	2,81	10.	2,50	16.
11. article about using DNA technology in police investigations	2,80	11.	2,43	23.
12. news texts about allergies and explorers	2,80	11.	3,17	1.
13. article about protecting beaches	2,73	13.	2,39	25.
14. motorbike advertisement	2,70	14.	2,33	29.
15. fable on father's over-protectiveness	2,69	15.	2,73	6.
16. bicycle assembly instructions for the buyer	2,69	15.	2,19	34.
17. brochure describing optician's work	2,67	17.	2,62	11.
18. instructions on preparing for a job interview	2,65	18.	2,92	2.
19. article on choosing running shoes	2,65	18.	2,65	9.
20. warranty instructions for a refrigerator	2,63	20.	2,28	31.
21. information sheet about a company's job centre	2,61	21.	2,36	28.
22. narrative about woman and panther trapped in Mississippi flood	2,60	22.	2,50	16.
23. floor plan of a library	2,58	23.	2,40	24.
24. information sheet about flu vaccination	2,55	24.	2,64	10.
25. editorial on problems that arise with new technology	2,50	25.	2,83	3.
26. graphs on nuclear waste	2,48	26.	2,38	26.
27. instructions for contacting employer	2,43	27.	2,48	19.
28. narrative about villagers' first encounter with movies	2,38	28.	2,38	26.
29. table from a report of an aid agency	2,32	29.	2,32	30.
30. telephone use instructions	2,29	30.	2,23	32.
31. a sheet from a catalogue and a mail-order form from an art store	2,20	31.	2,21	33.
32. graphs about Lake Chad's altitude and fauna	2,18	32.	2,67	7.
33. text on conquering the South Pole	2,09	33.	2,50	16.
34. extract from a play	2,09	33.	1,96	37.
35. flight timetable	1,84	35.	2,00	36.
36. tree diagram of structure of labour force	1,82	36.	2,44	22.
37. table of news agencies	1,78	37.	2,16	35.

The text with the most typical language for the girls was the article about NASA tests on drugged spiders. The second most typical language was that of the job application form. The third highest average language typicality value for girls was reached by three texts, namely the guidelines on etiquette for exchange

students, the argumentative texts about space research and the report on children's attitudes towards housework. The fourth highest average language typicality value was also reached by more than one text, namely the article on the rhinoceros and the information sheet for potential exchange students. All of the texts mentioned above included language that was clearly more typical for girls than for boys.

The least typical language as rated by the girls was that of the table of news agencies, the tree diagram of the structure of a labour force and the flight timetable. All three texts were non-continuous and appeared at the bottom of the other lists of PISA texts by typicality and interest, especially in the case of the girls. The play extract and the text on conquering the South Pole were also placed at the end of the list of PISA texts by language typicality. Apart from the play extract, which in the boys' assessment included the least typical language of all the PISA texts, all the texts mentioned above were rated more positively by boys than by girls.

The most authentic PISA texts for boys in terms of language typicality proved to be the two short news texts about allergies and explorers. The second most typical language for boys was that of the instructions on preparing for a job interview, a text which emerged in the girls' ratings as only the 18th most typical of the PISA texts in terms of language. The third most typical language among the PISA texts for the boys was that of the editorial on the problems emerging with the new fertilization technology. Again, girls rated the texts less positively since it was placed 25th. The fourth most typical language for the boys was that of the article about NASA tests on drugged spiders, which the girls had found the most typical. The fifth most typical language in the boys' ranking was that of the movie reviews.

The least typical PISA texts in terms of language for the boys were mainly the same texts as those deemed by the girls to have untypical language although the order in which the texts were placed varied. The least typical PISA text in terms of language was the play extract and the second least typical language was that of the flight timetable. The third least typical text in terms of language, the table of news agencies, was untypical for the girls as well. The fourth least typical language for the boys was that of the bicycle assembly instructions, and the fifth least typical that of the art store catalogue. The tree diagram of a country's labour force and the text on conquering the South Pole, which were among the five texts with the least typical language for girls, were rated clearly more positively by boys. The tree diagram was rated to have 22nd most typical language by the boys and the text on the South Pole the 16th most typical.

To summarize, once again the typicality of the PISA texts was deemed on average to be low although the top average language typicality values exceeded the value of 3,00 meaning 'typical'. It is noteworthy that the top average value given by the boys was higher than that given by the girls, who in most other categories were more positive in rating the texts. However, although in this category boys rated the typicality of language more positively than the girls did

at the extreme ends of the list of PISA texts by language typicality, the girls were overall more positive. For 22 texts out of 37 the average language typicality value given by the girls was higher than that given by the boys. In rating individual texts there were substantial differences over which texts contained typical language for the two genders. Common to both the boys' and the girls' lists were two texts among those with the most typical language and three among those with the least typical language. Otherwise there were differences in both the average typicality values and the rankings of the texts. At the bottom of the list of PISA texts by language typicality there were again more non-continuous texts than at the top of the list. At the top of the girls' list there was only one non-continuous text, the job application form, and at the top of the boys' list there were two, the job application form and the graphs on Lake Chad.

7.5.3 Gender differences in language interest

Again, the typicality of the PISA texts was greater than their interest for both gender groups. The highest average language interest values were lower than the values for language typicality for both girls and boys. Girls were slightly more positive in their ratings since their highest average interest value for language was 2,88 while for boys it was 2,79 (Table 24). The lowest average interest values for boys and girls were quite near each other. Girls were overall more positive in their ratings of the interest of the language used in the PISA texts since in 30 individual cases the girls' average interest value for the language was higher than that of the boys. Furthermore, on the girls' list of PISA texts by language interest there were 16 texts the average language interest value of which reached 2,50 but only 6 such texts on the boys' list.

According to the girls' ratings, the most interesting language was used in the argumentative texts about space research. In terms of language typicality these texts were also among the five most authentic texts for girls (see Table 23 on typicality of language by gender). The second most interesting language was that of the motorbike advertisement and the third most interesting language, with nearly the same average interest value, was that of the guidelines on etiquette for exchange students. Like the argumentative texts, the etiquette was also among the PISA texts with the most typical language for girls and was deemed more typical than interesting in terms of language. The motorbike advertisement, however, was a text with language more interesting than typical according to the girls participating in the study. The fourth most interesting language for the girls was that of the argumentative texts about graffiti, and the fifth most interesting was that of the article about NASA tests on drugged spiders. In four of the five cases mentioned above, the girls' average rating for language interest was higher than that of the boys. The article on NASA tests was the exception to the rule, since the boys' ratings resulted in higher average interest value for the language of the text.

TABLE 24 Interest of language of texts in the PISA 2000 reading literacy test by gender.

Text	Interest of language			
	Girls		Boys	
	mean	ranking	mean	ranking
1. argumentative texts about space research	2,88	1.	2,42	9.
2. motorbike advertisement	2,85	2.	2,54	5.
3. etiquette for exchange students	2,84	3.	2,56	4.
4. argumentative texts about graffiti	2,74	4.	2,28	17.
5. article about NASA tests on drugged spiders	2,69	5.	2,79	1.
6. article on rhinoceros	2,69	5.	2,17	22.
7. narrative about woman and panther trapped in flood	2,68	7.	2,30	16.
8. report on children's attitudes toward housework	2,68	7.	2,12	25.
9. article about using DNA technology in police investigations	2,68	7.	2,43	8.
10. information sheet for potential exchange students	2,62	10.	2,39	11.
11. news texts about allergies and explorers	2,60	11.	2,62	2.
12. article about protecting beaches	2,58	12.	2,13	24.
13. movie reviews	2,52	13.	2,48	7.
14. article on choosing running shoes	2,52	13.	2,32	13.
15. article about the quality of men's shirts	2,52	13.	2,58	3.
16. extract from a play	2,50	16.	2,00	29.
17. instructions on preparing for a job interview	2,48	17.	2,31	15.
18. fable on father's over protectiveness	2,42	18.	2,38	12.
19. job application form	2,37	19.	2,22	20.
20. editorial on problems that arise with new technology	2,32	20.	2,25	18.
21. information sheet about flu vaccination	2,32	20.	2,11	27.
22. table from a report of an aid agency	2,32	20.	2,00	29.
23. graphs on nuclear waste	2,31	23.	2,23	19.
24. brochure describing optician's work	2,30	24.	2,15	23.
25. floor plan of a library	2,23	25.	2,20	21.
26. narrative about villagers' first encounter with movies	2,23	25.	2,08	28.
27. instructions for contacting employer	2,15	27.	2,32	13.
28. graphs about Lake Chad's altitude and fauna	2,09	28.	2,42	9.
29. bicycle assembly instructions for the buyer	2,08	29.	1,88	32.
30. information sheet about a company's job centre	2,04	30.	2,00	29.
31. warranty instructions for a refrigerator	2,04	30.	1,68	35.
32. telephone use instructions	2,04	30.	1,77	33.
33. flight timetable	1,92	33.	1,60	37.
34. text on conquering the South Pole	1,87	34.	2,52	6.
35. a sheet from a catalogue and a mail-order form from an art store	1,85	35.	1,70	34.
36. table of news agencies	1,74	36.	1,68	35.
37. tree diagram of structure of labour force	1,64	37.	2,12	25.

The five texts with the least interesting language as found by the girls participating in the study were all given average language interest values below 2,00, meaning uninteresting' or 'boring'. The least interesting language for girls was that of the tree diagram of the structure of a country's labour force. The second least interesting language was that of the table of news agencies and the third least interesting text in terms of language was the art store catalogue. The

fourth least interesting language was that of the text on conquering the South Pole, a text which was clearly more interesting to boys than to girls (see below). The fifth least interesting PISA text in terms of language was the flight timetable, the only text among the five texts mentioned which included language more interesting than typical for the girls. The art store catalogue and the text on conquering the South Pole in particular were attributed markedly higher average language typicality values than language interest values by the girls. In terms of language interest alone, on the other hand, the text on the South Pole and the tree diagram were more interesting to boys than to girls.

Although boys rated the texts in terms of language interest more negatively than girls, which resulted in lower average language interest values, there were several texts in common among those found most interesting in terms of language by boys and girls. The most interesting PISA text for boys in terms of language was the article on NASA tests on drugged spiders. The text second most interesting in language for the boys, however, was not among the girls' top five or even top ten. The news texts about allergies that boys found to be second most interesting by language came only 11th on the girls' list although the average interest value given by the girls was nearly the same as that given by the boys. The third text in terms of language interest for boys was the article about the quality of men's shirts. The fourth and fifth most interesting PISA texts in terms of language interest for boys were texts that were also at the top of the girls' list of PISA texts by language interest. The fourth most interesting language was that of the guidelines on etiquette for exchange students and the fifth most interesting that of the motorbike advertisement. The average interest values of the texts were clearly lower than those given by the girls participating in the study.

The least interesting PISA texts in terms of language for boys were generally the same as for girls, apart from two exceptions. The least interesting language for boys was that of the flight timetable, which was also among the five least interesting texts in terms of language for girls. The second lowest average language interest value was shared by two texts, namely the warranty instructions for a refrigerator and the table of news agencies. The former text was placed 31st by the girls and the latter was the second least interesting in terms of language also for the girls. The fourth least interesting language for the boys was that of the art store catalogue. This text was ranked in nearly the same position on the girls' list of PISA texts by language interest but once again the average interest value given by girls was higher. The fifth least interesting language for the boys was that of the telephone use instructions.

The two texts that were among the least interesting in terms of language for girls but not among those uninteresting for boys were the tree diagram of a country's labour force and the text on conquering the South Pole. In both cases the language was clearly more interesting for the boys participating in the study. The tree diagram was placed 25th on the boys' list of PISA texts by language interest. Girls, however, found the text to be the least interesting in terms of language. The language of the diagram was more typical than

interesting to boys. The difference between the two genders was even greater over the text on the South Pole. The boys' ratings placed the text 6th on the list of PISA texts by language interest, while the girls found the language of the text quite uninteresting since their ratings gave it the fourth lowest average interest value. Both texts were more authentic to boys than to girls also in terms of topic, genre and layout.

To summarize, in terms of language interest there were clear differences between the two genders. Girls on average were more positive in rating language interest since 30 individual texts received higher average language interest values from the girls than from the boys. Some of the differences between interest values, however, were quite small. The top values were also higher in the girls' ratings. Overall, the language of the PISA texts was not considered to be very interesting by either gender since all the texts were attributed average language interest values under 3,00, meaning 'interesting', by both gender groups. Both among the most and least interesting PISA texts in terms of language were texts rated in common by boys and girls but also texts rated differently. Interesting texts in terms of language for both boys and girls were the article about NASA tests on drugged spiders, the guidelines on etiquette for exchange students and the motorbike advertisement. Equally uninteresting to both genders, on the other hand, were the flight timetable, the table of news agencies and the art store catalogue, all examples of non-continuous text. The texts rated most differently by boys and girls were the text on conquering the South Pole, the tree diagram of a country's labour force and the graphs about the fauna and altitude of Lake Chad, which were rated markedly more positively by boys than by girls. The difference in ratings where girls were more positive was greatest over the report on children's attitudes toward housework and in the play extract.

7.6   Summary of the results

7.6.1 General overview

The results of the study show that the overall authenticity of the PISA texts in terms of typicality and interest as attributes of authenticity was low. The most typical PISA text was given an average typicality value of 2,94 and the most interesting text an average interest value of only 2,75. Furthermore, the texts were usually deemed to be more typical than interesting since comparing the average typicality and interest values for each text revealed that 28 texts out of 37 were given a higher average typicality value than average interest value. It is to be noted, however, that in only 12 cases were the differences between average typicality and interest values statistically significant.

Looking at the results by topic and genre revealed that in these terms also the PISA texts were only moderately authentic. The top average values for topic

typicality and topic interest were slightly below 3,00, meaning 'typical' or 'interesting'. Most of the individual topics, however, were deemed more interesting than typical although in only 9 cases was the difference between average typicality and interest values statistically significant. The genres of the PISA texts were on average even less typical and interesting to the students participating in the study than the topics since the top genre typicality and interest values were 2,85 and 2,76 respectively. In the case of most texts the genre was deemed more typical than interesting. The differences between average typicality and interest values were, however, often small and only in 14 cases were they statistically significant.

In terms of layout and language also the PISA texts were only moderately authentic and on average more typical than interesting. The top value for layout typicality was 3,00, meaning 'typical', while the same value for layout interest was only 2,78. Furthermore, 25 texts were deemed more typical than interesting in terms of layout. The top value for language typicality was 3,00 but the top language interest value was 2,74. As many as 34 texts out of 37 were deemed more typical than interesting in terms of the language used in the text. The difference in average typicality and interest values for layout was statistically significant with 14 texts. For the language, on the other hand, the difference was statistically significant in 23 cases.

There were some clear gender differences both in overall typicality and interest as well as in terms of topic, genre, layout and language of the texts. Girls were overall more positive in their evaluation of the authenticity of the PISA texts since their ratings often resulted in higher average values than the ratings of the boys, both in higher top average values and in the number of individual texts attributed a higher average value than that given by the other gender. For overall interest and for genre typicality in particular the girls gave clearly higher ratings than the boys. However, the top average value for language typicality given by the boys (3,17) was higher than that given by the girls (3,08) although 22 individual texts were given a higher average language typicality value by the girls than by boys. Girls rated the genre typicality of the PISA texts to be the most authentic textual feature, since their top average genre typicality value was clearly higher (3,28) than any other average value in the study. The top average values for other textual features apart from the language interest were also around 3,00, meaning 'typical' or 'interesting'. Boys on the other hand found language typicality most authentic textual feature of the PISA texts as expressed by highest average values, since their top average language typicality value of 3,17 was the highest top average value given by boys.

Looking at the individual texts and their authenticity revealed that there were several texts found authentic (or inauthentic) by girls and boys in common, as well as texts that boys and girls rated very differently. These findings are summarized below.

7.6.2 Texts found authentic or inauthentic by girls and boys in common

A number of texts were authentic in terms of typicality and interest for both boys and girls. The short news texts about gene deficit linked to allergies and explorers in Antarctica and the movie reviews are examples which can be characterized as texts authentic to both genders in common (Table 25). These texts were often at the top of the rankings of the boys but the position in rank order was sometimes lower for girls even when the average typicality and interest values were nearly the same. The movie reviews, for instance, were given the same average genre typicality value by both genders (2,85) but this resulted in different ratings: the genre was placed ninth on the girls' list of PISA texts by genre typicality and second on the boys' list. As well as typical, this genre was also found interesting by both gender groups, but the movie reviews' topic and layout typicality were higher for boys than girls. The news texts were typical and interesting for both genders in terms of the typicality and interest of most of the textual features assessed but this genre was more typical and interesting for boys than girls. The language used was also found more typical by boys than girls.

The article on NASA tests on drugged spiders was also authentic for both genders. Overall, the article was more interesting than typical but it was nevertheless among the most typical texts in terms of genre, layout and language used. Its topic typicality, on the other hand, was lower and students' comments revealed that the title and the content of the article were unusual, which made it very interesting to read. The article about police use of DNA technology in their investigations was overall typical for both boys and girls, but it was also deemed equally interesting. The genre, however, was more typical and interesting for girls than boys. The layout of the article was found interesting by both genders.

The motorbike advertisement with a photograph of an old man and a poem, on the other hand, was more inauthentic in the sense that it was not found typical in the students' textual world. However, it was considered to be interesting by both genders in terms of the four textual features assessed. As shown by their comments, the message of the poem appealed to the students participating in the study.

The texts deemed authentic in terms of typicality and interest by both genders were all continuous ones, either expository or argumentative by text type. Three out of five texts represented a public reading situation, as defined in the PISA assessment framework.

TABLE 25 Texts authentic or inauthentic for boys and girls in common and their PISA text categories.

Text	text type	reading situation	text structure
Authentic texts			
news texts about allergies and explorers	expository	public	continuous
movie reviews	argumentative	private	continuous
article about NASA tests on drugged spiders	expository	public	continuous
article about police using DNA in investigations	expository	educational	continuous
motorbike advertisement	argumentative	public	continuous
Inauthentic texts			
table of news agencies	table	public	non-continuous
telephone use instructions	table	public	non-continuous
flight timetable	table	public	non-continuous

Boys and girls had views in common on what is not authentic as well. Both genders found that the table of news agencies, the flight timetable, and the instructions for using the telephone were inauthentic in terms of all the categories used in the study. Thus, both the topics and the genres as well as the layout and language of the texts were considered untypical and uninteresting. The texts were deemed irrelevant and unnecessary according to the comments students wrote. Students saw no need or purpose for reading the texts and the information the texts contained. It is noteworthy that these three texts considered consistently untypical and uninteresting by both genders were all non-continuous – tables – and originating in public reading situations, as defined in the PISA assessment framework.

Besides these texts being inauthentic for both boys and girls in every respect, there were other texts that were inauthentic in terms of certain textual features or either typicality or interest. The information sheet about a company's job centre was found inauthentic in terms of its topic, which was neither typical nor interesting for either gender group. This was the case also with the bicycle assembly instructions. Boys especially found the text untypical in terms of layout and language. The genre of the texts, however, was deemed more authentic than the topic. The genre of the floor plan of a library, on the other hand, was not found authentic by either of the gender groups. A notable exception in this regard was the layout of the floor plan, which proved to be one of the most interesting ones among the PISA texts. The warranty instructions of a refrigerator were a text that was inauthentic in terms of interest. Both gender groups found the topic and the genre uninteresting but slightly more typical. Boys especially deemed the language used in the warranty instructions uninteresting.

The catalogue of an art store and its mail-order form were considered to be untypical by both girls and boys. Furthermore, boys found the text uninteresting. The layout and the language in particular were inauthentic – both untypical and uninteresting – for boys. In general, students commented that the topics of the several non-continuous texts in particular were uninteresting and best suited for adults. Many of the texts deemed inauthentic by both genders were non-continuous, but there were also a few non-continuous texts among the most authentic texts. In terms of PISA reading situations forming rather a rough categorization of texts, most inauthentic texts represented a public reading situation, although educational (sheet of an art store catalogue) and occupational situations (the information sheet about a company's job centre) were also represented. Thus there is no association between the PISA text categories and text authenticity in the sense that texts found authentic by the students would represent only certain types of texts or certain reading situations. Even with many non-continuous texts found inauthentic there seems to be other factors than the text type explaining the degree of authenticity of the texts, the topic for instance.

7.6.3 Texts more authentic for boys than girls

A text that can clearly be characterized as a boys' text is the unit of two graphs describing the altitude and the fauna of Lake Chad (Table 26). This was placed ninth on the list of most typical texts by the boys but was placed only 28th on the corresponding list for girls. The interest of the graphs was greater among boys than girls since on the list of PISA texts by interest the boys placed it ninth and the girls 26th. The topic of the text was found more typical and interesting by boys than girls, unlike the girls, the boys also deemed the genre quite typical. In terms of layout and language typicality and interest, the graphs were among the nine top texts for boys but among the least authentic texts for girls except in terms of layout interest (19th most interesting for girls). A male student wrote: "*We often read these in our school work. Little text and clear pictures.*"

The editorial on ethical problems caused by the new fertilization technology also matched with the boys' textual world better than with that of the girls participating in the study. The boys ranked the text third on the list of PISA texts by typicality whereas the girls placed it 18th. Boys found the text to be relatively interesting (12th) but girls deemed the text quite uninteresting (21st). Overall the editorial was typical in terms of layout and language but less so in terms of genre. The text was found in general less interesting than typical although the topic was commented on as follows by male students: "*technology is slightly interesting*", "*the examples [in the text] are interesting*", "*the topic is interesting*".

TABLE 26 Texts more authentic for boys than girls by PISA text categories.

text	text type	reading situation	text structure
graphs about Lake Chad's altitude and fauna	chart/graph	public	non-continuous
editorial on problems arising with new fertilization technology	expository	public	continuous
text on conquering South Pole	expository	educational	continuous
article about quality of men's shirts	expository	private	continuous
instructions on preparing for a job interview	instructive	occupational	continuous
tree diagram of a country's labour force	diagram	educational	non-continuous

The text describing the conquest of the South Pole was also clearly more authentic for boys than girls. The boys deemed it 12th on the list of PISA texts by typicality but the girls ranked it fourth from bottom (34th) on the same list. Girls were not interested in the text at all and it came third from bottom on the girls' list of PISA texts by interest, whereas the boys ranked it the fifth most interesting PISA text. It seems that the topic especially was relatively typical and interesting for boys. Furthermore, the genre was found relatively interesting as was the layout with two columns of text, a map and a diagram, which was ranked as the third most interesting among the PISA texts by the boys.

The article about the quality of men's shirts was also clearly a boys' text. Boys rated the text more positively than girls in terms of topic and genre. In terms of language the text was more typical for girls than boys but the interest of the language was greater for boys than girls. Boys found the topic of the article interesting and the genre both typical and interesting. The instructions on preparing for a job interview also proved more authentic for boys than girls but the boys found it most authentic from the point of view of genre and language typicality. The instructions were not considered very interesting, nor even the topic of the text typical. The tree diagram of a country's labour force was another text more authentic for boys than girls. In fact, the text was rated among the least typical and interesting texts by the girls in terms of all four textual features covered in the study–topic, genre, layout and language. The boys, on the other hand, found the diagram both typical and interesting. The genre especially was found typical, but so was the layout to some extent.

The structure of the texts characterized as boys' texts was mainly continuous, but the graphs about Lake Chad and the tree diagram of a country's labour force were examples of non-continuous documents. However, there were also significant non-continuous elements in the text on conquering

the South Pole and the article on men's shirts. The situation for reading, as defined in PISA 2000 framework, represented by the texts which proved more authentic for boys than girls varies a great deal. Two texts are from public reading situation (the graphs on Lake Chad, and the editorial) and two from educational situation (the diagram of a labour force, and the text on the South Pole), but occupational (the instructions for a job interview) and private reading situations (the article on shirts) are also covered.

7.6.4 Texts more authentic for girls than boys

The most authentic text for girls in the PISA reading literacy test was the argumentative text unit on space research, which consisted of five texts written by young students (Table 27). The girls found the text to be both the most typical and the most interesting among the PISA texts. Girls emphasized in their comments that they find texts expressing opinions interesting and consequently they read them a lot. Moreover, the topic of the texts was also the most interesting and the most typical among the PISA texts as assessed by the girls, as was the layout and the language used in the texts. Despite the potentially appealing topic for boys they did not find the textual unit to be very authentic placing it 22nd on the list by typicality and 13th by interest. Another argumentative text unit expressing opinions was about graffiti. Despite the same appealing genre, the text was not equally authentic for the girls participating in the study. The text was of some interest to girls in its topic, but the genre and language used were found typical by boys also.

Although the boys did not find any of the literary texts authentic, there was one story among the PISA texts that the girls found authentic. This was the longest text in the PISA reading literacy test, a story that tells about a woman and a panther trapped in a Mississippi flood. The text was quite demanding to read because it contains none of the dialogue which often makes text varied and easier to read. However, the girls ranked the story 7th on the list of PISA texts by overall typicality. On the corresponding list for boys the story was placed only 21st. While the girls also found the text the most interesting literary text in the PISA test the boys also placed it 21st on their list of PISA texts by interest. Girls found the genre and the layout of the story to be typical and the genre also interesting.

The report on young people's attitudes towards housework was clearly more authentic for girls than boys. The girls ranked it the second most typical and sixth most interesting PISA text. The boys, however, deemed the report to be 20th on the list of PISA texts by typicality and 26th on the list by interest. Considering authenticity in terms of the four textual features assessed confirmed that this was more authentic for girls than boys since the girls consistently rated the text more positively than the boys in terms of typicality and interest for topic, genre, layout and language. The girls' ratings always placed the report among the eight most typical and interesting texts whereas the boys' ratings placed it 11th at best and usually below 20th place.

TABLE 27   Texts more authentic for girls than boys by PISA text categories.

text	text type	reading situation	text structure
argumentative texts about space research	argumentative	educational	continuous
narrative about a woman trapped in flood	narrative	private	continuous
report on children's attitude towards housework	charts/graphs	educational	non-continuous
information sheet for potential exchange students	expository	educational	continuous
etiquette for exchange students	instructive	educational	continuous
extract from a play	narrative	private	continuous
job application form	form	occupational	non-continuous
article about saving and protecting beaches	expository	public	continuous

There were two texts related to exchange students in the PISA 2000 reading literacy test and both of them were more authentic for girls than boys although the boys' average ratings of the texts were not very low either. The information sheet for potential exchange students was the more typical and interesting of the two texts. Both the topic and genre of the text were typical and interesting for girls but the topic was of some interest also to boys. The guidelines on etiquette for exchange students, on the other hand, had a topic and language that the girls especially found both typical and interesting. The topic was also typical for the boys but to a lesser degree, as was the case with language interest.

A literary text that was, if not the girls' favourite, still clearly more authentic for them than for the boys was the extract from a play on a romantic theme. Neither gender group found the play overall typical but girls clearly rated the text more positively than boys in terms of the interest of the topic. The genre in particular was found untypical by both genders. The boys ranked the topic as the least typical and the second least interesting. On the other hand, boys were more positive towards the genre in terms of its interest although they found the layout and the language used in the play to be among the least typical. The language of the play was deemed untypical also by girls although they found the language relatively interesting.

The job application form was another text more authentic for girls than boys, especially in terms of typicality since it was the third most typical PISA text in their assessment. In terms of interest, the text was rated more cautiously (14th) by the girls but still more positively than by the boys. Looking at individual textual features more closely revealed that girls particularly found the genre, the layout and the language of the text typical. Another text, the

article about saving and protecting beaches, was more authentic for girls than boys both in terms of typicality and the interest of the text. Although the article was not among the most authentic texts for them, appearing only halfway down the girls' lists of PISA texts, the girls rated it more positively than the boys on all the textual features under study.

Overall, the girls found a quite heterogeneous selection of texts authentic in terms of text typicality and interest. The job application form and the report on children's attitudes towards housework were the only texts containing a significant proportion of non-continuous elements that girls found authentic. Unlike boys girls also found some literary texts or at least texts in a narrative genre relevant to their lives. The narrative about the woman and the panther in the flood and to some extent the play extract were more appealing to girls than to boys. In terms of the reading situations defined in the PISA framework, girls found many educational texts authentic (e.g. students' opinions on space research, the report on housework, and the information sheet for exchange students). The private reading situation was also represented among the girls' PISA texts (the story about the woman trapped in the flood and the play extract). Public and occupational settings were represented to a lesser extent among the texts that the girls found more authentic than the boys. However, there were public texts among the texts authentic for both genders.

8   THE RELATIONS BETWEEN TYPICALITY AND INTEREST OF TOPIC, GENRE, LAYOUT AND LANGUAGE

In the case of most of the individual texts rated in the study, it was the same texts that ranked among the most typical and the most interesting ones regardless of the textual feature assessed. This raised the question of the students' ability to review and rate the textual features separately. Furthermore, it was found necessary to study the relation between typicality and interest as attributes of text authenticity in assessing the validity of the study. The questions considered were: to what extent typicality and interest cover different aspects of authenticity, to what extent they are associated with each other and to what extent the questionnaire designed to assess text typicality and interest succeeded in doing so.

The results of the factor analysis varied by text but some general trends appeared consistently. With most of the texts (32 out of 37) factor analysis resulted in two factors (initial eigenvalues >1.0) accounting for 59 to 89 per cent of the total variance. The loadings items had on the factors were generally quite high, and therefore only loadings greater than 0,4 have been reported. With 18 of the analyzes, the variables measuring *interest* of topic, layout, language and genre had high loadings on the one factor (hence called interest factor) and the variables measuring *typicality* of the same textual features had high loadings on the other factor (hence called typicality factor). Below are the matrix related to the article on the rhinoceros (Table 28), exemplifying this type of factor analysis, and the list of texts with two factors (Table 29).

TABLE 28   Rotated Factor Matrix: article on rhinoceros.

	Factor	
	interest	typicality
typicality: topic		,63
layout		,73
language		,88
genre		,82
interest: topic	,80	
layout	,81	
language	,83	
genre	,88	

TABLE 29   Texts with two factors in factor analysis.

article on rhinoceros
instructions for personnel
refrigerator warranty instructions
article on men's shirts
instructions on preparing for a job interview
information sheet on flu vaccination
job application form
brochure on optician's work
play extract
article on saving beaches
library floor plan
fable
table on aid agency's activities
opinions on graffiti
bicycle assembly instructions
information sheet to potential exchange students
table of news agencies
text on conquering South Pole

With the rest of the 14 texts belonging to this category the situation was slightly different since some variables had loadings on both factors. However, in most of these cases the loadings of the other (secondary) factor were clearly lower than those of the primary factor. For instance, the matrix below (Table 30) shows the factor loadings of the analysis related to the article on choosing running shoes. There are two variables that have loadings on both factors. The loadings, however, are clearly higher with the factor describing the aspect the particular variable was expected to have loadings on. Thus, the interest variables have high loadings on the interest factor – i.e. the factor all interest items have high loadings on. And vice versa, the typicality variables have high loadings on the typicality factor but the two interest variables having loadings on the typicality factor also have clearly lower loadings on that factor. A list of texts with results of this type in the factor analysis is presented in Table 31.

TABLE 30   Rotated Factor Matrix: article on choosing running shoes.

| | Factor | |
	interest	typicality
typicality: topic		,46
layout		,60
language		,48
genre		,67
interest: topic	,75	
layout	,82	
language	,75	,43
genre	,67	,41

TABLE 31   Texts with two factors: loadings on both factors.

graphs on nuclear waste
news texts
editorial
article on choosing running shoes
graph about Lake Chad
instructions on contacting an employer
article on drugged spiders
art store catalogue sheet
flight timetable
extract of a story about villagers' first encounter with movies
etiquette for exchange students

With three of the texts belonging to this category of two factors, several variables had relatively high loadings on both factors. This was the case with the movie reviews, the narrative about a woman and a panther trapped in a flood, and the motorbike advertisement. For instance, factor analysis related to the movie reviews resulted in the following matrix (Table 32), which shows nearly equally high loadings on both factors by the variables measuring typicality of layout and interest of the topic.

TABLE 32   Rotated Factor Matrix: movie reviews.

	Factor	
	typicality	interest
typicality: topic	,83	
layout	,59	,52
language	,83	
genre	,80	
interest: topic	,53	,46
layout		,78
language		,79
genre	,47	,57

Finally, there were five texts for which the factor analysis resulted in extracting only one factor. These texts were the reports on young people's attitudes to housework, the article on police use of DNA in their work, the telephone use instructions, the students' opinions on space research, and the tree diagram of a country's labour force. Below (Table 33) is a matrix related to the report on young people's attitudes to housework in different countries, which shows an example of this type of structure among the variables.

TABLE 33   Factor Matrix: report on housework.

	Factor
	1
typicality: topic	,75
layout	,68
language	,61
genre	,78
interest: topic	,79
layout	,85
language	,84
genre	,81

The different types of results obtained by factor analysis reveal that there are differences in the extent to which the typicality and interest variables are associated with each other. The first type of result, with two extracted factors with relatively high loadings, shows that typicality variables are associated with each other and interest variables with each other as there are relatively high correlations among them, but lower correlations across typicality and interest variables. The second type consists of cases with two factors extracted in the analysis but some items having loadings on both factors. In these cases, factor analysis shows, however, that one factor has clearly higher loadings than

the other, secondary, one. This is explained by the correlations between students' evaluations: usually in these cases the item that has low loadings also on the factor that was not assumed to have them has high correlations across typicality and interest items, specifically the other item measuring the characteristics of the *same textual feature*. Thus, for instance the typicality of language and the interest of language variables in the article on choosing running shoes are correlated to the extent that the interest of language variable has a loading on both typicality and interest factors (see Table 30). Nevertheless, the results of the factor analysis resemble those of the first type since the loadings of the interest of language variable are high on the interest factor and low on the typicality factor.

The third type of result in factor analysis is related to cases in which the factor analysis extracted two factors with spread out loadings, as described above (see e.g. Table 32). In these cases, there are several high correlations across typicality and interest variables with occasional lower correlations among typicality and interest variables. As these correlations across typicality and interest variables became more frequent and higher and sometimes even resulted in all eight item variables having nearly equally high correlations with each other, the factor analysis finally extracted only one factor (see e.g. Table 33).

All this means, that with most of the texts (29 out of 37) factor analysis supports the construct validity of the measurement in the sense that the variables designed to measure the typicality of textual features on the one hand and the interest of the textual features on the other, succeed in doing so. As typicality and interest variables are clustered according to the assumed factor structure – typicality items being strongly associated with each other and interest items with each other – they have measured separate aspects or attributes of students' evaluations of text authenticity. In this regard, theory, the practical realization of theory, and the results of the factor analysis are in line with each other. In terms of the content validity of the concept of authenticity, however, factor analysis does not provide any information, since on the basis of the data it is impossible to know how well the chosen attributes for authenticity (typicality and interest) represent the whole concept of authenticity. This aspect of the validity of the present study will be assessed on the basis of theoretical reasoning (see Chapter 9.3).

With the rest of the 8 PISA texts the situation in terms of typicality and interest as attributes of authenticity is somewhat different. In many cases even with the second type of result in factor analysis (Tables 30 and 31) there are some textual features the typicality and interest of which are highly correlated, causing some variables to have loadings on both factors extracted in the analysis. To a certain degree this would be as expected: a text must be to some extent interesting to become typical. The topic of a text, for instance, attracts readers to engage with the text if it interests them. Should certain types of texts – or topics – interest the reader he will choose those texts to read and the texts will become typical reading materials. However, with some extremely high correlations across typicality and interest variables it is necessary to ponder

whether the variables are in fact redundant, measuring nearly the same characteristic of the text. This is the case especially in cases where factor analysis resulted in the extraction of only one factor, but also where factor analysis yielded two factors with some variables having loadings on both factors. For instance, with the telephone use instructions the correlations between typicality of topic and the other variables vary with other typicality items from ,63 to ,70 and with interest items from ,56 to ,71. As the text was among the least typical and least interesting for whatever reason, this might suggest that the variables have measured a general attitude towards the text for which less specific variables would have been sufficient. On the other hand, the students' ratings are perfectly plausible since it is unlikely that instructions of the sort would be a typical text in the culture of 15-year-olds' or a text of great interest. Since the high correlations between the variables for this particular text are an indication of a genuine phenomenon, the explanation for the results of factor analysis needs to be sought text by text. The other four texts for which the factor analysis yielded one factor are texts that were rated to be either among the most typical and interesting ones in terms of several textual features or among the least typical and interesting ones. Therefore, in these cases there is a plausible explanation for the results of the factor analysis: the texts were rated similarly despite the textual feature or aspect of authenticity and thus the variables were highly correlated.

As the above interpretations of factor analysis have been from the point of view of the concept of authenticity, it is necessary to examine the results also from the point of view of the four textual features in order to see to what extent typicality and interest of topic, genre, layout and language are associated with each other and to what extent students were able to rate the textual features separately. With most of the texts – at least the 29 forming the two first groups of texts emerging from the factor analysis– the results of the factor analyzes suggest that the textual features are strongly associated in terms of typicality and interest but less across typicality and interest factors. This was as expected, as it is quite plausible that to a certain degree for instance the typicality of a text's topic, genre, layout and language are associated with each other. Genre especially was expected to be correlated to layout and language since they all represent textual features deriving from the contextual function of the text. However, as the correlations are often as high as ,71–,79 (as are the correlations between the interest of genre and other interest variables with the article on the rhinoceros) this suggests that they begin to be wholly redundant and therefore less detailed and less specific items would have been sufficient. Possibly the typicality and interest of the textual features chosen for the study are strongly associated, but nevertheless it might have been sufficiently specific for 15-year-olds in the time given to limit the review of text typicality and interest to for instance topic and genre.

It is noteworthy that in the cases where factor analysis resulted in one factor or two factors with spread out loadings, the eight variables in the study are more and more associated. Thus, there are relatively high correlations

between typicality and interest items, especially those dealing with the same textual feature. This suggests that with these texts the typicality and interest of textual features are strongly associated to the extent that items are nearly redundant. However, it is difficult to say what it is that these item variables then measure. Perhaps they measure some general attitude towards the texts. Since there are also relatively high correlations between all the variables, not only those measuring the typicality and interest of the same textual feature, this seems plausible. Nevertheless, this supports the conclusion that less specificity in terms of textual features would have been enough in the composition of the questionnaire. It can thus be concluded that reviewing the separate textual features is not authentic to 15-year-olds as they seem to respond to a text as a whole.

Overall, from the point of view of construct operationalization and the instrument this suggests that with most of the texts the questionnaire succeeds in measuring what it was designed to measure, sustaining the construct validity of the study. However, as the factor analyzes show differences with individual texts it is to be noted that with some texts the results suggest that unnecessarily specific variables have been used at least in terms of textual features, since either the typicality and interest of the topic, genre, layout and language of the text are strongly associated or students have been unable to review and rate the typicality and interest of the texts presented to them in terms of their textual features.

9   CONCLUSIONS AND DISCUSSION

9.1   Answering the research questions

In this chapter answers are put forward to the research questions of the empirical study and conclusions are drawn. A theoretical discussion of the results is presented thereafter.

1. Which textual features do readers find relevant to text authenticity?

The comments by the panel of eight expert readers during the pilot study phase revealed 11 different textual features relevant to text authenticity. These textual features were genre, context, content, structure, topic, layout, form, language, style, target group and text type. In terms of the number of thematic units representing each category of textual feature, the most significant of these to the expert readers on the panel were genre, context, content and structure. These textual features were referred to 89 – 136 times in the panellists' comments on the PISA texts. The textual feature commented on most often was genre (see Table 9, p. 80). Context was also commented upon frequently. In their comments about 'context' panellists referred to either context of appearance, context of use or lack of context. Lack of context was especially disturbing in the case of non-continuous texts appearing alone without any continuous text attached, as often occurs in authentic real-life reading events. 'Content' was also referred to quite often and it covered comments about the things and ideas presented in the text in detail. In addition, the pilot study data included many comments on the structure of the texts, referring mainly to the elements the texts consisted of.

The topic, layout and form of the texts were commented on relatively often, 59 – 61 times. 'Topic' referred to the issue the text deals with at a general level. 'Layout' referred to the appearance of the text, including comments about the use of space or fonts and font size etc. The category of 'form' included comments about the length of the texts or the format of non-continuous documents.

The language and style of the texts were also commented upon, each forming their own categories of thematic units. Comments on language (43 thematic units) referred to individual expressions and words used in the text or to the actual language used in the text (e.g. English shirt names) or to the fact that texts were translated. Comments about style (39 thematic units) referred to the way things are said in the text, and quite often these comments were also about the general tone of the text.

The data of the pilot study included some comments (16 examples) about the target group the texts were aimed at. In these comments adults and 15-year-olds were often contrasted. Finally, there were three comments about text type, defined on the basis of the linguistic criteria of Werlich (1983).

Clearly the most significant textual feature in evaluating text authenticity in terms of typicality was genre. This was further underlined by the large number of thematic units placed in the category of context, since some of the comments placed in this category included references to a context of appearance related to genre (e.g. magazines and newspapers). In contrast, text type was evidently an insignificant textual feature to the panellists, with the fewest thematic units (3) found in the content analysis. It can thus be concluded that the expert readers on the panel approached the texts in terms of extralinguistic features such as genre, context and layout rather than linguistic features such as text type classification. However, in international reading literacy assessments text are usually classified by text type rather than by genre. The results of the pilot study suggest that the level of genre should also be taken into account in defining the kinds of texts to be chosen for international reading literacy tests. Currently for instance in the PISA assessment framework the extralinguistic and functional characteristics of texts have been taken into account in defining situations for reading. The main criteria in defining reading situations are the writer's intended use for the text and who are the others involved in the reading event. (OECD 1999, p. 23.) As genre is defined on the basis of the communicative purpose of the text, the definition of situations for reading used in PISA in principle allows for the inclusion of text classification by genre together with the aspects of the context of appearance and the layout of the text. The four situations for reading, however, form rather rough a categorization to be applied in actual text classification.

On the basis of the pilot study data and theoretical considerations it was concluded that the four textual features most relevant to text authenticity were genre, topic, language and layout. The selection of these features was also influenced by practical reasons, especially the age of the students responding to the main study questionnaire and the time available for students to read and analyze the texts.

2. How authentic were the texts used in the PISA 2000 reading literacy test for Finnish 15-year-old students?

The results of the main study showed that text authenticity in terms of typicality and interest was low in respect of all the textual features assessed. At

best the students participating in the study found the texts used in the PISA 2000 reading literacy study typical and interesting. Most of the texts and most textual features, however, were found less than typical or interesting, and several texts and their textual features were even found untypical and uninteresting, receiving average typicality and interest values slightly above and below 2,00. It is also noteworthy that most of the PISA texts were found *more typical than interesting*.

The low text authenticity in terms of typicality and especially interest may in part derive from the conscious attempt to choose texts that would be acceptable and authentic in all the countries participating in the PISA 2000 reading literacy test. With a test that has as many participating cultures and languages as PISA, such an attempt is likely to result in a choice of texts that are neutral in many ways and they have in fact been remarked upon as bland and insipid. There are some topics that without further discussion are excluded from the test as too controversial. Therefore texts dealing with political and religious issues are rarely accepted for an international reading literacy test. The natural sciences and natural phenomena, on the other hand, form a field of issues and topics that are more easily accepted in different countries. In general, cultural sensitivity restricts the choice of authentic texts for a reading literacy test and in consequence the texts selected do not arouse any great enthusiasm in the readers. Overall the results of the present study suggest that the PISA texts do not represent the textual world of Finnish 15-year olds very well, at least in terms of interest. As the texts read by 15-year-olds cover a wide range of topics and genres from song lyrics to articles on deep sea diving with multimodal expressions of meaning, the texts used in PISA 2000 reading literacy test form only a pale reflection of the textual world of young people. This is true especially with the electronic texts that have practically been missing in the PISA test although they are a part of everyday reading events of 15-year-olds. The international option of assessing electronic reading available in the PISA 2009 assessment gives the opportunity to fill this gap between the PISA test and the everyday reality of readers.

It is noteworthy that the differences in average typicality and interest values between texts were often small and quite often not statistically significant. Therefore the order of the texts by typicality and interest is only suggestive. Nevertheless, the texts that were the most and least authentic for all the students participating stand out, and the focus of attention in analyzing the results has been on them. The PISA texts that stand out as authentic in terms of typicality and interest for the students participating in the study were the news texts about allergies and explorers, the movie reviews, the article about NASA tests on drugged spiders, the article about police using DNA in their investigations and the motorbike advertisement (see Table 25). Examined in terms of text and reading situation classification used in the PISA assessment framework, the only common denominator for these texts is that they are continuous type of texts. Otherwise the text types and formats as well as reading situations vary. This suggests that text type and situation for reading

do not necessarily offer any explanation for the typicality and interest of texts. On the other hand, it is obvious that these are all texts that originally appeared in newspapers or magazines. This was commented upon by students themselves in explaining their ratings. To conclude, the original context of appearance, which in part defines or is related to the genre of a text, seems to be of essential importance in texts deemed authentic in terms of typicality and interest.

Many texts originating in newspapers and magazines have been defined in the PISA classification of reading situations as public texts. This is not, however, the case with all these texts. Movie reviews, for instance, have been classified as texts originating in private situation for reading. Perhaps this was done on the basis of the intended use for the text. The article about the police using DNA in their investigations, on the other hand, was classified as an educational text, perhaps because the content of the text fits the idea of reading to learn. Both texts could equally well have been defined as public. Genre should receive more emphasis in classifying texts for international reading literacy assessment since this is found important by readers for text authenticity, and consequently reading situations could be defined according to the readers' context of use as well as the text's context of appearance, rather than on the basis of the writer's intended use as was done in PISA 2000 (OECD 1999).

The results of the present study – at least those related to the most authentic PISA texts – are consistent with the results obtained on the reading engagement of the 15-year-old students studied in PISA 2000 (Linnakylä 2002; OECD 2002; Linnakylä & Sulkunen 2005). The PISA 2000 and 2003 assessments showed that Finnish students frequently read newspapers and magazines, markedly more often than the average in OECD countries (Linnakylä 2002; Linnakylä & Sulkunen 2005). Furthermore, in defining and describing the reader profiles of 15-year-old students it was shown that even the least diversified readers frequently read magazines and fairly often read newspapers. The majority of Finnish 15-year-old readers can be characterized as diversified readers of shorter texts such as appear in newspapers, magazines and comic books. Fiction, on the other hand, is read less. (OECD 2002.) Taking all this into consideration the results of the present study are as expected. Since students often read newspapers and magazines, texts appearing in those contexts are bound to be typical and must also be interesting to some degree in order to be read in the first place. It has to be noted, however, that in an assessment situation the texts have been chosen for the test takers by the test designers. Therefore a high degree of interest cannot be readily assumed for the texts although it should be aimed at in selecting them. Since the data of the present study revealed that most PISA texts were more typical than interesting in terms of topic, genre, layout and language, the challenge is to find *more interesting* texts for the reading literacy test.

The classification of texts into continuous and non-continuous text types of the PISA assessment framework seems to some extent relevant in describing

the authentic texts of the PISA reading literacy test. Not only were the most authentic texts in terms of typicality and interest continuous in type but the texts deemed least authentic by both girls and boys, namely the table of news agencies, the telephone use instructions and the flight timetable, were non-continuous. These non-continuous texts were all tables originating in public reading situations, as defined in PISA. It would be short-sighted, however, to conclude that non-continuous texts, particularly tables, originating in public contexts are inauthentic for the students participating in the study. The comments provided by students answering the main study questionnaire reveal that the particular tables in question provided information students found irrelevant and unnecessary. Thus, it is especially the topic of these texts which seems to be inauthentic in terms of typicality and interest, although in fact these tables were among the least typical and interesting texts in terms of all the textual features assessed.

Furthermore, it is noteworthy that there were also texts with significant non-continuous elements that were still deemed relatively typical and interesting at least by one gender group, for instance the extract on conquering the South Pole in the boys' case, and the report on young people's attitudes towards housework in the girls'. This suggests that it was not only the topic of the particular tables mentioned above but also the lack of appropriate context that was found inauthentic. The comments presented by panellists during the pilot study phase echoed this view quite strongly since panellists often remarked that documents like these do not usually appear alone but with continuous text attached, which is why the lack of context was found disturbing. After all, young readers constantly encounter non-continuous elements in textbooks and in newspapers and magazines – the source of their favourite reading materials. Furthermore they will continue to encounter such documents throughout their studies and work life (Linnakylä et al. 2000). Therefore, as a text type and format a variety of non-continuous texts is an appropriate component for a reading literacy test. The topic of non-continuous texts, however, should be given careful consideration. There is no doubt that there are non-continuous texts the topics of which are of interest to 15-year-olds. That the lack of an appropriate context disturbs readers should also be taken into account in choosing texts for reading literacy tests. Decontextualizing texts necessarily alters text interpretation and thus text authenticity. This should thus be acknowledged in selecting texts for assessment purposes. The easiest way to ensure at least a somewhat relevant context for non-continuous texts is to place them in the wider context of continuous texts, as is the case in real life.

To conclude, the results of the present study suggest that context of appearance and genre seem to be one textual feature that must be considered in describing and defining the authenticity of texts. Certainly it is not a question of one textual feature determining text authenticity in terms of typicality and interest. Students' comments in the questionnaires reveal that with some texts it is the topic, in some the genre and in others the non-continuous format that their main attention is focused on and which explains their ratings.

Nevertheless, both interest of topic and relevant and meaningful context are characteristics of texts that should be emphasized in an assessment context.

3. Were the texts used in PISA 2000 reading literacy test equally authentic for Finnish 15-year-old boys and girls?

There were texts that were found equally authentic or inauthentic by both gender groups. The texts mentioned above (the news texts about allergies and explorers, the movie reviews, the article about NASA tests on drugged spiders, the article about police using DNA in their investigations and the motorbike advertisement) are good examples of texts found equally authentic for girls and boys. It is to be noted, however, that although the average typicality and interest values attributed by girls and boys may be different in individual cases these texts appeared at the top of rank order listings for both genders. On the other hand, there were also cases of clear gender differences in rating text authenticity in terms of typicality and interest and thus also texts that were not equally authentic for boys and girls. This was expected, since girls and boys are known to have somewhat different interests in choosing their reading materials. In the Finnish context, both boys and girls read newspapers, e-mail and web sites, as well as magazines, although magazines are read somewhat more often by girls than by boys. Comics are read considerably more by boys, and fiction is read more by girls than by boys. (Linnakylä 2002.)

The texts more authentic in terms of typicality and interest for girls than for boys were the argumentative texts about space research, the report on children's attitudes towards housework, the information sheet for potential exchange students, the etiquette for exchange students, the job application form, the narrative about a woman trapped in a Mississippi flood, the extract from a play and the article about saving beaches. Looking at these texts in terms of the PISA classification of texts and reading situations reveals that these texts form a very heterogeneous group. There are both continuous and non-continuous texts and furthermore texts represent many different text types and reading situations. Half of the texts found more authentic by girls than by boys mentioned above were drawn from educational situations for reading but students' comments on these texts reveal that there were other reasons for girls to find these texts typical and interesting. The argumentative texts about space research, for instance, were classified as educational texts probably because they were written by students at school on a given topic. However, girls commented that they like reading about other people's opinions and they frequently read opinions in newspapers, magazines and web pages. This was the case also with the report on children's attitudes towards housework. The most essential feature of that text for girls according to their own comments was that it was about other people's opinions and attitudes although the genres of the two constituent texts were different. The two texts related to exchange students, on the other hand, were found typical and interesting because of the topic, which is something familiar to 15-year-old students, as was reflected in students' own comments. The girls' ratings of the texts about exchange students

were consistent with their comments since the ratings revealed that the topic interest of the two texts was relatively high.

The most authentic literary PISA text for girls was the narrative about the woman in the flood although it was placed at the top of the rankings of PISA texts only in terms of its genre typicality, genre interest and layout typicality. Thus, girls' greater interest in fiction (Linnakylä 2002) was reflected here. Moreover, the fact that the main character was female as well as the length of the story could explain girls' greater interest in this particular text (see e.g. Taube & Munck 1996). The play extract, on the other hand, was not among the most authentic PISA texts by typicality and interest but it was *more* authentic to girls than to boys. Only the topic interest and language interest of this text reached average interest values of 2,50 in the girls' ratings. The results for these two texts suggest that girls do read fiction but not necessarily plays. In fact the genre typicality of the play was only 2,09 , i.e. near 2,00 meaning 'untypical'. The job application form was also more authentic for girls than boys. Girls found the genre to be relatively typical, as was the language and layout. Comments on the text revealed that forms in which names, addresses and other basic information is asked for are familiar to students but not so much the topic of applying for a job.

Texts more authentic to girls than to boys reflect the general reading interests of girls. Girls read heterogeneous reading materials in terms of genres and topics, and in addition to the kind of texts also authentic for boys girls favour especially texts expressing opinions and covering topics which belong to current issues of 15-year-old students' lives.

Although it might be concluded that boys were disadvantaged by texts more authentic to girls in terms of typicality and interest, there were also texts more authentic to boys than to girls, for instance the graphs about Lake Chad, the editorial on problems arising with the new fertilization technology, the text on conquering the South Pole, the article about the quality of men's shirts, the instructions on preparing for a job interview, and the tree diagram of a country's labour force. None of these texts were necessarily among the most authentic but still they were more authentic for boys than girls. As in the girls' case, boys' texts formed a very heterogeneous group when examined in terms of the PISA classification of texts and reading situations. There were both continuous and non-continuous texts and texts representing various text types and reading situations. It is noteworthy, though, that all the continuous texts more authentic for boys than girls were either expository (the editorial, the text on conquering the South Pole and the article about the quality of men's shirts) or instructive (instructions on preparing for a job interview). There were no narrative texts, for instance, that boys could have ranked among the most authentic PISA texts, or at least found more authentic than girls did. Some boys' texts also included significant non-continuous elements even when they were classified as continuous texts (the article about men's shirts, the text on the South Pole). The greater interest of non-continuous texts for boys is consistent with the studies that show that tasks attached to texts of that kind are relatively

easier for boys than tasks attached to other types of texts (Taube & Munck 1996; Roe & Taube 2003).

Since the texts that proved more authentic for boys than girls seem to have very little in common in terms of PISA text classification, the explanation must be looked for elsewhere. The data reveals that for instance non-continuous graphs about Lake Chad and the tree diagram of a country's labour force were authentic for boys especially in terms of genre typicality, layout typicality and language typicality. They were also somewhat interesting in topic and layout. This was echoed in students' comments about the two documents, in which they emphasized that diagrams and graphs are familiar from school, textbooks and magazines and newspapers, all of which are encountered and read relatively often by boys (Linnakylä 2002; OECD 2002). The ratings of the texts revealed that the textual features assessed in the study were more typical than interesting in these documents.

Both the editorial on problems caused by the new fertilization technology and the instructions for the job interview were texts that boys found in many respects typical but not so interesting. Especially the language, layout and genre of the texts were found relatively typical by boys. The comments students made about these texts, on the other hand, dealt with topic and content. The editorial was said to have an interesting topic by some students who noted that technology-related issues are interesting. Instructions for the job interview were familiar to students from youth magazines and students found the advice on the topic useful, necessary and interesting.

The text on conquering the South Pole, on the other hand, was a text that was more interesting than typical in terms of the average typicality and interest values for all four textual features assessed in the study. Students' comments plainly stressed that the story was interesting. One possible explanation for this could be the fact that the main characters were male explorers (Taube & Munck 1996). The article on men's shirts was also interesting in topic, genre and language. Furthermore, the genre was found typical, consistently with the reading interests of young students (Linnakylä 2002; OECD 2002). Students' comments revealed that the topic itself – men's shirts – could have been more interesting if it had been young people's clothes. On the other hand, the more precise content dealing with the relationship between branded clothes' quality and price was interesting. Also this kind of information relevant to consumers was found useful.

Based on the results concerning which texts are authentic for boys and girls in common and which are more authentic for either of the genders, some conclusions can be cautiously drawn about the characteristics of authentic texts for 15-year-old students. Although the number of students chosen for the study was small and thus the results cannot be generalized beyond the group of students participating the study, some suggestions can be made about the kind of texts authentic for 15-year-old Finnish students, since the results of the study are consistent with what is known about the reading interests and habits of 15-year-olds in general (Linnakylä 2002; OECD 2002; Linnakylä & Sulkunen 2005).

As could be expected, the students participating in the study found texts and especially genres sourced from newspapers and magazines authentic in terms of typicality and interest (news texts, articles, movie reviews). Furthermore, girls found various types of texts expressing people's opinions (the report on housework, the texts on space research) authentic. Some non-continuous texts were found typical but not equally interesting. For the girls, the job application form represented such a text, and for the boys the graphs about Lake Chad and the tree diagram. The girls also found the narrative on the woman trapped in a Mississippi flood to be a relatively authentic literary text, whereas the boys included none of the literary texts to be among the most authentic texts in the PISA test. However, a short fable with a male character (Taube & Munck 1996) was found more authentic by boys than by girls although it was not among the most authentic texts for boys.

Otherwise it seems that students find various types of texts on various topics authentic. The topics, however, need to be either generally and currently interesting or relevant to students of that age. Such topics for both genders were gene deficit linked to allergies and explorers in Antarctica, tests on drugged spiders and police use of DNA in their investigations, topics that reflect the general interest of science-related issues. Comparable topics for girls were student exchange, and to some extent the romantic story in the play about Amanda and the prince, while such topics only for boys, on the other hand, were technology as it appeared in the editorial and historical adventure in the text about the South Pole. In general, consumer information was found useful and interesting. The role of topic in assessing text authenticity was further underlined by the fact that the topics of some PISA texts were deemed inauthentic in terms of typicality and interest. Certain topics were considered to be irrelevant by the students and aimed at adults rather than their own age group. Such topics were the warranty of a refrigerator, bicycle assembly instructions and a company's job centre. Interestingly enough, the texts dealing with these topics had already been used in the International Adult Literacy Survey.

Overall, the PISA texts were only to a moderate degree authentic in terms of typicality and interest for the students participating in the study. Furthermore, girls were usually more positive in their use of the rating scale since they awarded higher top average values for typicality and interest. Thus, it is a particular challenge to find texts that are typical and above all interesting for boys in international reading literacy studies. The potential interest of the texts selected should be given more attention in order to motivate students and to provide a true reading experience for students taking the test. The most successful combination in terms of text interest seems to be a combination of typical genre and untypical and thus interesting topic. The results of the students' ratings and their explanations for their ratings of the PISA texts reveal that, for instance, the article on NASA tests on drugged spiders was successful in this respect. The motorbike advertisement was also found to be a text that successfully and in an unusual manner uses a poem representing soft values of

life to advertise a motorbike. Texts with untypical and surprising topics should thus be selected for the reading literacy test but also topics interesting and relevant to 15-year-olds. Although texts relevant to 15-year-olds in their future lives should be used in reading literacy tests (e.g. OECD 1999), this could be realized at the level of genre and text type and less at the level of topic. Topics should be closer to 15-year-olds than the warranty of a refrigerator or a company's job centre, to mention a few.

9.2 Theoretical and practical implications

Text authenticity as a group-level phenomenon

In the present study, text authenticity has been studied within the context of international reading literacy assessments, which has its own special demands and restrictions related to authenticity. The multicultural context of international reading literacy assessments as well as the requirements of equity and fairness in testing necessarily restricts the selection of authentic texts that are culturally relevant and typical, not to mention texts of individual preference. The texts selected for an international reading literacy test have to be authentic texts in the sense that they are acceptable, relevant and typical in all the participating countries and cultures. In an assessment such as PISA 2000, in which more than 30 countries participated, this results in selection of texts that are of global relevance and interest and also relatively neutral, without upsetting or stirring the passions of the readers. Within the restrictions of this particular context, text authenticity necessarily becomes a group level phenomenon with a group of test-takers of a certain age in focus and perhaps one national textual culture under scrutiny at a time.

However, text authenticity is ultimately an individual level phenomenon since it is something defined in the end by the individual reader (Bachman 1990; Widdowson 1979; Bachman & Palmer 1996; Garcia & Verville 1994; Weiss 1994). Nevertheless, individual readers are always in relation to the larger community (see e.g. Taylor 1992). Thus, although in the literature authenticity is seen as individual phenomenon, a matter of individual choice, it is not unreasonable to study it also at the group level since readers act as a part of cultural community of readers (see e.g. Taylor 1992). This view is consistent with social views of reading literacy that stress that the reading events of individual readers reveal and reflect practices that are socially and culturally patterned ways of using texts and literacy (see e.g. Barton 1994). This means that literacy practices are not only individual in nature. At the same time as individuals make use of their literacy practices, practices connect people, exist in the connections between people and involve shared cognitions that are represented in ideologies and social identities. (Barton & Hamilton 2000.) In a similar way, text authenticity may be socially and culturally patterned and

shared: shared textual experiences result in similar texts being authentic to a group of readers sharing the textual experience in common.

The issue of validity

In addition to the individual and group level dimensions of authenticity, there are also other apparently contradictory dimensions related to text authenticity in this particular context. One of them concerns the issue of validity. The demand for authentic texts in a reading literacy test is ultimately related to the issue of validity. Authentic tests reflecting real-life contexts and tasks are considered valid especially in terms of coverage since they can capture the whole range of students' competencies in actual real-life tasks (e.g. Huerta-Marcias 1995; Shohamy & Reves 1985; Wood 1993). Thus, authenticity is related to construct validity since the test results can only be generalized beyond the testing situation if the test reflects language use outside that situation (Bachman & Palmer 1996). Authenticity is seen as an important aspect of content validity as well, since the authenticity of test tasks can be seen in terms of their relevance to natural language use situations (Bachman and Palmer 1996; also Wu & Stansfield 2001). Face validity can also be regarded as relevant to test authenticity since it concerns the appearance of the test based on superficial examination (Linn 1993; Linn & Gronlund 1995). As authenticity is considered important from the point of view of the motivation of test-takers (e.g. Widdowson 1990), face validity is considered important since it helps to obtain the co-operation of the test-takers (Linn & Gronlund 1995).

On the other hand, the demand for fairness and equity, which is also related to test validity from the point of view of the comparability of the results, restricts the authenticity of texts. If the results of the test are to be compared internationally the test should cover only texts that are equally authentic or inauthentic to all. Due to this requirement individual and national preferences need to be overlooked, which results in the test "covering the transnational core" (Hamilton & Barton 1999) of reading literacy. The validity of the test is thus decreased since the test results are to be generalized only to transnational reading literacy events. The results cannot, strictly speaking, be generalized in reading literacy events that are relevant in a national or individual context alone. This limitation, however, is fully recognized within international reading literacy assessments, such as PISA 2000, that are not claimed to assess anything other than performance in the transnational reading literacy tasks.

Overall, in designing and compiling a reading literacy test that consists of authentic texts and tasks it is crucial to find a balance between what is authentic to individuals and what is authentic to a group of people, e.g. Finnish 15-year-olds. On the other hand, a balance must also be found between the demands for an authentic test covering a wide range of real-life reading events and the demands for equity and fairness of testing which are necessary for international comparison of results. Based on the results of the present study it can be cautiously concluded that the PISA reading literacy test validly does what it aims to do, that is assess reading literacy performance in transnational tasks,

but it somewhat lacks validity in the national culture of Finnish 15-year-olds. This is reflected in the low authenticity of the PISA texts in terms of typicality and interest as evaluated by a group of Finnish 15-year-olds. In particular the low interest of some of the PISA texts reveals that the selection of texts does not cover what is most meaningful to these 15-year-olds and overall it fails to cover the full range of authentic reading literacy events in their lives. The role of electronic reading, for instance, was almost completely neglected in the PISA 2000 test and yet Finnish youngsters use e-mail and the Internet frequently (Leino 2005), encountering different kinds of texts on various topics in electronic form. Luckily the international option for the assessment of electronic reading available in the PISA 2009 gives the opportunity to include also electronic texts in the assessment.

Text classifications in designing a reading literacy test

In the PISA assessment framework texts were primarily classified by format and text type, relying mainly on the work of Werlich (1983) (OECD 1999). The text typology of Werlich aims to describe universal text types whose structural differences derive from the forms and ranges of human cognition. Each text type is described in terms of a typical surface structure, i.e. sentence types and the relations between sentences in the text. Thus, the main attention in text type classification lies in the linguistic features of the text. The PISA assessment framework, however, also incorporates some functional or contextual criteria in its classification of situations for reading. Situations for reading are primarily defined in terms of the writer's intended use for the text but also other people and in addition the typical content related to the reading situations are taken into account. (OECD 1999.) In this study the readers found the functional characteristics of texts more significant than the linguistic criteria, a point which was reflected especially in the results of the pilot study.

There are problems with the text classifications used in PISA, both practical and theoretical. First of all, taking the writer's intended use of the text as a basis for text classification by reading situation may be practical but theoretically unsound. After all, the current social views of reading that PISA aims to reflect stress that the meaning is neither the writer's nor in the text but rather something constructed by the reader through social practices in the context of the text, the reader's prior knowledge and experience and the social setting. Texts reflect the culture and context around them but are also shaped by them. Texts are seen as mediators in social practices rather than carriers of determinate meanings. (See e.g. Rosenblatt 1994; Barton 1994; Barton et al. 2000; Gee 1990, 2000a, 2000b, 2003.) In the end, writers have very little say as to how and why readers read their texts, as readers are free to approach the text according to their own needs and purposes (see e.g. Rosenblatt 1994). Text classification by reading situation as done in PISA 2000 would be theoretically sounder if the starting point was the reader's purpose in reading the text rather than the writer's intended use. This would undeniably result in difficulties in

classifying texts, since the individual readers' purposes may vary with the same text, but the classification used in PISA was not able to avoid these difficulties either. With any classification of texts generalizations are needed and thus individual preferences need to some extent to be overlooked.

Secondly, the four reading situations specified in PISA were said to be based on the intended use of the texts, on the relation to others in the reading task and on the general contents of the reading material (OECD 1999, 23). Private reading, for instance, is described as reading carried out to satisfy the practical and intellectual interests of the individual. Educational reading, on the other hand, is reading for "acquiring information as part of a larger learning task". (OECD 1999, 23-24.) These descriptions do not seem to mention the intended use of the text but the general reasons for reading the texts as well as the context of reading. Thus, the use intended by the writer, which is said to be essential in the PISA text classification may in fact have a more marginal role. On the other hand, both approaches usually lead to the same classification. The movie reviews, for instance, were classified in terms of a private reading situation as was the article on the quality of men's shirts. Both can be and probably are read to satisfy individual interests, out of curiosity or to gain information on a topic of concerns. With these texts the intended use of the texts as well as the reader's reason for reading are usually the same. The article about the police using DNA in their investigations, on the other hand, was classified as an educational text. The article about police work certainly is very informative but so are other articles chosen for PISA, and they could all be read to satisfy personal curiosity without being part of a learning task such as educational texts are defined to be. Furthermore, one might ask how the intended use of the article on police work differs from that of the article on protecting beaches or the newspaper article about NASA tests on drugged spiders that were classified as public texts. Although the classification of texts into four situations for reading seems straightforward and clear the basis of it is theoretically problematic and to some extent so in practice as well.

Thirdly, apart from the theoretical problems and the controversial nature of text classification by reading situations, there are problems with the text type classification of the PISA texts. Classifying whole texts into text types in terms of the linguistic criteria defined by Werlich (1983) is problematic since whole texts rarely represent purely only one text type. Authentic texts are often heterogeneous and consist of sections that may have different functions. Thus different text types are used in different sections of the text. (See also Virtanen 1992.) Furthermore, there have been difficulties in distinguishing text type based on linguistic criteria and text classification based on functional criteria in PISA. For instance, the motorbike advertisement consisting of a poem in which the narrator tells how he would live his life again if given the chance to do so and a photograph of an old man was classified under the argumentative text type in PISA. Clearly, based on the function of the text it is argumentative. However, based on Werlich's linguistic criteria which the PISA classification of text types claims to follow, the text type used in the poem is narrative. Using

the terms of Virtanen (1992), the discourse type defined in terms of the function of the text is argumentative and the text type is narrative. In authentic texts it is relatively common that discourse type and text type do not coincide and, for example, the narrative text type may be used for purposes other than narration, often for argumentation or instruction (Virtanen 1992).

In view of these considerations it seems that the text classification used in PISA needs some clarification in terms of its theoretical basis as well as in terms of the relations between functional and linguistic criteria. Certainly the functional criteria that have already been used, as indicated above, need to be brought openly to the classification of texts. Text classification based on genre, which is defined in terms of functional and communicative criteria (Swales 1990; Bhatia 1993; Eggins & Martin 1997) would benefit international reading literacy studies in many ways. As the results of the present study suggest that genre is a textual feature relevant to readers from the point of view of text authenticity, classifying the texts to be selected for a reading literacy test by genre would also facilitate covering authentic texts in the test. It would be theoretically more consistent and more authentic to classify texts by genre which in fact covers many textual features, such as language used and layout. Furthermore, genre would be a useful concept for studying and describing cultural phenomena, or cultural authenticity, in texts. This does not, however, diminish the importance of linguistic analysis and text classification in selecting texts for a reading literacy test. Within genre approach linguistic analysis could be done as subordinate to genre of the text.

Based on the results of the present study as well as the observations on the PISA text classification set out above, the following framework is suggested as a basis for selecting authentic texts for international reading literacy tests (Figure 5). The framework has two continuums: one of time and the other for individual and cultural dimensions. The textual features used in the framework are topic and genre, which covers also extralinguistic features of texts, such as layout and context of appearance. The attributes of text authenticity in the framework are interest, relevance and typicality. Interest has individual emphasis and is usually related to the topic of the text. From this point of view it is essential to ask which topics are of interest to the reader. Typicality of text, on the other hand, is seen as a cultural phenomenon and is related primarily to the genre of the text. That is, the essential question to be asked is: which genres are – or will be–typical in the culture of this particular group of test-takers. Relevance, on the other hand, is an attribute of authenticity that can be either an individual or a cultural phenomenon. As relevance refers to the need and purpose the reader has for reading the text, it can be interpreted as deriving from the individual reader's personal interests and purposes in reading. On the other hand, the relevance of the text may also derive from culturally typical reading tasks that society imposes on its readers. The question to be asked here is: which genres do the reader need, i.e. what kind of functions are there for texts in the society/culture in question. The time dimension from the present time to the future emphasizes that it is important to consider texts that are of

172

interest and especially relevant and culturally typical not only currently but in the future lives of test-takers.
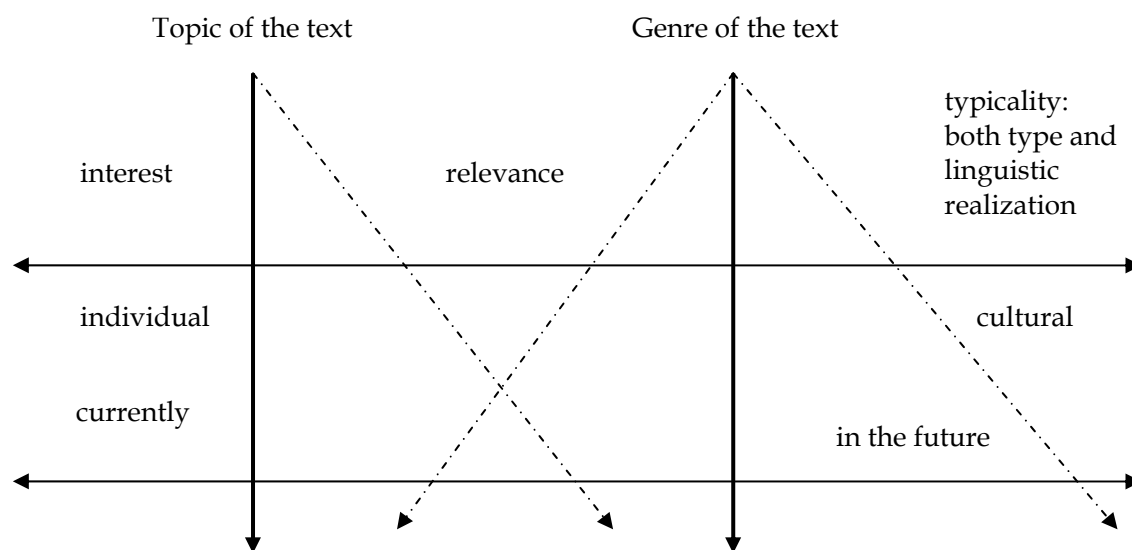
FIGURE 5   Framework for text authenticity in international reading literacy tests.

It should be noted that the continuums for the individual and cultural dimension and the time dimension of the framework do not necessarily correlate although the parallel lines in the figure might suggest otherwise. That is, although the topic of the texts is primarily related to the individual reader's current interests, this is not the case with all texts in every reading event. There are also texts with a topic of individual interest from the point of view of the future or topics of cultural interest. Topics of future interest, however, may also be considered relevant and thus interesting. This was the case, for instance, with some occupational texts in the PISA 2000 reading literacy test. The texts about applying for a job (the application form and the instructions for a job interview) were not considered very interesting right now but some students realized and stated that they contained information they would need in the future and for this reason the texts were considered to some degree interesting. Furthermore, although in the framework the genre of the texts is primarily related to socially relevant and culturally typical text types *now and in the future* this is not to exclude genres of individual interest. Individual reader's interests may also include certain genres, not only topics. Some readers, for instance, may find poems or fairy tales of special interest. The framework is to be read bearing in mind that ultimately it represents one way of classifying texts and is therefore necessarily characterized by compromises and generalizations.

In practice, the framework suggests that in selecting authentic texts for an international reading literacy test the main focus should be on the topic of the text and the genre of the text. Genre, which is defined in terms of purpose of communication (Swales 1990; Bhatia 1993; Eggins & Martin 1997), also includes

extra-linguistic features such as layout and context of appearance. The topics of the texts should be reviewed primarily from the point of view of individual readers' current interests or future relevance. Since international reading literacy assessments are large-scale studies the individual reader in this case refers to the prototype reader(s) representing the target-group of the study. Genre, on the other hand, should be reviewed primarily from the point of view of the test-takers' community and culture: firstly, genres chosen for the test should represent genres the command of which is required of the test-takers in real life either now or in the near future, and secondly, these genres and their linguistic realization should be typical in the national culture of the test-takers.

These two requirements in reviewing genre are practically two different sides of the same coin since texts a reader encounters are related to reading tasks typical to the culture and thus the texts are typical as well. Relevant texts become typical because such texts are read repeatedly. The emphasis here, however, is somewhat different since the notion of genres that are *culturally typical* refers also to the linguistic realization of the genre in the particular national culture. Since there is cross-cultural variation within the same genre common to e.g. industrialized cultures (see Bhatia 1993; Swales 1990) it is also necessary to review the organization of the genre as well as the linguistic realization of the moves (Swales 1990) of the genre in the participating countries when selecting texts for international reading literacy tests. In this way cultural typicality of the genres is ensured but also the equity and fairness of the selection of reading materials in the test. If the genre of a text is simply noted to be typical in the cultures of the participating countries, the fact may be overlooked that the linguistic realization of a common genre, e.g. the newspaper article, is different in different cultures (see also Harjunen 2007). This would result in an unequal and biased test. The texts – and the test – are hardly equivalent if some encounters, say, a news report in a typical linguistic realization and others read the same familiar genre in an untypical linguistic realization.

In selecting texts for an international reading literacy test it is essential that the texts are reviewed by individuals representing the target group of the test (see e.g. Wu & Stansfield 2001). Since text authenticity can be regarded as a cultural but also as an individual phenomenon, it is very difficult for those outside the sub-culture of the particular test-takers to assess text authenticity. Therefore it is to be recommended that the insight of the test-takers themselves should be brought in to the process of decision-making when texts are selected for the test.

Union and intersection

The clear difference between the textual worlds of the two genders shown by the results of the present study is hardly surprising. Gender difference in reading literacy performance as well as in the reading habits and interests of girls and boys has been verified in several studies of reading literacy, as it has in PISA (Elley 1992; Linnakylä 1995; OECD 2001; Linnakylä et al. 2002; OECD

2002; Roe & Taube 2003; Linnakylä et al. 2005). This difference has – or should have–its implications in selecting the texts for a reading literacy test.

Within the context of international assessments of educational achievement it is often questioned whether to limit the assessment to the intersection of the curricula of the participating countries or to cover the union of the curricula of the participating countries (Linn 2002; Porter & Gamoran 2002). The emphasis is usually on the national cultures of the countries participating in the study. The same question, however, could be posed emphasizing the two genders. It should be noted that in the reading literacy assessment, as in all the assessment areas in PISA, the test design was not related to school curricula. Test design was based rather on a common understanding of the knowledge and skills that students need in their lives now and in the future (OECD 1999). The question of limiting content coverage should thus be posed in terms of the textual culture of the target group under assessment: should the test cover the intersection or the union of texts authentic in the test-takers culture.

In terms of gender, the test should include texts authentic for both genders but also texts that are authentic for boys and for girls separately. Opting for the intersection of the textual worlds of girls and boys would mean that only texts authentic for both genders would be selected for the test. Since there are clear differences in which texts are authentic for boys and for girls this would necessarily result in too narrow a sample of texts from the point of view of the construct to be assessed. According to the definition of reading literacy adopted in PISA 2000 a wide selection of real-life texts and reading literacy tasks should be covered in the test. This would not be necessarily the case if only texts authentic for both genders were selected.

Covering the union of texts authentic to girls and to boys separately, on the other hand, would probably be too expansive to be practical. Therefore it is suggested in this study that the reading literacy test in international assessments should include texts authentic for both genders (the intersection) but also a sample of texts authentic for girls and a sample of texts authentic for boys. In this way the test would be valid since it would include a wide selection of texts, which is consistent with the construct definition in the case of PISA, while it would also be fair and equal for both genders.

One aspect of test authenticity that requires attention here is the way authentic texts should be used in testing. It is essential that the tasks related to authentic texts in a reading literacy test should also be authentic, i.e. simulating the real-life use of texts. Task authenticity is excluded from the focus of the present study but in the test design phase of a reading literacy assessment programme it is important to review also task authenticity. One could even say that ultimately it is the tasks simulating the use of texts that define the text authenticity since in testing situation they create the need for reading the text. The genre approach which this study suggests should be included in the process of selecting authentic texts in part covers also the use of the texts, as genre is defined in terms of the purpose of the text (e.g. Swales 1990). Thus, in

reviewing task authenticity it is recommended that the tasks related to the text should be compared with the real-life use of the text and other texts of a similar kind (see e.g. Bachman & Palmer 1996).

In an international reading literacy assessment such as PISA that is not based on curriculum analysis, the question of the content coverage of tests is related not only to the sub-culture of genders as described above but also to national cultures as a whole, that is, including the school curriculum and culture but not limited to it. This is related to the question of culturally typical genres mentioned in the framework for text authenticity in international reading literacy assessments (Figure 5). It is imperative that the texts selected for international reading literacy assessments should be equally authentic, e.g. typical, in the national cultures of participating countries. In practice, the texts selected for these assessments derive from the intersection of the range of texts of each participating country and its culture at best and thus the test covers only the cross-national core of literacy (Hamilton & Barton 1999). However, it has been argued that the intersection rarely is the reality and that test content and item-formats are biased towards Anglo-American culture (see e.g. Rowan 2002). In addition to the common texts (the intersection) there have been suggestions that international assessments should cover also country-specific content (see e.g. Porter & Gamoran 2002), in this case texts and tasks. Country-specific items, however, would be very difficult to include in a comparison of results due to concerns about bias. Still, countries participating in e.g. PISA need to be active in submitting test material, both texts and tasks, to the organizing consortium in order to avoid the domination of only one culture.

Overall, in attempting to cover a wide selection of real-life reading situations and thus also real-life texts it is always necessary to make compromises since it is impossible to cover all possible situations. The construct definition of each reading literacy study as well as the assessment framework is the starting point of test design in describing what is measured (e.g. Luoma 2001). Since boys and girls have at least in part such different textual worlds test designers need to approach the selection of texts also in terms of gender, and in international assessments in terms of national culture. It is suggested in the present study that a compromise between the intersection of texts common to both genders and the union of all texts authentic to either gender should be found. Texts found authentic in common should form the core of the reading literacy test but the test's fairness and equity in gender terms should be ensured by including also texts authentic for girls and boys separately. In terms of national cultures, covering the intersection texts – genres – common to the participating countries is challenging enough in test design. Moreover, covering the union of the texts authentic in all the countries participating in the test would be impractical, and in any case country-specific items would be probably biased towards the originating country's respondents. Thus, the results based on these items would be difficult to include in a comparison of the results. It is noteworthy that the test should always reflect the construct and the assessment

framework and thus any compromises in test coverage should be the result of careful consideration.

Does text authenticity matter in reading literacy assessment? And to whom?

This question is provoked by the study of Lewkowicz (1997; 2000), who suggested that test authenticity may be more important to researchers than to test-takers. In her study, Lewkowicz gave students two different language tests: one was a 90-item multiple-choice test presumed inauthentic by the researcher and the other an English for Academic Purposes (EAP) test perceived as "reasonably authentic" (p. 53). After the tests the students were asked about their perceptions of the advantages and disadvantages of the test type and their opinion on which test was a better indicator of their ability to use English in an academic context. Relatively few students in the study referred to issues related to authenticity in describing the advantages and disadvantages of the test types. The tendency was for students performing in the top third to be more likely to find the EAP test better at assessing their ability to use English in an academic context. Lewkowicz concluded that authenticity is not necessarily important to test-takers. However, authenticity in Lewkowicz's study was defined merely as the "real-life resemblance" of the test tasks although the interactional component of authenticity was acknowledged in her review of the background literature. There were students in her study who referred to the question of interest (the multiple-choice test not being interesting), which can be interpreted as an authenticity-related issue as has been done in the present study. The interest of test material is likely to enhance genuine and authentic interaction between test and test-takers and therefore should be regarded as an attribute of authenticity (see e.g. Bachman 1990; Bachman & Palmer 1996).

Based on Lewkowicz's (2000) results it is impossible to draw conclusions on the importance of the authenticity of a test to test-takers. Neither does the present study support conclusions on that subject since the focus of the present study was restricted to text authenticity and the *importance* of text authenticity *to test-takers* was not studied. As Spolsky (1985; see also Widdowson 1990) points out, test-takers tend to play along in a test situation, and therefore it is quite possible that they accept also less authentic tests and carry out the test tasks accordingly. As Lewkowicz (2000) herself states, tests consisting of inauthentic language material may still be treated as authentic tests.

From the point of view of the test-takers it has been suggested that test authenticity improves motivation (Widdowson 1990) and is potentially linked to performance in the test (Bachman & Palmer 1990). Empirical evidence for the latter may be difficult to find but nevertheless authenticity is important from the point of view of at least the face validity of the test. Thus Lewkowicz may be right in suggesting that authenticity is more important to researchers than test-takers, but so it should be. From the point of view of the present study it is essential that authenticity is taken to be an important aspect of test construction by test-designers and researchers in order to ensure test validity and thus the

generalizability of the test results, not to mention fairness and equity in testing (Bachman 1990; Bachman & Palmer 1996; Huerta-Marcias 1995; Shohamy & Reves 1985; Wood 1993). Valid tests also ultimately benefit the test-takers although not necessarily by directly improving their test performance. This is the case especially in an assessment such as PISA that assesses students' performance in and for real-life contexts.

Pedagogical implications

The present study covers themes that are relevant to instructional contexts as well. As described in Chapter 4.2, the aim of authentic literacy instruction is to enculturate students in the culture of reading and writing and the practices involved. Social views on literacy have stressed the different uses and functions different texts have and resulted in demands for the use of authentic texts in literacy instruction (e.g. Pitkänen-Huhta 1999; Street 1996). Authentic texts in reading literacy instruction familiarize students with the large variety of literacy activities associated with the large variety of different purposes and texts. During such a process of enculturation students learn to know different texts and at the same time the contextualized uses of the texts together with the discoursal rules and literacy practices intertwined and reflected in them. The essence of such instruction is in preparing students for real-life challenges beyond the school context (see also Luukka 2004b).

The present study encourages the recognition that text authenticity is ultimately both an individual and a socio-cultural phenomenon. Therefore, it is important in instruction to take into account different interests and preferences students may have in terms of texts. It is especially important to acknowledge that there are important everyday texts that are common to all young readers, such as newspaper articles, but also texts that are not necessarily equally authentic for both genders. Particularly boys, who do not adjust to school discourse as easily as girls (e.g. Brozo 2002; Smith & Wilhelm 2002; Taylor 2004), need to have successful learning experiences in their mother tongue. Since overall in Finland boys' self-confidence as learners of the mother tongue and in reading literacy has been found to be lower than that of girls (Välijärvi et al. 2002), it must be stressed that boys also need positive learning experiences to gain greater self-confidence and joy in reading. One crucial way of supporting boys' literacy is that boys themselves should have the opportunity to choose texts within their own range of interests (Taylor 2004; Luukka & Leiwo 2004). Boys are for instance interested in non-fiction texts, such as the newspaper and magazine articles that they frequently read. Therefore, it should be acknowledged that boys do read plenty of diversified reading materials but that this type of reading is not necessarily valued at school. School reading is inauthentic for boys especially in the sense that they do not see the relevance of the literacy activities of school to real-life contexts. The school view of literacy cannot remain isolated from the lives of the students but it should be widened to appreciate literacy practices and events in out-of-school contexts as well (e.g. Brozo 2002; Luukka 2004a). This means acknowledging that boys do actually

read a great diversity of texts for various purposes and that they do thus show a competence in reading literacy in contexts and for purposes relevant to them. Like all readers they choose texts according to their needs and interests. Unlike the case with girls, these needs and interests are not necessarily considered important at school but they are relevant, that is to say authentic, to boys.

The new national framework curriculum implemented in Finland since 2004 provides good opportunity to widen the conception of texts used in mother tongue instruction. In the opening chapter of the mother tongue curriculum it is stated that the conception of text behind the framework curriculum is wide and includes spoken and written texts as well as fiction and factual texts, texts consisting of pictures and photographs, audio material, graphic presentations and any combinations of these (FC 2004, p. 23; see also Luukka 2003). Implementing this conception of texts in practical instruction makes room not only for texts that include the visual elements favoured by boys (e.g. Taylor 2004) but also for non-continuous documents such as commonly appear in newspapers and magazines and also typical in work related literacy events (see Linnakylä et al. 2000).

One type of texts increasingly significant for Finnish youngsters is electronic texts that appear, for instance, on the Internet and in e-mail or SMS messages. Electronic texts combine several meaning making systems (Cope & Kalantzis 2000) in the same textual entities and thus require multiple skills of interpretation. Including electronic texts in reading literacy instruction brings real-life relevance to the classroom, especially for boys. Real-life relevance in this case benefits also girls, who can easily become absorbed in the world of fiction and thus out of touch with the textual skills required in real-life contexts, e.g. work life (e.g. Leiwo 2005). The idea of authentic reading literacy instruction is not, however, to use only texts that are already typical and thus familiar in students' experience. Rather the idea is to use authentic texts as a bridge to inauthentic texts in widening students' textual horizons (see e.g. Kauppinen 2005). Furthermore, in learning and developing new skills authentic texts may motivate students.

To conclude, emphasizing the socio-cultural aspect of text authenticity in reading literacy instruction in practice means that students face in instruction a large variety of real-life texts and thus different kinds of texts, their purposes and the literacy practices typical in their society become familiar to them. On the other hand, the individual aspect of text authenticity in instruction stresses that students' own interests should be the starting point in the process of enculturating them into the world of texts and literacy practices prevailing in the culture of their society.

9.3 Assessing the quality of the study

In this chapter, the validity of the present study is examined. As the study employs a mixed methodology in which qualitative and quantitative data analysis are used in the same study (Teddlie & Tashakkori 2003), the criteria for evaluating the quality of the study used in both quantitative and qualitative approaches have to be taken into account. Since the aim of the study is to discover and describe rather than test a theory or hypothesis (Morse 2003) the qualitative approach is dominant, despite the strong presence of quantitative analysis in the main study phase.

The quality of the study can be evaluated in terms of its validity which in general refers to the adequacy and appropriateness of the interpretations and use of the results. Validity is currently seen as a unitary concept emphasizing the consequences of the study and should always be evaluated with regard to a specific use of the results. (Linn 1993; Linn & Gronlund 1995; Messick 1989a.) Despite being a unitary concept, validity is often evaluated on the basis of various kinds of evidence. Evidence has to be obtained covering content, the relationship of the results to other measures, the construct and the consequences of the uses and interpretations of the results (Linn & Gronlund 1995). Furthermore reliability, defined as the consistency of the results, is considered a necessary, albeit in itself insufficient, condition for validity (e.g. Linn & Gronlund 1995).

Validity and reliability are typically characteristics used in evaluating quantitative studies. In evaluating studies employing qualitative methods, different criteria of varying kind are used. It is noteworthy, however, that quality criteria within the qualitative approach echo to some extent concerns similar to those of the validity criteria within quantitative approach. The labels for the criteria as well as the emphasis, however, differ. One set of criteria that help to evaluate the quality of conclusions within the qualitative approach are *objectivity/confirmability*, *reliability/dependability*, *credibility/authenticity* (also referred to as internal validity), *transferability* (external validity) and *utilization/application* (Miles & Huberman 1994). Since the present study is dominated by the qualitative approach, it is evaluated employing the above criteria. However, the validity and reliability issues characteristic of the evaluation of quantitative studies are emphasized when appropriate.

Objectivity/confirmability

The objectivity/confirmability of the study refers to the transparency of the study process. Furthermore, the position of the researcher as well as possible biases requires to be explicated. (Miles & Huberman 1994.) This criterion thus emphasizes the quality of description of the procedures used in the study, since the transparency of the procedures followed derives to a great extent from an open and detailed description of the study and all of its phases (Miles &

Huberman 1994). This not only improves the quality of the study but also helps the readers to understand its orientation and to determine for themselves its quality.

It has been attempted in the present study to meet the demands set by this criterion by *explicit and detailed description and reasoning* for each phase of the study. First of all, the *context of the study* is described in the 'Introduction' section in order to make clear to the readers the exact focus of the study with its emphasis and limitations. The decisions regarding the methods chosen were to a great extent affected by the context and aims of the study. Secondly, the *research questions* were precisely set out in the early phase of the thesis. The research questions were restated in the 'Research questions and methods' section in which the methods of the study were also described and the reasons behind the decisions made explained. The aim of this section was to *provide readers with all the relevant information* regarding the choice of methods used as well as the implementation of the methods. The data gathering was therefore described in detail, including the development of the questionnaires and information about the respondents as well as the instructions given to the students answering the questionnaire in the main study. Examples of the questionnaires used in the pilot and main study phases were given in Appendices 1 and 4 in order to give readers themselves a chance to determine the adequacy and appropriateness of the questionnaire. The data analysis was described in detail especially for the pilot study, in the case of which content analysis of the panellists' comments was used. Examples of typical comments were included to illustrate the data.

To give readers a more comprehensive picture of the data as well as the analysis and categories found in the content analysis, an example sheet of the content analysis was provided and the categories found in the analysis were set out in Appendices 2 and 3. The results of the content analysis were reported and extensively illustrated with examples from the data. This was done in order to give readers the opportunity to evaluate the correctness of the content analysis. Since the selection of the four textual features for the main study questionnaire was made on the basis of the pilot study as well as the literature review, the considerations behind the decision were fully set out for readers. As the criterion of objectivity/confirmability demands explicit description also in terms of "backstage" information (Miles & Huberman 1994, p. 278), the practical considerations relevant to the decision about the four textual features, e.g. the age of the main study respondents, were made explicit in the study report and discussed in terms of the theoretical framework of the study.

The description of the procedures used in the study also included explicit discussion of the *role of the researcher* found relevant in Miles's & Huberman's (1994) criteria for both objectivity and reliability. The researcher was aware of her prior knowledge and experience necessarily effecting the categories found in the content analysis. Furthermore, the description of the pilot study panellists was detailed and included discussion of the possible effect of the panel consisting only of females on the results of the panel review thus showing that

researcher was *aware of the possible sources of bias* and has weighed the bias issues.

On the other hand, there were also parts of the description of the study that could have been improved. For instance, the group of 15-year-olds answering the main study questionnaire could have been described in more detail as some background information was gathered and could have been used. Furthermore, the calculation of the typicality and interest indices for each text were described rather briefly in the study. This part of the study could have been elaborated further.

Reliability/dependability

Reliability/dependability as a criterion of quality is used to evaluate consistency and carefulness in the conduct of the study. To meet this criterion the research questions must be clearly stated and methods of inquiry chosen congruent with them as well as connected to theory. Additionally, the researcher's role and status must be explicitly described, and analytic constructs clearly specified. Furthermore, this criterion requires that data be collected using appropriate methods (settings, time, respondents etc.) and coding checks made. (Miles & Huberman 1994.)

As noted above, the research questions were explicitly stated right at the beginning of the study as they derive from the context and aims of the study described in the 'Introduction' in Chapter 1. Furthermore, the research questions were restated in the 'Research questions and methods' section of the study. In choosing and refining the methods, *coherence between theory, research questions and methods used* (Miles & Huberman 1994) was essential. For instance, both in the pilot and main phases of the study methods of data gathering were chosen bearing in mind the value and status a reader gives to a text, since text authenticity can validly be determined only by readers themselves (e.g. Widdowson 1979). This was to emphasize that typicality and interest are in this study taken to be attributes of authenticity, not qualities related to texts only. For instance, in developing the pilot study questionnaire that was used as a basis for the panel review, the panellists first reviewed the authenticity of the PISA texts in terms of their familiarity to 15-year-old students (the PISA target population). Since on the basis of both practical experience and theory it was argued that readers are most competent to judge text authenticity from their own perspective (e.g. Widdowson 1979), in the pilot study itself the panellists reviewed the authenticity of the PISA texts from their own perspective in terms of the texts' typicality in their own lives. This approach to reviewing text authenticity was considered more valid since it reflected the theory better.

Previous research as well as the literature on methodology (e.g. Banerjee & Luoma 1997; Bachman 2000) were referred to in explaining the usefulness of the methods chosen for the specific purposes of the present study. There were also practical considerations in choosing and developing the methods of data gathering. For instance, the panel review by 8 expert readers was carried out on the basis of a questionnaire rather than e.g. an interview, in order to give

panellists a chance to respond at their own pace after finishing the coding of the open-ended responses of the PISA reading literacy test. The main study data, on the other hand, was gathered using a questionnaire because the number of respondents (240) made the possibility of interviewing the students impractical. The aim was to develop methods of data gathering that were theoretically sound and yet practical.

Reliability/dependability requires that constructs are clearly specified (Miles & Huberman 1994). The core construct studied in the present study is text authenticity. Since the study concentrates on text authenticity in the very specific context of international reading literacy assessments, the limitations of the context necessarily affected the way the construct was defined and operationalized. The construct of text authenticity was discussed and defined in the literature review. The construct was operationalized as typicality and interest in the study since as such 'authenticity' would not have been accessible to Finnish readers, at least not to 15-year-olds. The degree to which operationalizing the concept in terms of typicality and interest was successful in achieving content validity is to be evaluated below. However, the aim has been once again to find a solution that is theoretically sound yet practical. Practicality was considered essential in ensuring the accessibility of the concept of authenticity to the respondents participating in the study.

In the pre-pilot phase of the study, text authenticity was reviewed by individual readers but only from the point of view of text typicality. 'Typicality' was chosen for use in the study since it captures two essential aspects of text authenticity: that of being typical and familiar in the reader's culture and that of resembling real-life texts. (See e.g. Valencia et al. 1994; Harris & Hodges 1995.) However, since it was found that another essential part of authenticity is overlooked if only 'typicality' is used, another additional attribute of authenticity was chosen in order to operationalize authenticity in the main study phase. This attribute was 'interesting'. As made evident in the literature review presented in Section 4.1, authentic text is not only typical in the reader's culture but also engaging, rich and interesting for the individual reader (Garcia & Verville 1994; Kapinus et al. 1994; Weiss 1994). As expressed by Widdowson (1990), reader authenticates a text by taking an interest in it. Since typicality and interest as an attributes of a text may not and need not coincide, it was seen as important to include them both in the main study questionnaire presented to students. However, both 'typicality' and 'interest' should have been used already in the pilot study questionnaire since the textual features mentioned in reviewing text interest might have been different from those mentioned in reviewing text typicality, or at least the emphasis might have been different. This weakens the consistency of the study, since the operationalizing of text authenticity was not consistent in different phases of the study. This inconsistency affects the reliability of the study and necessarily affects the credibility of the results.

As for reviewing authenticity within the context of the present study is concerned, the contradiction between authenticity as an individual

phenomenon and in large-scale assessments as a group-level phenomenon was discussed in Chapter 9.2. It has been argued that it is possible to study text authenticity as a group-level phenomenon since individual readers act as a part of cultural community of readers (see e.g. Taylor 1992). Just as with reading literacy events and practices (see Barton 1994), text authenticity may also be socially and culturally patterned and shared: shared textual experiences result in similar texts being authentic to a group of readers sharing the textual experience in common. Thus, there is theoretical support for the decision to study text authenticity as a group-level phenomenon.

In terms of theoretical consistency, however, there is a limitation in the present study that has not been explicitly discussed in previous chapters. Narrowing the focus of the present study on text authenticity without the contexts of tasks attached to the texts is somewhat inauthentic itself. It has been acknowledged that texts are a form of language use (e.g. Karvonen 1995) and in a testing situation the tasks define the way texts are used. Thus it can be argued that what makes a text authentic in this context is the way it is used – i.e. tasks. Therefore, in reviewing and studying text authenticity a more authentic and valid approach is to include the context of use, in a testing situation the tasks, in the process as well.

The description of the procedures used in the study also included explicit discussion of the *role of the researcher*, which is found important in Miles's & Huberman's (1994) criteria for both objectivity (above) and reliability. The two requirements are somewhat overlapping since they both suggest explicit statement of any possible bias introduced by the researcher. (Other types of bias will be considered in evaluating objectivity.) In the present study, the researcher was aware of her prior knowledge and experience necessarily affecting the categories found in the content analysis during the pilot study. Moreover, some categories such as those of topic and content and of language and style were in part rather close to each other. Based on the judgment of the researcher these categories were kept apart but other analysts could have made a different decision. Furthermore, the role of the researcher was crucial in the main study, since the instructions and examples given by the researcher to the students answering the questionnaire might have had an effect on the way they interpreted the questions. On the other hand, in the case of the textual features studied the effect was found positive in ensuring consistency in students' interpretations of the meaning of the textual features. This was not, however, explicitly stated in the study report and thus weakens the objectivity/confirmability of the study.

As already stated, it was aim of the researcher to explain methodological decisions in terms of theoretical consistency but also taking into account practical considerations that were usually due to the age of the respondents in the main study phase. The operationalizing of text authenticity was affected by similar considerations. The selection of the textual features for the main study questionnaire was based on the content analysis in the pilot study phase and on the theoretical considerations presented in the relevant literature but also on

practical considerations. The textual features chosen for the questionnaire had to be accessible to students at the age of 15 in order to give them a chance to review the textual feature within the given time of 45-minute lesson. In this time students had to skim on average 3 or 4 texts and response to questions. It was obvious that some students would have needed more time, since not everyone completed the questionnaires. This was clearly a weakness in terms of the reliability of the study. Another weakness relating to the reliability of the data gathering was the fact that two groups of students were instructed by a person other than the researcher, who herself gathered the data on the other occasions. Due to the illness of the researcher a substitute took over the data gathering for the two final sessions. Although thorough guidance was given to her concerning the instructions to be provided to the students, the quality of the data obtained was somewhat different in these students' questionnaires. This may also have been due to the timing of the session– the last lesson of the week on Friday afternoon. However, it would have been wiser to reschedule the meeting with these students.

Furthermore, evaluating the conduct of the study afterwards shows that it would have been wise to collect more data and thus increase the number of respondents, i.e. reviewers per text, to the extent that more sophisticated quantitative methods of analysis could have been used in the main study. In this way the differences between the indices of typicality and interest for the two genders could have been tested for statistical significance. Thus also the credibility of the findings would have been greater (see below).

The Requirement of coding checks included in the criterion of reliability/dependability reflects the same concerns over reliability as within the traditional view of validity in the quantitative approach, i.e. consistency in the results. With the present study, the issue most concerns the content analysis carried out in the pilot study. The content analysis was repeated after one year to double *check the analysis*. During this reanalysis some refinements were made to the categories. This was done by the researcher herself, although it could have been another person altogether, which would probably have increased the reliability and credibility of the analysis.

Credibility/authenticity

Credibility/authenticity is also referred to as the internal validity of the study which brings about the plausibility of the findings. The findings of the study must be plausible and convincing, and the use of triangulation is recommended in order to achieve a more complete understanding of the phenomenon. Furthermore, this criterion requires considering the extent to which the measurements used reflect the constructs, areas of uncertainty are identified, findings have been replicated in other parts of the database than the one they originated in, and the conclusions were considered accurate by the original informants (Miles & Huberman 1994.) Many features relevant to reliability/dependability are relevant also to the credibility of the findings and conclusions. For instance, the measurements used and the extent to which they

reflect the construct were discussed above under the criterion of reliability/dependability. The same discussion of the issue is relevant to the criterion presently under consideration, since the quantification procedure and the methods of gathering data were designed bearing in mind the need for theoretical consistency. This increases both the reliability and credibility of the findings but there are also weaknesses in the quantification that have been brought to the attention of readers above.

The extent to which the parameters used reflect the construct was also treated in Chapter 8 with factor analysis summarising the relations between the variables. On the basis of the factor analysis it was concluded that for most texts the main study questionnaire succeeded in measuring both typicality and interest, thus supporting the construct validity of the study. However, as factor analysis showed differences with regard to individual texts it is to be noted that in some cases the results suggest that unnecessary specific variables were used, at least in respect of textual features, since either the typicality and interest of the topic, genre, layout and language of a text are strongly associated or students were unable to review and rate the typicality and interest of the texts in terms of the textual features presented to them. It evidently would be more authentic to review texts as a whole instead of reviewing and analysing separate textual features.

As far as the content validity of the quantification used and the conclusions based on the results obtained thereby are concerned, a more theoretical approach is required to evaluate the extent to which typicality and interest cover the field of the concept of authenticity. Based on the literature review, it was concluded that typicality and interest cover important aspects of text authenticity. However, it was not assumed that they would cover all the important aspects of the concept. For instance, the relevance of the text or of the information in the text was not covered in the study design at all. However, some of the students' comments in the main study questionnaire revealed that it is not always the interest of the text that is essential but rather its relevance, deriving from the need for reading it. Therefore, relevance has been included in the framework for text authenticity in international reading literacy assessments (Figure 5) in order to increase the content validity of that particular conclusion of the study, even though it was not included in the study design.

The findings of the present study seem convincing in the light of other studies. Recent international reading literacy assessments, such as PISA 2000 and 2003, have provided information not only about reading literacy performance but also about reading engagement and the reading interests of 15-year-old students in the participating countries. The PISA texts found most authentic in the present study were texts originally appearing in magazines and newspapers, such as are frequently read by Finnish 15-year-olds. (See e.g. OECD 2001, 2002; Linnakylä 2002; also Taube & Munck 1996; Roe & Taube 2003.)

The credibility of the results of the present study has been strengthened by methodological triangulation. It has been argued that using mixed methods

increases the validity of a study since by applying multiple methods to the same phenomenon a more complete understanding can be achieved. Furthermore, the weaknesses of any one method can be compensated for by another method, and this also increases the validity of the study. (See e.g. Frechtling & Sharp 1997; Teddlie & Tashakkori 2003.) Although the qualitative approach has been dominant in this study, a quantitative approach is also present, especially in the main study phase. Along with the indices for typicality and interest, students' comments were included as illustrative examples supporting the numerical findings. This also increased the authenticity of the findings of the study, as did including panellists' comments in presenting the results of the content analysis of the pilot study phase.

In the main study indices for typicality and interest were calculated as means, which do not reveal individual differences in text authenticity. Furthermore, the differences between indices were often quite small and often not statistically significant, which means that they can be regarded only as suggestive. A rank ordering of the PISA texts has been used but this hardly solves the problem since it too is only suggestive due to the small differences in the indices on the basis of which the texts have been ranked. Nevertheless, certain texts stand out and the results obtained for these cases can be seen as reliable and valid. This also applies to the gender differences that were quite clear in the case of some texts. In presenting the results of the study, the main attention has been given to cases that clearly stand out in order to avoid emphasizing differences and results that are not necessarily there. Overall, a larger number of respondents, especially per text, would have enabled the use of more sophisticated statistical methods of analysis. This would have increased the credibility of the findings as, for instance, the differences in indices between gender groups could also have been tested for statistical significance. On the other hand, the weaknesses of the method have been explicitly brought to the attention of the readers of the study report, which increases both the objectivity/confirmability and the credibility of the findings.

Although presenting the students' and panellists' comments increases the authenticity of the study and seems to support the findings from the quantitative data, these comments could have been used even more. For instance, in the main study comments were presented only in reporting overall authenticity in terms of typicality and interest as well as topic and genre typicality and interest. This was because students' comments were in general quite short and vague. Most of them repeated and echoed the terms used in the question and thus contributed little further information to the study. Language and layout were rarely commented upon at all by students. On the other hand, some students commented at greater length and their comments were used as illustrative examples of the students' point of view.

The findings of the study have not been presented to the original informants nor have they been discussed with them. Thus, the credibility of the findings cannot be evaluated from the informants' point of view.

Transferability

Transferability, which has also been referred to as external validity, is a criterion concerned with evaluating the conditions of generalizability of the findings. This criterion for the quality of conclusions requires consideration of whether the original sample of persons (or processes etc.) is described sufficiently well to permit comparisons with other samples, whether threats to generalizability have been considered, whether the findings have been reduplicated in other studies and whether the report suggests further testing of the findings. (Miles & Huberman 1994.)

In the present study, the generalizability of the findings needs to be considered focusing principally on the main study phase. This question has been addressed in the study report but is briefly summarized here. Due to the small number of respondents, and especially the limited number of reviews per text, and the fact that the respondents made up a convenience sample selected without obeying the principles of probability sampling, the results are not to be generalized beyond the group of students participating in the study. However, since the results were consistent with the results of the PISA 2000 reading literacy assessment, in which reading engagement was also studied, it is possible to suggest that the results of the present study to some degree may represent the general views of 15-year-olds. It should be noted that bold generalizations would be irresponsible, since there is no evidence to support them, only speculations to be presented at the level of the educated guess. Furthermore, the weaknesses of the methods of gathering and analyzing data presented above need to be taken into account, since the reliability and credibility of the results certainly affects their generalizability as well.

Transferability needs to be evaluated in the present study also in terms of the constructs used in the study. As discussed above, in the pilot study, in which the aim was to find the textual features constituting of text authenticity for readers, expert panellists reviewed text authenticity in terms of text typicality only. In the main study, however, another aspect of text authenticity was also included in the questionnaire–'interest'. The textual features chosen for the main study questionnaire (topic, genre, layout, language) were thus relevant to expert readers in reviewing text typicality, and it was *presumed* that the same textual features would be relevant also in reviewing text interest in the main study phase. From the point of view of transferability, however, one needs to ask whether the textual features are transferable from the context of typicality to that of the interest. Furthermore, one needs to ask whether the textual features found relevant by expert readers are relevant to 15-year-olds, i.e. are textual features transferable from one group of respondents to another.

To speculate on the first question, one might expect that if panellists had reviewed the PISA texts also in terms of text interest the emphasis on particular textual features emerging in the data might have been different from that found in the content analysis now. However, it is unlikely that there would have been any new textual features that were not among those found in the content analysis of the present study at all. Furthermore, the differences in emphasis

would not have necessarily had any effect on the selection of textual features for the main study since it was influenced not only by the results of the content analysis but also by the literature review and practical considerations in terms of the respondents' age and time available.

To respond to the second question, it needs to noted that the aim of the main study was not to find out which textual features are relevant to 15-year-olds in terms of text authenticity but to find out how authentic the PISA texts were for them in terms of typicality and interest of topic, genre, layout and language. Thus, the issue of the transferability of textual features from one context and set of respondents to another is somewhat irrelevant although it deserves discussion. If the relevance or importance of the textual features selected from the point of view of 15-year-olds had been the focus of examination, it is probable that textual features commented upon by the students would have been different from those relevant to expert readers who possess the knowledge and tools for analyzing texts to a greater extent than 15-year-olds.

Utilization/application

The utilization/application criteria for assessing the quality of conclusions concern the extent to which the results of the study and the conclusions drawn reach their audience and form the basis for innovative practices and further assessment studies (Miles & Huberman 1994). While the study was being carried out presentations were given at in national and international conferences as well as presentations and training with instructional emphasis at various events organized under the themes of the present study. In the early phases of the study these presentations had a theoretical emphasis but later, as the study proceeded, the findings of the pilot study and the main study were also reported. Furthermore, articles based on the present study have been published, mainly in books about Finnish students' and adults' reading literacy performance and their profiles as readers (e.g. Sulkunen 2004a and 2004b; also Sulkunen 2002).

One channel for distributing the findings of the present study to a wider audience is certainly the present publication itself. It will probably to some extent reach researchers and test developers within the field of international reading literacy assessments – the group likely to benefit most from the study. Certainly the language of the study report is not an obstacle to that. However, further presentations will be made and articles published in order to gain a wider audience and greater visibility for the findings of the study. Presentations will be given at national and international conferences, and articles will be published both in scientific forums but also in more popular forums targeted at teachers, for instance, as the idea of the authentic text fits in with the current interest in authentic literacy instruction in the Finnish national framework curriculum for mother tongue instruction (FC 2004).

The findings of the study are applicable in the context of the designing of international reading literacy assessments. In Chapter 9.2 certain new practices

have been suggested for the selection of the texts for such reading literacy tests in the form of a framework for text authenticity in international reading literacy assessments (Figure 5). Suggestions have also been made about covering the textual worlds of both genders in reading literacy tests in that it is recommended that the overlapping textual worlds authentic to both genders in common should be complemented with texts belonging solely – or mainly – to boys' or girls' textual worlds. The extent to which these recommendations will be taken into account remains to be seen, but there is certainly constant interest in equity of testing. Firstly, of course, they need to be made known within the community of researchers.

Within this chapter (9.3) the quality of the study has been evaluated employing the criteria presented by Miles and Huberman (1994). The criteria cover a wide range of issues and perspectives to be taken into account although they are to some degree overlapping, which results unavoidably in repeating the same points with reference to different criteria. The criteria used are intended for evaluating the quality of qualitative studies but they also cover points in common with the traditional views of validity used in evaluating quantitative studies, for instance reliability and generalizability. Since the present study employs mixed methods with a predominantly qualitative approach, emphasis has been placed on the importance of content validity of the concept of text authenticity as well as on the reliability and generalizability of the results, thus echoing the concerns of more traditional views of validity in evaluating the quality of the present study. By ways of conclusion it may be observed that whatever the shortcomings and weaknesses revealed in the gathering and analysis of the data it has been the aim of the researcher to explicitly describe the possible shortcomings and the limitations of the study and to provide readers with a reasoned account of the decisions made during the process with reference to theory and previous studies. There are many things that could have been done differently but the decisions made were carefully thought out on the basis of the knowledge and understanding available at the time and taking into account the practical resources at the disposal of the researcher.

On the basis of the present study and its findings, but also limitations and shortcomings, further studies on the topic are suggested in Chapter 9.4 below.

9.4   Suggestions for future research on authentic texts

The present study includes a theoretical review on the subject of text authenticity and provides information on the authenticity of the texts used in the PISA 2000 reading literacy assessment from the point of view of Finnish 15-year-old students. Based on these, the study provides new perspectives on the selection of authentic texts for reading literacy tests. The framework for text authenticity in international reading literacy tests (Figure 5) offers a synthesis of

these perspectives. The framework also gives a central role to the concept of genre and to functional criteria in text classification in place of text type with its purely linguistic criteria, which has traditionally been used in classifying texts in an assessment context. The study and its theoretical and practical implications have also opened up new perspectives and prompted suggestions for future research on text authenticity.

As the present study sought to provide insight into Finnish 15-year-old students' views on the authenticity of the texts used in the PISA 2000 reading literacy test, it concentrates on the cultural authenticity of those texts in one particular culture from among all the countries and cultures participating in PISA. Moreover the sub-cultures of the two genders are in focus. Since text authenticity is strongly linked to the culture and textual world of readers, similar studies within other cultures would be of great interest. Thus, it is suggested that a comparative study with a similar design should also be made in other countries and cultures to reveal the cultural differences in the authenticity of the PISA texts. However, such a comparative study could be implemented using more attributes of text authenticity, for instance, including relevance along with typicality and interest, and with less specific textual features being presented to the students evaluating text authenticity. Furthermore, including the task authenticity to the study design would result in more authentic and valid approach covering also the text use in testing situation.

Another issue that should be taken into consideration both in designing international reading literacy tests and in further research is that of the decontextualizing and recontextualizing of authentic texts. The results of the pilot study showed that the panellists found some texts inauthentic in terms of typicality because they were taken out of context and thus became untypical and sometimes incomprehensible. The panellists' comments concerned the same non-continuous texts that the students in the main study found consistently untypical and uninteresting, e.g. the table of news agencies, and the flight timetable. However, there were also texts found typical and interesting which had significant non-continuous elements, suggesting that it is not the non-continuous format that makes certain texts inauthentic but other, more complex features. One of these is an irrelevant and uninteresting topic. Another feature related to the phenomenon is that the non-continuous texts in question appeared in the reading literacy test without a meaningful context. Thus, they became nothing more than isolated fragments of (unnecessary) information. Because these texts have been decontextualized, important interpretational clues have been missed. (Leontjev 1973.)

In the terms of genre and register theory based on Halliday's systemic functional grammar (e.g. Martin 1993; Martin & Rothery 1993; Eggins and Martin 1997), in decontextualizing non-continuous texts the dimensions of situational context, mode, field and tenor have been deleted. Furthermore, there are very few linguistic elements left in the text to express the textual, ideational and interpersonal meanings that reflect the mode, field and tenor of the

situational contexts of the text. In continuous texts these dimensions are always to some extent reflected in the linguistic choices used. With non-continuous texts, which include relatively little written text, the dimensions of situational context are revealed in the context of use and/or in the continuous text attached. Field and tenor in particular are dimensions of context from which the purpose for reading a text derives. For instance, a flight timetable is a text whose field presupposes the intention of travelling by air and whose tenor presupposes the role of passenger or travel-agent. With a decontextualized timetable the text itself – or other elements in the testing context–should be able to construct, say, the role of passenger for the reader. Even so, from the readers' point of view the lack of an authentic situational context means that there is limited opportunity for genuine interaction with the text, since without the relevant purpose for reading the reader is just an outside observer of the text (e.g. Leontjev 1973).

In order to shed light on the issue of decontextualizing texts and recontextualizing them in a test, further research is suggested on the effect decontextualizing has on the text and on readers' perception of and interaction with it. Furthermore, it is necessary to study to what degree the original context of appearance and use of texts should be and can be simulated in a testing situation. Since test-takers tend to play along in a testing situation, they may equally accept less authentic tests and undertake the test tasks accordingly (Spolsky 1985; see also Widdowson 1990). Nevertheless, it is possible to increase the authenticity of texts in a test by simulating the context. In some reading literacy assessments this has been aimed at by printing newspaper-style (IALS) or magazine-style (PIRLS) booklets in which texts to be read appear. In PISA it has been attempted to increase authenticity by providing students with instructions including a description of the text's context of use and by aiming at making the tasks as authentic as possible. However, no study has been made of how successful these strategies are from the point of view of the reader.

Further research is also recommended on the cultural and linguistic differences affecting texts that represent transnational, global genres. As has been suggested earlier in this study, genres exhibit variation across cultures. Due to globalization there are many socially defined tasks that are common in several cultures. Thus the genres represented by the texts used to accomplish these tasks are also common. Although some of these genres are very uniform, e.g. in an academic context, there also occurs cross-cultural variation within particular genres common to many cultures. The rhetorical structures of languages differ, and thus the linguistic realization of the same moves can also differ, not to mention the linguistic strategies applied. (Bhatia 1993; Swales 1990; see also Harjunen 2007.) This cross-cultural variation poses a serious threat to equity in testing since texts – and the test – are hardly equivalent if others read, say, a news report with typical linguistic realization and others read the same familiar genre yet with untypical linguistic realization.

The texts selected for an international reading literacy test should therefore be under scrutiny in each participating country not only in terms of

their topic and genre but also by their linguistic realization. To provide the basis for this there is an increasing need for research on different genres in different cultures; more specifically, the rhetorical structure and the linguistic strategies and devices used in global genres in different cultures need to be studied. This kind of research has already been carried out for some genres: in the Finnish context alone studies related to genre have been made of news texts (Kunelius 1993, 1996; Pietilä 1993; Reunanen 1991, 1993), press releases (Kankaanpää 2001, 2006; Komppa 2006), textbooks (Karvonen 1995), recipes with a historical perspective (Taavitsainen 2001, 2006) and several types of oral discourse (e.g. on service encounters see Ventola 2006). Genre studies of most benefit to international reading literacy assessments would focus on global genres within a comparative perspective.

To summarize, the further research recommended here is related to the cultural aspect of text authenticity and a comparative perspective is suggested for both studies on the authenticity of the texts used in the PISA 2000 reading literacy test in other participating countries and cultures and for studies focusing on the linguistic realization of global genres in different cultures. The suggested research on the linguistic and other symbolic realizations of genres also brings in a theme that is always current in international assessments, namely equity in testing. Equity in testing is a necessary condition for making valid comparisons of results in an international context. This aspect of equity, it is suggested, should be considered on the level of the text within the framework of genre, which emphasizes the functional and all symbolic characteristics of texts while at the same time enabling the study of the construction of textual, ideational and interpersonal meanings in the language of texts (e.g. Eggins & Martin 1997). This theoretical emphasis also provides the conceptual tools for studying the effect decontextualizing texts has on texts and their readers. Overall, the suggested research would continue the task of the present study in contributing to more valid international reading literacy tests, balanced in terms of both equity and authenticity.

TIIVISTELMÄ

Tekstien autenttisuus kansainvälisissä lukutaidon arviointitutkimuksissa: PISA 2000.

Johdanto

Viime vuosina toteutetuissa kansainvälisissä lukutaidon arviointitutkimuksissa autenttisuus on ollut esillä yhtenä niistä kriteereistä, joilla määritellään luku-taitokokeeseen valittavia tekstejä ja niihin liitettäviä tehtäviä. Myös OECD:n PISA–arviointiohjelmassa (Programme for International Student Assessment), jossa joka kolmas vuosi arvioidaan 15-vuotiaiden nuorten taitoja lukutaidon, matemaattisen osaamisen ja luonnontieteellisen osaamisen alalla, autenttisuus on ollut keskeisellä sijalla koetta laadittaessa. PISA-arvioinnissa on pyritty sii-hen, että sekä luettavat materiaalit että niihin liitettävät tehtävät olisivat mah-dollisimman autenttisia eli todellista elämää vastaavia (OECD 2003).

 Sen lisäksi että autenttisten tekstien tulisi olla todellista elämää vastaavia eli todellisen elämän tekstejä (esim. Harris & Hodges 1995), tekstien autentti-suuteen on liitetty myös monia muita ominaisuuksia, jotka liittyvät usein toi-siinsa. Autenttiset tekstit ovat ensinnäkin aitoja tekstejä, jotka on laadittu aitoja viestinnällisiä tarkoituksia varten. Tällaisia tekstejä ei siis ole laadittu varta vasten opetus- tai arviointitarkoituksiin eikä niitä myöskään ole näitä tarkoituk-sia varten muutettu esimerkiksi kieliasultaan tai sisällöltään yksinkertaisem-miksi. (Widdowson 1978; Harris & Hodges 1995; Lewkowicz 2000.) Tämän lisäksi autenttiset tekstit ovat merkityksellisiä, relevantteja ja tyypillisiä lukijan kokemuspiirissä (esim. Valencia ym. 1994; Framework and Specifications for PIRLS Assessment 2001). Myös tekstien kiinnostavuus lukijalle ja arviointikontekstissa niiden sopivuus ovat tekstien autenttisuutta määrittäviä tekijöitä (esim. Garcia & Verville 1994; Kapinus et al. 1994; Weiss 1994).

 Näiden kriteerien myötä tekstien autenttisuutta tarkasteltaessa keskiöön nousee lukija: viime kädessä lukija itse määrittelee, mitkä tekstit ovat hänelle autenttisia, ei kukaan muu (Widdowson 1979; Bachman 1990; Garcia & Verville 1994; Weiss 1994; Bachman & Palmer 1996). Olennaista on, että yksilöllä on lu-kemistilanteessa todellinen viestinnällinen tarkoitus tekstin lukemiselle. Tästä tarkoituksesta tai tarpeesta nousevat luonnollisena seurauksena myös tekstin autenttisuuteen liittyvät ominaisuudet, erityisesti kiinnostavuus, merkitykselli-syys, tyypillisyys ja relevanssi. Tekstien autenttisuuden viitekehyksessä nämä ominaisuudet liittyvät lukijan ja tekstin väliseen suhteeseen, eivät pelkästään tekstiin (Widdowson 1990).

 Tekstien autenttisuudella on erityisesti opetus- ja arviointikontekstissa myös yhteisöllinen puolensa, joka on yhdenmukainen sosiokulttuurisen luke-miskäsityksen kanssa. Tämän käsityksen mukaan kirjoittamista ja lukemista ei tarkastella erillisinä taitoina. Pikemminkin puhutaan tekstitaidoista (literacy), jotka kattavat kaiken teksteihin liittyvän toiminnan. Teksteihin liittyvä toiminta on olennaisesti toimintaa sosiaalisessa ympäristössä. Tässä lähestymistavassa

yksilö nähdään aina osana yhteisöä, jolla on yhteisiä sosiaalisesti määrittyneitä käytänteitä. Tällaisia käytänteitä ovat myös lukukäytänteet, jotka ovat kulttuurisidonnaisia tapoja hyödyntää tekstiä. (Barton 1994; Pitkänen-Huhta 1999; Gee 2000a.) Näin ollen yksittäisen yksilöllisen lukutapahtuman taustalla on aina kulttuurisesti määrittyneet lukukäytänteet. Sillä mikä näyttäytyy yksilöllisenä valintana, on myös sosiaalinen ulottuvuutensa. Tästä näkökulmasta tekstitaitojen oppiminen ja omaksuminen on sosiaalistumista yhteisön käytänteisiin ja sääntöihin. Tällöin äidinkielen opetuksessa painotetaan autenttisia tekstejä, joiden avulla oppilaat voivat tutustua erilaisiin tekstikäytänteisiin, teksteihin ja niihin liittyviin viestinnällisiin tarkoituksiin ja tekstikonventioihin (e.g. Pitkänen-Huhta 1999; Street 1996; Luukka 2004a). Opetuksen tavoitteena on valmistaa ja valtauttaa oppilaita vastaamaan todellisen elämän haasteisiin ja toimimaan todellisen elämän tekstitilanteissa (esim. Luukka 2004b; Luukka & Leiwo 2004). Yhteiskunta ja yhteisö osaltaan määrittävät sen, millaiset tekstit ovat lukijalle autenttisia nyt ja tulevaisuudessa.

Lukutaidon arvioinnin kontekstissa tekstien autenttisuus liittyy oppilaiden vastaamismotivaatioon, mutta ennen kaikkea kokeen validiuteen. Kokeeseen osallistujien kannalta heille merkityksellisten ja heitä kiinnostavien tekstien käytön lukutaitokokeessa on katsottu lisäävän koetilanteen mielekkyyttä ja kohottavan vastausmotivaatiota (Widdowson 1990). Kokeen validiuden näkökulmasta tekstien autenttisuudella on yhtymäkohtia sekä käsitevalidiuteen, sisältövalidiuteen että ilmeisvalidiuteen. Osa käsitevalidiutta on tarkastella, missä määrin esimerkiksi lukutaitokokeesta saatuja tuloksia voi yleistää koskemaan kokeen ulkopuolista tilannetta. Jos koe ei heijasta todellisen elämän kielenkäyttöä – esimerkiksi lukemistilanteita –, ei sen tuloksiakaan voi yleistää koskemaan koetilannetta laajempaa kontekstia. (Bachman & Palmer 1996; Wu & Stansfield 2001.) Autenttisista materiaaleista koostuvan kokeen katsotaan siis olevan validi, koska se tavoittaa oppilaiden osaamisen koko kirjon todellisen elämän viestintätilanteissa (ks. myös Huerta-Marcias 1995; Shohamy & Reves 1985; Wood 1993). Autenttisuus on myös olennainen osa sisältövalidiutta, sillä kokeen tehtävien autenttisuus voidaan määritellä niiden relevanssina todellisen elämän lukemistilanteiden kannalta (Bachman & Palmer 1996; Wu & Stansfield 2001). Lisäksi koemateriaalien autenttisuus liittyy ilmeisvalidiuteen, sillä se viittaa kokeen uskottavuuteen ja relevanssiin vastaajien näkökulmasta ja parantaa siten vastaajien halukkuutta vastata kokeeseen.

Nykyinen validiuskäsitys painottaa käsitevalidiuden ensisijaisuutta, mutta toisaalta ottaa huomioon entistä laajemmin myös kokeen ja arvioinnin kontekstia ja muun muassa arvioinnin sosiaalisia seurauksia (Messick 1989a). Kansainvälisen lukutaitokokeen kontekstissa tekstien autenttisuuden vaatimus on perusteltavissa kokeen ja arvioinnin validiudella, mutta toisaalta kansainvälisen arvioinnin validiuteen liittyvä vaatimus tulosten vertailukelpoisuudesta rajoittaa tekstien autenttisuutta. Kokeeseen voidaan valita vain kaikille osallistujille yhtä autenttisia tai yhtä epäautenttisia tekstejä, jolloin kansallisesti tyypilliset ja yksilöille relevantit tekstit rajataan kokeesta pois. Kokeen tekstivalikoima kattaa tällöin parhaimmillaankin vain kaikille osallistujamaille yhteisiä

tekstitapahtumia. Tuota tekstivalikoimaa on syytä tarkastella kriittisesti kansallisen kulttuurin näkökulmasta. Kuinka validi esimerkiksi PISAn lukutaitokoe on suomalaislukijoiden näkökulmasta?

Tutkimusongelmat, menetelmät ja aineisto

Tutkimuksen tarkoituksena oli selvittää PISA 2000 –lukukokeessa käytettyjen tekstien autenttisuutta suomalaisnuorten kulttuurissa. Tutkimuskysymykset olivat seuraavat: 1. Mitkä tekstien piirteet ovat lukijalle relevantteja tekstien autenttisuuden näkökulmasta? 2. Kuinka autenttisia PISA 2000 –lukukokeen tekstit ovat 15-vuotiaille suomalaisnuorille? 3. Ovatko PISA 2000 –lukukokeen tekstit yhtä autenttisia 15-vuotiaille suomalaistytöille ja –pojille?

Tutkimus koostui kahdesta eri vaiheesta. Esitutkimusvaiheessa 8-henkinen paneeli arvioi PISA-tutkimuksen lukutaitokokeessa olleiden tekstien autenttisuutta niiden tyypillisyytenä sekä numeerisesti että kirjallisesti. Tämän vaiheen tarkoituksena oli saada selville, mihin tekstien piirteisiin panelistit kiinnittivät huomiota perustellessaan arvioitaan. Varsinainen esikokeen tutkimusaineisto koostui siis panelistien kirjallisista kommenteista ja perusteluista. Aineiston analyysissä hyödynnettiin sisällönanalyysiä (Kyngäs & Vanhanen 1999; Ryan & Bernard 2000). Analyysin tulosten ja lähdekirjallisuuden perusteella päädyttiin siihen, että neljä tekstipiirrettä–aihe, tekstilaji, ulkoasu ja kieli–olivat keskeisiä tekstien autenttisuudelle.

Päätutkimusvaiheessa joukko 15-vuotiaita suomalaisnuoria (yhteensä 240 nuorta) vastasi kyselyyn, jonka laatiminen pohjautui osaksi esitutkimusvaiheen tuloksiin. Kyselyyn vastanneet nuoret arvioivat PISA-tekstien autenttisuutta aiheen, tekstilajin, ulkoasun ja kielen tyypillisyytenä ja kiinnostavuutena. Numeeristen arvioiden lisäksi (asteikko 1 – 4) nuoria pyydettiin perustelemaan arvionsa kirjallisesti. Numeeristen arvioiden perusteella laskettiin kullekin tekstipiirteelle ja koko teksteille tekstikohtaiset tyypillisyys- ja kiinnostavuusindeksit, jotka pohjautuivat oppilaiden antamien numeeristen arvioiden keskiarvoihin ja joiden perusteella tekstit asetettiin tyypillisyys- ja kiinnostavuusjärjestykseen. Keskiarvojen erojen tilastollista merkitsevyyttä testattiin verrannollisten parien t-testillä (Holopainen & Pulkkinen 2003). Tyttöjen ja poikien osalta keskiarvojen eroja ei testattu vastaajien pienen lukumäärän vuoksi. Oppilaiden kirjallisia kommentteja käytettiin kuvaamaan ja havainnollistamaan numeerisia tietoja.

Päätutkimusvaiheessa käytettyjen muuttujien välistä yhteyttä tutkittiin laskemalla muuttujien välisiä korrelaatiokertoimia (Pearsonin korrelaatiokerroin). Korrelaatiot tyypillisyys- ja kiinnostavuusmuuttujien kesken olivat korkeita mutta selvästi matalampia näiden muuttujien välillä, minkä vuoksi muuttujien välisiä suhteita tutkittiin edelleen eksploratiivisen faktorianalyysin avulla (Darlington 1997). Useimpien tekstien osalta faktorianalyysi tuotti odotuksenmukaisesti kaksi faktoria. Tyypillisyysfaktorille latautuivat tyypillisyysmuuttujat ja kiinnostavuusfaktorille kiinnostavuusmuuttujat. Joukossa oli kuitenkin tekstejä, joiden kohdalla jotkut muuttujat latautuivat kummallekin fakto-

196

rille, sekä tekstejä, joiden kohdalla faktorianalyysi tuotti vain yhden faktorin. Näin ollen useimpien tekstien kohdalla kysely onnistui mittaamaan kiinnostavuutta ja tyypillisyyttä erillisinä autenttisuuteen liittyvinä piirteinä. Toisaalta vaikutti siltä, että tekstipiirteiden osalta kysely oli 15-vuotiaille liian yksityiskohtainen, koska tyypillisyysmuuttujien keskinäiset korrelaatiot ja kiinnostavuusmuuttujien keskinäiset korrelaatiot olivat korkeita. Vaikutti siltä, että nuoret eivät olleet kyenneet arvioimaan tekstejä niiden piirteitä eritellen, vaan pikemminkin kokonaisuutena.

Tulokset

Tutkimuksen tulokset osoittivat, että PISA 2000 –lukukokeessa käytettyjen tekstien tyypillisyys ja kiinnostavuus oli kokeeseen osallistuneiden nuorten mielestä suhteellisen alhainen. Korkeimmatkin keskimääräiset tyypillisyys- ja kiinnostavuusindeksit jäivät alle kolmen, joka käytetyllä asteikolla (1 – 4) merkitsee tyypillistä tai kiinnostavaa tekstiä. Useimpien tekstien kohdalla tyypillisyysindeksit olivat lisäksi korkeampia kuin kiinnostavuusindeksit, vaikka erot eivät aina olleetkaan tilastollisesti merkitseviä.

Myöskään tekstien aiheet ja niiden tekstilajit eivät olleet kyselyyn vastanneiden oppilaiden mielestä kovin tyypillisiä tai kiinnostavia. Tekstien aiheiden kiinnostavuusindeksit olivat kuitenkin useimpien tekstien kohdalla korkeampia kuin tyypillisyysindeksit. Tekstilajien kohdalla tilanne oli päinvastainen. Teksteissä käytetyn kielen ja niiden ulkoasun tyypillisyys arvioitiin kyselyn perusteella korkeammaksi kuin samojen tekstipiirteiden kiinnostavuus, mutta kaiken kaikkiaan tekstien kieli ja ulkoasu eivät olleet kovin autenttisia.

Koska oppilaiden arvioiden perusteella lasketut tyypillisyys- ja kiinnostavuusindeksit olivat keskiarvoja ja tekstien väliset erot olivat sangen pieniä, on tuloksiin suhtauduttava varovaisesti. Tekstien keskinäistä tyypillisyys- ja kiinnostavuusjärjestystä onkin pidettävä vain viitteellisenä. Tästä huolimatta tekstien asettaminen tyypillisyys- ja kiinnostavuusjärjestykseen paljasti tiettyjen tekstien kohdalla kaavamaisuutta: samat tekstit olivat tyypillisimpien ja kiinnostavimpien tekstien joukossa tekstipiirteestä riippumatta, samoin samat tekstit olivat tekstipiirteestä riippumatta epätyypillisimpiä ja vähiten kiinnostavia. Tämän lisäksi tyttöjen ja poikien välillä oli joidenkin tekstien kohdalla selviä ja tekstipiirteestä riippumattomia eroja. Toki oli myös tekstejä, joiden kohdalla tyttöjen ja poikien arviot tyypillisyydestä ja kiinnostavuudesta osuivat yksiin. Tällaisia kummallekin sukupuolelle autenttisia tekstejä olivat muun muassa lyhyet uutistekstit, lyhyet elokuva-arvostelut, uutisartikkeli NASAn huumatuilla hämähäkeillä tekemistä kokeista, lehtiartikkeli poliisin käyttämistä DNA-tutkimuksista ja moottoripyörämainos. Kumpikaan sukupuoli ei pitänyt autenttisena taulukkoa, johon oli listattu tietotoimistoja, puhelimen käyttöohjeita tai lentoaikatauluja.

Kummallekin sukupuolelle yhteisesti autenttisten tekstien lisäksi tulokset paljastivat useita tekstejä, jotka olivat autenttisia vain joko pojille tai tytöille. Tällaisia pojille autenttisia PISA-lukukokeen tekstejä olivat graafiset kuviot

Tsad-järven pinnankorkeuden vaihtelusta, pääkirjoitus uuden teknologian aiheuttamista eettisistä ongelmista, asiateksti Etelänavan valloituksesta, lehtiartikkeli miesten paitojen laadusta, ohjeet työhaastatteluun valmistautumiseen ja puudiagrammi työvoiman rakenteesta. Nämä tekstit eivät välttämättä aina olleet juuri niitä tyypillisimpiä tai kiinnostavimpia, mutta kuitenkin selvästi autenttisempia pojille kuin tytöille. Tytöille autenttisia tekstejä olivat puolestaan mielipidekirjoitukset avaruustutkimuksesta, tarina tulvan saartamaksi joutuneesta naisesta, raportti lasten asenteista kotitöitä kohtaan, tiedote vaihto-oppilaaksi hakeville, käytösohjeet vaihto-oppilaille, näytelmä prinssistä ja Amandasta, työpaikan hakulomake ja lehtiartikkeli rantojen suojelemisesta.

Pohdinta

Tutkimuksen tulosten perusteella voi tehdä useita johtopäätöksiä kansainvälisiin lukukokeisiin valittavista teksteistä ja niiden valintakriteereistä. Koska tekstien autenttisuus liittyy olennaisilta osin kokeen validiuteen, on pyrkimys valita autenttisia todellisen elämän tekstejä lukukokeeseen tärkeä. Kansainvälisissä arvioinneissa usein kuitenkin painottuu autenttisuuden yhteisöllinen ulottuvuus ja käytännön syistä siihen liittyvä yksilöllinen puoli jää sivuosaan. On selvää, että yksilöllisiä eroja tekstien kiinnostavuudessa ja tyypillisyydessä on hankala huomioida, kun yksittäisiä vastaajia kokeessa on tuhansia, kansainvälisesti jopa satoja tuhansia. Koska yksittäisillä lukijoilla usein kuitenkin on yhteisiä kokemuksia ja tekstitapahtumia saman yhteisön jäseninä, tulee tekstien autenttisuutta määritettäessä arviointikontekstissa ottaa huomioon ainakin kulttuurin keskeisten alaryhmien kiinnostuksen kohteet.

Koska tutkimuksen tulokset selvästi vahvistavat aiempien tutkimusten tuloksia (esim. OECD 2002; Linnakylä 2002; Linnakylä & Sulkunen 2005) siitä, että tyttöjen ja poikien tekstimaailmat ovat osin erilaiset, on keskeistä, että lukukokeen tekstien joukossa on paitsi kummallekin sukupuolelle autenttisia tekstejä myös vain tytöille ja vain pojille autenttisia tekstejä. Näin kokeen validius säilyy siten, että koe kattaa mahdollisimman laajasti todellisen elämän tekstejä ja lukutapahtumia, eikä kokeen tekstivalikoima muodostu liian kapeaksi. Näin väistämättä kävisi, jos kokeeseen valittaisiin vain kummallekin sukupuolelle yhteisesti autenttisia tekstejä. Erityisenä haasteena lukukokeen tekstejä valittaessa voi pitää poikia kiinnostavien ja siten heitä motivoivien tekstien löytämistä.

Kansainvälisessä lukutaidon arvioinnissa tekstien autenttisuuteen liittyvä kulttuurinen puoli korostuu nimenomaan kansallisten kulttuurien näkökulmasta. Tavoitteena on, että koetulosten vertailukelpoisuuden takaamiseksi kokeessa olevat tekstit olisivat yhtä autenttisia tai yhtä epäautenttisia kaikissa osallistujamaiden kulttuureissa. Tämä vaatimus luonnollisesti rajoittaa tekstivalikoimaa siten, että kaikki kansallisesti arkaluontoiset aiheet ja epätyypilliset tekstit on jätettävä pois. Arvioinnin validiuden näkökulmasta tämä tarkoittaa sitä, että koetulokset eivät välttämättä ole yleistettävissä osallistujamaiden kansallisesti tyypillisiin tekstitapahtumiin, elleivät nämä ole tavallisia muissakin

maissa ja siten edustettuna kokeessa. Sama rajoitus koskee myös yksilöllisiä tekstitapahtumia. Tästä näkökulmasta PISAn lukukoe vaikuttaa tekstivalikoimaltaan suomalaislukijalle suhteellisen epäautenttiselta: tekstivalikoima kattanee kohtuullisen hyvin yhteiset, globaalit tekstitapahtumat ja sisältänee kaikille osallistujamaille yhtä (epä)autenttisia tekstejä, mutta yksittäiset suomalaislukijat ja kulttuurisesti keskeiset alaryhmät, kuten tytöt ja pojat, eivät näytä löytävän kokeesta heille tyypillisiä ja kiinnostavia tekstejä.

Arvioitaessa lukukokeeseen ehdolla olevien tekstien autenttisuutta on syytä muistaa, että vaikka teksti olisi tyypiltään relevantti ja tyypillinen jokaisen osallistuvan maan kulttuurissa, ei teksti välttämättä silti tarkemmin katsottuna ole autenttinen kielellisiltä valinnoiltaan. Vaikka tekstin viestinnällinen tarkoitus ja siten tekstilaji – genre – olisikin eri maissa sama, voivat kielelliset valinnat johtaa esimerkiksi tyyliltään ja rakenteeltaan erilaiseen tekstiin. Tekstien arviointi vaatii siis tarkkuutta, mutta ennen kaikkea genretietoutta eli tietoutta siitä, millaisia eri genret ovat rakenteeltaan ja kielellisiltä keinoiltaan kussakin kulttuurissa.

Tekstien autenttisuuden arvioimiseksi tutkimuksessa ehdotetaan viitekehystä, jossa yhtäältä tarkastellaan tekstin aiheen kiinnostavuutta ja sen relevanssia kohderyhmän näkökulmasta ja toisaalta tekstilajin–genren–kulttuurista tyypillisyyttä sekä kansallisen kulttuurin että sen alakulttuurien piirissä. Aiheen kiinnostavuuden arvioinnissa korostuu lukijoiden nykyhetki, mutta aiheen relevanssia arvioitaessa myös kohderyhmän tulevaisuuden tekstinäkymät tulee ottaa huomioon. Genren tyypillisyyden arviointi toisaalta mahdollistaa sekä tekstin edustaman tekstilajin tyypillisyyden arvioimisen että tekstin rakenteen ja siinä käytettyjen kielellisten keinojen tyypillisyyden arvioimisen. Tässä kunkin osallistujamaan edustajat ovat keskeisessä asemassa oman kulttuurinsa ja sen tekstien parhaina tuntijoina.

REFERENCES

Anastasi, A. 1986. Evolving concepts of test validation. Annual Review of Psychology, 37, 1 – 15.

Angoff, W. H. 1988. Validity: an evolving concept. In H. Wainer & H. I. Braun (eds) Test validity. Hillsdale: Erlbaum.

APA 1974. American Psychological Association 1974. Standards for educational and psychological tests. Washington DC: APA.

Arffman, I. 2004. Kansainvälisten lukukokeiden kääntäminen ja kulttuurinen sovittaminen. In P. Linnakylä, S. Sulkunen & I. Arffman (eds) Tulevaisuuden lukijat. Suomalaisnuorten lukijaprofiileja. PISA 2000. Jyväskylä: University of Jyväskylä, Institute for Educational Research, 49–71.

Bachman, L. F. 1990. Fundamental considerations in language testing. Oxford: Oxford University Press.

Bachman, L. F. & Palmer, A. S. 1996. Language testing in practice: Designing and developing useful language tests. Oxford: Oxford University Press.

Bachman, L. F. 2000. Modern language testing at the turn of the century: assuring that what we count counts. Language Testing, 17 (1), 1 – 42.

Banerjee, J. & Luoma, S. 1997. Qualitative approached to test validation. In C. Clapham & D. Corson (eds) Encyclopedia of Language and Education. Volume 7: Language Testing and Assessment. Dordrecht: Kluwer Academic Publishers, 275 – 287.

Barton, D. 1994. Literacy: An introduction to the ecology of written language. Oxford: Blackwell Publishers.

Barton, D. & Hamilton, M. 2000. Literacy practices. In D. Barton, M. Hamilton & R. Ivanic (eds.). Situated literacies. Reading and writing in context. London: Routledge.

Beach, R. & Hynds, S. 1991. Research on reader response. In R. Barr, M. L. Kamil, P. Mosenthal & P. D. Pearson (Eds), Handbook of Reading Research (vol. 2). New York: Longman, 453–489.

Berkenkotter, C. & Huckin, T. N. 1995. Genre knowledge in disciplinary communication: Cognition/culture/power. Hillsdale: Erlbaum.

Bhatia, V. 1993. Analyzing genre. Language use in professional settings. Harlow: Longman, Pearson Education.

Bhatia, V. 2004. Worlds of written discourse: a genre-based view. London: Continuum.

Biber, D. 1989. A typology of English texts. Linguistics 27, 3 – 43.

Bonnet, G. 2002. Reflections in a critical eye [1]: on the pitfalls of international assessment. Assessment in Education, 9 (3), 387 – 398.

Bopry, J. 1999. A discussion of authenticity with examples taken from portfolios on the web. In P. Linnakylä, M. Kankaanranta & J. Bopry (eds) Portfolioita verkossa – Portfolios on the web. Jyväskylä: Institute for Educational Research, 78 – 96.

200

Borg, W. R. & Gall, M. D. 1989. Educational research. Introduction. (5th edition) New York: Longman.

Breen, M.P. 1985. Authenticity in the language classroom. Applied Linguistics, 6, 1.

Brown, J. S., Collins, A. & Duguid, P. 1989. Situated cognition and the culture of learning. Educational Researcher, 18 (1), 13–20.

Brozo, W.G. 2002. To be a boy, to be a reader: Engaging teen and preteen boys in active literacy. Newark: International Reading Association.

Bruner, J. 1986. Actual minds, possible worlds. Cambridge, MA: Harvard University Press.

Callaghan, M., Knapp, P. & Noble, G. 1993. Genre in practice. In B. Cope & M. Kalantzis (eds) The powers of literacy: A genre approach to teaching writing. London: Falmer Press.

Canale, M. & Swain, Merrill 1980. Theoretical bases of communicative approaches to second language teaching and testing. Applied Linguistics 1 (1), 1–47.

Collins, A. 1994. Goal-based scenarios and the problem of situated learning: A commentary on Andersen Consulting's design of goal-based scenarios. Educational Technology, 34 (11), 30 – 32.

Collins, A., Brown, J.S. & Newman, S.E. 1994. Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In L.B. Resnick (ed.) Knowing, learning and instruction. Hillsdale: Erlbaum.

Cope, B. & Kalantzis, M. 1993. Introduction: How a genre approach to literacy can transform the way writing is taught. In B. Cope & M. Kalantzis (eds) The powers of literacy: A genre approach to teaching writing. London: Falmer Press, 1–21.

Cope, B. & Kalantzis, M. (eds.) 2000. Multiliteracies. Literacy learning and the design of social futures. London: Routledge.

Cosgrove, Judith 2003. A comparison of the PISA 2000 assessment and curricula for reading and mathematics in Ireland. Paper presented at National Project Managers' Meeting, Prague, December 8, 2003.

Davies, A. 1988. Communicative language testing. In A. Hughes (ed.) Testing English for university study: ELT documents 127. Oxford: Modern English Publications.

Darlington, R. B. 1997. Factor Analysis. [On-line] Available at:
http://comp9.psych.cornell.edu/Darlington/factor.htm

Dillon, D. R., O'Brien, D. G. & Heilman, E. E. 2000. Literacy research in the next millennium: From paradigms to pragmatism and practicality. Reading Research Quarterly, 35 (1). Newark: International Reading Association.

Dudley-Evans, T. 1994. Genre analysis: an approach to text analysis for ESP. In M. Coulthard (ed.) Advances in written text analysis. London: Routledge, 219–232.

Dunbar, S. B., Koretz, D. M. & Hoover, H. D. 1991. Quality control in the development and use of performance assessment. Applied Measurement in Education, 4, 289 – 303.

Edelson, Daniel C. 1998. Realising authentic science learning through the adaptation of scientific practice. In B. J. Fraser & K. G. Tobin (eds) International Handbook of Science Education.. London: Kluwer Academic, 317–331.

Eggins, S. & Martin, J. R. 1997. Genres and registers of discourse. In T. van Dijk (ed.) Discourse as Structure and Process. London: Longman, 230–256.

Elley, B. W. 1992. How in the world do students read? IEA.

Fairclough, N. 1989. Language and power. London: Longman.

Fairclough, N. (ed.) 1992. Critical language awareness. London: Longman.

FC 2004 = National framework curriculum for the Finnish comprehensive school. 2004. Helsinki: National Board of Education.

Framework and specifications for PIRLS assessment 2001. 2000. International Study Centre, Lynch School of Education, Boston College.

Frechtling, J. & Sharp, L. (Eds.) 1997. User-friendly handbook for mixed method evaluations. The National Science Foundation (USA). Westat. Available at http://www.nsf.gov/pubs/1997/nsf97153/start.htm

Frederiksen, J. R. & Collins, A. 1989. A systems approach to educational testing. Educational Researcher, 18 (9), 27 – 32-

Garcia, M. W. & Verville, K. 1994. Redesigning teaching and learning: The Arizona student assessment program. In S. W. Valencia, E. H. Hiebert & P. P. Afflerbach (eds) Authentic reading assessment: Practices and possibilities. Newark: International Reading Association, 228–246.

Gee, J. P. 1990. Social linguistics and literacies: Ideology in discourses. New York: Falmer press.

Gee, J. P. 2000a. Discourse and sociocultural studies in reading. In M. L. Kamil, P. B.Mosenthal, P. D. Pearson & R. Barr (eds), Handbook of Reading Research (vol. 3). New Jersey: Erlbaum, 195–207.

Gee, J. P. 2000b. The new literacy studies: from 'socially situated' to the work of the social. In D. Barton, M. Hamilton & R. Ivanic (eds.). Situated literacies. Reading and writing in context. Routledge, London, 180–196.

Gee, J. P. 2003. Opportunity to learn: a language-based perspective on assessment. In D. Johnson & G. Kress (eds) Assessment in Education, 10 (1), 27–46.

Goodman, K. S. 1985. Unity in reading. In H. S. Singer & R. B. Ruddell (eds) Theoretical models and processes of reading. (3rd edition.) Newark: International Reading Association, 813–840.

Gottlieb, M. 1997. A peek into portfolio practices. In A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (eds) Current developments and alternatives in language assessment – Proceedings of LTRC 96. Jyväskylä: University of Jyväskylä and University of Tampere.

Gough, P. B. 1985. One second of reading: Postscript. In H. Singer & R. B. Ruddell (Eds), Theoretical models and processes of reading (3rd ed.). Newark: International Reading Association, 687–688.

Greene, J. C. & Caracelli, V. J. 2003. Making paradigmatic sense of mixed methods practice. In A. Tashakkori & C. Teddlie (eds) Handbook of mixed methods in social and behavioural research. Thousand Oaks, CA: Sage.

Halliday, M. A. K. 1978. Language as social semiotic. London: Edward Arnold.

Halliday, M. A. K. 1985. An introduction to functional grammar. London: Edward Arnold.

Hambleton, R. K. 2002. Adapting achievement tests into multiple languages for international assessments. In A. C. Porter & A. Gamoran (eds) Methodological advances in cross-national surveys of educational achievement. Washington, DC: National Academy Press, 58–76.

Hamilton, M. & Barton, D. 1999. The international adult literacy survey: What does it really measure? Centre for Language in Social Life. Working Paper Series 107. Literacy Research Group. Lancaster University.

Hancock, J., Turbill, J & Cambourne, B. 1994. Assessment and evaluation of literacy learning. In S. W. Valencia, E. H. Hiebert & P. P. Afflerbach (eds) Authentic reading assessment: Practices and possibilities. Newark: International Reading Association, 46–62.

Hansen, J. 1994. Literacy Portfolios: Windows on potential. In S. W. Valencia, E. H. Hiebert & P. P. Afflerbach (eds) Authentic reading assessment: Practices and possibilities. Newark: International Reading Association, 26–40.

Harjunen, E. 2007. Kurkistus kansainväliseen lukutaitotutkimukseen: aidon oloisia lukutehtäviä käännöksen käännöksistä? In S. Grünthal & E. Harjunen (eds) Näkökulmia äidinkieleen ja kirjallisuuteen. Tietolipas 217. Helsinki: Suomalaisen kirjallisuuden seura, 201–224.

Harman, H. H. 1967. Modern factor analysis. Chicago: University of Chicago Press.

Harris, T.L. & Hodges, R.E. (eds) 1995. The literacy dictionary. The vocabulary of reading and writing. International Reading Association.

Heath, S. B. 1980. The functions and uses of literacy. Journal of Communication. Winter, 123–133.

Hiebert, E. H., Valencia, S. W. & Afflerbach, P. P. 1994. Definitions and perspectives. In S. W. Valencia, E. H. Hiebert & P. P. Afflerbach (eds) Authentic reading assessment: Practices and possibilities. Newark: International Reading Association, 6–21.

Holopainen, M. & Pulkkinen, P. 2003. Tilastolliset menetelmät. Helsinki: WSOY.

Honebein, P. C., Duffy, T. M. & Fishman B. J. 1993. Constructivism and the design of learning environments: Context and authentic activities for learning. In T.M. Duffy, J. Lowyck and D.H. Jonassesn (eds) Designing environments for constructive learning. NATO ASI Series F: Computer and system sciences 105. Berlin: Springer-Verlag, 87 – 108.

Huerta-Marcias, A. 1995. Alternative assessment: responses to commonly asked questions. TESOL Journal, 5, 8 – 11.

Huhta, A. 1993. Teorioita kielitaidosta – Onko niistä hyötyä testaukselle? In S. Takala (ed.) Oral proficiency and its assessment. University of Jyväskylä.

Institute for Educational Research. Publication series B. Theory into practice 77.

Kádár-Fulöp, J. 1985. The reading survey. Evaluation in Education 9, 117–153.

Kankaanpää, S. 2001. From letters to news reports: Diachronic changes in Finnish municipal press releases 1979 – 1999. In W. Vagle & K. Wikberg (eds) New direction in Nordic text linguistics and discourse analysis: Methodological issues. Oslo: Novus Forlag, 229–242.

Kankaanpää, S. 2006. Hallinnon lehdistötiedotteiden kieli. Helsinki: Suomalaisen Kirjallisuuden Seura.

Kapinus, B. A., Collier, G. V. & Kruglanski H. 1994. The Maryland school performance assessment program: A new view of assessment. In S. W. Valencia, E. H. Hiebert & P. P. Afflerbach (eds) Authentic reading assessment: Practices and possibilities. Newark: International Reading Association, 255–276.

Karvonen, P. 1995. Oppikirjateksti toimintana. Suomalaisen Kirjallisuuden Seura, Helsinki.

Kauppinen, A. 2005. Kieli – selviytymisstrategia. Virittäjä 1/2005, 83 – 86.

Kern, R. 1995. Redefining the boundaries of foreign language literacy. In C. Kramsch (ed.) Redefining the boundaries of language study. Boston: Heinle & Heinle, 61–98.

Kirsch, I. & Mosenthal, P. 1989 – 1991. Understanding documents. A monthly column. Journal of Reading. Newark, DE: International Reading Association.

Kohonen, V. 1997. Authentic assessment as an integration of language learning, teaching, evaluation and the teacher's professional growth. In A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (eds) Current developments and alternatives in language assessment – Proceedings of LTRC 96. Jyväskylä: University of Jyväskylä and University of Tampere.

Komppa, J. 2006. Tiedotteen rakenteen potentiaalit. In A. Mäntynen, S. Shore & A. Solin (eds) Genre – tekstilaji. Tietolipas 213. Helsinki: Suomalaisen Kirjallisuuden Seura, 303–326.

Kress, G. 1993. Genre as social process. In B. Cope & M. Kalantzis (Eds) The powers of literacy: A genre approach to teaching writing. London: Falmer Press, 22–37.

Krippendorf, K. 1980. Content analysis: An introduction to its methodology. Beverly Hills, CA. Sage.

Kunelius, R. 1993. Uskottavuuden anatomia. Kova uutinen, genre ja kansalainen. Tiedotustutkimus, 16 (2), 33 – 45.

Kunelius, R. 1996. The news, textually speaking. University of Tampere. Acta Universitatis Tamperensis 520.

Kyngäs, H. & Vanhanen, L. 1999. Sisällön analyysi. Hoitotiede 11 (1), 3 – 12.

LaBerge, D. & Samuels, S. J. 1985. Toward a theory of automatic information processing in reading: Updated. In H. Singer & R. B. Ruddell (Eds), Theoretical models and processes of reading (3rd ed.). Newark: International Reading Association, 719–721.

Langer, J. 1995. Envisioning literature. Newark: International Reading Association.

Leino, Kaisa 2005. Tietotekniikan käyttö. In P. Kupari & J. Välijärvi (eds) Osaaminen kestävällä pohjalla. PISA 2003 Suomessa. University of Jyväskylä, Institute for Educational Research & Ministry of Education, 173–182.

Leiwo, M. 2005. Miten äidinkielen opetusta voisi tarkastella aksiologisesti ja emansipatorisesti? Puhe ja kieli 25 (4), 175 – 185.

Leontjev, A. A. 1973. Some problems in learning Russian as a foreign language. Soviet Psychology, 11 (4).

Lewkowicz, J. A. 1997. Authentic for whom? Does authenticity really matter? In A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (eds) Current developments and alternatives in language assessment – Proceedings of LTRC 96. Jyväskylä: University of Jyväskylä and University of Tampere.

Lewkowicz, J. A. 2000. Authenticity in language testing: some outstanding questions. Language Testing 17 (1), 43 – 64.

Lie, S. & Roe, A. 2003. Unity and diversity of reading literacy profiles. In S. Lie, P. Linnakylä & A. Roe (eds) Northern lights on PISA. Unity and diversity in the Nordic countries in PISA 2000. University of Oslo, Department of teacher education and school development, 147–157.

Lier, L. van 1996. Interaction in the language curriculum. Awareness, autonomy and authenticity. Harlow: Longman.

Linn, R. L. 1993. Educational measurement. American Council on Education. ORYX Press.

Linn, R.L. 2002. The measurement of student achievement in international studies. In A. C. Porter & A. Gamoran (eds) Methodological advances in cross-national surveys of educational achievement. Washington, DC: National Academy Press, 27–57.

Linn, R. L., Baker, E. L. & Dunbar, S. B. 1991. Complex, performance-based assessment: Expectations and validation criteria. Educational Researcher, 20 (8), 5 – 21.

Linn, R. L. & Gronlund, N. E. 1995. Measurement and assessment in teaching. (7th edition.) New Jersey: Englewood Cliffs.

Linnakylä, P. 1995. Lukutaidolla maailmankartalle. Kansainvälinen lukutaitotutkimus Suomessa. The IEA study of reading literacy in Finland. Jyväskylän yliopisto. Kasvatustieteiden tutkimuslaitos.

Linnakylä, P. 2002. Nuorten lukemisaktiivisuus ja lukuharrastus. In J. Välijärvi & P. Linnakylä (eds) Tulevaisuuden osaajat. PISA 2000 Suomessa. Koulutuksen tutkimuslaitos & Opetushallitus, 141–166.

Linnakylä, P., Kupari P., Törnroos, J. & Reinikainen, P. 2005. Sukupuolierot lukutaidossa, matematiikan ja luonnontieteiden osaamisessa sekä ongelmanratkaisussa. In P. Kupari & J. Välijärvi (eds) Osaaminen kestävällä pohjalla. PISA 2003 Suomessa. Ministry of Education & Institute for Educational Research, 105–114.

Linnakylä, P., Kupari P. & Reinikainen, P. 2002. Sukupuolierot lukutaidossa sekä matematiikan ja luonnontieteiden osaamisessa. In J. Välijärvi & P. Linnakylä (eds) Tulevaisuuden osaajat. PISA 2000 Suomessa. Institute for Educational Research & National Board of Education, 73–88.

Linnakylä, P., Malin, A., Blomqvist, I. & Sulkunen, S. 2000. Lukutaito työssä ja arjessa. Second International Adult Literacy Survey in Finland. National Board of Education & Institute for Educational Research.

Linnakylä, P. & Sulkunen, S. 2002. Millainen on suomalaisten nuorten lukutaito? In J. Välijärvi & P. Linnakylä (eds) Tulevaisuuden osaajat. PISA 2000 Suomessa. Institute for Educational Research & National Board of Education, 9–39.

Linnakylä, P. & Sulkunen, S. 2005. Suomalaisnuorten lukutaito ja –harrastus. In P. Kupari & J. Välijärvi (eds) Osaaminen kestävällä pohjalla. PISA 2003 Suomessa. Ministry of Education & Institute for Educational Research, 37–64.

Linnakylä, P. & Välijärvi, J. 2005. Arvon mekin ansaitsemme. Kansainvälinen arviointi suomalaisen koulun hyväksi. Opetus 2000. Jyväskylä: PS-kustannus.

Luoma, S. 2001. What does your test measure? Construct definition in language test development and validation. University of Jyväskylä, Centre for Applied Language Studies.

Luukka, M.-R. 2003. Tekstitaituriksi koulussa. Virittäjä 3/2003, 413 – 419.

Luukka, M.-R. 2004a. Tulevaisuuden tekstitaitureita? In M.-R. Luukka & P. Jääskeläinen (eds) Hiiden hirveä hiihtämässä: hirveä(n) ihana kirjoittamisen opetus. Äidinkielen opettajain liiton vuosikirja XLVIII. Helsinki: Äidinkielen opettajain liitto, 111 – 119.

Luukka, M.-R. 2004b. Genre-pedagogiikka: askelia tekstitaitojen jatkumolla. In M.-R. Luukka & P. Jääskeläinen (eds) Hiiden hirveä hiihtämässä: hirveä(n) ihana kirjoittamisen opetus. Äidinkielen opettajain liiton vuosikirja XLVIII.Helsinki: Äidinkielen opettajain liitto, 145 – 160.

Luukka, M.-R. & Leiwo, M. 2004. Opetussuunnitelma äidinkielenopetuksen suunnannäyttäjänä? In K. Sajavaara & S. Takala (eds) Kielikoulutus tienhaarassa. University of Jyväskylä, Centre for Applied Language Studies, 13–36.

Marzano, R. J. 1994. Commentary on Hansen's literacy portfolios: Windows on potential. In S. W. Valencia, E. H. Hiebert & P. P. Afflerbach (eds) Authentic reading assessment: Practices and possibilities. Newark: International Reading Association, 41–45.

Martin, J. R. 1993. A contextual theory of language. In B. Cope & M. Kalantzis (Eds) The powers of literacy: A genre approach to teaching writing. London: Falmer Press, 116–136.

Martin, J. R. & Rothery, J. 1993. Grammar: Making meaning in writing. In B. Cope & M. Kalantzis (Eds) The powers of literacy: A genre approach to teaching writing. London: Falmer Press, 137–153.

Maxcy, S. J. 2003. Pragmatic threads in mixed methods research in the social sciences: the search for multiple modes of inquiry and the end of the philosophy of formalism. In A. Tashakkori & C. Teddlie (eds) Handbook of mixed methods in social and behavioural research. Thousand Oaks, CA: Sage.

McNamara, T. P., Miller, J. D. & Bransford, D. L. 1991. Mental models and reading comprehension. In R. Barr, M. L. Kamil, P. Mosenthal & P. D. Pearson (Eds), Handbook of Reading Research (vol. 2). New York: Longman, 490–511.

McQueen, J. & Mendelovits, J. 2003. PISA reading: cultural equivalence in a cross-cultural study. Language Testing 20 (2), 208 – 224.

Messick. S. 1989a. Validity. In R. L. Linn (ed.) Educational measurement. (3rd edition.) Washington DC: American Council on Education and National Council on Measurement in Education, 13–103.

Messick, S. 1989b. Meaning and values in test validation: The science and ethics of assessment. Educational Researcher, 18 (2), 5 – 11.

Messick, S. 1994. The Interplay of evidence and consequences in the validation of performance assessments. Educational Researcher, 23 (2), 13 – 23.

Messick, S. 1995. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50 (9), 741 – 749.

Miles, M. B. & Huberman, A. M. 1994. An expanded sourcebook. Qualitative data analysis. London: Sage publications.

Moessner, L. 2001. Genre, text type, style, register: A terminological maze? European Journal of English Studies, 5 (2), 131 – 138.

Morse, J. M. 2003. Principles of mixed methods and multimethod research design. In A. Tashakkori & C. Teddlie (eds) Handbook of mixed methods in social and behavioural research. Thousand Oaks, CA: Sage.

Moss. P. 1992. Shifting conceptions of validity in educational measurement: Implications for performance assessment. Review of Educational Research, 62 (3), 229 – 258.

Moss, P. 1994. Can there be validity without reliability? Educational Researcher, 23 (2), 5 – 12.

Newmann, F. M., Marks, H. M. & Gamoran, A. 1996. Authentic pedagogy and student perfomance. American Journal of Education, 104, 280–312.

Newmann, F. 1998. Comment on Terwilliger's Semantics, psychometrics, and assessment reform: A close look at "authentic" assessments. Educational Researcher, 27 (6), 19 – 20.

Nummenmaa, T., Konttinen, R., Kuusinen, J. and Leskinen, E. 1996. Tutkimusaineiston analyysi. Porvoo: WSOY.

OECD 1997. Literacy skills for the knowledge society. Further results from the International Adult Literacy Survey. 1997. Human Resources Development Canada & OECD.

OECD 1999. Measuring student knowledge and skills. A new framework for assessment. Organization for Economic Co-operation and Development.

OECD 2000. Literacy in the information age. Final report of the International Adult Literacy Survey. OECD & Statistics Canada.

OECD 2001. Knowledge and skills for life. First results from PISA 2000. Organization for Economic Co-operation and Development.

OECD 2002. Reading for change. Performance and engagement across countries. Results from PISA 2000. OECD.

OECD 2003. The PISA 2003 assessment framework – mathematics, reading, science and problem solving knowledge and skills. OECD.

Onwuegbuzie, A.J. & Teddlie, D.L.G. 2003. Typology of analytical and interpretational errors in quantitative and qualitative educational research. Current issues in Education [on-line], 6 (2). Available: http://cie.ed.asu.edu/volume6/number2/

Palincsar, A.S. 1989. Less charted waters. Educational Researcher, 18 (2), 5 – 7.

Pearson, P. D. 1994. Commentary on Weiss's California's new English – language arts assessment. In S. W. Valencia, E. H. Hiebert & P. P. Afflerbach (eds) Authentic reading assessment: Practices and possibilities. Newark: International Reading Association, 218–227.

Pietilä, V. 1993. Ikkunako maailmaan? Uutisgenre ja uutisen todellisuusvaikutelma. Tiedotustutkimus, 16, 46 – 58.

PISA National project manager's manual for PISA 2000. Australian Council for Educational Research. Unpublished working document.

Pitkänen-Huhta, A. 1999. Vieraalla kielellä lukemisen tutkimus. In K. Sajavaara & A. Piirainen-Marsh (eds) Kielenoppimisen kysymyksiä. University of Jyväskylä, Center for the Applied Linguistics, 259–288.

Pollari, P. 1997. Could portfolio assessment empower EFL learners? Portfolios in the teaching of English as a foreign language in Finnish upper secondary school. In A. Huhta, V. Kohonen, L. Kurki-Suonio & S. Luoma (eds) Current developments and alternatives in language assessment – Proceedings of LTRC 96. Jyväskylä: University of Jyväskylä and University of Tampere.

Porter, A.C. & Gamoran, A. 2002. Progress and challenges in large-scale studies. In A. C. Porter & A. Gamoran (eds) Methodological advances in cross-national surveys of educational achievement. Washington, DC: National Academy Press.

Prais, S. J. 2003. Cautions on OECD's recent educational survey (PISA). Oxford Review of Education, 29 (2), 139–163.

Reunanen, E. 1991. Merkitysympäristö ja uutisgenren säännöt. Tampere University: Publications A76.

Reunanen, E. 1993. Journalismin genret kontekstien konteksteina. Tiedotustutkimus, 16, 21 – 32.

Richards, J., Platt, J. & Weber, H. 1985. Longman Dictionary of Applied Linguistics. Longman.

Roe, A. & Taube, K. 2003. Reading achievement and gender differences. In S. Lie, P. Linnakylä & A. Roe (eds) Northern lights on PISA. University of Oslo, 21 – 38.

208

Rosenblatt, L. 1994. The transactional theory of reading and writing. In R. B. Ruddell, M. R. Ruddell & H. Singer (Eds), Theoretical models and processes of reading (4th edition.). Newark: International Reading Association, 1057–1092.

Rowan, B. 2002. Large-scale, cross-national surveys of educational achievement: Promises, pitfalls, and possibilities. In A. C. Porter & A. Gamoran (eds) Methodological advances in cross-national surveys of educational achievement. Washington, DC: National Academy Press, 321–349.

Ruddell, R. B. & Speaker, R. 1985. The interactive reading process: A model. In H. Singer & R. B. Ruddell (eds), Theoretical models and processes of reading (3rd ed.). Newark: International Reading Association, 751–793.

Rumelhart, D.E. 1980. Schemata: The building blocks of cognition. In R. J. Spiro, B. C. Bruce & W. F. Brewer, Theoretical issues in reading comprehension. Perspectives from cognitive psychology, linguistics, artificial intelligence and education. New Jersey: Erlbaum.

Ryan, G. W. & Bernard, H. R. 2000. Data managements and analysis methods. In N. K. Denzin & Y. S. Lincoln (eds) Handbook of qualitative research. (Second edition.) Thousand Oaks, CA: Sage.

Saukkonen, P. 1984. Mistä tyyli syntyy? Juva: WSOY.

Scribner, S. & Cole, M. 1981. The psychology of literacy. Cambridge, MA: Harvard University Press.

Shavelson, R. J., Baxter, G. P. & Gao, X. 1993. Sampling variability of performance assessment. Journal of Educational Measurement, 30, 215 – 232.

Shohamy, E. & Reves, T. 1985. Authentic language tests: where from and where to? Language Testing, 2 (1), 48–58.

Smith, M.W. & Wilhelm, J.D. 2002. "Reading don't fix no chevys": Literacy in the lives of young men. Portsmouth: Heinemann.

Spence-Brown, R. 2001. The eye of the beholder: authenticity in an embedded assessment task. Language Testing, 18 (4), 463 – 481.

Spolsky, B. 1985. The limits of authenticity in language testing. Language Testing 2, 31 – 40.

Standards 1999. Standards for educational and psychological testing. American Educational Research Association, American Psychological Association and National Council on Measurement in Education.

Street, B. 1996. Multiple perspectives on literacies: Critical literacies and the new literacy studies. Paper presented in the third domains of literacy – conference. September 7. 1996. University of London. Institute of Education.

Sulkunen, S. 2002. Suomalaisten aikuisten vapaa-ajan lukemistottumukset. Lukukokeen laatijan näkökulma. In E. Mattila & E. Olkinuora (eds) Koulutus, kasvatus ja arviointi. Kasvatustieteellisiä arviointitutkimuksia koulutuksen kehittämiseksi. Turun yliopiston kasvatustieteiden tiedekunnan julkaisuja A:198, 43–57.

Sulkunen, S. 2004a. Millaiset tekstit ovat nuorten mieleen? Lukukoetekstien autenttisuus suomalaisnuorten kulttuurissa. In Linnakylä, Sulkunen & Arffman (eds). Tulevaisuuden lukijat. Suomalaisnuorten lukijaprofiileja. PISA 2000. University of Jyväskylä, Institute for Educational Research, 5 – 31.

Sulkunen, S. 2004b. PISAn tekstit opetussuunnitelman näkökulmasta. In Linnakylä, Sulkunen & Arffman (eds). Tulevaisuuden lukijat. Suomalaisnuorten lukijaprofiileja. PISA 2000. University of Jyväskylä, Institute for Educational Research, 33 – 48.

Swales, J. M. 1990. Genre analysis. English in academic and research settings. Cambridge: Cambridge University Press.

Taavitsainen, I. 2001a. Changing conventions of writing: The dynamics of genres, text types, and text traditions. European Journal of English Studies, 5 (2), 139 – 150.

Taavitsainen, I. 2006. Genredynamiikkaa aikaperspektiivissä. In A. Mäntynen, S. Shore & A. Solin (eds) Genre – tekstilaji. Tietolipas 213. Helsinki: Suomalaisen Kirjallisuuden Seura, 266–282.

Tashakkori, A. & Teddlie, C. 1998. Mixed methodology: Combining quantitative and qualitative approaches. Applied Social Research Methods Series (vol 46). Thousand Oaks, CA: Sage.

Taube, K. & Munck, I. 1996. Gender differences at the item level. In H. Wagemaker, K. Taube, I. Munck, G. Kontogiannopoukou-Polydorides & M. Martin (eds) Are girls better readers? Gender differences in reading literacy in 32 countries. Amsterdam: IEA, 53–77.

Taylor, C. 1992. The ethics of authenticity. Cambridge, MA: Harvard University Press.

Taylor, D.L. 2004. "Not just boring stories": Reconsidering the gender gap for boys. Journal of Adolescent and Adult Literacy, 48 (4). Newark: International Reading Association, 290 – 298.

Teasdale, A. & Leung, C. 2000. Teacher assessment and psychometric theory: a case of paradigm crossing? Language Testing, 17 (2), 163–184.

Teddlie, C. & Tashakkori, A. 2003. Major issues and controversies in the use of mixed methods in the social and behavioural sciences. In A. Tashakkori & C. Teddlie (eds) Handbook of mixed methods in social and behavioural research. Thousand Oaks, CA: Sage.

Terwilliger, J. 1997. Semantics, psychometrics, and assessment reform: A close look at "authentic"assessments. Educational Researcher, 26 (8), 24–27.

Trask, R. L. 1997. Student's dictionary of language and linguistics. London: Arnold.

Valencia, S. W., Hiebert, E. H. & P. P. Afflerbach 1994. Realizing the possibilities of authentic assessment: Current trends and future issues. In S. W. Valencia, E. H. Hiebert & P. P. Afflerbach (eds) Authentic reading assessment: Practices and possibilities. Newark: International Reading Association, 286–300.

Valencia, S. W. & Place, N. A. 1994. Literacy portfolios for teaching, learning, and accountability: The Bellevue literacy assessment project. In S. W. Valencia, E. H. Hiebert & P. P. Afflerbach (eds) Authentic reading assessment: Practices and possibilities. Newark: International Reading Association, 134–156.

Valencia, S. & Pearson, P. D. 1987. Reading assessment: Time for a change. The reading teacher, 40 (8), 726–732.

Ventola, E. 2006. Genre systeemis-funktionaalisessa kielitieteessä. Esimerkkinä asiointitilanteet. In A. Mäntynen, S. Shore & A. Solin (eds) Genre – tekstilaji. Tietolipas 213. Helsinki: Suomalaisen Kirjallisuuden Seura, 96 – 121.

Virtanen, T. 1992. Issues of text typology: Narrative – a basic type of text? Text 12 (2), 293 – 310.

Vygotsky, L. 1978. Mind in society. Cambridge, MA: Harvard University Press.

Välijärvi, J., Linnakylä, P., Kupari, P., Reinikainen, P. & Arffman, I. 2002. Finnish Success in PISA – and some reasons behind it. Institute for Educational Research, University of Jyväskylä. National Board of Education.

Välijärvi, J., Kupari, P. & Linnakylä, P. 2002. Oppilaat opiskelunsa laadun ja määrän säätelijöinä. In J. Välijärvi & P. Linnakylä (eds) Tulevaisuuden osaajat. PISA 2000 Suomessa. Institute for Educational Research & National Board of Education, 122–140.

Weiss, B. 1994. California's new English – language arts assessment. In S. W. Valencia, E. H. Hiebert & P. P. Afflerbach (eds) Authentic reading assessment: Practices and possibilities. Newark: International Reading Association, 197–217.

Werlich, E. 1983. Text grammar of English. Heidelberg: Quelle & Mayer.

Widdowson, H.G. 1978. Teaching language as communication. Oxford: Oxford University Press.

Widdowson, H.G. 1979. Explorations in applied linguistics. Oxford: Oxford University Press.

Widdowson, H. G. 1990. Aspects of language teaching. Oxford: Oxford University Press.

Wiggins, G. 1989. A true test: Toward more authentic and equitable assessment. Phi Delta Kappan, November, 703–713.

Wiggins, G. 1998. Comment on Terwilliger's Semantics, psychometrics, and assessment reform: A close look at "authentic" assessments. Educational Researcher, 27 (6), 20–21.

Wood, R. 1993. Assessment and testing. Cambridge: Cambridge University Press.

Wu, W. M. & Stansfield, C. W. 2001. Towards authenticity of tasks in test development. Language Testing, 18 (2), 187 –206.

APPENDIX 1   Example of the questionnaire presented to panellists in 2000.

MOVIE REVIEWS (booklet 1, page 20)

How **typical** is this text of the kinds of texts that you generally read?

5	4	3	2	1	5 = very typical
					4 = fairly typical
					3 = partly typical, partly untypical
					2 = fairly untypical
					1 = very untypical

Explain your rating. What familiar or strange features are there in the text?

APPENDIX 2   Example of the content analysis of the data consisting of written comments by panellists.

Text identification	Panellist identification	Written comment	Reduced thematic units	Final categories
Movie reviews R245	SJ	Movie reviews are usually these longer ones or just a couple of lines. There are movie reviews either in the culture section in a newspaper (the longer ones) or in the TV-section in which case they are shorter. I just realized that to review text typicality one needs to know the context. Where was the text [originally]? Otherwise it is hard to say.	..are longer.. ..are in the culture section in a newspaper or in the TV-section.. ..needs to know the context…	form context of appearance
Movie reviews R245	AK	I read a lot of movie reviews. These are just rather short ones. In newspapers, not to mention movie magazines, reviews are much longer.	…read movie reviews... …are short…are longer… In newspapers..magazines..	genre form context of appearance
Movie reviews R245	KL	Short summary of the plot. Typical in magazines and on the back of books.	..summary of the plot… …in magazines and on the back of books…	structure context of appearance
Movie reviews R245	OL	There are movie reviews in every possible newspaper and magazine, it is thus typical (also) in Finnish culture. The language in the reviews is typical of authentic reviews, for instance "The director fails to make this comedy...even minimally entertaining.."	…in newspaper..and magazine… …language is typical of reviews…	context of appearance language
Movie reviews R245	RM	I read a lot of movie reviews but these are not typical in my opinion. The short summary of the plot reminds me of the plot summaries on the back of video cassettes and reviewing the film is very minimal. I have read much more	…read movie reviews… …summary of the plot…reviewing is minimal…much more	genre structure

		detailed reviews. On the other hand, some video reviews e.g. in Gloria are short like this.	detailed… ...are short... ...in Gloria are like this...	form context of appearance
Movie reviews R245	SN	The reviews could be from a TV-guide. In a very typical manner movies have been introduced in a sentence or two. It is probably most important to introduce the film makers, actors and other facts. Short reviews usually include stars rating the film that are missing here. In a longer article-like review there is typically thematic description of topic and plot, which is totally missing in this review, naturally.	…from a TV-guide… …introduced by a sentence or two…introduce film makers…include stars…thematic description of topic and plot… …short reviews…in a longer article-like…	context of appearance structure form
Movie reviews R245	AR	The texts seem very familiar. I read texts like this almost daily in a newspaper. The topic is typical since I am interested in movies. Also the length of the text is familiar; short reviews. The language and style also seems familiar– typical reviews. Based on short extracts like this one decides what to watch and what not.	…daily in a newspaper… …topic is typical… …length of the reviews… …language seems familiar… …style seems familiar… …short …like this one…	context of appearance topic form language style form

APPENDIX 3   Terminology used in the content analysis.

Topic:	The topic or theme the text is about.
Content:	What is said in the text. In contrast to topic, thematic units referring to content describe and comment in detail on something that is said in the text.
Genre:	The type of text based on the purpose of the text. Often includes references to the context of appearance or use.
Text type:	Text type based on linguistic features only. Usually includes references to Werlich's text types, such as description, narration, etc.
Context:	Refers to thematic units commenting on the text's context of appearance and/or context of use (i.e. the purpose of the text) and/or lack of context. In cases where the context has been commented upon in the context of genre, such a thematic unit has been categorized under 'genre'. If only the context has been commented upon, the category used is 'context'.
Form:	Refers to the length of the text or the type of text in the case of non-linear text ("this table is familiar but I have never read tables on this topic").
Structure:	The elements text consists of.
Language:	Thematic units commenting on the choice of words, (un)familiar expressions, quality of expressions. Refers also to thematic units commenting on the use of foreign language ("Why is this text in Finnish if it is from a hotel in Istanbul?").
Style:	Thematic units commenting on the way things have been said in the text ("This text represents Reader's Digest–style").
Target group:	Thematic units with references to the target group the text might be aimed at; often adults and 15-year-olds are contrasted.
Lay out:	The appearance of the text: the choice of font, the use of bold letters, the quality of pictures, etc.

APPENDIX 4   Example of the main study questionnaire.

MOVIE REVIEWS (page 20 of the PISA booklet)

How **typical is** this text in your life, at school and in your free time,

	Very Untypical	Untypical	Typical	Very typical
a) in terms of its topic?	$\square_1$	$\square_2$	$\square_3$	$\square_4$
b) in terms of its lay out?..	$\square_1$	$\square_2$	$\square_3$	$\square_4$
c) in terms of its language?...	$\square_1$	$\square_2$	$\square_3$	$\square_4$
d) in terms of its genre (e.g. news report, review, poem, textbook)?	$\square_1$	$\square_2$	$\square_3$	$\square_4$

Explain your ratings. What familiar or strange features are there in the text?

How **interesting** is this text in your opinion

	Very boring	Boring	Interesting	Very interesting
a) in terms of its topic?.	$\square_1$	$\square_2$	$\square_3$	$\square_4$
b) in terms of its lay out?...	$\square_1$	$\square_2$	$\square_3$	$\square_4$
c) in terms of its language?	$\square_1$	$\square_2$	$\square_3$	$\square_4$
d) in terms of its genre (e.g. news report, review, poem, textbook))?	$\square_1$	$\square_2$	$\square_3$	$\square_4$

Explain your ratings. What interesting or boring features are there in the text?

APPENDIX 5   Rotation of texts in main study questionnaires.

TABLE 1   Rotation of text clusters in main study questionnaires.

Questionnaire	Text cluster	Text cluster
1	1	2
2	2	3
3	3	4
4	4	5
5	5	6
6	6	7
7	7	1
8	8	9
9	9	8

TABLE 2   Texts included in text clusters and number of respondents per text.

	Text identification	Respondents, total	Respondents, boys	Respondents, girls
Cluster 1	Job application form	51	24	27
	Argumentative texts on graffiti	53	26	27
	Article about rhinoceros	50	24	26
	Article about protecting beaches	49	23	26
Cluster 2	Movie reviews	54	27	27
	Graphs about nuclear waste	53	26	27
	Bicycle assembly instructions	52	26	26
	Information sheet about a company's job centre	49	25	24
	Warranty instructions for a refrigerator	49	25	24
Cluster 3	Floor plan of a library	51	25	26
	Narrative about woman and panther trapped in flood	52	27	25
Cluster 4	Brochure describing optician's work	54	26	28
	Motorbike advertisement	51	24	27
	Article about a quality of men's shirts	51	24	27
	Information sheet for potential exchange students	49	23	26
Cluster 5	Table of news agencies	52	25	27
	Article about NASA texts on drugged spiders	50	24	26
	A sheet from an art store catalogue	50	24	26
	Fable on father's over-protectiveness	48	22	26
	Flight timetable	47	21	26
Cluster 6	Narrative about villagers' first encounter with movies	52	26	26
	Report on children's attitude toward housework	50	25	25
	Article about using DNA in police investigations	48	23	25

	Telephone use instructions	46	22	24
Cluster 7	Etiquette for exchange students	51	26	25
	Table from a report of an aid agency	50	25	25
	Argumentative texts about space research	51	26	25
	Text on conquering the South Pole	47	24	23
Cluster 8	Article about choosing runner shoes	49	26	23
	Instructions on preparing for a job interview	49	26	23
	Editorial on problems arising with new fertilization technology	46	24	22
	Instruction for contacting employer	46	25	21
	News texts about allergies and explorers	44	24	20
Cluster 9	Information sheet about a flu vaccination	49	27	22
	Extract from a play	48	26	22
	Tree diagram of structure of labor force	47	25	22
	Graphs about Lake Chad's altitude and fauna	46	24	22

218

APPENDIX 6 Examples of texts used in the PISA 2000 reading literacy test (Finnish version).

graffiti

Kiehun kiukusta, kun koulun seinää puhdistetaan ja maalataan jo neljättä kertaa, jotta graffiteista päästäisiin eroon. Luovuus on ihailtavaa, mutta itseään voisi toteuttaa myös tavoilla, jotka eivät aiheuta yhteiskunnalle ylimääräisiä kustannuksia.

Miksi pilaatte nuorison maineen maalaamalla graffiteja kiellettyihin paikkoihin? Ammattitaiteilijatkaan eivät ripusta taulujaan katujen varsille, vai mitä? Sen sijaan he etsivät rahoitusta ja hankkivat mainetta luvallisilla näyttelyillä.

Mielestäni rakennukset, aidat ja puistonpenkit ovat taideluomuksia sinällään. On todella säälittävää pilata arkkitehtuuria graffiteilla, ja kaiken lisäksi maalausmenetelmä tuhoaa otsonikerrosta. En todellakaan ymmärrä, miten nämä kriminaalitaiteilijat viitsivät, sillä heidän "taideteoksensa" vain poistetaan näkyvistä kerta toisensa jälkeen.

Helga

Makuasioista ei voi kiistellä. Yhteiskunta on täynnä viestintää ja mainostamista. Yritysten logoja, kauppojen nimiä. Suuria tyrkyttäviä julisteita katujen varsilla. Ovatko ne hyväksyttäviä? Yleensä kyllä. Ovatko graffitit hyväksyttäviä? Jonkun mielestä kyllä, toisen mielestä ei.

Kuka maksaa graffitien hinnan? Kuka lopulta maksaa mainosten hinnan? Oikein. Kuluttaja.

Ovatko plakaattien laittajat kysyneet sinulta pystytyslupaa? Eivät. Pitäisikö graffitimaalarien sitten kysyä? Eikö kyse ole vain viestinnästä - omasta nimestä, jengien nimistä ja suurista kadunvarsiteoksista?

Ajatellaanpa ruutu- ja raitakuviovaatteita, jotka ilmestyivät kauppoihin muutama vuosi sitten. Tai hiihtoasuja. Niiden kuviot ja väritykset on suoraan varastettu kukkivista betoniseinistä. On varsin huvittavaa, että nämä kuvioinnit ja väritykset hyväksytään ja niitä jopa ihaillaan, mutta samalla samantyylisiä graffiteja kauhistellaan.

Taiteella on kovat ajat.

Sofia

Alla oleva puudiagrammi osoittaa erään maan työvoiman tai "työikäisen väestön" rakenteen. Vuonna 1995 maan kokonaisväestö oli noin 3,4 miljoonaa.

Työvoiman rakenne 31. maaliskuuta 1995 päättyneenä vuonna (000)[1]


```
                     ┌─────────────────────────┐
                     │   Työikäinen väestö [2]  │
                     │        2656,5            │
                     └─────────────────────────┘
              ┌──────────────┴───────────────────────────────┐
   ┌─────────────────────────┐                    ┌─────────────────────────────┐
   │  Työvoimaan kuuluvat    │                    │ Työvoimaan kuulumattomat[3] │
   │  1706,5       64,2%     │                    │  949,9          35,8%       │
   └─────────────────────────┘                    └─────────────────────────────┘
        ┌────────────┴─────────────┐
 ┌──────────────────┐     ┌──────────────────┐
 │    Työlliset     │     │    Työttömät     │
 │  1578,4   92,5%  │     │  128,1    7,5%   │
 └──────────────────┘     └──────────────────┘
    ┌──────┴──────────┐        ┌──────┴──────────┐
┌──────────────┐ ┌──────────────┐ ┌──────────────┐ ┌──────────────────┐
│ Kokoaikaiset │ │ Osa-aikaiset │ │    Hakee     │ │ Hakee osa-aikaista│
│1237,1  78,4% │ │ 341,3  21,6% │ │ kokoaikaista │ │      työtä        │
└──────────────┘ └──────────────┘ │    työtä     │ │  26,5     20,7%   │
                                   │ 101,6  79,3% │ └──────────────────┘
                                   └──────────────┘
              ┌──────┴──────┐
       ┌──────────────┐ ┌──────────────┐
       │    Hakee     │ │    Ei hae    │
       │ kokoaikaista │ │ kokoaikaista │
       │    työtä     │ │    työtä     │
       │ 23,2   6,8%  │ │ 318,1  93,2% │
       └──────────────┘ └──────────────┘
```
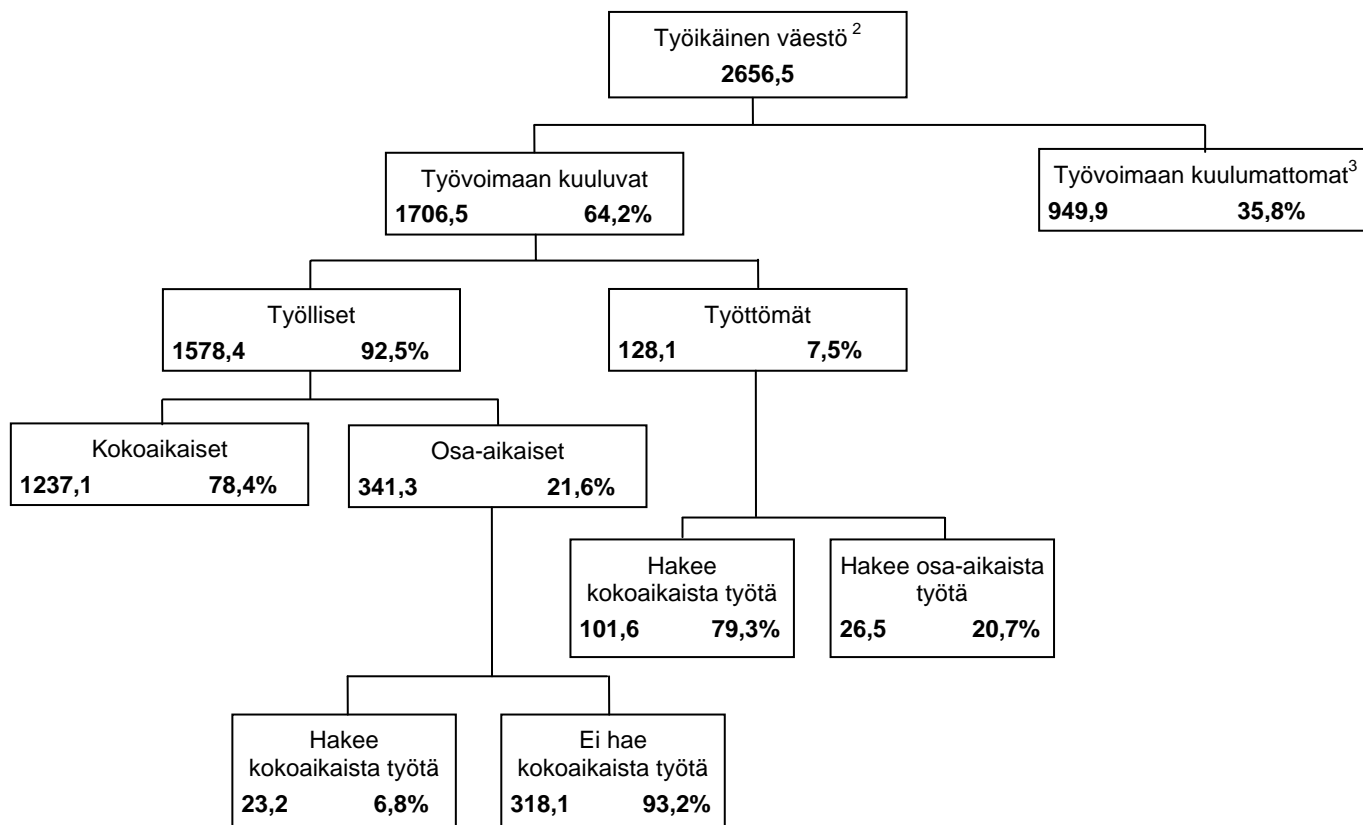
Huomautukset

1. Henkilöiden lukumäärät ilmoitetaan tuhansina (000).

2. Työikäinen väestö määritellään henkilöinä, jotka ovat 15 ja 65 ikävuoden välillä.

3. "Työvoimaan kuulumattomat" henkilöt ovat niitä, jotka eivät aktiivisesti hae työtä ja/tai eivät ole käytettävissä työhön.

TŠADJÄRVI

Kuvio 1 kuvaa Saharassa Pohjois-Afrikassa sijaitsevan Tšadjärven pinnankorkeuden vaihteluita. Tšadjärvi katosi kokonaan noin vuonna 20 000 eKr., viimeisimmän jääkauden aikana. Noin 11 000 eKr. se syntyi uudelleen. Nykyään sen pinta on suunnilleen samalla tasolla kuin se oli 1 000 jKr.
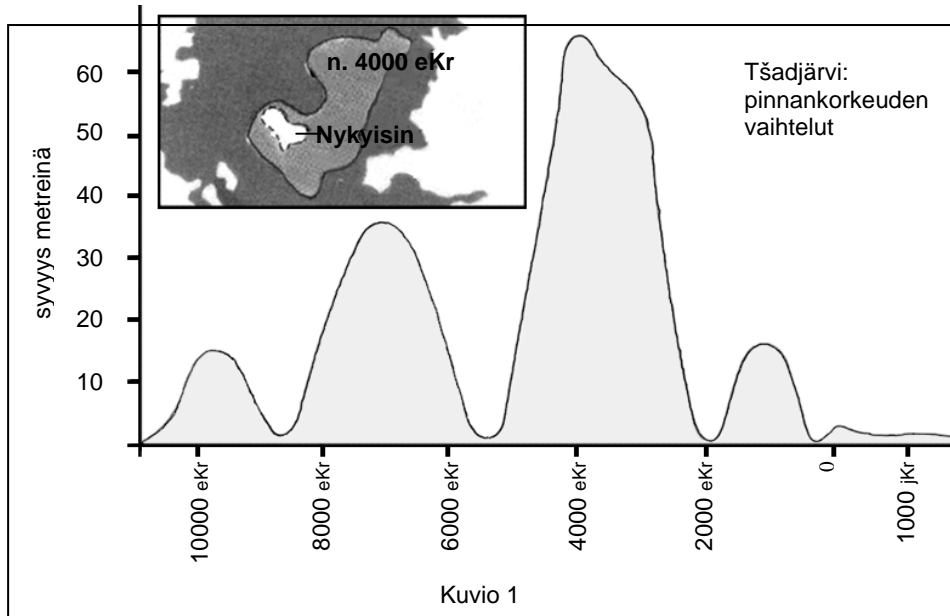
Kuvio 1

Kuvio 2 esittää saharalaisia kalliopiirroksia (luolien seinistä löydettyjä ikivanhoja piirroksia tai maalauksia) ja eläimistön vaihtelua.

Kuvio 2
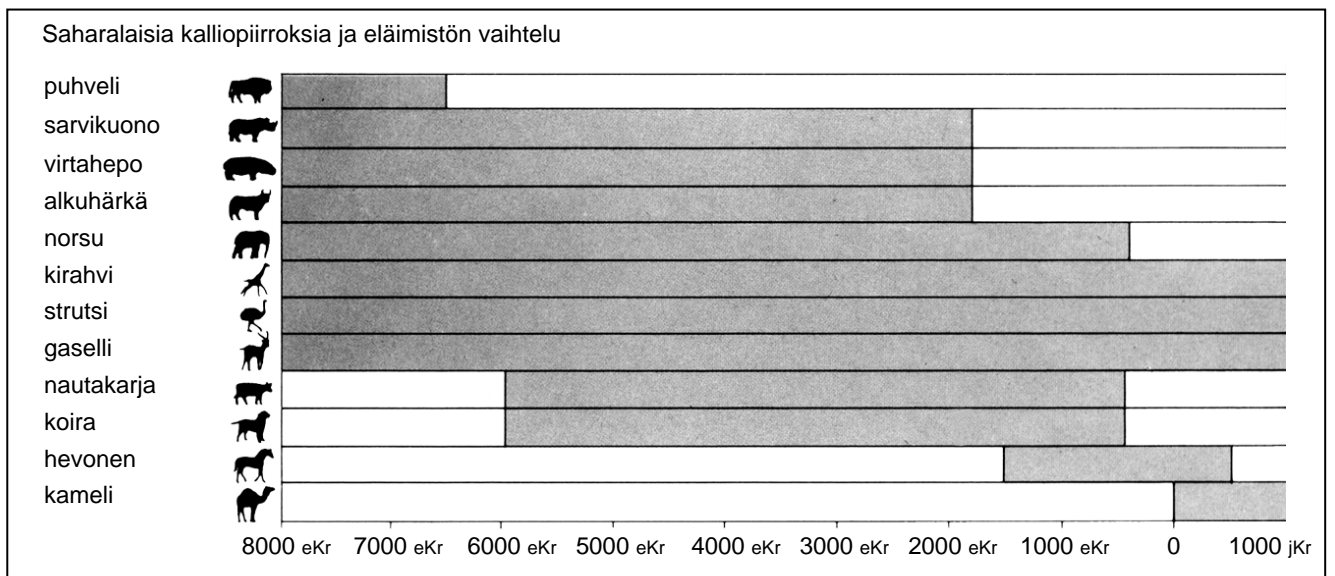
APPENDIX 7   Factor analysis matrices.

Matrices of texts resulting with two factors in factor analysis:

Rotated Factor Matrix: job application.

	Factor	
	interest	typicality
typicality: topic		,70
layout		,92
language		,61
genre		,81
interest: topic	,83	
layout	,82	
language	,77	
genre	,82	

Rotated Factor Matrix: opinions on graffiti.

	Factor	
	interest	typicality
typicality: topic		,52
layout		,79
language		,64
genre		,64
interest: topic	,66	
layout	,65	
language	,77	
genre	,80	

Rotated Factor Matrix: article on rhinos.

	Factor	
	interest	typicality
typicality: topic		,63
layout		,73
language		,88
genre		,82
interest: topic	,80	
layout	,81	
language	,83	
genre	,88	

Rotated Factor Matrix: article on saving beaches.

	Factor	
	typicality	interest
typicality: topic	,61	
layout	,85	
language	,83	
genre	,58	
interest: topic		,53
layout		,64
language		,82

Rotated Factor Matrix: bicycle assembly instructions.

	Factor	
	typicality	interest
typicality: topic	,49	
layout	,77	
language	,81	
genre	,73	
interest: topic		,69
layout		,64
language		,73
genre		,60

Rotated Factor Matrix: information sheet on a company's job centre.

	Factor	
	typicality	interest
typicality: topic	,53	
layout	,73	
language	,69	
genre	,82	
interest: topic		,69
layout		,62
language		,47
genre		,77

Rotated Factor Matrix: warranty instructions for a refrigerator.

	Factor	
	typicality	interest
typicality: topic	,56	
layout	,67	
language	,54	
genre	,94	
Interest: topic		,55
layout		,68
language		,75
genre		,54

Rotated Factor Matrix: library floor plan.

	Factor	
	interest	typicality
typicality: topic		,69
layout		,71
language		,48
genre		,77
interest: topic	,73	
layout	,72	
language	,70	
genre	,88	

Rotated Factor Matrix: brochure about optician's work.

	Factor	
	typicality	interest
typicality: topic	,53	
layout	,58	
language	,82	
genre	,72	
interest: topic		,67
layout		,59
language		,70
genre		,63

Rotated Factor Matrix: article on men's shirts.

	Factor	
	typicality	interest
typicality: topic	,68	
layout	,74	
language	,67	
genre	,77	
interest: topic		,58
layout		,53
language		,76
genre		,73

Rotated Factor Matrix: information sheet to potential exchange students.

	Factor	
	interest	typicality
typicality: topic		,50
layout		,62
language		,75
genre		,73
interest: topic	,76	
layout	,75	
language	,91	
genre	,83	

Rotated Factor Matrix: table on newsagencies.

	Factor	
	typicality	interest
typicality: topic	,88	
layout	,73	
language	,74	
genre	,84	
interest: topic		,66
layout		,76
language		,79
genre		,69

Rotated Factor Matrix: fable.

	Factor	
	interest	typicality
typicality: topic		,74
layout		,74
language		,73
genre		,74
interest: topic	,72	
layout	,74	
language	,84	
genre	,85	

Rotated Factor Matrix: table on aid agency's activities.

	Factor	
	interest	typicality
typicality: topic		,66
layout		,81
language		,80
genre		,61
interest: topic	,77	
layout	,71	
language	,75	
genre	,72	

Rotated Factor Matrix: text on conquering South Pole.

	Factor	
	interest	typicality
typicality: topic		,74
layout		,90
language		,72
genre		,85
interest: topic	,81	
layout	,78	
language	,85	
genre	,87	

Rotated Factor Matrix: instructions for a job interview.

	Factor	
	interest	typicality
typicality: topic		,60
layout		,71
language		,72
genre		,67
interest: topic	,89	
layout	,70	
language	,80	
genre	,78	

Rotated Factor Matrix: information sheet on flu vaccination.

	Factor	
	interest	typicality
typicality: topic		,72
layout		,78
language		,64
genre		,67
interest: topic	,80	
layout	,71	
language	,85	
genre	,75	

Rotated Factor Matrix: extract of a play.

	Factor	
	typicality	interest
typicality: topic	,84	
layout	,85	
language	,77	
genre	,82	
interest: topic		,76
layout		,77
language		,85
genre		,63

Matrices of texts resulting with two factors, some variables having loadings on both factors:

Rotated Factor Matrix: article on drugged spiders.

	Factor	
	interest	typicality
typicality: topic	,41	,42
layout		,94
language		,79
genre		,79
interest: topic	,70	
layout	,87	
language	,76	
genre	,80	

Rotated Factor Matrix: sheet of an art store catalogue.

	Factor	
	typicality	Interest
typicality: topic	,70	,51
layout	,65	
language	,77	
genre	,93	
interest: topic		,52
layout		,70
language		,74
genre		,86

Rotated Factor Matrix: flight time table.

	Factor	
	interest	typicality
typicality: topic		,85
layout	,40	,87
language		,81
genre		,86
interest: topic	,84	,44
layout	,79	
language	,90	
genre	,91	

Rotated Factor Matrix: story of villagers' first encounter with movies.

	Factor	
	typicality	interest
typicality: topic	,63	
layout	,89	
language	,75	
genre	,69	
interest: topic		,79
layout	,41	,58
language		,69
genre		,71

Rotated Factor Matrix: graphs on nuclear waste.

	Factor	
	interest	typicality
typicality: topic		,67
layout		,66
language		,75
genre	,42	,67
Interest: topic	,67	
layout	,68	
language	,72	
genre	,81	

Rotated Factor Matrix: etiquette for exchange students.

	Factor	
	typicality	interest
typicality: topic	,82	
layout	,90	
language	,43	,41
genre	,71	
interest: topic		,69
layout		,61
language		,80
genre	,46	,76

Rotated Factor Matrix: article on choosing runner shoes.

	Factor	
	interest	typicality
typicality: topic		,46
layout		,60
language		,48
genre		,67
interest: topic	,75	
layout	,82	
language	,75	,43
genre	,67	,41

Rotated Factor Matrix: news on allergies and explorers.

	Factor	
	typicality	interest
typicality: topic	,86	
layout	,84	
language	,88	
genre	,62	,45
interest: topic		,74
layout		,61
language		,77
genre		,74

Rotated Factor Matrix: editorial on problems caused by new fertilization technology.

	Factor	
	typicality	interest
typicality: topic	,82	
layout	,85	
langauge	,85	
genre	,68	,45
interest: topic		,63
layout		,61
language		,69
genre		,87

Rotated Factor Matrix: graphs on Lake Chad's altitude an fauna.

	Factor	
	interest	typicality
typicality: topic		,78
layout		,87
language		,75
genre	,44	,67
interest: topic	,86	
layout	,80	
language	,74	
genre	,70	

Rotated Factor Matrix: instructions on contacting an employer.

	Factor	
	interest	typicality
typicality: topic		,66
layout		,78
language		,72
genre		,69
interest: topic	,69	
layout	,59	,41
language	,87	
genre	,85	

Matrices of texts resulting with two factors in factor analysis, variables having high loadings on both factors:

Rotated Factor Matrix: movie reviews.

	Factor	
	typicality	interest
typicality: topic	,83	
layout	,59	,52
language	,83	
genre	,80	
interest: topic	,53	,46
layout		,78
language		,79
genre	,47	,57

Rotated Factor Matrix: story of a woman trapped in the flood.

	Factor	
	interest	typicality
typicality: topic		,49
layout		,87
language	,43	,62
genre		,79
interest: topic	,65	
layout	,58	
language	,92	
genre	,70	,51

Rotated Factor Matrix: motor bike advertisement.

	Factor	
	interest	typicality
typicality: topic		,69
layout		,71
language		,68
genre	,41	,53
interest: topic	,78	
layout		
language	,63	,49
genre	,91	

Matrices of texts resulting with one factor in factor analysis:

Factor Matrix: report on attitudes towards household tasks.

	Factor
typicality: topic	,75
layout	,69
langauge	,61
genre	,78
interest: topic	,79
layout	,85
language	,84
genre	,81

Factor Matrix: article on police using DNA in investigations.

	Factor
typicality: topic	,72
layout	,72
language	,79
genre	,75
interest: topic	,82
layout	,83
language	,84
genre	,89

Factor Matrix: telephone use instructions.

	Factor
typicality: topic	,80
layout	,70
language	,79
genre	,75
interest: topic	,84
layout	,82
language	,83
genre	,85

Factor Matrix: argumentative texts on space research.

	Factor
typicality: topic	,77
layout	,74
language	,78
genre	,74
interest: topic	,76
layout	,75
language	,76
genre	,86

Factor Matrix: tree diagram on a country's labor force.

	Factor
typicality: topic	,63
layout	,86
language	,77
genre	,79
interest: topic	,81
layout	,87
language	,82
genre	,89