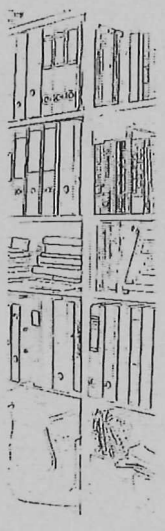
Katholieke Universiteit Leuven

Faculteit Psychologie en Pedagogische Wetenschappen

Departement Pedagogische Wetenschappen



Doctoraatsprojecten

Quantitative and qualitative analysis of reading errors and strategy use in elementary school children assessed by an automated word decoding assessment tool

Cleuren, Leen

16 april 2007

Promotor : Prof. Dr. P. Ghesquière

FACULTEIT PSYCHOLOGIE EN PEDAGOGISCHE WETENSCHAPPEN
DEPARTEMENT PEDAGOGISCHE WETENSCHAPPEN
CENTRUM VOOR ORTHOPEDAGOGIEK



KATHOLIEKE
UNIVERSITEIT
LEUVEN

Quantitative and Qualitative Analysis of Reading Errors and Strategy Use in Elementary School Children Assessed by an Automated Word Decoding Assessment Tool

---

Doctoral Research Proposal
Submitted by **Leen Cleuren**, Research Assistant
Centre for Disability, Special Needs Education and Child Care

March 2007

Promotor	Prof. Dr. Pol Ghesquière
Copromotor	Prof. Dr. Ir. Hugo Van hamme

Begeleidingscommissie 16 april 2007 - Ped. Wetenschappen

Quantitative and Qualitative Analysis of Reading Errors and Strategy Use in Elementary School Children Assessed by an Automated Word Decoding Assessment Tool.

Summary of the doctoral research proposal submitted by Leen Cleuren, research assistant at the Centre for Disability, Special Needs Education and Child care.
January 2007
Promotor: Prof. Dr. Pol Ghesquière
Copromotor: Prof. Dr. Ir. Hugo Van hamme

The present research proposal forms part of a research project called SPACE, which stands for SPeech Algorithms for Clinical and Educational applications. One of its aims is to develop an automated speech recognition-based tool for the assessment of word decoding skills in elementary school children (grade 1 to 4). By improving speech recognition technology in order to make it suitable as a supporting technology for an automated reading tutor, the SPACE project wants to address all disadvantages of traditional word decoding assessment such as the fact that it is time-consuming, its evaluation suffers from examiner bias, and reaction times are measured inaccurately. By means of a speech recognizer, the oral reading assessment can become automated and objective, and accurate measures of the child's reading performance can be obtained.

The *first part* of this research proposal deals with the development of a new computerized test battery for the assessment of children's word decoding skills that consists of a real word reading test (WRT), a pseudo word reading test (PWRT), and a story reading test (SRT). Both the WRT and PWRT contain three lists of respectively 40 1-syllable, 40 2-syllable, and 40 3- or 4-syllable real words or pseudowords. The SRT consists of 9 graded texts stories ranging from AVI 1 to AVI 9. Different versions of the tests have been evaluated, adapted and rewritten; the current version was computer-assessed from 256 regular elementary school children, as such providing us a final evaluation of the test items, and the data needed for the scaling, standardization, reliability assessment and validation of the test battery.

Because of the computerized way of assessment, we have at our disposal detailed information about the timing of stimulus presentation, timing and length of the reader's utterances, and inter-word-latencies (i.e. the time that elapses between words that a student reads). With respect to the WRT and PWRT, timing of stimulus presentation and timing and length of the reader's utterances will be used in attempt to automatically detect the child's reading strategy. With respect to the SRT we will try to automatically capture reading fluency growths by calculating average time-per-(correct) syllable and inter-word latency times.

The *second part* of this research proposal deals with the analysis of the quantitative and qualitative development in reading errors and reading strategies in elementary school children (grade 1 to 4) with and without reading disabilities. On the one hand, we want to increase our understanding of the reading process; features of oral reading errors reveal what the reader chose as the basis for his response in a particular instance and thereby reveal the strategies readers use to identify words. On the other hand, reading error analysis is needed in order to identify a set of likely reading errors that eventually can be modeled in the speech recognizer. If an automated reading tutor would listen for every possible phoneme sequence in place of a correct word, this would result in too many recognition errors, given the limited accuracy of current speech recognizers. Beside this, if the reading tutor would be able to classify reading errors and reading strategies himself, this would make an automatized qualitative analysis of a child's reading performance possible.

To classify oral reading errors and strategies of regular elementary school children (n = 256) and children attending special education of type 8 (n = 150), both assessed by our test battery, we developed a classification scheme containing 48 reading error categories and 9 reading strategy categories. We then want to analyze the relative influence of particular child-related (i.e. *Chronological age*, *Reading ability*, *IQ-reading ability discrepancy*), and reading task-related (i.e. *Relative passage difficulty*, *Pseudowords*, *Context*, *Word length*) variables on reading error occurrence and strategy use. Additionally, we will investigate whether it is possible to automatically distinguish guessing (a reading style characterized by fast and global reading containing lots of reading errors) from spelling (a reading style characterized by slow and fragmented inaccurate reading) behavior on the basis of reading errors and reading rate, or on the basis of reaction and production times in reading.

The *last part* is the methodology chapter that first discusses the selection of participants, the specific data that are collected, and the development of the CHOREC database (CHildren's Oral Reading Corpus), a database of recorded, transcribed and annotated children's oral reading and oral reading errors and strategies. Here, special attention goes out to the assessment of annotation consistency which is a necessary evaluation since the annotation of each sound file is largely done manually by different transcribers. Then, to conclude, the different analyses we want to carry out to answer our research questions are elaborated.

Table of Contents

I. Introduction I-4

A. The SPACE project: SPeech Algorithms for Clinical and Educational applications I-4

B. Situating the research proposal against the SPACE project I-4

1. Traditional oral reading assessment I-4

2. Computerized oral reading assessment I-5

II. Development of a new computerized test battery for the assessment of children's word decoding skills II-7

A. Rationale II-7

B. Achievements II-8

1. Development of a computerized reading tutor assessment platform II-8

2. Development of a test battery to assess elementary school children's word decoding skills II-8

C. Research objectives II-9

1. Scaling, standardization, reliability assessment, and validation of the test battery II-9

a) Scaling and standardization of the WRT and PWRT II-9

b) Reliability assessment and validation of the WRT and PWRT II-10

2. Looking for an alternative measure to capture reading fluency: inter-word latency and time-per-syllable II-10

III. Analysis of the quantitative and qualitative development in reading errors and reading strategies in elementary school children (grade 1 to 4) with and without reading disabilities III-12

A. Rationale III-12

B. Achievements III-13

C. Research objectives III-14

1. Influence of child-related variables on reading error occurrence and strategy use III-14

a) Influence of chronological age III-14

b) Influence of reading ability III-16

c) Influence of IQ-reading ability discrepancy III-16

2. Influence of task-related variables on reading error occurrence and strategy use III-17

a) Influence of relative passage difficulty III-17

b) Influence of pseudowords III-18

c) Influence of context III-18

d) Influence of word length III-18

3. Automatically distinguishing decoding from guessing behavior III-19

IV. Methodology IV-21

A. Data collection IV-21

1. Participants IV-21

a) Phase I IV-21

b) Phase II IV-22
2. Data IV-22
a) Phase I + II IV-22
b) Phase II IV-24
3. Data annotation and formal check of the annotations IV-24
4. Assessment of annotation consistency IV-25
5. Relevant data extraction IV-25
B. Data analyses IV-25
1. Assessment of annotation consistency IV-25
2. Standardization, scaling, reliability assessment and validation of the WRT and PWRT IV-26
a) Speed-accuracy trade-off in the WRT and PWRT IV-26
b) Standardization of the WRT and PWRT IV-26
c) Reliability assessment and validation of the WRT and PWRT IV-27
3. Looking for an alternative measure to capture reading fluency: inter-word latency and time-per-syllable IV-27
4. Influence of child-related and task-related variables on reading error occurrence and strategy use .IV-28
a) Operationalization of the independent variables IV-28
b) Operationalization of the dependent variables IV-28
c) Analyses IV-29
(1) Effects on Speed and Accuracy IV-29
(2) Effects on Reading error occurrence and Reading strategy use IV-30
5. Automatically distinguishing decoding from guessing behavior IV-30
V. Preliminary Overview Research Output V-31
VI. References VI-32
VII. Appendices VII-40
A. Current version of the word reading test, pseudoword reading test and story reading test VII-40
B. Classification of errors VII-45
C. Paper VII-50
D. Poster VII-55

I. Introduction

The present research proposal forms part of a research project called SPACE that aims to develop an automated speech recognition-based tool for the assessment of word decoding skills in elementary school children (grade 1 to 4), and a speech recognition- and speech synthesis-based reading tutor. Therefore, this introductory section briefly presents the SPACE project. Then, the next paragraph explains how this research proposal can be situated *against* the SPACE project, and concisely considers one of the project's core concepts which is of great importance for this proposal: computerized assessment of word decoding skills.

A. The SPACE project: SPeech Algorithms for Clinical and Educational applications

In the SPACE project[1], which is funded by the Flemish government (IWT SBO project), five research groups cooperate: ESAT (K.U.Leuven), ELIS (U.Gent), ETRO (V.U.Brussel), ORTHO (K.U.Leuven), and COM (U.Antwerpen). The main goals of the project are: (1) to improve speech recognition and speech synthesis technologies to make them suitable as supporting technologies for educational and clinical applications, and (2) to demonstrate the benefits of speech technology based tools for these applications, namely for the development of an automated reading tutor (used for assessment and remediation) and for the development of a tool to measure intelligibility of pathological speech (e.g. dysarthria).

In both envisaged applications, one of the basic steps is the subjective evaluation of the user's speech by a teacher or clinician. This is a time consuming step, whose result cannot be replicated and which suffers from examiner bias (the more familiar one is with the speech of a person, the easier it is to understand that person). If one could automate the speech assessment by means of a speech recognizer, one could obtain objective measures of the user's performance. Moreover, in the case of a reading tutor, if the tool can really listen to a child reading aloud, and if natural, highly intelligible and phonetically correct speech can be synthesized, it is also possible to build interactive tools so that the reading tutor can act as a fluent reader model and is able to provide immediate and appropriate feedback to the learning child.

B. Situating the research proposal against the SPACE project

The current research proposal can mainly be situated within the project's aim to develop an automated reading tutor that is able to automatically assess a child's reading level. Therefore, we specifically consider some aspects of traditional and computerized assessment in the next two paragraphs before moving on to the two main topics of this research proposal.

1. Traditional oral reading assessment

In Flanders, elementary school children's progress is regularly assessed in order to detect early learning difficulties such as reading disabilities. With respect to the child's reading development, much attention is paid to the assessment of word decoding skills, as these are crucial for adequate reading.

---

[1] URL: www.esat.kuleuven.be/psi/spraak/projects/space

To accurately and fluently decode printed words in an orthographic language (such as Dutch for example), the reader uses both knowledge of grapheme-phoneme-correspondence rules *and* knowledge of specific word spellings (e.g. of irregular words and polyphonic words) (Gough & Wren, 1998; Nunn, 1998). At the same time, it is important to understand that adequate reading is more than mere recognition of the printed word. More specifically, reading can be seen as a product of decoding (or 'technical' reading) and comprehension. This view is called 'the simple view of reading' (Hoover & Gough, 1990). Nevertheless, although there is much more to reading than word decoding, there is never reading without decoding. When we talk about assessing a child's 'technical' reading level, this means that we specifically focus on a child's mere decoding skills without paying attention to the comprehension aspect.

In the literature, different kinds of instruments to assess one's word decoding skills have already been proposed. A well known way to assess one's word decoding skills is reading aloud (computer presented or paper presented) real words or pseudowords. Besides this, stories (of graded difficulty level) are often presented to the child to evaluate its reading skills beyond the word level. Examples of well-known and commonly used real word tests in Flanders are the 'One-Minute-Test (Brus & Voeten, 1995)', the 'Three-Minutes-Test (Verhoeven, 1995)' and the 'Tests of LVS-CSBO (Dudal, 1997)'. An example of a, in Flanders, commonly used pseudoword test is the 'Klepel (van den Bos, Spelberg, Scheepstra & De Vries, 1994)'. At the sentence level, a well appreciated test in Flanders is the 'AVI-toets (Visser, van Laarhoven & ter Beek, 1994)'. Another way to assess word decoding skills is by means of a lexical decision task where individual words are presented to the child on a computer monitor and the child is asked to classify the word as being a real word or a pseudoword.

2. Computerized oral reading assessment

Although the quality of the paper-and-pencil instruments mentioned above isn't questioned, it is clear that the administration of these tests is very time consuming. Another disadvantage of this way of assessment is that, although usually performed by experienced examiners, the evaluation suffers from examiner bias. Thirdly, the manual use of a chronometer is not a very accurate way to keep track of reaction times.

Computer applications for psycho-educational assessment have primarily consisted of scoring programs designed to convert raw scores into standard scores or percentiles, and of word-processing programs for writing reports. Other applications have included programs designed to analyze, integrate, store or graph assessment data. These applications have focused on computer use *after* the psychological-assessment data have been gathered (Evans, Tannehill & Martin, 1995). A Dutch example of such a so-called expert system is 'Dyslexpert (Blonk, van den Bercken & de Bruyn, 1996)'.

Although such applications will remain important, the greatest potential of computers is their use *during* assessment (Evans et al., 1995). Examples of computerized instruments that want to assess one's reading level are, for example, CRAI (Irvin, 1987), COTAL (van der Leij, Smeets & van Daal, 1990), CARD (Carver, 1991), CAAS (Cisero, Royer, Marchant III & Jackson, 1997), CONDALS (van Aarle & van den Bercken, 1999) and Leeskist (Seys, Van Biervliet & Van Biervliet, 2000).

By improving speech recognition technology in order to make it suitable as a supporting technology for an automated reading tutor, the SPACE project wants to address all disadvantages of traditional word decoding

assessment. By means of a speech recognizer, the oral reading assessment can become automated and objective *and* accurate measures of the child's reading performance can be obtained.

In the next two sections, our research proposal is elaborated. First, we discuss our achievements and plans with respect to the development of a new test battery to assess word decoding, and a second section deals with our plans to analyze children's reading errors and their strategy use behavior.

II. Development of a new computerized test battery for the assessment of children's word decoding skills

A. Rationale

Since we want to develop a *computerized* test battery, we have to take into account the fact that children are asked to read computer presented text instead of traditional paper presented text. It is clear that these two presentation modes differ largely, and that, when using a computer set-up, the influence of screen size, screen quality, screen resolution, display rate, reader-text distance, text-background contrast, posture and orientation of the head, text splitting, etc. on reading fluency and comprehension needs to be taken into account (Belmore, 1985; Bridgeman, Lennon, & Jackenthal, 2003; Chen & Lin, 2004; Dillon, 1992; Dillon, Richardson, & McKnight, 1990; Garland & Noyes, 2004; Gould & Grischkowsky, 1986; Gould et al., 1987; Kerr & Symons, 2006; Muter & Maurutto, 1991; Shieh & Lin, 2000; Wang & Chen, 2003).

Moreover, for the printing of paper presented text, different standards (for letter size, color, contrast, font, line length, etc.) to optimize readability exist; however, for the optimal presentation of text on a computer screen, such standards are still lacking (Bernard, Chaparro, Mills & Halcomb, 2003; De Bruijn, De Mul & Van Oostendorp, 1992; Dyson, 2004; Dyson & Haselgrove, 2001; Garcia & Caldera, 1996; Gould & Grischkowsky, 1986; Hall & Hanna, 2004; Shieh & Lin, 2000; Wang & Chen, 2003). Additionally, since we intend to assess *children's* word decoding skills (instead of adults'), we have to be aware of how computer-related characteristics have their influence on children's reading particularly, and on the attractive power the reading material has for children from grade 1 to 4. Characteristics that seem to be very determining are letter size and type (Bernard, Chaparro, Mills & Halcomb, 2002), but until now little research on children has been carried out.

As a result, since new standards are needed, we aim to develop a whole new test battery for the assessment of children's word decoding skills. At the same time, we try to address some objections often raised with respect to the 'Klepel (van den Bos et al., 1994)' and the 'AVI-toets (Visser et al., 1994)'. These objections include the fact that the pronunciation of particular pseudowords of the 'Klepel' causes confusion, and second, the fact that some stories of the 'AVI-toets' contain vocabulary that sounds very northern Dutch and are therefore not suitable for the Flemish population. At the same time we chose to develop both a real word and a pseudoword test which consist of three lists of words of different length (instead of 1 list containing words of increasing difficulty). This enables us to get a better view of the child's reading performance on a particular level of word difficulty. To conclude, we want to construct a real word and pseudoword test in which each child reads the same number of words so that the time a child needs to read the list varies.

B. Achievements

1. Development of a computerized reading tutor assessment platform

At ESAT[2], a reading tutor platform has been developed with which the test reading material is currently presented to the reader[3] and recordings of the reading performance are carried out. The platform contains a teacher module and a student module. In the teacher module (fig. 1), test items (or in the future, exercises) can be assigned to a particular student whereas in the student module (fig. 2) the actual assessment (or tutoring) happens. Recently, this platform has been extended with an automatic speech recognition server so that recognition can be done 'online'; in the future a speech synthesis server will be added in order to be able to give appropriate feedback and guidance.

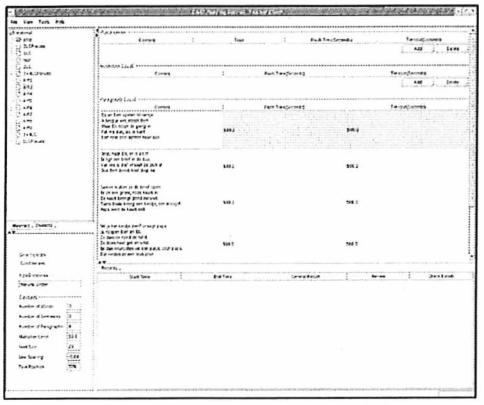


Fig. 1: Teacher module



Fig. 2: Student module

2. Development of a test battery to assess elementary school children's word decoding skills

We have developed three new tests to assess a child's word decoding skills. More specifically, we constructed a real word reading test (WRT), a pseudoword reading test (PWRT) and a story reading test (SRT). Both the WRT and the PWRT contain three lists of respectively 40 1-syllable, 40 2-syllable and 40 3- or 4-syllable real words or pseudowords. The SRT consists of 9 graded text stories (vocabulary of 538 distinct word forms), ranging from AVI 1 to AVI 9.

A detailed description of how first versions of these tests are constructed and evaluated, can be found in the internal report D4.1, downloadable on www.esat.kuleuven.be/psi/spraak/projects/space (click on 'reports', username: guest, password: ECAPS). These different steps imply:

1. Careful construction and selection of the test items.
2. Evaluation of the first version of the WRT, PWRT and SRT by means of a computerized try-out in November 2005 on 60 children attending a regular elementary school in Leuven (grade 2 to 4). Mean reading times and errors were calculated.

[2] Department of Electrical Engineering, Faculty of Engineering, K.U.Leuven.

[3] The child is asked to read accurately and fluently the words and pseudowords or sentences that are presented one by one (or paragraph by paragraph) on the computer screen. The presentation of a new word, pseudoword or paragraph is controlled by the examiner by pressing a key on the computer's keyboard.

3. Evaluation of how the SRT stories' difficulty level might be subjectively experienced. This is done by means of an inquiry that asked a group of reading specialists[4] to score the difficulty level of each text. All 9 stories were, printed in a similar lay-out (same letter size) presented to 25 reading specialists who were asked to intuitively assign an AVI level to each story. From their judgments, mean assigned AVI levels and Cohen's weighted kappa's between each rater and the intended ratings were calculated.
4. Development of a second version of the tests: changed word order of the WRT and PWRT seemed necessary as well as small adaptations to the SRT.
5. Second evaluation of the SRT stories' difficulty level by again presenting the texts to another group of reading specialists. This time, all texts were presented in a similar lay-out containing same letter size, same use of capital letters, and same paragraph structure (so that every new sentence begins on a new line). The reason for changing lay-out characteristics was because the first sample of judgers seemed to enormously rely on the presence of capital letters and paragraph structure. Therefore, because we really wanted to have an indication of subjective difficulty *besides* these controllable factors, we changed the presentation lay-out for the second group of judgers. Again, mean assigned levels were compared and Cohen's weighted kappa's were calculated. However, this evaluation didn't result in new adaptations of the test battery.

C. Research objectives

As explained above, after several versions of the tests were evaluated, adapted and rewritten, the current version was obtained (Appendix A). This version was computer-assessed from 256 children, as such providing us a final evaluation of the test items and providing us with the data for the research objectives outlined below:

1. Scaling, standardization, reliability assessment, and validation of the test battery

The scaling, standardization, reliability assessment and validation of the SRT will *not* form part of this doctoral research project because of its complexity, which is due to the fact that one reading level scale must be constructed. Every story forms part of this scale but does not cover it entirely. The eventual objective is that it is sufficient to administer only a few stories to be able to determine the reading level of the student. Therefore it is important that the 9 stories gradually rise in difficulty level. For the use of the stories by diagnosticians, the scale needs to be divided in 11 parts, from 'lower than AVI 1' to 'higher than AVI 9'. The limits of this scale must be translated to reading times for each story. At the same time, a maximally allowed number of errors must be determined for each story (cf. Visser, 1997).

a) Scaling and standardization of the WRT and PWRT

As is the case in most - if not all - tests which are used in the context of educational measurement, the WRT, PWRT and SRT are partially speed and partially accuracy tests. On the one hand, these are speed tests without a time limit, allowing the subjects to complete the test. On the other hand, accuracy matters too and must there-

[4] Reading specialists are speech therapists, *(ortho)pedagogen* (Flemish terminology), (remedial) teachers, *zorgcoördinatoren* (Flemish terminology), etc. who are experienced in the assessment and scoring of Visser's AVI-test (1997).

fore also be taken into account (Jansen, 1997). As in many types of perceptual-motor tasks, there is a speed-accuracy trade-off, which is a trade-off between how fast a task can be performed and how many mistakes are made in performing the task. That is, a user can either perform the task very fast with a large number of errors or very slowly with very few errors. When asked to perform a task as well as possible, people will apply various strategies that may optimize speed, optimize accuracy, or combine both. For this reason, comparing the performance of two users cannot be done on the basis of speed or accuracy alone, but both values need to be known and the relationship between speed and accuracy needs to be mapped out (Dennis & Evans, 1996).

It is clear then that a measure is needed which captures both speed *and* accuracy. In case of the WRT and PWRT, this measure will serve as the basis for the development of test standards and for the analysis of reading performance distribution within each grade and sex.

b) Reliability assessment and validation of the WRT and PWRT

After having constructed a new assessment instrument, we want to register whether it is a reliable and valid one. In other words, we want the WRT and PWRT to measure consistently and measure what we aim to measure with it (i.e. word decoding ability).

Research Questions

---

1. What measure can be used to capture both speed and accuracy in the WRT and PWRT? How can the accuracy-speed trade-off be characterized?
2. How can the development of scores on the three different lists of the WRT and PWRT be characterized between grades and genders?
3. Are the WRT and PWRT reliable?
4. Are the WRT and PWRT valid instruments?

---

2. Looking for an alternative measure to capture reading fluency: inter-word latency and time-per-syllable

Traditionally, in the paper-and-pencil word decoding assessment instruments being used (such as 'One-Minute-test', 'Three-Minutes-Test' and 'Klepel'), a child is given a fixed amount of time and asked to read a list of words or pseudowords as quickly and as clearly as possible. Afterwards, the total amount of correctly read words is calculated. In case of the AVI-test, the child is asked to read a story as quickly and as clearly as possible; afterwards, the time needed to complete the story and the total amount of errors made is calculated.

Because of the computerized way of assessment, we have other and more detailed information at our disposal, especially with respect to the timing of stimulus presentation, timing and length of the reader's utterance and inter-word latencies (i.e. the time that elapses between words that a student reads). With respect to the WRT and PWRT, timing of stimulus presentation and timing and length of the reader's utterances will be used in an attempt to automatically detect whether a child is reading by unit-by-unit decoding or by direct word recognition (see further on, p. IV-18).

With respect to the SRT, we don't have at our disposal detailed information about the presentation timing of each word since the story is presented paragraph by paragraph. However, we can try to automatically capture reading fluency growths by calculating the average *time-per-(correct) syllable*, or number of (correct) syllables per time unit. Additionally, information on *inter-word latency times* can be extracted too from the recorded data we have gathered. As such, time-per-syllable and inter-word latency could be alternative measures to assess a student's reading ability (reading fluency) and could even be useful measures for tracking the student's improvements automatically without the need for the explicit administration of a test battery. Moreover, in the future, using automatic speech recognition to analyze latency changes between *successive encounters* of words in the same or different exercise contexts will provide new data about how fluency grows (Beck, Jia & Mostow, 2004; Jia, Beck & Mostow., 2002). This kind of research can be related to the research carried out by Cucchiarini and colleagues (Cucchiarini, Strik, & Boves, 1998; 2000; 2002) that tried to determine whether expert speech fluency ratings of read and spontaneous speech can be predicted on the basis of automatically calculated temporal measures of speech quality such as speech rate, articulation rate, number and length of paused, number of dysfluencies, etc.

Research Questions

---

1. Can time-per-(correct) syllable or number of (correct) syllables per time unit be used as a measure for reading fluency?
2. Can inter-word latency in the SRT be used as a measure for reading fluency?

III. Analysis of the quantitative and qualitative development in reading errors and reading strategies in elementary school children (grade 1 to 4) with and without reading disabilities

A. Rationale

The ultimate goal of researchers studying reading disabilities (RD) is to obtain an accurate characterization of the phenotypic performance pattern of children with RD. Therefore, the analysis of reading errors has long attracted the attention of reading researchers. Many have considered it as a useful way to increase our understanding of the reading process (of good, average and weak readers) and as a basis for making decisions about classroom instruction (Chinn, Waggoner, Anderson, Schommer & Wilkinson, 1993; Leu, 1982). That is, features of oral reading errors reveal what the reader chose as the basis for his response in a particular instance and thereby reveal the strategies readers use to identify words; it is one of the few observable manifestations of processing (Biemiller, 1970; Chiappe & Siegel, 1999; Goodman, 1969; Leu, 1982; Stuart & Coltheart, 1988; Thomson, 1986; Weber, 1970).

In the beginning of oral reading error research, the emphasis was mainly put on the visual basis of the process and on orthographic information influencing oral reading errors (Moscicki & Tallal, 1981). There was no clearly articulated framework for analyzing oral reading errors. As a result, surface level error categories were used providing little insight into the cognitive or linguistic processes central to reading. More specifically, oral reading error behavior was described in terms of the frequency of substitutions, omissions, insertions, repetitions, gross and partial mispronunciations, and inversions. Inaccurate responses were regarded as indicative of a visual perceptual deficit in the reader, rather than as evidence of the reader's level of linguistic competence (e.g. Bakker, 1965; for a review see Weber, 1968).

In the late 1960s, the nature of oral reading error analysis fundamentally changed. A number of individuals began to view errors in relation to a theoretical framework that was based on the linguistic sources of information available to the reader (e.g. Barr, 1972; Biemiller, 1970, 1979; Blumenthal, 1980; Clay, 1966; Cohen, 1974-75; Goodman, 1969; Y. Goodman, 1972; Miller & Isakson, 1978; Thomson, 1978; Weber, 1970). Investigators stopped counting error types and began using categories related to the information sources available in the text (syntactical, semantic and graphical information). Reading was then put within the perspective of language in general, and was placed as a process secondary to language (see Leu, 1982, for a review). Oral reading error analyses then differed in the information categories emphasized (Reitsma, 1992). From then on, oral reading errors are often referred to as 'miscues', suggesting that they are mistakenly cued by the cognitive and linguistic systems of readers as they interact with written text, and attention was drawn to syntactically acceptable and unacceptable errors, semantically acceptable and unacceptable errors and graphically similar and dissimilar errors (Moscicki & Tallal, 1981; Leu, 1982). From then on, this theoretical framework has been maintained (e.g. A. Adams, 1990; Blaxall & Willows, 1984; D'Angelo & Mahlios, 1983; Francis, 1984; Hard-

ing, 1984; Harding, Beech & Sneddon, 1985; Leslie, 1980; Moscicki & Tallal, 1981; Willich, Prior, Cumming & Spanos, 1988).

More recently, following greater awareness of the importance of word level factors, such as meta-phonological awareness and word recoding skills in fluent reading (e.g., Gough, Juel & Griffith, 1992; Perfetti, 1985; Stanovich, 1986), researchers have also attempted to develop more detailed accounts of single word reading errors (as was already done by Moscicki & Tallal, 1981). It has been argued, for example, that single word reading errors which are phonetically similar to target words are important as they may suggest the use of processes involving grapheme to phoneme correspondences, whereas pronunciations which are phonologically dissimilar to target words suggest that other processes may have been used to generate word pronunciations (e.g., Chiappe & Siegel, 1999; Coltheart & Leahy, 1992; Laing, 2002; McGuiness, 1997; Reitsma, 1992; Savage, Stuart & Hill, 2001; Sprenger-Charolles & Siegel, 1997; Stuart & Coltheart, 1988).

Within the scope of the SPACE project, we distinguish reading errors versus reading strategies. First, reading error analysis is needed in order to identify a set of likely reading errors that eventually can be modeled in the speech recognizer. Here we want to combine the above mentioned *surface* level error categories *and* the more detailed account of errors at the *phoneme level.* If an automated reading tutor would listen for every possible phoneme sequence in place of a correct word, this would result in too many recognition errors, given the limited accuracy of current speech recognizers (Mostow, Beck, Winter, Wang & Tobin, 2002). Besides this, if the reading tutor would be able to classify reading errors himself, this would make an automatized qualitative analysis of a child's reading performance possible. Secondly, the analysis of the used strategy to come to the correct or incorrect reading aloud of a word can be situated within the above mentioned analysis of influencing *linguistic* information. At that level we are not particularly interested in whether a reading error actually occurred or not, but in the strategy used during reading (whether successfully or not).

B. Achievements

To classify oral reading errors we developed a classification scheme containing 48 categories based on the categories proposed by Bakker (1965). Adjustments are made based on Dumont (1994), Visser et al. (1994), Tal and Siegel (1996), Billiaert and Grysolle (1998) and our own data. Reading errors can occur at four different levels: at the level of the paragraph, the sentence, the word, or the grapheme. At the paragraph level, errors include omissions or repetitions of a whole line or sentence. At the sentence level, a part of a sentence can be omitted or repeated; a word can be mistakenly inserted; a word can change places with another word in the sentence; a word can be substituted for a word being a synonym of that word; or a word can be substituted for a word that is semantically related to the context the original word is embedded in. At the level of the word, reading errors are generally characterized by errors directly resulting from an incorrect compensatory decoding strategy. A decoding strategy can result in letter-by-letter or syllable-by-syllable reading without (correctly) synthesizing; and a direct word recognition strategy can result in the substitution of a word or pseudoword by another word or pseudoword respectively, or the substitution of a pseudoword by a real word. Four main categories of reading errors can be distinguished at the grapheme level: sequential errors, substitution errors, dele-

tion errors, and insertion errors. Sequential errors include changing letter order of adjacent and nonadjacent letters. Substitution errors include vowel substitutions, consonant or consonant cluster substitutions, and vowel-consonant substitutions. Deletion and insertion errors are characterized by vowel deletions c.q. insertions, consonant or consonant cluster deletions c.q. insertions, and syllable deletions c.q. insertions (Appendix B).

With respect to the classification of reading strategies, 4 strategies for arriving at the correct reading of words can be distinguished from 5 not leading to the correct reading of words (9 in total). Strategies for arriving at the correct reading of words, besides correct direct word recognition within the first trial, are: (1) repeating a correctly directly recognized word once or more; (2) partially or completely correctly or incorrectly spelling out (decoding) a word before correctly synthesizing it (revealing a spelling strategy); (3) incorrect direct word recognition in the first trial but correct direct recognition in the final trial (possibly revealing a guessing strategy). Strategies that do not result in the correct reading of words, besides incorrect direct recognition within the first trial (possibly revealing a guessing strategy), are: (1) partially or completely correctly or incorrectly spelling out (decoding) and incorrectly synthesizing a word or not synthesizing it at all (revealing a spelling strategy); (2) correct direct word recognition within the first trial but then 'correcting' it wrongly in a second trial (possibly revealing a guessing strategy); (3) mistakenly omitting or inserting a word; (4) asking for a partial or complete prompt of a word before carrying on reading it.

C. Research objectives

In order to obtain a more accurate characterization of the phenotypic performance pattern of children with reading disabilities and in order to provide the speech recognizer with a better model of likely oral reading errors, our objective is to analyze the relative influence of child-related and reading task-related variables on reading error occurrence and strategy use.

1. Influence of child-related variables on reading error occurrence and strategy use

a) Influence of chronological age

As intuitively expected, more proficient (older) readers read more accurately. Correspondence between the target word and what has actually been read, becomes larger as reading skills develop (e.g., Biemiller, 1970; Francis, 1984; Harding et al., 1985; Moscicki & Tallal, 1981; Verhoeven & van Leeuwe, 2003; Weber, 1970). We can take two perspectives to look at this natural development.

A *first perspective* emphasizes that probably the type of *surface level errors* doesn't change with development. Some do not (e.g., Tal & Siegel, 1996), while others do find relative differences between readers of different age with respect to one of the surface level error categories (e.g. A. Adams, 1990). What does change is the extent to which different sources of information have their influence on oral reading errors, resulting in the fact that developing readers go through different stages of *information use*[5]. Biemiller (1970), expanding the findings of Clay (1968) and Weber (1970), noted that within the first year of instruction there was a shift from the predominate use of context to a non response phase, and then to an increased attention to cues given by first, second and last letters of a word. This was interpreted as a signal of the child's growing awareness of

[5] As mentioned by Biemiller (1979), this debate parallels the debate regarding sight-word versus phonics instruction.

letter function while he was yet unable to express it. When the use of graphic information becomes consolidated, there is a return of the use of context information. Miller and Isakson (1978), however, question this early use of contextual constraints and according to them, evidence shows that there is a gradually growing context use. Moreover, for first graders instructed by a *phonic* method[6], the picture seems to be different (Cohen, 1974-75; Harding et al., 1985). At the level of the *phoneme*, differences between younger and older readers have been noticed too. A study of Moscicki and Tallal (1981) showed that younger readers (6.11 years) make significantly more errors on digraphs than graphs and on graphs in the final position than older readers (8.7 years).

A *second and slightly different way* to look at the development of reading errors and strategy use is from the perspective of Ehri's developmental model of reading (Ehri & McGormick, 1998). She expanded the model of Frith (1985, 1986) that distinguishes three phases within the word decoding learning process: a logographic, alphabetic and orthographic phase. Initially, Ehri adopted this division (Ehri, 1991) but later on, she added a fourth phase between the logographic and alphabetic phase: the partial-alphabetic phase (Ehri, 1995). She also renamed the four phases to pre-alphabetic, partial alphabetic, full alphabetic and consolidated alphabetic (Ehri, 2005; Ehri & McGormick, 1998):

- The *pre-alphabetic phase* characterizes sight word learning at the earliest period. Because children know little about the alphabetic system, they do not form letter-sound connections to read words. If they read words at all, they do it by remembering selected visual features. For example, they might remember *look* by the two eyeballs in the middle, or *dog* by the tail at the end, or *camel* by the humps in the middle (Gough et al., 1992).
- Children progress to the *partial alphabetic phase* when they learn the names or sounds of alphabet letters and use these to remember how to read words. However, they form connections between only some of the letters and sounds in the words, often only the first and final letter sounds, which are easier to detect, for example the letters *s* and *n* to read *spoon* (Savage et al., 2001). They may confuse words such as *spoon* and *skin* having the same boundary letters.
- Children become *full alphabetic phase* readers when they can learn sight words by forming complete connections between letters in spellings and phonemes in pronunciations. This is possible because they know the major grapheme-phoneme correspondences and they can segment pronunciations into phonemes that match up to the graphemes they see.
- The *consolidated phase* emerges as full phase readers retain increasingly more sight words in memory. As they become familiar with letter patterns that recur in different words, the grapheme-phoneme connections in these words become consolidated into larger units.

It is clear that every phase will induce other kinds of error occurrence and strategy use. A disadvantage that has been stated as a point of critique is that Frith's model assumes a particular class of teaching methods, based on

[6] A method used in the teaching of reading, based on recognition of the relationships between individual letters, or groups of letters, and sounds.

explicit phonics instruction (Ramus, 2004).

b) Influence of reading ability

Until now, research didn't provide us yet with evidence that quantitative and/or qualitative differences between normal and poor readers can be expected with respect to *surface level reading categories.* Poor reading children make more errors than normal, reading-level matched control groups but the differences do not reach statistical significance. The types of errors made by poor readers and normal, reading level matched controls are practically indistinguishable (Baddeley, Logie & Ellis, 1988; Bakker, 1965; D'Angelo & Mahlios, 1985; Tal & Siegel, 1996).

According to Biemiller (1979), there exist two different groups of theorists with respect to the use of *linguistic information* in reading. One group of theorists argues that able readers use context more effectively than poor readers while using less and less graphic input; and that effective reading involves perceiving meaning rather than decoding to sound (Au, 1977; Blaxall & Willows, 1984; Cohen, 1974-75; Goodman, 1969; Laing, 2002; Leslie, 1980; Willich et al., 1988). However, evidence is mixed regarding the view that differences in the use of contextual information account for differences in reading ability. Weber (1970) for example found that both good and weak first-grade-readers use constraints of preceding grammatical context. Other contrary evidence is given by Cohen (1974-75), Biemiller (1978, 1979), Thomson (1978), Juel (1980), Stanovich, West & Feeman (1981), Harding (1984) and Harding et al. (1985). A second group of theorists suggest that able readers differ from less able readers in making more effective use of graphic information. In comparison to able readers, poor readers are characterized by for example less graphically similar substitution errors (Weber, 1970), worse graphophonic similarity scores (Adams, 1990; Harding, 1984) and later development of strategies emphasizing graphic information when learning to read (Biemiller, 1970). Contrary evidence for this proposal is given by Au (1977), Clay (1969) and Cohen (1974-75).

c) Influence of IQ-reading ability discrepancy

As pointed out by Leu, since the middle of the 1960's, the analysis of oral reading errors (OREs) has been mainly used to investigate issues related to differences in age or reading ability level (see Leu, 1982, for a review). As such, this kind of research plays an important role in the discussion about whether or not an IQ-discrepancy definition can be used to characterize the phenotypic performance pattern of children with reading difficulties (RD) and whether or not RD can be viewed as being the result of a specific phonological deficit or a general developmental lag. For instance, if the processes lying at the basis of RD of a IQ-discrepant poor reader do not differ from those of a IQ non-discrepant poor reader (and therefore both groups suffer from the same

do not differ from those of a IQ-non-discrepant poor reader (and therefore both groups suffer from the same phonological deficit), then we shouldn't expect there to be qualitative differences in nature of produced OREs and the specific graphemes/phonemes involved in these errors. If it could be demonstrated that a specific phonological deficit distinguishes IQ-discrepant from IQ-non-discrepant poor readers, then research has to show whether this difference can be translated in qualitative and/or quantitative differences in OREs. Furthermore, if we assume that there is no discrepancy between the two groups of poor readers, then it is important to know whether the reading development of children with RD qualitatively differs from that of younger ability-matched normally achieving readers. If there is a qualitative difference, one can ask oneself if this difference will show itself in the kind of OREs made between these two groups of readers.

Until now, research didn't provide us yet with evidence that quantitative and/or qualitative differences between IQ-discrepant and IQ-non-discrepant poor readers can be expected with respect to *surface level reading categories.* They seem to make the same number of errors and the same types of errors in their misreadings (Baddeley et al., 1988; Bakker, 1965; D'Angelo & Mahlios, 1983; Tal & Siegel, 1996). This suggests that reading disability should be defined on the basis of poor reading alone and calls into question the value of using the discrepancy between IQ and reading achievement as a justification for separating reading disabled from generally poor readers. However, other authors do claim to have found error categories such as insertions (Goodman, 1976), omissions (Thomson, 1978) and consonant additions (Werker, Bryson & Wassenberg, 1989) that do distinguish between normal, IQ-discrepant and/or IQ-non-discrepant poor readers.

Research Questions

---

1. How does the occurrence of oral reading errors and the use of reading strategies quantitatively and qualitatively change from grade 1 to 4? Can we distinguish different developmental phases?
2. Are there quantitative and/or qualitative differences to be found between good, average and poor (both IQ-discrepant and IQ-non-discrepant) readers with respect to oral reading errors and strategy use?
3. Can we find any patterns in the relation between reading strategy use and reading error occurrence?

---

2. Influence of task-related variables on reading error occurrence and strategy use

a) Influence of relative passage difficulty

Adequate attention should be paid to the effect of relative passage difficulty on error type and strategy use since it is possible to hypothesize that some apparent differences between different ability groups in many studies are actually generated by their different responses to the same difficulty of text materials (Adams, 1990; Blaxall & Willows, 1984; Leu, 1982). Research did find that difficult stories produce more graphically similar errors and fewer contextually acceptable errors while easy stories have the reverse effect (Adams, 1990;

errors and fewer contextually acceptable errors while easy stories have the reverse effect (Adams, 1990;

Biemiller, 1979; Blaxall & Willows, 1984). However, according to some authors, the difficulty level of text is not a factor in the production of differential rates of basic miscues, types of substitution errors, or reversal errors (Adams, 1990); whereas others did find differences (Blaxall & Willows, 1984). Another point of attention, according to Blaxall & Willows (1984), is that the effect text difficulty has on reading, is not the same for good, normal and poor readers. And a final remark comes from Biemiller (1979) who postulates that children taught using 'meaning emphasis' curricula, shift towards increased use of graphic information when confronted with reading material in which words are harder to identify.

b) Influence of pseudowords

The reading of pseudowords is viewed as the most pure way to measure phonological decoding skills because these words can not be decoded using a direct or lexical route. To read pseudowords, one relies more on phonological processes than to read regular or irregular real words; an indirect sublexical route is followed (Coltheart, 1978). It is now well established that, compared with chronological age controls, children with reading disabilities (RD) display deficits in pseudoword reading, reflecting the difficulty they experience in mapping alphabetic symbols to their corresponding sounds. However, it has been more difficult to determine whether these children with RD display deficits in phonological decoding when compared with a younger group of children without such disabilities, who are matched on reading level (Siegel, 1993; Snowling, 1991; Tal & Siegel, 1996; Wimmer, 1996). At the same time, there's research that doesn't confirm these findings (Szeszulski & Manis, 1987). Other research related to this topic is done by Baddeley et al. (1988), Coltheart & Leahy (1992) and Stanovich & Siegel (1994).

c) Influence of context

Goodman (1965) found that children are able to recognize considerably more words when the words are presented in sentences than in isolation. When real words or pseudowords are presented without context, the user can only use graphic information from the (pseudo)word itself to decode this (pseudo)word using grapheme-phoneme-correspondence rules. When words are presented in a context however, there is contextual information defining the syntactical and semantic restrictions for that particular word (Miller & Isakson, 1978). It seems plausible that the presence of context has a great influence on the nature of the produced oral reading errors. This influence appears from the fact that adults make considerably less errors when reading connected text than when reading series of unrelated words (Reitsma, 1992). Another example of the influence of syntax is that, when a word gets substituted for another word, grammaticality is often preserved (although one has to take into account that this phenomenon is observed too in studies analyzing reading errors in individual words only (cf. Potter, 1980)) (Goodman, 1969; Weber, 1968). More research on the influence of context has been done by Allington & McGill-Franzen (1980), Gough & Wren, 1998, Shankweiler & Liberman (1972), Leslie (1980), West, Stanovich, Feeman & Cunningham (1983), Stanovich, Cunningham & Feeman (1984), and Ben-Dror, Pollatsek & Scarpati (1991).

d) Influence of word length

Word features such as word length and word complexity determine the ease of reading of a particular child. Children make more reading errors when reading longer or more complex words. With respect to word complexity, a study of Moscicki and Tallal (1981) shows that readers from grade 1 and grade 3 make significantly more errors when reading digraphs then when reading monographs. Besides this, part of speech (content versus function words), imageability, and frequency of a word are known to influence the reading performance too (Baddeley et al., 1988).

Research Questions

1. What is the influence of relative passage difficulty on reading error occurrence and strategy use, and how can the interaction with chronological age and reading ability be characterized?
2. What is the influence of the nature of the reading task (real words, pseudowords, words in isolation, and words in context) on reading error occurrence and strategy use, and how can the interaction with chronological age and reading ability be characterized?
3. What is the influence of word length on reading error occurrence and strategy use, and how can the interaction with chronological age and reading ability be characterized?

---

3. Automatically distinguishing decoding from guessing behavior

If children knew only a single strategy for performing a given task, there would be no need to determine how they choose among multiple strategies, much less to examine individual differences in how they choose among the strategies. However, people do use diverse strategies which differ in their accuracy, in the amounts of time needed for execution, and in their memory demands. Wise choices of strategies allow people to meet situational demands and to overcome limited knowledge and processing resources (Siegler, 1988).

However, with respect to reading, there is the major observable distinction between guessing and spelling behavior, which refers to differences in reading style (fast-inaccurate vs. slow-accurate)[7]. Although both strategies can be useful for any beginning reader, children with RD often use them too excessively. When children are guessing, their reading is characterized by a fast and global reading style. Errors are made, such as omissions, additions, substitutions, letter reversals, false word identifications, and other word mutilating errors. When a child is decoding, reading is slow and fragmented, because in that case the identification of words is mainly based on an elaborate grapheme-to-phoneme translation process. It results in accurate reading in that it leaves the ultimate reading response intact (van der Schoot, Licht, Horsley, & Sergeant, 2002). At the word recognition level, the slow-accurate-fast-inaccurate dichotomy has been associated with indirect- versus direct-word approach. In the indirect or phonological route, word identification is attained trough generation of a phonological representation, formed by the stepwise translation of graphemes into phonemes (reading by decoding). The direct or lexical route does not require an intermediate phonological code, since the use of specific orthographic codes enables direct access to word memory (reading by direct word recognition) (Coltheart, 1978; van der Schoot et al., 2004).

To classify children's behavior as guessing or decoding, often criteria (which include measures of dichotic listening) initially developed by Bakker (1981) have been used which cover about 60-65% of the variability in the reading of dyslexics. Grace and Spreen (1994), however, reported that a classification based on reading speed and reading errors is more reliable than Bakker's original method (van der Schoot et al., 2002).

---

[7] Often, a dual-subtype model of dyslexia is based on this dichotomy between guessing and spelling readers such as Bakker's (1979, 1981) L and P types, van der Leij's (1983) guessers and spellers type, Boder's (1970, 1973) dysphonetic and dyseidetic readers, Lovett's (1984) accuracy and rate disabled readers and Mitterer's (1982) whole-word and recoding subtypes.

Analysis of predominant reading strategies can be of great importance for situating readers within a developmental model of reading (Ehri, 2005) or for discriminating readers with reading disabilities from those without (Coltheart, Rastle, Perry, Langdon & Ziegler, 2001). If a reading tutor could do this automatically, analysis of oral reading strategies could give more insight into the quality of a child's reading performance. Therefore, because we want to let our speech recognizer try to automatically detect the reading style a particular child in a particular reading task adopts, we search for objective measures that can be used as a basis to make this qualitative reading behavior analysis. As proposed by Grace and Spreen (1994), reading speed and reading errors could function as these observable measures. Besides this, however, we wonder whether reaction and production times are appropriate too (cf. supra, p. III-9). Additionally, we want to explore if we have other cues (e.g. particular kinds of reading errors) at our disposal we can use to predict a child's reading strategy use.

Research Questions

---

1. Is it possible to automatically distinguish guessing from spelling behavior on the basis of reading errors and reading rate?
2. Can timing of stimulus presentation, and timing and length of a reader's utterances be linked to a particular reading strategy?
3. Are there other cues we can use to predict the reading strategy a child is using?

IV. Methodology

A. Data collection

In spite of the availability of the text being read, automatic speech recognition within the context of reading assessment and instruction is a very challenging task due to reading-related and child-related developmental processes. For example, due to a child's variable maturation, articulatory competencies differ and due to variable word decoding skills, oral reading of novice readers or readers with reading disabilities can be fraught with oral reading errors. In order to improve the speech recognizer's ability to accurately detect reading errors, it is necessary to characterize reading behavior statistically and to provide the recognizer with a model that contains information on the nature and prevalence of likely reading errors. To achieve this goal, the CHOREC (CHildren's Oral REading Corpus) database of recorded, transcribed and annotated children's oral reading and oral reading errors and strategies is being developed (Cleuren, Duchateau, Sips, Ghesquière & Van hamme, 2006). At the same time, however, this database offers us the data we need to carry out the research proposed in this research proposal. With respect to the analysis of reading errors and strategies it even offers us the unique possibility to analyze reading behavior of a rather large group (400 children in total).

1. Participants

Ideally, the research sample is chosen as a representative subgroup of the whole population (in this case: (a) 1st – 4th grade regular elementary school children, and (b) children attending special schools of type 8[8]) so that results can be generalized to this population and the research's external validity is guaranteed. However, because our instruments ask for an individualized assessment, representativeness wasn't feasible; therefore pragmatic considerations were influential when selecting the participating schools.

a) Phase I

The CHOREC database contains reading sessions of Dutch (mother tongue)-speaking regular elementary school children (grade 1-4); until now 256 children are recorded (approximately 75 hours of speech)[9]. Children were selected from grade 1 to 4, meaning that they were 6 to 10 years old. Table 1 represents the number of participating boys and girls for each school and grade.

		Boys	Girls	
1st grade	Mozaïek	8	8	
	HHC	10	8	**59**
	HH	4	8	
	Boven-Lo	6	7	
2nd grade	Mozaïek	6	9	
	HHC	11	7	**69**

[8] Schools that offer individually adapted education to children with severe learning difficulties.

[9] Data was gathered during a three-week period (February 2006) in 4 different regular elementary schools: Gemeentelijke Basisschool Mozaïek in Humbeek, Vrije Basisschool Heilig-Hartcollege in Heist-Op-Den-Berg, Vrije Gem. Lagere Oefenschool Heilig Hart in Heverlee and Vrije Basisschool Boven-Lo in Kessel-Lo. For all children, parental consent was obtained.

	HH	7	7	
	Boven-Lo	13	9	
3rd grade	Mozaïek	7	6	
	HHC	9	10	**64**
	HH	7	10	
	Boven-Lo	5	10	
4th grade	Mozaïek	5	6	
	HHC	8	8	**64**
	HH	14	8	
	Boven-Lo	4	11	
	TOTAL	**124**	**132**	**256**

Table 1: Numbers of participating Phase I-children

b) Phase II

At this moment, CHOREC additionally contains reading sessions of 27 children[10] with reading difficulties. In the near future the database will be extended by including at least another 123 children with reading difficulties or dyslexia[11]. Schools that will be asked to participate are schools for special education of type 8.

2. Data

a) Phase I + II

On the one hand, the data collection encompasses a *questionnaire for the parents and the teachers*. For both the regular school-children (Phase I) and the children in special education (Phase II), the teacher is asked what reading instruction method is used at his or her school, and for each child: whether reading difficulties are present, what the child's AVI-level is, and whether the child has ever doubled or passed over a grade. The parents (for both types of children, Phase I + II) were asked to indicate their child's name, birth place and date, nationality, residence and postal code, previous residence and for how long the child has lived there, mother tongue, other languages spoken by the child, and whether reading difficulties are present or not. Parents in phase II are additionally asked to indicate both parents' educational level. For this, the International Standard Classification of Education (ISCED) system, developed by the UNESCO (OECD, 1999) is used which offers us a marker of the family's socio-economic status.

On the other hand, the data collection consists of making *high-quality two-channel recordings* using the ESAT Reading Tutor assessment platform (Cleuren et al., 2006) of children reading aloud our newly developed test material, and of administering the One-Minute-Test (OMT), the Klepel and the AVI-test to each child. A 30-page data collection protocol is written to make sure this collection is done uniformly by all examiners involved so that the internal reliability of this research can be assured. The protocol includes the letters used to send to the school and to the parents, an explanation of how to specify settings for the Reading Tutor and how

[10] Data was gathered during a one-week period (December 2006) in Eureka Onderwijs, Kessel-Lo (Leuven).

[11] The current definition of dyslexia in Flanders and the Netherlands emphasizes that the diagnosis depends on the observation of severe reading and spelling problems (< Pc 10) *and* on the observation that these problems are persistent and resistant to the usual teaching methods and remedial efforts. In this research we will rely on the official diagnosis established by a particular child's school.

to handle the recording hardware (laptop, amplifier, microphones) so that high quality recordings can be guaranteed. Further, the document explains what reading material each child is supposed to read and in what order, and how the recordings should be organized, saved and back-upped.

For each child, 3 counterbalanced sessions are needed to complete the recordings. In the first session, the OMT, the Klepel and the AVI-test are administered. In this same session, one of the computerized reading tests (WRT, PWRT, SRT) is administered too. In the second and third session, the other 2 computerized tests are administered. Each child reads minimally one and maximally three real word lists and pseudoword lists, and minimally one and maximally four text stories depending on that particular child's reading level.

- *One-Minute-Test* (Brus & Voeten, 1995): The child is asked to read one list (B) consisting of 1-syllable, 2-syllable, and 3- and 4-syllable words (gradually rising in difficulty) for 1 minute. Words are presented in four columns on a piece of paper. The child is asked to read the words as fluently and as accurately as possible.
- *Klepel* (van den Bos et al., 1994): The child is asked to read one list (B) consisting of 1-syllable, 2-syllable, and 3- and 4-syllable pseudowords (gradually rising in difficulty) for 2 minutes. Pseudowords are presented in four columns on a piece of paper. The child is asked to read the pseudowords as fluently and as accurately as possible.
- *AVI-test* (Visser et al., 1994): Based on the results on the OMT and based on the AVI-level assigned by the school [12], the child is asked to read 1 story out of 9 stories of graded difficulty level (AVI 1 – 9 (B) to check whether the child reads that story at control, instructional or frustrational level.
- *WRT*: The child is asked to read three different lists of 1-syllable, 2-syllable, and 3- and 4-syllable words respectively. Words are presented one-by-one on a computer screen by means of an examiner pushing the enter button. The child is asked to read the words as fluently, as accurately and as clearly as possible.
- *PWRT*: The child is asked to read three different lists of 1-syllable, 2-syllable, and 3- and 4-syllable pseudowords respectively. Pseudowords are presented one-by-one on a computer screen by means of an examiner pushing the enter button. The child is asked to read the pseudowords as fluently, as accurately and as clearly as possible.
- *SRT*: Based on the results on the AVI-test[13], the child is asked to read 2 or 3 stories out of 9 stories of graded difficulty level (AVI 1 – 9). The stories are presented paragraph by paragraph on a computer screen by means of an examiner pushing the enter button. The child is asked to read the story as fluently, as accurately and as clearly as possible.

[12] First, an AVI-level is assigned based on the OMT-score. Then, this AVI-level is compared with the AVI-level assigned by the school. When these levels correspond, this is the level that will be assessed. When these levels differ 1 level from each other, the highest level is assessed; when these levels differ more than 1 level; the average level of these 2 is calculated and assessed.

[13] Which SRT-stories the child has to read is based on whether the AVI-story was read on control, instructional or frustrational level.

b) Phase II

The data-collection for the children in special education will additionally include a measure of intelligence in order to be able to distinguish between IQ-reading discrepant children and IQ-reading non-discrepant children. To do so we will administer Raven's Progressive Matrices (Raven, Court & Raven, 1992), a collective non-verbal intelligence test measuring spatial reasoning.

3. Data annotation and formal check of the annotations

The recorded reading sessions have to be transcribed and annotated manually by means of a customized version of 'Praat' (http://www.Praat.org/), a free-source tool for speech analysis and synthesis (developed by Boersma & Weenink, Institute of Phonetic Sciences, University of Amsterdam). This tool provides the possibility to attach a text-grid (containing different tiers) to the speech sound. As such, each tier provides another layer of descriptive information about the speech sound the particular tier is attached to.

The 8 annotation tiers used in the CHOREC database include both information on utterances directly resulting from the child's efforts to read what is presented on the computer screen, as well as information on background noise and reading task-related and -unrelated unforeseen utterances made by both the child and the examiner. Each tier (except for tier 7 and 8) originally contains segment boundaries corresponding to the timing of new stimulus presentation. In case of the WRT and PWRT, boundaries indicate the presentation of a new real word or pseudoword; in case of the SRT, they indicate the presentation of a new paragraph (Cleuren et al., 2006). Figure 3 gives a screen shot of an annotation example.

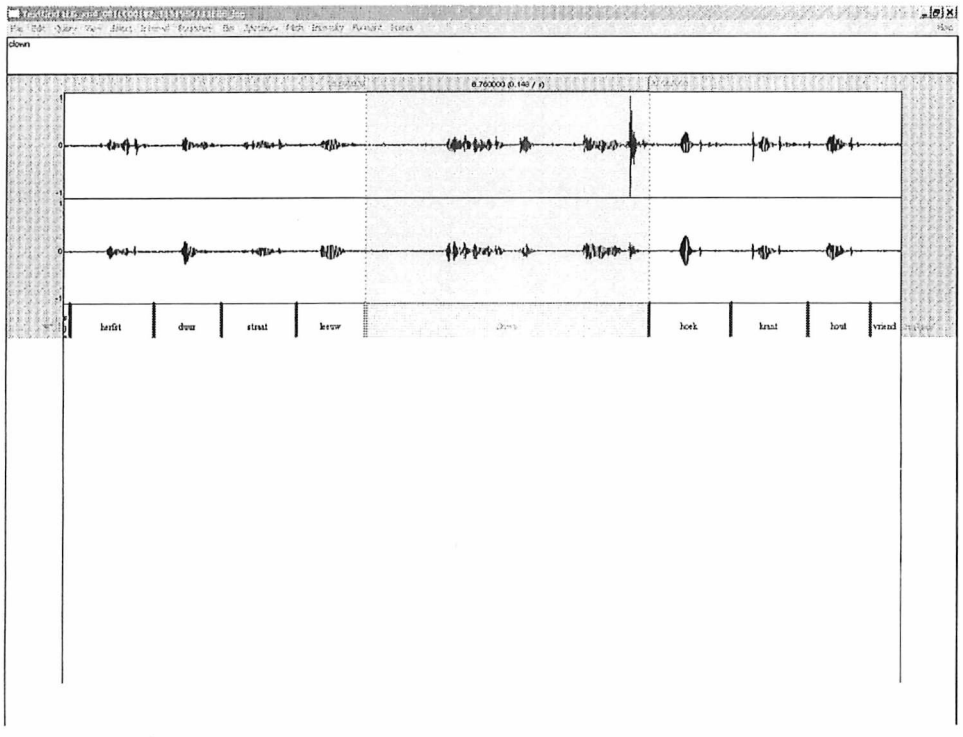


Fig. 3: Screenshot of ‘Praat’

For convenience reasons, the segmentation and annotation/labeling will be carried out in two phases. In a first phase, all tiers except for tier 5 and 6, will be annotated. In a second phase, only for those words the child hesitates or makes a reading error, tier 5 (labeling of oral reading strategies) and 6 (labeling of oral reading errors) will be completed. A clearly written 40-page long detailed annotation protocol makes sure that annotation is done uniformly by all transcribers.

Then, based on the annotation protocol we will carry out a formal check to detect all possible detectable annotation procedure errors such as: number of entities in the different tiers, correctness of phonetic transcription adaptations etc. These checks will be made possible by means of scripts written in Python[14]. Corrections, however, will be made manually.

4. Assessment of annotation consistency

It is clear that, since the annotation of each sound file will be largely done manually by different transcribers, an analysis of transcriber consistency in segmentation and labeling is recommended in order to assure the research's internal reliability. To be able to assess consistency, a specific percentage of the corpus will be annotated by more than one transcriber. Then, to obtain a measure of consistency, we will calculate pair wise inter-transcriber agreement scores to express segmentation agreement, phonetic transcription agreement and reading error and strategy labeling agreement (cf. Pitt, Johnson, Hume, Kiesling & Raymond, 2005; Raymond et al., 2002).

5. Relevant data extraction

The data coming from the questionnaire for the teachers and the parents, together with the child's scores on the OMT, Klepel and AVI-test are assembled in a data matrix for each school. The audio-data, on the other hand, are digitally stored as wave-files whereas data about timing of stimulus presentation, data with respect to timing and duration of the child's utterances (obtained by means of the forced alignment technique), and data about the accuracy of the reading performance, about occurring reading errors and used reading strategies is available in text-file format (Praat-file). It is clear than that, except for the excel-data, data is not directly available but needs to be extracted from the Praat-files by means of Python scripts.

B. Data analyses

In this paragraph, we will describe what kind of statistical analyses will be carried out in order to answer the above formulated research questions, and we will indicate how the variables used in these analyses are operationalized.

1. Assessment of annotation consistency

To be able to assess annotation consistency, approximately 10 % of the total corpus will independently be labeled by 3 transcribers (as advised by Neuendorf, 2002). This is a large percentage when compared to other speech database annotations, e.g. only 0.3 % of the Buckeye Corpus was independently labeled by 4 transcrib-

[14] Python® is a dynamic object-oriented programming language that can be used for many kinds of software development. It offers strong support for integration with other languages and tools, and comes with extensive standard libraries (www.python.org).

ers (Pitt et al., 2005). Then, to obtain a measure of consistency, pair wise inter-transcriber agreement scores will be calculated to express segmentation agreement, phonetic transcription agreement and reading error and strategy labeling agreement. First, phonetic transcription, reading error labeling[15] and strategy labeling agreement will be measured by counting the number of labeling agreements for all pairs of transcribers. For example, when one labeler (T1) disagrees with the others that do agree with each other (T2 and T3), then there is one transcriber pair who agrees with each other (T2-T3); the number of transcriber pairs that disagrees with each other is two (T1-T2, T1-T3). This means that only one out of three possible pairings of transcriptions show agreement, resulting in an agreement rate of 33% (i.e. agreement = agree / (disagree + agree)[16]). Another measure of inter-judge agreement is the *kappa* statistic (Cohen, 1960); it controls for chance and varies between 0 and 1. Secondly, we want to compare transcriber's temporal placement of boundaries by calculating the average deviation of boundary locations (cf. Pitt et al., 2005).

2. Standardization, scaling, reliability assessment and validation of the WRT and PWRT

For the analyses presented in this paragraph, only data of regular elementary school children will be used (Phase I).

a) Speed-accuracy trade-off in the WRT and PWRT

First, we want to characterize the frequency distribution for speed[17] and accuracy separately (and check whether these distributions can be approximated by a normal distribution), and map out the correlation between speed and accuracy for each test (each list) and sex in each grade. Then, for each of the three tests (and lists), we want to know how the speed-accuracy trade-off can be characterized for each sex and in each grade. Therefore, we will carry out a multivariate analysis of variance (MANOVA) with grade and sex as factors, and speed and accuracy (number of errors) as dependent variables.

b) Standardization of the WRT and PWRT

In case of the WRT and PWRT, a possible test score to base the test's standards on would be the accuracy score: the number of correctly read words (with a maximum of 40). However, response times[18] (or mean time-per-item score) will possibly better discriminate between subjects (especially in an orthographic transparent language such as Dutch) and can be a measure too to characterize word decoding skills. Important here however is that there is an accuracy-time trade-off that must be taken into account. A child that reads fast but reads it all incorrectly will have an outstanding score on a time-per-item measure. To capture this, we will explore 3 differently calculated scores. Score 1 will be calculated as a mean time-per-correct item score. This comprises that the total response time is divided by the number of correctly read words. Score 2 will be obtained by dividing the mean time-per-item score by the error percentage (cf. van den Bos, 2003). Score 3 has been proposed

[15] Before measuring the agreement of adjudged error category, we want to know agreement scores for the mere detection of an error in general.
[16] A less conservative method of calculating agreement is the method of Shriberg and Lof (1991) where agreement is defined as agree / ((disagree/2) + agree).
[17] Speed is characterized as the sum of all reaction and production times. We can not use the overall time score anymore since this score is largely influenced by the examiner's 'pressing the key'-behavior.
[18] The total response time per item is the sum of the reaction and production time for that particular item.

by Townsend and Ashby (1983) and combines speed response times (RT) and accuracy scores in the formula: RT / (1-% error)[19].

To start, we want to obtain descriptive statistics for the measures proposed above (score 1, 2 and 3). More specifically, we are interested in their frequency distribution for each grade, sex, task, and list within each task (and check whether these distributions can be approximated by a normal distribution), and want to map out the correlati

on between the different measures. Then, based on these distributions, we will discriminate 5 reading

performance groups (A: best performing 25%, B: above average performing 25%, C: below average performing 25%, D: far below average performing 15%, E: worst performing 10%) which will serve as the basis for the calculation of standards, and as a basis to determine cut-off scores for each list. We also want to analyze the relations between the scores on the three different lists and their development over time. The outcomes of these analyses can give evidence with respect to the reliance of the individual word and pseudoword lists at different assessment moments and with respect to the development of technical word decoding skills.

c) Reliability assessment and validation of the WRT and PWRT

To test the reliability of the WRT and PWRT, we could investigate the internal consistency by splitting up the test in as much parts as there are items. Then, correlations between item-scores themselves and between the item-scores and the total score could be analyzed by means of Cronbach's alpha. However, because the factor 'speed' largely attributes to the accuracy performance on each of these tests, the method of item analysis isn't appropriate (van den Bos et al., 1994). If we would have had a parallel test-version or the opportunity to retest children, parallel-forms and test-retest reliability would have been a good alternative. This, however, is not the case.

Evidence of content validity is already obtained for each test since all items were carefully selected (and judged by experts) before being included in the particular reading test (cf. p. II-7). Then, each test's construct validity is measured, which is the way these tests are related to other 'related' tests measuring the same reading skills (i.e. congruent construct validity). We want to evaluate whether the results on each of the tests come up to the expectations we have on the basis of theory. This will be investigated by correlating the results on the WRT and the PWRT with the results on the One-Minute-Test (Brus & Voeten, 1995) and Klepel (van den Bos et al., 1994) respectively.

3. Looking for an alternative measure to capture reading fluency: inter-word latency and time-per-syllable

With respect to the SRT we are interested in the inter-word latency times. For each child and each task, we will calculate mean inter-word latency times and characterize this measure's frequency distribution for each task, grade and sex (and check if its distribution can be approximated by a normal distribution). Then, we want to map out the correlations between this measure and accuracy, speed and AVI-level respectively. Secondly, we will investigate whether time-per-syllable or time-per-correct-syllable can be a valid measure too as an indica-

[19] Alternatively, we will investigate the possibility to use speed and accuracy performance separately in a first stage (and calculate a performance score for both probably normally distributed variables separately), and then combine these two scores to a new normally distributed variable.

tion of a particular child's reading level by correlating these measures with accuracy, speed and AVI-level respectively. These correlations will be investigated in both data from regular elementary school children (Phase I), and data from children attending special schools (Phase II).

4. Influence of child-related and task-related variables on reading error occurrence and strategy use

a) Operationalization of the independent variables

In this study, the independent variables we will use in our analyses are: school, chronological age, reading ability, Reading performance, IQ and IQ-discrepancy, gender, reading task, reading subtask, and passage difficulty. These are operationalized as follows:

- *School* (S01, S02, S03, S04, S05,...): 4 regular elementary schools and at least 3 special schools participate.
- *Chronological age* (expressed by birth year): Since regular elementary school children only participated if they did not double or pass over a grade; grade corresponds to one single birth year. Thus, all children within the same grade have the same chronological age. For children attending special schools however, grade will often not be linked to birth year.
- *Reading ability* (RD, NRD): Children attending schools for special education all have reading difficulties (RD). Most children attending regular elementary schools don't have reading difficulties (NRD).
- *Reading performance* (expressed by raw score on the OMT): Children's performance on the One-minute-test will serve as an indication of their reading performance.
- *IQ* (expressed by raw score on Raven) and *IQ-discrepancy* (DIS, NDIS): If a child's reading level is lower than what can be expected on the basis of his general cognitive abilities (as measured by Raven's progressive matrices, Raven et al., 1992), then there is a discrepancy (DIS); otherwise there is no discrepancy (NDIS).
- *Gender* (M, F): Males versus females.
- *Reading task* (WRT, PWRT, SRT): Word reading test, pseudoword reading test and sentence reading test.
- *Reading subtask* (WRT1, WRT2, WRT3, PWRT1, PWRT2, PWRT3, SRT1, SRT2, SRT3, SRT4, SRT5, SRT6, SRT7, SRT8, SRT9): Both the word reading test (WRT) and the pseudoword reading test (PWRT) consist of 3 lists of 1-syllable, 2-syllable, and 3+4-syllable words or pseudowords respec-
- *Passage difficulty*

tively. The SKT stories are graded in difficulty level from AVT1 to AVT9.

(expressed by a passage difficulty score): For every subtest for every child a passage difficulty score will be obtained, based on the number of errors per 100 words.

b) Operationalization of the dependent variables

The dependent variables are: speed, accuracy, reading error occurrence and reading strategy use. A potentially disturbing variable, namely nature of reading instruction, is controlled for since all participating school use an analytic-synthetic approach.

- *Speed*: The time (in seconds) needed to complete a reading task, calculated as the sum of all reaction and production times.
- *Accuracy*: Total number of errors made in a reading task.
- *Reading error occurrence*: Frequency of occurrence of the error categories. On the basis of the data, we will reduce the 48 error categories to maximally 10 error categories.
- *Reading strategy use*: Frequency of occurrence of each of the 9 reading strategy categories.

c) Analyses

Our aim is to analyze the individual and interaction effects of these different independent variables on *Speed*, *Accuracy*, *Reading error occurrence* and *Reading strategy use*.

(1) Effects on Speed and Accuracy

To investigate the effects on *Speed* and *Accuracy*, we will make use of multivariate analyses of covariances (MANCOVA). Table 2 summarizes the analyses we want to carry out.

Independent variables	Data
School *Chronological age* *Gender* *Reading performance* *Reading task* *Reading subtask* *Passage difficulty*	Phase I
Chronological age *Gender* *Reading ability* *Reading performance* *Reading subtask*	Phase I + II
School *Chronological age* *Gender* *Reading performance* *IQ-discrepancy* *IQ* *Reading subtask*	Phase II

Table 2: summary of analyses to investigate the effect on *Speed* and *Accuracy*

(2) Effects on Reading error occurrence and Reading strategy use

To gain more insight in the effects on *Reading error occurrence* and *Reading strategy use*, we will look at reading error distribution, reading strategy distribution, and reading strategy efficiency for each of the relevant independent variables: *Reading ability*, *IQ-reading ability discrepancy*, *Passage difficulty*, *Reading task*, and *Reading subtask*. Reading error and reading strategy distribution will be determined by calculating the relative frequency with which a reading error occurs or a strategy is used. Strategy efficiency will be determined by calculating the mean response time with which each strategy is executed.

Additionally, to investigate if we can find any patterns in the relation between reading strategy use and reading error occurrence, we will look at the reading error distribution for each reading strategy.

5. Automatically distinguishing decoding from guessing behavior

Every time a new word or pseudoword is presented on the screen, we can keep track of the time needed for a child to start his or her utterance (reaction time) and the time needed for the utterance itself (production time). It seems natural that a long reaction time combined with a short production time corresponds to an internal decoding strategy and a short reaction time combined with a long production time corresponds to an external decoding strategy. In contrast, a short reaction and production time corresponds to correct decoding or decoding by guessing. We want to analyze whether these types of decoding strategies can be detected automatically by making use of reaction and production times. We will do this by comparing these features with the reading strategies manually assigned by the transcribers. Following Grace and Spreen (1994) however, we will also investigate whether reading strategy use can be predicted on the basis of reading errors and reading speed. Then, regression analysis will be used to explore if there are other features (e.g. reading error categories) we can use to predict the reading strategy use of a particular child.

V. Preliminary Overview Research Output

Paper:

Cleuren, L., Duchateau, J., Sips, A., Ghesquière, P. and Van hamme H. (2006). Developing an Automated Assessment Tool for Children's Oral Reading. Proceedings of the 9th International Conference on Spoken Language Processing, Pittsburgh, USA, pp. 817-820 (Appendix C).

Poster presentation:

Developing an Automated Assessment Tool for Children's Oral Reading. 9th International Conference on Spoken Language Processing, Pittsburgh, USA, September 17-21 2006 (Appendix D).

VI. References

Adams, M.J. (1990). *Beginning to read: Thinking and learning about print.* Cambridge, MA: MIT Press.

Adams, A. (1990). The oral reading errors of readers with learning disabilities: Variations produced within the instructional and frustrational ranges. *Remedial and Special Education, 12*(1), 48-55.

Allington, R.L., & McGill-Franzen, A. (1980). Word identification errors in isolation and in context: apples vs. oranges. *Reading Teacher*, 795-800.

Au, H.K.P. (1977). Analyzing oral reading errors to improve instruction. *Reading Teacher, 31*(1), 46-9.

Baddeley, A.D., Logie, R.H., & Ellis, N.C. (1988). Characteristics of developmental dyslexia. *Cognition, 29*, 197-228.

Bakker, D.J. (1965). Leesstoornissen: een foutenanalyse [Reading disabilities: an error analysis]. *Nederlands Tijdschrift voor Psychologie, 20*(3), 173-183.

Bakker, D.J. (1979). Hemispheric differences and reading strategies: Two dyslexias? *Bulletin of the Orton Society, 29*, 84-100.

Bakker, D.J. (1981). A set of brains for learning to read. In K.C. Diller (Ed.). *Individual differences and universals in language learning aptitude* (pp.56-71). Rowley, MA: Newbury House.

Barr, R.C. (1972). The influence of instructional conditions on word recognition errors. *Reading Research Quarterly, 7*(3), 509-529.

Beck, J.E., Jia, P., & Mostow, J. (2004). Automatically assessing oral reading fluency in a computer tutor that listens. *Tech. Inst. Cognition and Learning, 1*, 61-81.

Belmore, S.M. (1985). Reading computer-presented text. *Bulletin of the Psychonomic Society, 23*(1), 12-14.

Ben-Dror, I., Pollatsek, A., & Scarpati, S. (1991). Word identification in isolation and in context by college dyslexic students. *Brain and Language, 40*, 471-490.

Bernard, M.L., Chaparro, B.S., Mills, M.M., & Halcomb, C.G. (2002). Examining children's reading performance and preference for different computer-displayed text. *Behaviour and Information Technology, 21*(2), 87-96

Bernard, M.L., Chaparro, B.S., Mills, M.M., & Halcomb, C.G. (2003). Comparing the effects of text size and format on the readability of computer-displayed Times New Roman and Arial text. *International Journal of Human-Computer Studies, 59*, 823-835.

Biemiller, A. (1970). The development of the use of graphic and contextual information as children learn to read. *Reading Research Quarterly, 6*(1), 75-96.

Biemiller, A. (1978). Relationships between oral reading rates for letters, words, and simple text in development of reading-achievement. *Reading Research Quarterly, 13*(2), 223-253.

Biemiller, A. (1979). Changes in the use of graphic and contextual information as functions of passage difficulty and reading achievement level. *Journal of Reading Behavior, 11*(4), 307-318.

Billiaert, E., & Grysolle, R. (1998). *LVS-CSBO leerlingvolgsysteem, Lezen: analyse en handelen [LVS-CSBO pupil's monitoring system for reading: analysis and treatment].* Leuven: Garant.

Blaxall, J., & Willows, D.M. (1984). Reading ability and text difficulty as influence on second graders' oral reading errors. *Journal of Educational Psychology, 76*(1), 330-341.

Blonk, A.M., van den Bercken, J.H.L., & de Bruyn, E.E.J. (1996). Evaluation of DYSLEXPERT: A comparison of a knowledge-based system with experienced clinicians in the diagnosis of dyslexia. *Computers in Human Behavior, 12*(4), 567-580.

Blumenthal, S.H. (1980). A study of the relationship between speed of retrieval of verbal information and patterns of oral reading errors. *Journal of Learning Disabilities, 13*(9), 42-44.

Boder, E. (1970). Developmental dyslexia: A new diagnostic approach based on the identification of three subtypes. *The Journal of School Health, 40*, 289-290.

Boder, E. (1973). Developmental dyslexia: A diagnostic approach based on three atypical reading-spelling patterns. *Developmental Medicine and Child Neurology, 15*, 663-687.

Bridgeman, B., Lennon M.J., & Jackenthal, A. (2003). Effects of screen size, screen resolution, and display rate on computer-based test performance. *Applied Measurement in Education, 16*(3), 191.205.

Brus, B.T., & Voeten, B.J. (1995). *Eén-Minuut-Test Vorm A en B. Verantwoording en handleiding [One-Minute-Test. Manual]*. Nijmegen: Berkhout.

Carver, R.P. (1991). Using letter-naming speed to diagnose reading disability. *Remedial and Special Education, 12*(5), 33-43.

Chen, M.-T., & Lin, C.-C. (2004). Comparison of TFT-LCD and CRT on visual recognition and subjective preference. *International Journal of Industrial Ergonomics, 34*, 167-174.

Chiappe, P. & Siegel, L.S. (1999). Phonological awareness and reading acquisition in English- and Punjabi-speaking Canadian children. *Journal of Educational Psychology, 91*(1), 20-28.

Chinn, C.A., Waggoner, M.A., Anderson, R.C., Schommer, M., & Wilkinson, I.A.G. (1993). Situated actions during reading lessons: A microanalysis of oral reading error episodes. *American Educational Research Journal, 30*(2), 361-392.

Cisero, C.A., Royer, J.M., Marchant III, H.G., & Jackson, S.J. (1997). Can the computer-based academic assessment system (CAAS) be used to diagnose reading disability in college students? *Journal of Educational Psychology, 89*(4), 599-620.

Clay, M. (1966). *Emergent reading behaviour*. Unpublished PhD thesis, University of Auckland.

Clay, M. (1968). A syntactic analysis of reading errors. *Journal of Verbal Learning and Verbal Behavior, 7*, 434-438.

Clay, M. (1969). Reading errors and self-correction behaviour. *British Journal of Educational Psychology, 39*, 47-56.

Cleuren, L., Duchateau, J., Sips, A., Ghesquière, P. and Van hamme H. (2006). Developing an Automated Assessment Tool for Children's Oral Reading. Proceedings of the 9th International Conference on Spoken Language Processing, *Read-*

ing Research Quarterly, 10

sessment Tool for Children's Oral Reading. Proceedings of the 9th International Conference on Spoken Language Processing, Pittsburgh, USA, pp. 817-820.

Cohen, A.S. (1974-75). Oral reading errors of first grade children taught by a code emphasis approach. (4), 616-650.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37-46.

Coltheart, M. (1978). Lexical access in simple reading tasks. In G. Underwood (Ed.), *Strategies of information processing*. New York: Academic Press.

Coltheart, V., & Leahy, J. (1992). Children's and adult's reading of nonwords: Effects of regularity and consistency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18*(4), 718-729.

Coltheart, M, Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: A dual route cascade model of visual word recognition and reading aloud. *Psychological Review, 108*, 204-256.

Cucchiarini, C., Strik, H., & Boves, L. (1998). Quantitative assessment of second language learner's fluency: an automatic approach. Proceedings of the 5th International Conference on Spoken Language Processing, Sydney, Australia, pp. 2619-2622.

Cucchiarini, C., Strik, H., & Boves, L. (2000).Quantitative assessment of second language learners' fluency by means of automatic speech recognition technology. *Journal of the Acoustical Society of America, 107*(2), 989-999.

Cucchiarini, C., Strik, H., & Boves, L. (2002). Quantitative assessment of second language learners' fluency: comparisons between read and spontaneous speech. *Journal of the Acoustical Society of America, 111*(6), 2862-2873.

D'Angelo, K. & Mahlios, M. (1983). Insertion and omission miscues of good and poor readers. *The Reading Teacher*, 778-782.

De Bruijn, D., De Mul, S., & Van Oostendorp, H. (1992). The influence of screen size and text layout on the study of text. *Behaviour and Information Technology, 11*(2), 71-78.

Dennis, I., & Evans, J.St.B.T. (1996). The speed-error trade-off problem in psychometric testing. *British Journal of Psychology, 87*, 105-129.

Dillon, A. (1992). Reading from paper versus screens: a critical review of the empirical literature. *Ergonomics, 35*(10), 1297-1326.

Dillon, A., Richardson, J., & McKnight, C. (1990). The effects of display size and text splitting on reading lengthy text from screen. *Behaviour and Information Technology, 9*(3), 215-227.

Dudal, P. (1997). *Leerlingvolgsysteem lezen: Toetsen 1-2-3. Basisboek [Pupil's Monitoring System for Reading: Grade 1-2-3]*. Leuven: Vrije PMS-Centra & Garant.

Dumont, J.J. (1994). Rotterdam: Lemniscaat.

Dyson, M.C. (2004). How physical text layout affects reading from screen (6), 377-393.

Dyson, M.C., & Haselgrove, M. (2001). The influence of reading speed and line length on the effectiveness of reading from screen. , 585-612.

Ehri, L.C. (1991). Development of the ability to read words. In R. Barr, M. Kamil, P. Mosenthal, & P. Pearson (Eds.), Handbook of Reading Research Volume II (383-417). New York: Longman.

Ehri, L.C. (1995). Phases of development in learning to read words by sight. *Journal of Research in Reading, 18*(2), 116-125.

Ehri, L.C. (2005). Learning to read words: Theory, findings, and issues. *Scientific studies of reading, 9*(2), 167-188.

Ehri, L.C., & McGormick, S. (1998). Phases of word learning: implications for instruction with delayed and disabled readers. *Reading & Writing Quarterly, 14*(2), 135-163.

Evans, L.D., Tannehill, R., & Martin, S. (1995). Children's reading skills: A comparison of traditional and computerized assessment. *Behavior Research Methods, Instruments, & Computers, 27*(2), 162-165.

Francis, H. (1984). Children's knowledge of orthography in learning to read. *British Journal of Educational Psychology, 54*, 8-23.

Frith, U. (1985). *Beneath the surface of developmental dyslexia.* In K.E. Patterson, J.C. Marshall & M. Coltheart (Eds.), 301-330. London: Lawrence Erlbaum.

Frith, U. (1986). A developmental framework for developmental dyslexia. *Annals of Dyslexia, 36*, 69-81.

Garcia, M.L., & Caldera, C.I. (1996). The effect of color and typeface on the readability of on-line text. *Computers and Industrial Engineering, 31*(1/2), 519-524.

Garland, K.J., & Noyes, J.M. (2004). CRT monitors: Do they interfere with learning? *Behaviour and Information Technology, 23*(1), 43-52.

Goodman, K.S. (1965). A linguistic study of cues and miscues in reading. *Elementary English, 42*(6), 639-643.

Goodman, K.S. (1969). Analysis of oral reading miscues: applied psycholinguistics. *Reading Research Quarterly, 5*(1), 9-30.

Goodman, K.S. (1976). What we know about reading. In O.D. Allen and D.J. Watson (Eds.) Findings of research in miscue analysis: Classroom implications. Urbana Illinois U.S. National Council of Teachers of English (NCTE).

Goodman, Y.M. (1972). Reading diagnosis – qualitative or quantitative? *The Reading Teacher, 26*(1), 32-37.

Gough, P.B., Juel, C., & Griffith, P.L. (1992). Reading, spelling and the orthographic cipher. In P.B. Gough, L. Ehri, & R. Treiman (Eds.). *Reading acquisition* (pp. 35-48). Hillsdale, NJ: Erlbaum.

Gough, P.B., & Wren, S. (1998). The decomposition of decoding. In C. Hulme & M. Joshi (Eds.), *Reading and spelling: Development and disorders* (pp. 19-32). Mahwah, NJ: Erlbaum.

Gould, J.D., Alfaro, L., Barnes, V., Finn, R., Grischkowsky, N., & Minuto, A. (1987). Reading is slower from CRT displays than from paper: Attempts to isolate a single-variable explanation. *Human Factors, 29*(3), 269-299.

Gould, J.D., & Grischkowsky, N. (1986). Does visual angle of a line of characters affect reading speed? *Human Factors, 28*

Hall, R.H., & Hanna, P. (2004). The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention. *Behaviour and Information Technology, 23*(3), 183-195.

Harding, L.M. (1984). Reading errors and style in children with a specific reading disability. *Journal of Research in Reading,* 7(2), 103-112.

Harding, L.M., Beech, J.R. & Sneddon, W. (1985). The changing pattern of reading errors and reading style from 5 to 11 years of age. *British Journal of Educational Psychology, 55*, 45-52.

Hoover, W.A., & Gough, P.B. (1990). The simple view of reading. *Reading and Writing, 2*(2), 127-160.

Irvin, J.L. (1987). Computer-based reading assessment instrument (CRAI) – Blanchard, J.S.. *Reading Teacher, 41*(1), 92-94.

Jansen, M.G.H. (1997). The Rasch model for speed tests and some extensions with applications to incomplete designs. *Journal of Educational and Behavioral Statistics, 22*(2), 125-140.

Jia, P., Beck, J.E. & Mostow, J. (2002). Can a reading tutor that listens, use inter-word latency to assess a student's reading ability? *ITS 2002 Workshop on Creating Valid Diagnostic Assessments, San Sebastian, Spain*, 23-32.

Juel, C. (1980). Comparison of word identification strategies with varying context, word type, and reader skill. *Reading Research Quarterly, 15*(3), 358-376.

Kerr, M.A., & Symons, S.E. (2006). Computerized presentation of text: effects on children's reading of informational material. *Reading and Writing, 19*, 1-19.

Laing, S.P. (2002). Miscue analysis in school-age children. *American Journal of Speech-Language Pathology, 11*, 407-416.

Leslie, L. (1980). The use of graphic and contextual information by average and below-average readers. *Journal of Reading Behavior, 12*(2), 139-149.

Leu, D.J. (1982). Oral reading error analysis: A critical review of research and application. *Reading Research Quarterly, 3*, 420-437.

Lovett, M.W. (1984). A developmental perspective on reading dysfunction: Accuracy and rate criteria in the subtyping of dyslexic children. *Brain and Language, 22*, 171-186.

McGuiness, D. (1997). *Why our children can't read and what we can do about it: A scientific revolution in reading.* New York: Free Press.

Miller, J.W. & Isakson, R.L. (1978). Contextual sensitivity in beginning readers. *The Elementary School Journal, 78*(5), 325-331.

Mitterer, J.O. (1982). There are at least two kinds of poor readers: Whole word readers and recoding poor readers, *Canadian Journal of Psychology, 36*, 445-461.

Moscicki, E.K. & Tallal, P. (1981). A phonological exploration of oral reading errors. *Applied Psycholinguistics, 2*, 353-367.

Mostow, J., Beck, J., Winter, S.V., Wang, S. & Tobin, B. (2002). Predicting oral reading miscues. *Proc. of the*

1221-1224.

Muter, P., & Maurutto, P. (1991). Reading and skimming from computer screens and books: the paperless office revisited? *Behaviour and Information Technology, 10*(4), 257-266.

Neuendorf, K.A. (2002). *The content analysis guidebook.* Thousand Oaks, CA: Sage.

Nunn, A. (1998). *Dutch orthography: A systematic investigation of the spelling of Dutch words.* The Hague: Holland Academic Graphics.

OECD (1999). *Classifying educational programmes: Manual for ISCED-97 implementation in OECD countries.* Paris: OECD.

Perfetti, C.A. (1985). *Reading ability*. New York: Oxford University Press.

Pitt, M.A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W. (2005). The Buckeye corpus of conversational speech: labeling conventions and a test of transcriber reliability. *Speech Communication, 45*, 89-95.

Potter, F. (1980). Miscue analysis: a cautionary note. *Journal of Research in Reading, 3*(2), 116-128.

Ramus, F. (2004). The neural basis of reading acquisition. In M.S. Gazzaniga (Ed.), *The Cognitive Neurosciences* (3rd ed.) (815-824). Cambridge, MA: MIT Press.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests.* Chicago: The University of Chicago Press.

Raven, J.C., Court, J.H., & Raven, J. (1992). *Standard progressive matrices.* Oxford: Oxford Psychologists Press.

Raymond, W.D., Pitt, M., Johnson, K., Hume, E., Makashay, M., Dautricourt, R., & Hilts, C. (2002). An analysis of transcription consistency in spontaneous speech from the Buckeye corpus. Proceedings of the International Conference on Spoken Language Processing, pp. 1125-1128.

Reitsma, P. (1992). Invloed van leesfrequentie op fouten als teken van de ontwikkeling van een orthografisch geheugen [The influence of reading frequency on errors as a sign of orthographic memory development]. *Nederlands Tijdschrift voor de Psychologie, 47*, 223-234.

Savage, R., Stuart, M. & Hill, V. (2001). The role of scaffolding errors in reading development: Evidence from a longitudinal and a correlational study. *British Journal of Educational Psychology, 71*, 1-13.

Seys, J., Van Biervliet, J., & Van Biervliet, P. (2000). *Leeskist: een diagnostisch en remediërend softwarepakket voor aanvankelijk lezen in het eerste leerjaar [Leeskist: a diagnostic and remedial software package for beginning reading in first grade].* Wommelgem: Van In.

Shankweiler, D., & Liberman, I.Y. (1972). Misreading: a search for causes. In J.F. Kavanagh & I.G. Mattingly (Eds.), Language by ear and by eye (pp. 319-329). Cambridge, MA; MIT Press.

Shieh, K.-K., & Lin, C.-C. (2000). Effects of screen type, ambient illumination, and color combination on VDT visual performance and subjective preference. *International Journal of Industrial Ergonomics, 26*, 527-536.

Shriberg, L.D., and Lof, G. L. (1991). Reliability studies in broad and narrow phonetic transcription. *Clinical Linguistics and Phonetics, 5*, 225-279.

Siegel, L.S. (1993). The development of reading. *Advances in Child Development and Behavior, 24*, 63-97.

Siegler, R.S. (1988). Individual differences in strategy choice: good students, not-so-good students, and perfectionists. *Child Development, 59*, 833-851.

Snowling, M.J. (1991). Developmental Reading Disorders. *Journal of Child Psychology and Psychiatry, 32*(1), 49-77.

Sprenger-Charolles, L., & Siegel, L.S. (1997). A longitudinal study of the effects of syllabic structure on the development of reading and spelling skills in French. *Applied Psycholinguistics, 18*(4), 485-505.

Stanovich, K.E. (1986). Cognitive processes and the reading problems of learning disabled children. Evaluating the assumption of specificity. In J.K. Torgesen & B.Y.L. Wong (Eds.), Psychological and educational perspectives on learning disabilities. Orlando, FL: Academic Press.

Stanovich, K.E., Cunningham, A.E., & Feeman, D.J. (1984). Relation between early reading acquisition and word decoding with and without context: A longitudinal study of first-grade children. *Journal of Educational Psychology, 76*(4), 668-377.

Stanovich, K.E., & Siegel, L.S. (1994). Phenotypic performance profile of children with reading disabilities: A regression-based test of the phonological-core variable-difference model. *Journal of Educational Psychology, 86*(1), 24-53.

Stanovich, K.E., West, R.F., & Feeman, D.J. (1981). A longitudinal study of sentence context effects in second-grade children: tests of an interactive-compensatory model. *Journal of Experimental Child Psychology, 32*, 185-199.

Stuart, M., & Coltheart, M. (1988). Does reading develop in a sequence of stages? *Cognition, 30*(2), 139-81.

Szeszulski, P.A., & Manis, F.R. (1987). A comparison of word recognition processes in dyslexic and normal readers at two reading-age levels. *Journal of Experimental Child Psychology, 44*, 364-376.

Tal, N.F. & Siegel, L.S. (1996). Pseudoword reading errors of poor, dyslexic, and normally achieving readers on multisyllable pseudowords. Applied Psycholinguistics, 17, 215-232.

Thomson, M. (1978). A psycholinguistic analysis of reading errors made by dyslexics and normal readers. *Journal of Research in Reading, 1*(1), 7-20.

Thomson, M. (1986). Some theoretical and practical aspects of developmental dyslexia. *Bulletin of the British Psychological Society, 39*, A99-A99.

Townsend, J.T., & Ashby, F.G. (1983). *Stochastic modeling of elementary psychological processes.* New York: Cambridge University Press.

van Aarle, E.J.M., & van den Bercken, J.H.L. (1999). The development of a knowledge-based system supporting the diagnosis of reading and spelling problems (II). *Computers in Human Behavior, 15*, 693-712.

van den Bos, K.P. (2003). De relatie tussen Lees- en Benoemsnelheid bij Leerlingen in het Basis- en Voortgezet Onderwijs. *Pedagogische Studiën, 80*, 288-308.

van den Bos, K.P., Spelberg, H.C.L., Scheepstra, A.S.M., & De Vries, J.R. (1994). *De Klepel: Pseudowoordentest [The Klepel: Pseudoword Test].* Nijmegen: Berkhout.

van der Leij, D.A.V. (1983). *Ernstige leesproblemen [Serious reading problems].* Doctoral dissertation. Lisse, the Netherlands: Swets & Zeitlinger.

van der Leij, A., Smeets, H., & van Daal, V.H.P. (1990). De ontwikkeling van de Computergestuurde Toets voor Automatisering van het lezen [Development of the computerized test for the automatization of reading]. *Pedagogische Studiën, 67*, 163-178.

van der Schoot, M., Licht, R., Horsley, T.M., Aarts, L.T., van Koert, B., & Sergeant, J.A. (2004). Inhibitory Control During Sentence Reading in Dyslexic Children. *Child Neuropsychology, 10*(3), 173-188.

van der Schoot, M., Licht, R., Horsley, T.M., & Sergeant, J.A. (2002). Fronto-central dysfunctions in reading disability depend on subtype: guessers but not spellers. *Developmental Neuropsychology, 22*(3), 533-564.

Verhoeven, L. (1995). *Drie-Minuten-Toets [Three-Minutes-Test]*. Arnhem: Cito.

Verhoeven, L. & van Leeuwe, J. (2003). Ontwikkeling van decodeervaardigheid in het basisonderwijs [Development of word decoding ability in elementary school]. *Pedagogische Studiën, 80*, 257-271.

Visser, J. (1997). *Lezen op AVI-niveau. De schaling, normering en validering van de nieuwe AVI-toetskaarten [Reading at AVI-level. Scaling, standardization and validation of the new AVI test stories]*. PhD thesis, Katholieke Universiteit Nijmegen.

Visser, J., van Laarhoven, A., & ter Beek, A. (1994). *AVI Toetspakket [AVI Test Package]*. 's Hertogenbosch: KPC.

Wang, A.-H., & Chen, C.-H. (2003). Effects of screen type, Chinese typography, text/background color combination, speed, and jump length for VDT leading display on users' reading performance. *International Journal of Industrial Ergonomics, 31*, 249-261.

Weber, R.-M. (1968). The study of oral reading errors: A survey of the literature. *Reading Research Quarterly, 4*(1), 96-119.

Weber, R.-M. (1970). A linguistic analysis of first-grade reading errors. *Reading Research Quarterly, 5*(3), 427-451.

Werker, J.F., Bryson, S.E., & Wassenberg, K. (1989). Toward understanding the problem in severley disabled readers, Part II: Consonant errors. *Applied Psycholinguistics, 10*, 13-30.

West, R.F., Stanovich, K.E., Feeman, D.J., & Cunningham, A.E. (1983). The effect of sentence context on word recognition in second- and sixth-grade children. *Reading Research Quarterly, 19*(1), 6-15.

Willich, Y., Prior, M., Cumming, G. & Spanos, T. (1988). Are disabled readers delayed or different? An approach using an objective miscue analysis. *British Journal of Educational Psychology, 58*, 315-329.

Wimmer, H. (1996). The nonword reading deficit in developmental dyslexia: evidence from children learning to read German. *Journal of Experimental Child Psychology, 61*, 80-90.

VII. Appendices

A. Current version of the word reading test, pseudoword reading test and story reading test

1. Word reading test (WRT)

	1 SYL	2 SYL	3 + 4 SYL
1	klas	water	omdraaien
2	mooi	mevrouw	poppenwagen
3	koe	drogen	konijnenhok
4	arm	winkel	elastiekje
5	groei	auto	ruziemaken
6	strand	schouders	teddybeer
7	bed	verhaal	dierentuin
8	eerst	koning	paddenstoelen
9	voor	moeilijk	verstoppertje
10	draai	speelplaats	wasmachine
11	sjaal	drinken	fototoestel
12	herfst	hoofdpijn	toiletpapier
13	duur	regen	vrachtwagen
14	straat	vliegtuig	buurmannen
15	leeuw	stoppen	vogelkooi
16	clown	opnieuw	olifant
17	hoek	gooien	schommelen
18	krant	sneeuwen	iedereen
19	hout	moeder	schoenenwinkel
20	vriend	liedje	knutselen
21	gauw	potlood	ophangen
22	chips	fietsbel	verjaardag
23	groen	vinger	sprookjesboek
24	feest	dichtbij	tandenborstel
25	reis	meisje	lucifer
26	jas	chauffeur	slaapkamer
27	huis	muziek	achterdeur
28	paard	waarom	ziekenhuis
29	vijf	scheuren	nieuwsgierig
30	muts	lawaai	afblijven
31	nieuw	zwemmen	kabouter
32	kind	vuurwerk	washandje
33	bang	appel	sneeuwwitje
34	oog	cola	goeiendag
35	zacht	kussen	vakantie
36	schoen	eerste	limonade
37	plas	circus	autorijden
38	neus	kleuren	eindelijk
39	knoop	voetbal	familie
40	plank	vlinder	chocolade

2. Pseudoword reading test (PWRT)

	1 SYL	2 SYL	3 + 4 SYL
1	fook	millen	ometuif
2	sooi	vocil	toejietsen
3	dau	geutel	oonwijlen
4	ant	sjobel	jattesiet
5	beeg	ierpieuw	nurudien
6	sprunt	walaan	stoenydaas
7	hool	erke	deuveltek
8	larst	haweel	juitonie
9	vout	saarweng	gevijdel
10	zwoei	gevicht	sidowaan
11	fam	eemde	spekkeraai
12	rachts	bepoud	wachteniek
13	vaap	orstalk	verpierik
14	sprieuw	veten	nappegreeuw
15	keng	gefouw	mantaroen
16	swoers	vurpaand	schielendaspen
17	doer	nizung	crobeklunker
18	plirt	fiewon	kabbestepen
19	jien	kneurem	verwarig
20	blard	vawaai	ooiebiekje
21	guul	strellen	fandelig
22	hoekt	zwieten	jalekrewen
23	neeuw	foetbans	smoralij
24	noork	oonste	zeekvlachine
25	vid	muider	kanaroe
26	zans	grijnken	toineetlijgen
27	leum	schielstaug	meitsegrok
28	haans	prilsood	kantelogsten
29	spaai	vloender	ondermind
30	sjalt	milste	choporatie
31	heik	veurder	zennebral
32	sank	kloeien	ijraspangen
33	roen	ulen	blottenduuf
34	frijk	orponk	girdofhaalder
35	eem	schodig	tobbermoeit
36	schard	ijpo	poentalschouden
37	grek	menuur	havedil
38	dron	spreikje	verbrakkertje
39	snaaf	hiffreeuw	gerauwejaak
40	stuid	wooien	hapeneren

3. Story reading test (SRT)

AVI 1

feest voor jan

nog een dag of vier.
dan is jan zes jaar.
ik geef je iets, zegt lies.
weet je al wat je wilt?

dat weet jan wel.
een bal of een beer.
een pen of een boek.
of een vis in een kom.

wat een lijst, roept lies.
ik kies wel iets.
en steek het in een doos, zegt ze.

nu is het feest.
jan is zes jaar.
hij heeft een kroon.
en er is taart en snoep.

lies geeft de doos aan jan.
geen bal of beer, geen pen of boek.
maar een kom met een vis.
zes jaar zijn, is heel fijn.

AVI 2

waar is de schoen?

het brood is op.
de kaas is ook bijna op.
mama zegt: ik ga naar de winkel.
ik haal een brood bij de bakker.
en daarna koop ik kaas.
wie gaat er mee?

ik ga mee, roept tom.
en ik haal poes, zegt an.
want poes gaat graag naar buiten.
mama stopt haar in een mand.

an staat al klaar en wacht.
ze heeft haar jas aan.
maar waar is tom naar toe?
tom zoekt zijn schoen.
waar is die schoen toch?

ik tel tot tien, roept mama.
tom zoekt en zoekt maar door.
waar blijf je? vraagt mama.
de schoen ligt onder zijn bed.
snel loopt hij de gang in.
oef, net op tijd.

AVI 3

Goed nieuws

Els en Bert spelen tikkertje.
Ik krijg je wel, roept Bert.
Maar Els loopt de gang in.
Pak me dan, als je kan!
Bert rent snel achter haar aan.

Stop, hijgt Els, er is post.
Er ligt een brief in de bus.
Van wie is die? vraagt ze zich af.
Ook Bert denkt heel diep na.

Samen maken ze de brief open.
Er zit een grote, roze kaart in.
De kaart brengt goed nieuws.
Tante Hilde kreeg een kindje, een meisje!
Papa leest de kaart ook.

Wil je het kindje zien? vraagt papa.
Ja, roepen Bert en Els.

AVI 4

Lekke band

Het is zondag en Joris verveelt zich.
Dan hoort hij de bel.
Joris loopt naar de deur en doet open.

Er staat een meisje op de stoep.
Kan iemand een handje helpen? vraagt ze.
Mijn auto heeft een lekke band.

Ook moeder en vader komen nu kijken.
Vader wil de band wel vervangen.
Kijk maar goed, Joris, zegt hij stoer.
Eerst neemt hij de wieldop weg.
Dan haalt hij een krik uit de koffer.
Zo kan hij één kant van de auto optillen.
De lekke band schroeft hij los.
Dan kan de nieuwe band erop.
Ten slotte gaat de dop weer op het wiel.

De handen van vader zijn erg zwart.

Ze dansen rond de tafel.
Ze doen heel gek en wild.
En dan knutselen we een pakje, zegt papa.
Dat vinden ze een leuk plan.

AVI 5

Verliefd

Tim is onze buurman.
Hij houdt van grapjes en is altijd vrolijk.
Wanneer we voetballen, doet hij graag mee.

Maar nu is hij verliefd.
Nu heeft hij geen zin meer in spelen.
Hij denkt alleen maar aan haar
en loopt met zijn hoofd in de wolken.
Ik denk dat dat normaal is,
als je pas verliefd bent.
En dat het een heel fijn gevoel is.

We stellen hem wel honderd vragen.
Ben je al met haar naar de film geweest?
Is ze knap en speelt ze graag voetbal?
Wanneer zien we haar een keer?

Haar naam is Liesbet, vertelt hij.
Ze is heel sportief, maar ze kan niet zwemmen.
Ze speelt wel goed basketbal.
En net als ik, houdt ze van muziek.

Tim vertelt dat ze niet kan lopen, ze zit in een rolstoel.
Haar twee benen zijn verlamd.
De bus nemen of ergens aanbellen,
is voor haar niet gemakkelijk.

Dan verklapt hij dat ze snel zullen trouwen.
En dat Liesbet onze buurvrouw wordt.
Dan zullen we haar vaak genoeg kunnen zien!

Op zijn wang zit een veeg.
Maar dat is niet erg.
Joris kijkt hem vol trots aan.
En van moeder krijgt hij een zoen.
Blij rijdt het meisje weer naar huis.
Zo was het toch geen saaie dag!

AVI 6

Het winnende gedicht

Er is een cameraploeg op school vandaag.
Op de speelplaats staat een grote vrachtwagen.
Dat komt omdat het de week van de poëzie is.
Elk kind uit onze klas schreef daarom een gedicht.
En er is één gedicht dat wint.

In onze klas zijn twee mannen druk bezig.
Er staat een grote lamp en de vloer ligt vol met kabels.
Er komt nog iemand binnen met een luidspreker.
Hij zet hem op het bureau van de juf.
Dan gaat de lamp aan.
De reporter vertelt over de week van de poëzie.
Hij zegt dat één gedicht uitkiezen moeilijk was.

Toch is er slechts één winnaar.
Het gedicht dat wint, is dat van Anneleen.
De hele klas klapt voor haar.
Ze wordt een beetje rood.
Voor de camera leest ze haar gedicht aan ons voor.
Wanneer ze klaar is, klappen we opnieuw.
Iedereen staat om haar heen.

Die avond kijken we naar het nieuws.
We luisteren hoe Anneleen het gedicht voorleest.
Ze spreekt heel mooi en haar hoofd lijkt niet zo rood.
Wat zijn we trots op haar!

AVI 7

Het leven bij de indianen

Vroeger leefden er heel veel soorten indianen. Die woonden in het bos, de bergen of in de woestijn. Soms waren het nomaden. Dat zijn mensen die geen vaste plaats hebben om te wonen. Ze moesten hun kamp steeds verplaatsen. Dat deden ze om de bizons te volgen. Die hadden ze nodig om te kunnen leven.

Voor ons is lezen en schrijven heel belangrijk. Maar voor de indianen was dat niet zo. Voor hen moest een jongen een goede jager en krijger worden. Een meisje moest goed kunnen koken. En dat leerden ze niet op school.

De vader, de grootvaders en de ooms vertelden de kleine jongen alles over de jacht en de oorlog. Ze toonden

Meisjes deden niet mee aan de jacht of de oorlog. Ze leerden koken en het huishouden doen. Dat zou later immers hun taak zijn. Daarom hielpen ze hun moeder al heel jong met fruit plukken en hout zoeken. Ze zorgden ook vaak voor hun kleine broertjes en zusjes. Zo maakten ze zich klaar voor hun rol als moeder.

Had jij graag als indiaan geboren willen zijn? Of zou je je klas en je school dan missen?

AVI 8

De maan

De maan is net zo oud als de aarde. Dat betekent dat ze vijf miljard jaar oud is. Net als de aarde is zij ontstaan uit gas en stof. Daaruit zijn ook de zon en alle planeten ontstaan.

De maan is het hemellichaam dat zich het dichtst bij de aarde bevindt. Sommige wetenschappers denken dat ze van de aarde is losgerukt. Eigenlijk weten we het niet zeker. Het is in elk geval wel zo dat de maan in de buurt van de aarde is ontstaan. Dat blijkt bijvoorbeeld uit bepaalde stoffen die ook op de aarde voorkomen.

Wanneer je 's avonds naar de maan kijkt, lijkt het net of je een gezicht kunt zien. Dat komt door de donkere vlekken op de maan. Die noemen we kraters. Dat zijn plekken waar lang geleden door de ruimte zwevende brokken zijn ingeslagen. Met een verrekijker kan je ze nauwkeurig bekijken.

Op de maan is er geen lucht. De ruimtevaarders die er rondlopen, moeten dus ruimtepakken aanhebben. Daarin zit de lucht die ze nodig hebben om te ademen. Ook moet het pak hen beschermen tegen de zon. Die kan in de ruimte fel schijnen omdat er geen lucht en wolken zijn. Het kan tot meer dan honderd graden warm worden op de maan!

AVI 9

De postduif

Mijn opa heeft wel twintig duiven. Hij geeft ze allemaal een naam. Ze wonen in een hok in de tuin. Elke dag krijgen ze granen, zaden en peulvruchten te eten. Die hebben ze nodig om voldoende energie te hebben en om de temperatuur van hun lichaam op peil te kunnen houden.

Opa vertelde me dat er veel soorten duiven bestaan. Eén van hen is de postduif. Dat is een duif die een heel grote afstand in één keer kan afleggen en gemakkelijk de weg naar huis terugvindt. Daar heeft die duif geen kompas of landkaart voor nodig. Hij weet de weg door naar de zon te kijken!

Daarom werden deze duiven in de tijd van de oude Romeinen en Grieken gebruikt om berichten te versturen. Toen had men immers nog geen auto of vliegtuig om post mee te vervoeren. De duiven droegen dan kleine berichtjes met zich mee in een kokertje aan hun poot.

Nu zorgt de postbode dat je brief in de juiste bus terechtkomt. Toch zijn veel mensen in Nederland en België lid van een postduivenvereniging. We noemen ze duivenmelkers. Ze gebruiken hun duiven alleen nog voor de sport. De duiven worden goed getraind om mee te kunnen doen aan wedstrijden. Als je duif als eerste thuiskomt, dan heb je gewonnen. En dan win je geld of een andere leuke prijs!

B. Classification of errors

FOUTE LEESSTRATEGIE	
decoderend lezen	
(a)	lettergrepig spellen zonder nadien (juist) te synthetiseren
(b)	letter spellen zonder nadien (juist) te synthetiseren
(c)	delen van het woord (letters, lettergrepen, andere woorddelen) werden afzonderlijk goed gelezen zonder nadien (juist) te synthetiseren
(d)	
directe woordherkenning	
(e)	een bestaand woord wordt vervangen door een ander bestaand woord dat grafemisch gelijkend is
(f)	een bestaand woord wordt vervangen door een ander bestaand woord dat semantisch gelijkend is, bv. poes/kat
(g)	een pseudowoord wordt vervangen door een ander pseudowoord dat grafemisch gelijkend is
(h)	een pseudowoord wordt vervangen door een bestaand woord dat grafemisch gelijkend is (enkel indien het HELE woord een bestaand woord is)
(i)	
(j) **fouten niet te benoemen**	

SEQUENTIE	
(1)	omwisselen van letters die vlak naast elkaar staan, bv. droon/doorn
(2)	omwisselen van letters die door minstens 1 letter van elkaar gescheiden zijn, bv. stol/slot
(3)	
SUBSTITUTIE	
klinker	
(4)	substitutie van klinkers, bv. bal/bel
(5)	verwisselen van korte en lange klanken
(6)	fouten tegen leenvocalen, bv. 'au' in 'chauffeur' / 'oi' in 'toiletpapier'
(7)	tweeklank niet als dusdanig uitspreken, maar als aparte klanken
(8)	
medeklinker	
(9)	spiegelen van medeklinkers, meestal voorkomend bij b/d/p/q/g, t/f, m/w
(10)	verwisselen van een stemhebbende medeklinker met zijn stemloze variant of omgekeerd, bv. /x//g, s/z /, f/v, t/d, p/b
(11)	verwisselen van visueel op elkaar lijkende letters, bv. c/G, m/n, k/h, t/l, l/k, r/z, v/w, y/j, h/b
(12)	maken van fouten tegen stemassimilatie, meerbepaald vervangen van /N/ door /nx/, /nG/, /Nx/ of /NG/, vervangen van /Nk/ door /nk/ of /N/, fouten tegen stemassimilatie tussen 's' en 'j' in 'sj'

SEQUENTIE	
(1)	omwisselen van letters die vlak naast elkaar staan, bv. droon/doorn
(2)	omwisselen van letters die door minstens 1 letter van elkaar gescheiden zijn, bv. stol/slot
(3)	
SUBSTITUTIE	
klinker	
(4)	substitutie van klinkers, bv. bal/bel
(5)	verwisselen van korte en lange klanken
(6)	fouten tegen leenvocalen, bv. 'au' in 'chauffeur' / 'oi' in 'toiletpapier'
(7)	tweeklank niet als dusdanig uitspreken, maar als aparte klanken
(8)	
medeklinker	
(9)	spiegelen van medeklinkers, meestal voorkomend bij b/d/p/q/g, t/f, m/w
(10)	verwisselen van een stemhebbende medeklinker met zijn stemloze variant of omgekeerd, bv. /x//g, s/z /, f/v, t/d, p/b
(11)	verwisselen van visueel op elkaar lijkende letters, bv. c/G, m/n, k/h, t/l, l/k, r/z, v/w, y/j, h/b
(12)	maken van fouten tegen stemassimilatie, meerbepaald vervangen van /N/ door /nx/, /nG/, /Nx/ of /NG/, vervangen van /Nk/ door /nk/ of /N/, fouten tegen stemassimilatie tussen 's' en 'j' in 'sj'

(13)	maken van fouten tegen leenwoorden/leenklanken, meerbepaald vervangen van /S/ door /x/, /k/ of /G/ in bv. chocolade, machine; vervangen van /x/ door /k/; vervangen van /si/ door /ti/ in bv. vakantie; kiezen van de foute variant /s/ of /k/ voor het grafeem 'c'; foutief de grafemen 'ch' uitspreken als /tS/
(14)	verwisselen van een essentiële letter voor een willekeurige andere bv. kjop/knop
(15)	
klinker-medeklinker	
(16)	verwisselen van een medeklinker door een klinker of omgekeerd die visueel op elkaar lijken, bv. u-n / i-j
lettergreep	
(17)	substitutie van een lettergreep door een andere lettergreep of letter
DELETIE	
klinker	
(18)	deletie van willekeurige klinker
(19)	
medeklinker	
(20)	weglaten van de meervouds-n of –s of de vervoegings-t of -n
(21)	deletie van r of l in een medeklinkercluster
(22)	deletie van 'n' of 'g' in 'ng' of 'n' of 'k' in 'nk'

(23)	deletie van een willekeurige medeklinker, bv. rei/reis, netje/netjes
(24)	
lettergreep	
(25)	deletie van het achtervoegsel EN of ER, bv. kinder/kinderen, schoenwinkel/schoenenwinkel
(26)	deletie van een willekeurige lettergreep, bv. Gert/Gerrit
(27)	
andere	
(28)	de laatste poging is het begin van het woord maar het kind geeft dan op (of krijgt de kans niet meer om zich te verbeteren)
(29)	fouten als gevolg van het niet hernemen van het volledige woord bij correctie
INSERTIE	
klinker	
(30)	insertie van een willekeurige klinker
(31)	
medeklinker	
(32)	onjuist toevoegen van meervouds –n of –s, of vervoegings-t of -n
(33)	ten onrechte lezen van 'sch' wanneer er maar 1 of 2 van de drie grafemen gedrukt staan
(34)	insertie van 'n' voor 'g' of 'k', of 'g' of 'k' na 'n' waardoor het foneem /N/ of /Nk/ ten onrechte gelezen wordt

(35)	insertie van een willekeurige medeklinker, bv. poest/poes, gristeren/gisteren
(36)	
lettergreep	
(37)	willekeurig toevoegen van de achtervoegsels ER of EN, bv balker/balk
(38)	invoeging van een willekeurige lettergreep
(39)	

C. Paper

Cleuren, L., Duchateau, J., Sips, A., Ghesquière, P. and Van hamme H. (2006). Developing an Automated Assessment Tool for Children's Oral Reading. Proceedings of the 9th International Conference on Spoken Language Processing, Pittsburgh, USA, pp. 817-820 (Appendix C).

D. Poster

Developing an Automated Assessment Tool for Children's Oral Reading. 9th International Conference on Spoken Language Processing, Pittsburgh, USA, September 17-21 2006.