Reliability, Validity and Sensitivity to Growth

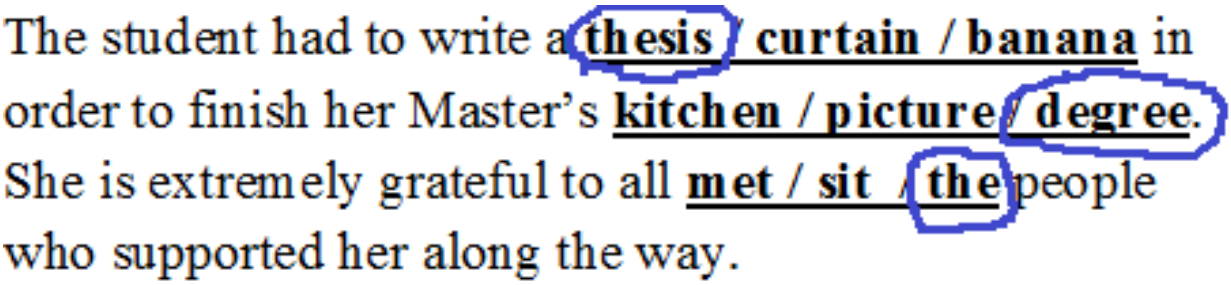
To what extent are CBM *Maze* scores indicators of general reading proficiency for students enrolled in Dutch PrOschools?

Master Thesis

Education and Child Studies

Faculty of Social Science

Leiden University



E. L. A. Koers

s0637033

Supervisor: Prof Dr. C. A. Espin

Second reader: Zsofia K. Takacs

Date: 16 July 2013

Table of Contents

Abstract

The purpose of this study was to examine to what extent CBM *Maze* scores were indicators of general reading proficiency for students enrolled in Dutch vocational secondary education (PrO-schools). Three research questions were addressed regarding the reliability, the validity and the sensitivity to growth of the *Maze* scores. The participants were 96 students (44 girls; 51 boys; 1 unknown sex) across four school years enrolled in a Dutch secondary school for practical education. The relation between the CBM *Maze* scores and two criterion variables (a reading comprehension test *Nieuwsbegrip* and School Grades for Dutch) was examined. In addition, beginning-and-end-of-year mean *Maze* scores were examined in order to determine whether the scores were sensitive to growth. Positive and moderately to strong correlations provide support for the reliability of the *Maze* scores. The present data, however, provide only minimal support for their validity. Finally, results suggest the *Maze* scores were sensitive to growth.

Introduction

"In today's world the ability to read is a valued and vital skill" (Cain, 2010, p.2). People are continuously surrounded by print and must engage in reading on a regular, daily basis. Reading is a fundamental tool that is necessary if one is to achieve success in today's society (Snow, Burns & Griffin, 1998). Reading opens up educational and employment opportunities and provides access to the essentials of everyday life, such as buying food, obtaining healthcare or even exercising one's right to vote.

1.1 Learning to read

In order to become a skilled reader, children need to acquire two broad skills: "the ability to decode the individual words on the page and the ability to comprehend text" (Cain, 2010, p.26). The decoding process requires understanding the phoneme-grapheme relationships and translating printed words into spoken language. Comprehension involves understanding the meaning of words in isolation and in context (Mercer & Pullen, 2009), and building coherent mental representations of text while reading (van den Broek, Young, Tzeng & Linderholm, 1999).

In The Netherlands, children typically first learn to read at about the age of six. Reading instruction begins with teaching children the process of decoding by applying the alphabetic principle. At this stage, children learn the systematic relationship between letters and sounds and the fact that letters and sounds go together to form words (Nicholson, 1997). In Dutch primary schools, the level of decoding ability is indicated by so-called AVI-levels, which range from AVI-M3 to AVI-Plus (Stichting Cito, 2012). Children are expected to gradually achieve higher AVI reading levels as they move through their school careers. For example, by the end of the last grade in primary school, children are expected to have reached the level AVI-Plus (Stichting Cito, 2012).

The ability to decode and fluently read words, also referred to as *word-reading*, is an important prerequisite for *reading comprehension* (Perfetti, 1977). A child who struggles to read words will generally struggle to understand the meaning of the text (Cain, 2010). In addition, research has demonstrated high correlations, often $r = .70$ or greater, between children´s word reading ability and their understanding of what is read (Vellutino, Tunmer, Jaccard & Chen, 2007). Consequently, it is of fundamental importance that word reading skills be taught at the beginning stages of reading instruction for children to become skilled, comprehensive readers.

1.2 Reading difficulties among Dutch (secondary) school students

The majority of Dutch children learn to read relatively quickly without encountering any severe problems. Unfortunately, not every child learns to read at the desired pace. While the skilled reader learns to read relatively effortlessly, about 10% - 15% of the Dutch primary school children undergo great difficulty in their attempts to master this vital skill in their first year of reading instruction (groep 3) (Inspectie van het Onderwijs, 2005). Their laborious and inaccurate decoding of words results in slow and inefficient access to meaning. As a consequence, memory and cognitive processes are primarily devoted to decoding words rather than to extracting meaning from text, and extracting meaning is the ultimate goal of learning to read (Perfetti, 1977; Cain, 2010). By way of illustration, research carried out by the Dutch School's Inspectorate (Inspectie van het Onderwijs) in 2007 and 2008 demonstrated that approximately 25% of the primary school children in group 8 (final year at primary school) did not achieve the final AVI reading level and left primary schools with – on average - a two year lag (reading level AVI-M6-E6; group 6).

Vernooy (2006) suggested that insufficient reading levels are related to inadequate reading instruction and a lack of effective reading interventions for poor readers within the Dutch primary schools. Consequently, it was not surprising when later research indicated many reading difficulties often persisted into secondary school (Inspectie van het Onderwijs, 2007). For example, a study which involved 40 schools demonstrated that first year Dutch Vmbo-students (one of the lower levels of secondary education in The Netherlands) achieved lower results regarding text comprehension than the national average (Inspectie van het Onderwijs, 2011a). Similarly, Hacquebord, Linthorst, Stellingwerf and de Zeeuw (2004) found that 25% of the Vmbo-students (Bbl & Kbl) had severe problems in independently reading and understanding texts in their schoolbooks. Although these lower results regarding text comprehension for Vmbo-students might not be seen as surprising – given the fact that these students are often enrolled in a lower level of secondary school because of lower reading levels - end-of-year reading results did not show any *improvement* compared to results at the beginning of the school year (Inspectie van het Onderwijs, 2011a).

According to research carried out by the Programme for International Student Assessment (PISA, 2009), 14.3% of the Dutch secondary school children have been shown to read at a level that reflects an inability to read and understand simple texts. Not surprisingly, these levels have been found to correspond to levels of low literacy (Inspectie van het Onderwijs, 2011a). The majority of these children are mostly found in so-called PrO-schools (Dutch: *Praktijkonderwijs*) and pre-vocational secondary education (Vmbo Bbl & Kbl)

(Inspectie van het Onderwijs, 2011a). Thus, it is not without reason that people refer to this situation as a *'reading crisis'* (Vernooy, 2006).

1.3 A 'Reading Crisis': the importance of monitoring reading performance

Past research has proven the importance of good reading skills as a prerequisite for school success and future participation in society (Vernooy, 2006). Participation in our society entails participation in our so-called *knowledge-based economy*. A knowledge-based economy refers to the central role that knowledge plays in economic development and growth (Organisation for Economic Cooperation and Development (OECD), 1996). An economy of this kind is heavily dependent on a labor force that has acquired a range of knowledge and skills and that is capable of continuously adapting itself through 'life-long learning'. Life-long learning requires the acquisition of fundamental basic skills such as *literacy* and numeracy (European Commission, 2007). *¨Building on these skills, an individual should be able to access, gain, process and assimilate new knowledge and skills"* (European Commission, 2007, p.8).

In accordance with the concept of life-long learning, learning to read in itself is inevitably viewed as a life-long process (Commission for Reading of the National Council of Teachers of English, 2004). However, as students grow older, the amount of reading they do decreases (Stichting Lezen en Sectorinstituut Openbare Bibliotheken, 2013). Furthermore, Willms and Murray (2007) demonstrated that literacy levels drastically diminish once poor readers give up reading. Besides, research carried out by Kamil (2003) has shown that a majority of secondary school teachers do not feel responsible for the reading skills of their students. They feel they should concentrate on the content of their classes and not on reading skill development (Vernooy, 2006).

Research conducted by Biancarosa & Snow (2005) has highlighted two major reasons why the improvement of reading skills of weak readers within secondary schools is so important. First, students are expected to possess advanced reading skills which are necessary to process the content of a wide array of sometimes complex school subjects. Second, students avoid reading which, in turn, ensures a further decline in their level of reading. In addition, students show less motivation to become better readers and lack interest in reading books related to their school subjects (Vernooy, 2006).

In summary, the magnitude of the so-called *'reading crisis'* should be taken seriously. A large part of our future national labour force is to be manned by this group of young people (Vernooy, 2009). Considering the importance of being able to read in today's society and

economy, the development and implementation of effective interventions to enhance secondary school students' poor reading levels has become a high priority within the Dutch government (Rijksoverheid, 2012). A majority of students are in need of intensive reading interventions and in turn, their teachers are in need of an easy tool to measure and evaluate the efficacy of their interventions on students' learning. The use of Curriculum-Based Measurement (CBM), an approach for assessing the growth of students in basic skills including reading, may be the answer to this problem.

1.4 Curriculum Based Measurement (CBM)

CBM originated in 1977 (Deno and Mirkin) in an attempt to test the effectiveness of a specific special education intervention model. Since 1985, however, the use of CBM has extended its objectives to monitoring individual progress and the progress of the class as a whole and using those data to evaluate the effectiveness of students' instructional programs (Wayman, Wallace, Wiley, Tichá & Espin, 2007). CBM has established itself as an alternative to commercial standardized tests and to informal observations within the American school system (Deno, 1985).

The fundamental idea behind CBM is that it uses a systematic set of procedures for indexing students' academic competence inside the school's curriculum (Deno, 1985). Standardized tests in basic academic skills (including reading) are frequently administered to monitor a student's progress. The corresponding results inform teachers as to whether a particular instruction or intervention is working properly (Fuchs & Fuchs, 1992). Consequently, teachers can decide whether certain instructional changes are needed to fit the individual student's needs (Tichá, Espin & Wayman, 2009).

Reading is an example of a basic academic skill that can be monitored through the use of CBM. In many studies, the so-called *Maze tasks* are used as brief assessments of reading performance to measure students' reading progress and to evaluate the effectiveness of instructional programs (National Center on Student Progress Monitoring, 2010). A *Maze* is a multiple-choice cloze task that students complete while silently reading a short passage (Shinn, Deno & Espin, 2000). Every first sentence of the passage stays intact, and subsequently every seventh word is deleted and replaced with three word choices. One of the alternatives is the correct word choice and the two incorrect word choices are the distractors. Distractors are designed to be easily distinguished from the correct choice (Fuchs & Fuchs, 1992). The student's score is typically the number of words chosen correctly in two or three minutes (Tolar, Barth, Francis, Fletcher, Stuebing & Vaughn, 2012).

The use of *Maze* tasks as a progress-monitoring tool is interesting to teachers because multiple data points can be generated to graph growth across a school year and the *Mazes* can be administered in a group format. Thus, teachers are able to frequently assess the reading proficiency of their entire class within minutes (Wiley & Deno, 2005). *Maze* tasks may also have the potential to measure the level of reading comprehension. More specifically, Fuchs, Fuchs and Maxwell (2002) suggested that the ability to make a correct word choice in a *Maze* task may reflect language-based processes (such as the use of background knowledge and inference making) that are involved in building coherent mental models of the text while reading. Furthermore, capable readers generally understand the grammar and the meanings of the words as they are used in texts (Cain, 2010). Students with reading difficulties, however, do not comprehend what they read well enough to choose words based on semantic and syntactic accuracy. Reading assessment through the use of the *Maze* scores may help identify these students and monitor changes in their reading behaviors as the result of instruction or practice.

1.5 Technical adequacy of the Maze scores

Although the use of the CBM *Maze* scores sounds promising and appealing, the following question arises: *Is there sufficient scientific proof for their use as reliable and valid indicators of general reading proficiency?*

Since the 1980s, numerous research studies on CBM of reading have been carried out in attempts to examine the efficacy of the *Maze* scores (Wayman et al., 2007). As a result, evidence has been provided for their technical adequacy, including the reliability and validity, as indicators of general reading proficiency. In order to help interpret the strength of reliability and validity coefficients from previous CBM research studies in reading, Wayman et al. (2007) randomly chose the following guidelines: strong relations (r = .70 and above); moderate relations (r = .50 to .70) and weak relations (r < .50) (Wayman et al., 2007). Several results will be discussed below.

Across various CBM research studies, the *Maze* score has been shown to be a reliable and valid measure of reading ability that measures skills such as decoding and comprehension (Fuchs & Fuchs, 1992; Shin et al., n, 2000; Wiley & Deno, 2005; Tichá et al., 2009; Wayman et al., 2007). For instance, in a study conducted by Shin et al. (2000), the reliability of the *Maze* was measured by assessing the reading performance of 43 second graders (7-8 years old) over a school year. Ten different *Maze* passages were collected monthly via the computer, using different forms of the task. Correlations between monthly *Maze* scores

ranged from .69 to .91 (*M* = .81) (Shin et al., 2000). Also, the *Maze* scores reflected improvement of both individual and group student performance over the year and were positively related to later reading performance on a standardized reading test (Shin et al., 2000). Similarly, research conducted by Wiley and Deno (2005) provided evidence (*r* = .73) that *Maze* measures were predictive of primary school children's performance on a standardized, Curriculum-Based reading comprehension test, the *Minnesota Comprehension* Assessment (MCA) in reading. The children were asked to complete comprehension tasks, including identification of the main idea in a text and making predictions based on information in the passage.

Although most evidence for the use of *Maze* scores as indicators of general reading proficiency stems from research conducted across primary schools, some proof has also been provided for its use among secondary school students. For example, a study carried out among 236 eighth-grade students (Espin, Wallace, Lembke, Campbell & Long, 2010) revealed that the *Maze* task (progress measured at 2, 3, and 4 minutes) was a reliable and valid predictor of performance on the *Minnesota Basic Standards Test* in reading (.70 and above). In accordance, Tichá et al. (2009) observed high reliability and validity coefficients for *Maze* task measures among secondary school students. In that study, reading performance of 35 eighth-grade students was assessed weekly over a time period of 10 weeks. Results revealed the *Maze* measures had good reliability and validity as indicators of performance on the *Minnesota Basis Skills Test* in reading (MBST; a states standard test in reading) and the *Woodcock-Johnson III Test of Achievement Broad Reading Cluster* (WJ-III; assessed letter-word identification, reading fluency and passage comprehension) (Tichá et al., 2009). Most relations observed were .80 and above. Correlation coefficients between the *Maze* selection and the WJ-III ranged from .86 and .88, suggesting high criterion validity.

In addition to the studies regarding the reliability and validity of the *Maze* task, a review by Fuchs and Fuchs (1992) revealed that teachers rated the *Maze* task as an efficient tool to measure general reading proficiency, enabling the assessment of multiple aspects of reading, including decoding ability, fluency and comprehension.

1.6 Current research

Although extensive research has been conducted concerning the reliability and validity of the CBM *Maze* scores as indicators of general reading proficiency within the American school system, no CBM research in reading has been carried out with regard to the Dutch school system. Because the development and implementation of effective monitoring tools and

interventions to enhance students' poor reading levels has become a high priority within the Dutch government (Rijksoverheid, 2012), this study will focus on the effectiveness of the CBM *Maze* scores as indicators of general reading proficiency among Dutch secondary school students. More specifically, special attention will be paid to the use of the *Maze* task by the lowest achieving students enrolled in vocational training programs at PrO-schools (*Praktijkonderwijs*).

The main question of this study is: *Are the Maze scores indicators of general reading proficiency for students enrolled in Dutch vocational secondary education (PrO-schools)?* The following research questions should assist in finding the answer.

1.7 Research Questions

Research Question 1: *What is the alternate-form reliability of CBM Maze scores as indicators of general reading proficiency for students in vocational secondary education?*

Subquestion 1.1: What is the correlation between adjacent CBM *Maze* scores?

Hypothesis A: Positive correlations are expected between adjacent maze tasks, because *Maze* tasks are expected to assess general reading proficiency.

Hypothesis B: Correlations are expected to increase gradually as students complete more and more *Maze* tasks due to an expected increase of the variability in reading scores. Students are expected to improve at different rates.

Research Question 2: *What is the validity of the CBM Maze scores as indicators of general reading proficiency for students in vocational secondary education?*

Subquestion 2.1: What is the relation between the *Maze* scores and the scores on the criterion test for reading comprehension (*Nieuwsbegrip*)?

Hypothesis C: A strong correlation is expected between the *Maze Pretest* scores and the criterion test scores for *Nieuwsbegrip*, based on the assumption that the *Maze* scores are indicators of general reading proficiency.

Hypothesis D: A significant mean difference in *Maze* scores is expected between students in the AA and A *Nieuwsbegrip* levels. Based on the assumption that the *Maze* scores are valid, higher *Maze* scores are expected for the group of level A participants than for the group of level AA participants.

Subquestion 2.2: What is the relation between the *Maze* scores and year in school?

Hypothesis E: The assumption is that older students will have significantly higher scores on the *Maze*, because they are better readers than younger students.

Subquestion 2.3: What is the relation between the *Maze* scores and the school grades in the subject Dutch?

*Hypothesis F***:** A strong relation is expected between the *Maze Posttest* scores and the school grades in Dutch, based on the assumption that school grades in Dutch reflect a general language and reading proficiency.

Research Question 3: *Are the CBM Maze scores sensitive to growth?*

Hypothesis G: The mean difference between the *Maze Pretest scores* and the *Maze Posttest* scores is expected to be significant, based on the assumption that the students general reading proficiency will improve over time.

Methods

2.1 Background information: The Dutch school system

In the Netherlands, children normally leave primary school at around the age of 11 or 12. At the end of primary school (group 8), the vast majority of schools conduct an achievement test (*Cito-toets, NIO-toets* or *Schooleindonderzoek*) which is designed to help in the decision making regarding the type of secondary education best suited to a pupil. Secondary education is compulsory until the age of 16 and is offered at several levels. The choice for the right level of secondary education is not only based on the test scores, but also is heavily dependent on the recommendation of the school teacher and the opinion of the pupil and his or her parents. A choice is made for either *practical education* (Praktijkonderwijs: vocational training), *Vmbo* (preparatory middle-level applied education) which consists of four different levels), *Havo* (higher general secondary education) or *Vwo* (pre-university education). Vmbo programs combine general and vocational education, after which pupils can continue in senior secondary vocational education and training (Mbo) lasting one to four years (http://www.kempel.nl/DeKempel/Documents/EducationSystemInTheNetherlands.pdf, obtained on 4th February 2013). According to research carried out by the Programme for International Student Assessment (PISA, 2009), approximately 50% - 60% of the secondary school students in the Netherlands are enrolled in a level of Vmbo. Havo and Vwo are the two programs of general education that grant admission to higher education.

2.2 Research Setting

This study was conducted as part of a larger research study which focused on the reliability and validity of an individual progress monitoring tool called Curriculum-Based Measurement (CBM) at the Dutch secondary school level. In order to obtain sufficient results across various secondary school levels, the participation of a large secondary school was required. A public school in a large city (± five-hundred-thousand people) in the Netherlands (Zuid-Holland) participated in the larger research study. The school offered different levels of secondary school education across three school sites in the city. These school levels consisted of education tailored to second-language learners (*School site 1*), Vmbo (BBL & KB), Mavo (Vmbo TL), Havo and Vwo (*School site 2*) and practical education (*School site 3*). Students from various ethnic and cultural backgrounds attended the school.

The school site that participated in the study reported here was the practical education site (*School site 3*). Practical education is meant for the group of lower (lowest) achieving

students and primarily consists of vocational training. It is tailored to students who would otherwise not be able to obtain a Vmbo-diploma. Moreover, practical education is a form of on-the-job training aimed at allowing students to enter the job market directly. Nonetheless, a sufficient level of reading is still important in order to function adequately within the job setting and to gain access to the essentials of everyday life such as buying food or obtaining healthcare.

2.3 Participants

A total of 127 students across four years of the school participated in the study. Due to absences and incomplete information, students (including one complete school year) were excluded from the study. The final research sample (N = 96) consisted of 44 girls (45.8%) and 51 boys (53.1%); of one student the sex was not reported. Students were born in the Netherlands (40.6%), in countries where Dutch was one of the national languages (9.4% percent), in Eastern-Europe and Russia (8.3%), in Non-western and Mediterranean countries (26%) and for the rest of the students their country of birth was unknown (13.5%). At the onset of the study, the average age of the students was 14.55 years (*SD* = 1.17; *Min* = 12; *Max* = 17). An overview of the average age per year in school (March 2012) is shown in Table 1.

Table 1

Descriptives average student age per school year (March 2012).

	M	*SD*	*Min*	*Max*
Year 1 (N = 22)	13.09	.526	12	14
Year 2 (N = 24)	14.17	.761	13	16
Year 3 (N = 36)	15.19	.624	14	17
Year 4 (N = 13)	15.92	.760	15	17

2.4 *Measures*

The following measures were used in the study: a CBM progress monitoring task (the *Maze*), a test for reading comprehension (*Nieuwsbegrip*) and *School Grades in Dutch*.

Maze (predictor variable). The *Maze* tasks were texts with an approximate length of 400 words. As a rule, every first sentence of the passage stayed intact, and subsequently every seventh word was deleted and replaced with three word choices. One of the alternatives was the correct word choice and the two others were distractors. The three word choices were

underlined in bold print and, in order to preserve continuity for the reader, were not split at the end of the sentence (a paragraph of a *Maze* task is shown in Figure 1).

Figure 1

Example of a Maze paragraph

Vrije tijd is eigenlijk precies wat er staat: namelijk de tijd dat je vrij bent. Maar wat wordt daar nu precies **mee / hak / lui** bedoeld? Met vrije tijd wordt de **knal / hobo/ tijd** bedoeld dat je niet hoeft te **inhoud / werken / eigeel**, niet naar school hoeft en geen **huishoudelijk / kurkentrekker / verschillende** werk hoeft te doen.

At the onset of the study (March 2012), all students were asked to complete the first two *Maze* texts on one day in order to establish an average baseline score. These texts, Task 1 and Task 2 (*Maze Pretest*), were the same for all the students in the study. Subsequently, the teachers administered one *Maze* text per week in counterbalanced order. In June 2012, the students completed Task 18 and Task 19 (*Maze Posttest*) on one day, creating an average end score. The tasks in the *Maze Posttest* (Task 18 and Task 19) were identical to the tasks in the *Maze Pretest* (Task 1 and Task 2).

Students were given two minutes to read each passage and choose as many correct words as possible by circling their word choice. In addition, they were asked to draw a line behind the last word they read after two minutes of reading.

Mazes were scored for the number of correct word choices. Maze scoring was stopped when three consecutive incorrect word choices were made. In that case, the participant's reading score was the total number of correct word choices made before the three consecutive incorrect word choices.

Nieuwsbegrip (criterion variable). *Nieuwsbegrip* is a comprehensive reading method with a special focus on the use of current news topics. It was designed and implemented in Dutch schools by the CED-groep Rotterdam, a national expertise center for support and innovation in the educational field. The research department of the CED-groep is responsible for the development of classroom materials and teaching methods. *Nieuwsbegrip* is an example of one of their products.

Nieuwsbegrip is designed to promote students' reading motivation in a meaningful way, both in elementary and secondary schools (CED groep, 2013). Unfortunately, there is no

evidence available yet with regard to the reliability and validity of the scores from the test as indicators of reading comprehension. These efforts are currently underway.

Since its implementation in 2005, many schools have incorporated the *Nieuwsbegrip* method into their curriculum. By way of illustration, in 2009, more than 3000 primary schools in the Netherlands were using the method (www.edux.nl, obtained on 8th of June 2013). In addition, according to SLO (2012), a national expertise center for curriculum development, *Nieuwsbegrip* meets five important learning objectives (as described by the Dutch government in 2006) for high quality education in reading comprehension. For example, *Nieuwsbegrip* requires students to draw important information from informative and instructive texts and assists them in comparing information and authors' opinions in different texts.

Nieuwsbegrip can be administered at several levels (AA, A, B, C, D), depending on the students current level of reading. According to the CED groep (2013), level AA is suitable for primary school children in grade 4 (AVI-E3 – AVI-E4; 7 years old) and level A is suitable for grade 5 and grade 6 (AVI-E4-AVI-E6; 8 to 9 years old). Furthermore, CED groep recommends the use of level A to measure the level of reading comprehension for students enrolled in practical (vocational) education.

The students' scores on reading comprehension were assessed at three time points throughout the school year of 2011-2012. Due to a lack of results at several points in time, the decision was made to only use the *Nieuwsbegrip* scores obtained in March 2012. Those scores were compared with scores on the *Maze Pretest* (also March 2012) in order to examine the concurrent validity of the *Maze* task.

The students were accidentally given a different version of the *Nieuwsbegrip* reading test than was originally intended. Instead of completing a supposedly broad reading comprehension test (measuring a wide array of skills needed to be able to comprehend a text), students were given a so-called Nieuwsbegrip '*strategy-test'*. The purpose of the strategy test is to pinpoint which reading strategies are not sufficiently developed in students and aids teachers in determining which additional instruction is needed for each student. In short, the scores on the strategy test do not reflect a general level of reading proficiency.

Three levels of the *Nieuwsbegrip* (strategy) reading tests (AA, A and B) were used in this study. The choice to administer a particular level was based on the teachers' personal opinion regarding their students' level of reading comprehension and was decided by the school. The students were asked to read 3 texts and answer 15 questions. One point was given for each correct answer (*Maximum score*: 15). Thirty-five students completed a level AA test

(*Mean score* = 5.71; *SD* = 2.037 *Min* = 2.0; *Max* = 9.0) and thirty-eight students completed a level A test (*Mean score* = 6.16; *SD* = 2.455 *Min* = 0.0; *Max* = 10.0). Level B reading results were excluded from the study due to too few scores.

School Grades in Dutch (criterion variable). In the Netherlands, the traditional grading scale is from 1 through 10, where 1 is the lowest and 10 the highest grade. The pass mark for a single subject is usually 5.5 or 6 (Nuffic, 2006). The most common grades in secondary education are 6 and 7. Grades 1 through 1 to 4 are rarely given, and the same is true for grades 9 and 10 (Nuffic, 2006).

The school grades used in this study reflect the individual level and progress regarding the participants' general Dutch language proficiency. According to the school, no clear criteria were used by the teachers to determine individual grades. Moreover, the height of a grade appeared to be largely influenced by the degree of students' school commitment and the teachers' personal opinion regarding their students' level and progress.

Grades from three different school periods were obtained via the school teachers. The grades achieved at the end of the third term (June 2012) were used in this study to measure the relationship (criterion validity) with the achieved scores on the end-of-year *Maze Posttest* tasks. This relationship was examined separately within each school year, because the same grade achieved across different school years might represent different levels of performance. For example, a grade of '6' for a student in Year 1 probably reflects a different level of proficiency than the same grade for a Year 3 student. An overview of the descriptives regarding the variable *School Grades in Dutch* is given for each school year in Table 2.

Table 2

Descriptives School Grades in Dutch achieved in June 2012 for each school year

	M	SD	Min	Max
Year 1 (N = 11)	6.68	.72	5.50	7.50
Year 2 (N = 23)	6.87	.98	5.00	8.00
Year 3 (N = 34)	7.12	.78	5.00	8.50
Year 4 (N = 12)	6.17	.44	5.50	7.00

2.5 Data analysis

In order to answer the research question regarding the *reliability* of the *Maze* scores, adjacent correlations were computed between all consecutive *Maze* tasks.

With regard to the *validity* of the *Maze* scores as indicators of general reading proficiency, correlations were computed between the scores on the *Maze Pretest* and the scores on the criterion test for reading comprehension *Nieuwsbegrip* for both Level AA and Level A. For both measures, the scores were obtained within the same time period (March 2012) and were therefore compared to measure the *concurrent validity* of the *Maze* task. Secondly, a one-way ANOVA was carried out to examine whether there was a difference in mean scores on the *Maze Pretest* between the students from the four different school years. LSD pairwise comparisons were computed to reveal which *Maze Pretest* mean scores between specific school years were significantly different. Third, in order to examine the relationship between the criterion variable *School Grades in Dutch* and the *Maze* tasks, correlations were computed between the school grades achieved in June 2012 and the *Maze Posttest* (also administered in June 2012).

Finally, in order to determine whether the Maze scores were sensitive to growth, a t-test was computed between the mean scores on the *Maze Pretest (Maze Task 1 an Task 2)* and the *Maze Posttest (Maze Task 18 and Task 19).*

Results

3.1 Reliability Maze scores

In Table 1, descriptive statistics are presented for the *Maze* tasks. Initially, 19 *Maze* tasks were to be administered. However, due to a high non-response rate, the following tasks were removed from the study: Task 13 (*N* = 26, Task 14 (*N*=17), Task 15 (*N*=1), Task 16 (*N*= 0) and Task 17 (*N*=0). Task 1 and Task 2 were conducted on the same day, as were Task 18 and Task 19. These pairs of *Maze* tasks formed the *Maze Pretest* (Task 1 and Task 2) and the *Maze Posttest* (Task 18 and Task 19).

Table 1 displays the mean scores on the different *Maze* tasks. Means ranged from *15.30* to *22.03* correct word choices. The skewness ranged from *-1.33* to *.770.* Relatively high skewness scores were found for *Task 5 (-1.33)* and *Task 8 (.770).*

Table 1

Task (Maze Texts) descriptives (N =96)

Task	*N*	*Missings*	*M*	*SD*	*Skewness*	*Min*	*Max*
1	89	7	15.30	7.20	.061	1	32
2	87	9	16.08	7.50	-.053	0	32
3	88	8	16.90	7.56	.027	0	33
4	83	13	17.92	7.42	-.068	0	38
5	81	15	18.94	7.73	-1.33	0	37
6	83	13	18.01	8.99	.171	0	41
7	84	12	20.45	8.67	-.056	0	42
8	87	9	19.71	9.03	.770	1	56
9	87	9	19.62	9.81	.273	0	48
10	79	17	19.15	9.31	-.217	0	37
11	72	24	20.86	10.53	.185	0	46
12	71	25	19.20	10.11	-.077	0	44
18	77	19	20.99	8.53	-.230	2	41
19	76	20	22.03	8.09	-.341	0	40

Note: Tasks 1 & 2 were conducted on the same day, as were Tasks 18 & 19. Task 1 and Task 2 (*Maze Pretest*) were given again as a posttest (*Maze Posttest*) at the end of the study, and are labeled as Tasks 18 & 19.

To determine the reliability of the *Maze* scores as a indicators of general reading proficiency, correlations were calculated between adjacent scores on the *Maze* tasks (e.g. Task 1 + Task 2, Task 2 + Task 3, Task 3 + Task 4 etc.).

Recall that the *Maze Pretest* (Task 1 and Task 2) and the *Maze Posttest* (Task 18 and Task 19) were each administered on the same day. With regard to all the other *Maze* tasks, *Maze* tasks were administered one week apart. Reliability coefficients are reported in Table 2. Correlations ranged from r = .61 to r = .89. All correlations were significant. The highest correlations were found between Tasks 1 and 2 (*Maze Pretest*) (r = .89) and Tasks 18 and 19 (*Maze Posttest*) (r = .86). Recall that these pairs of tasks were administered on the same day. The lowest correlations were found between Tasks 5 & 6 (r = .65) and Tasks 6 & 7 (r = .61).

Table 2

Bivariate correlations adjacent Maze tasks

	N	*r*
Task 1&2	87	.89
Task 2&3	86	.72
Task 3&4	76	.77
Task 4&5	76	.77
Task 5&6	77	.65
Task 6&7	78	.61
Task 7&8	80	.67
Task 8&9	83	.68
Task 9&10	77	.77
Task 10&11	72	.67
Task 11&12	67	.71
Task 18&19	75	.86

Note: All results are significant $p < .001$

3.2 Validity Maze scores

The second research question addressed the validity of the *Maze* scores. More specifically: *What is the validity of the Maze scores as indicators of general reading proficiency for students enrolled in vocational secondary education?* Three analyses were carried out in an attempt to answer this question.

First, the relation between the *Maze* scores and the scores on the criterion test for reading comprehension (*Nieuwsbegrip)* was examined. For this analysis, scores on the *Maze Pretest* were used. The descriptives on the *Maze Pretest* are reported in Table 3.

Table 3

Descriptives Maze Pretest

	Maze Pretest (March 2012)
N Valid	87
Missing	9
Mean	**15.839**
Std. Deviation	7.045
Skewness	.053
Minimum	1.0
Maximum	31.0

Recall that two levels of the *Nieuwsbegrip* test were given; levels AA (lower) and A (higher). Mean scores on the *Nieuwsbegrip* test for these levels are reported in Table 4.

Table 4

Descriptives Nieuwsbegrip grouped by Level AA and Level A participants

	Level AA	*Level A*
N Valid	35	38
Mean	**5.71**	**6.16**
Std. Deviation	2.037	2.455
Skewness	-.340	-.516
Minimum	2.0	0
Maximum	9.0	10

Table 5 reports Pearson's correlation coefficient between the scores on the *Maze Pretest* and the scores on the *Nieuwsbegrip* test for both levels (AA and A).

Table 5

Correlation Maze Pretest and scores Nieuwsbegrip Level AA and Level A

	Scores Nieuwsbegrip Level AA *r*	Scores Nieuwsbegrip Level A *r*
Maze Pretest	.346* N = 30	.417** N = 37

Note: *Correlation is significant at $p < .05$ (1-tailed). **Correlation is significant at $p < .01$ (1-tailed).

The achieved scores on *Maze Pretest* are positively related to the scores on *Nieuwsbegrip Level AA and A*. Although the correlations are significant, the strength of the relations are weak ($r < .50$). The correlation between the scores on the *Maze Pretest* and the *Nieuwsbegrip* test appear to be higher for the *Level A* participants ($r = .417$) than the *Level AA* participants ($r = .346$).

As a second analysis for the validity of the *Maze* scores, mean scores on the *Maze Pretest* for students in the AA and A *Nieuwsbegrip* level were compared. If the *Maze* scores are valid, we would expect higher *Maze* scores for the group of level A participants. The mean scores reported in Table 6 reveal a difference of approximately 1 correct word choice between the two groups; however these differences were not significant ($t(65) = -.760$, $p = .450$), perhaps due in part to the large standard deviations, especially for the level AA students.

Table 6

Descriptives Maze Pretest grouped by Level AA and Level A participants (Nieuwsbegrip)

	Level AA	*Level A*
N Valid	30	37
Mean	**14.517**	**15.878**
Std. Deviation	8.915	5.657
Skewness	.303	.206
Minimum	1.0	4.5
Maximum	31.0	29.5

As a third analysis, the differences in mean scores between the four consecutive school years were examined. Table 7 provides an overview of the mean scores on the *Maze Pretest* broken down by year in school. Inspection of the mean scores reveals a general pattern of increase in scores across the school years, except for Year 2.

Table 7

Overview of mean Maze Pretest scores broken down by year in school

	Year 1	*Year 2*	*Year 3*	*Year 4*
N Valid	22	21	31	13
Missing	0	3	6	0
Mean	**14.227**	**12.595**	**16.935**	**21.192**
Std. Deviation	6.841	5.813	7.255	5.345
Skewness	.506	-.293	-.218	.800
Minimum	4.0	1.0	2.5	15.0
Maximum	27.5	21.5	29.5	31.0

In order to examine the reliability of the differences in scores on the *Maze Pretest* between the four school years, a one-way-ANOVA was carried out. Overall, a significant effect was observed for school year ($F = (3, 86) = 5.316$, $p = .002$). Follow up analyses (LSD) revealed significant mean differences between years 1 and 4, years 2 and 3, and years 2 and 4 (Table 8).

Table 8

Mean Differences Maze Pretest for each School Year

Difference in school year	*Mean difference*	*Standard Error*	*p*
Year 1 and Year 2	1.632	2.003	.418
Year 1 and Year 3	2.708	1.831	.143
Year 1 and Year 4	**6.965***	2.298	**.003**
Year 2 and Year 3	**4.340***	1.856	**.022**
Year 2 and Year 4	**8.597***	2.318	**.000**
Year 3 and Year 4	4.257	2.170	.053

Note: *. The mean difference is significant at $p < .05$.

As a final analysis regarding the validity of the *Maze* scores, the relations between the scores on the *Maze* tasks and the school grades for the subject Dutch were examined within each school year. For this analysis, scores on the *Maze Posttest* were used. The descriptives on the *Maze Posttest* are reported in Table 9. Correlations between scores on the *Maze Posttest* and the mean school grades in Dutch within each school year are reported in Table 10.

Table 9

Descriptives Maze Posttest

	Maze Posttest
N Valid	78
Missing	18
Mean	**21.09**
Std. Deviation	8.14
Skewness	-.123
Minimum	3.0
Maximum	41.0

Table 10

Correlations mean School Grades in Dutch and Maze Posttest within school year

School year	*Mean Grade Dutch*	*SD*	*r*
Year 1 (N = 10)	6.682	.717	-.274 (*p* = .222)
Year 2 (N = 23)	6.870	.980	.197 (*p* = .183)
Year 3 (N = 32)	7.118	.779	**.338* (*p* = .029)**
Year 4 (N = 12)	6.167	.444	-.131 (*p* = .685)

Note: *. Correlation is significant at p < .05 (1-tailed).

Results revealed only one significant relation (school year 3) (r = .338, p < .05). Although the correlation is significant, the strength of the relationship is weak (r < .50).

Because the large majority of students had grades of around 7, the variability in the criterion variable was quite small. Thus, a second analysis was conducted in which mean different scores on the *Maze Posttest* were compared for students who had *passed* the subject Dutch (grade > 8.0) or who had *failed* the subject (grade < .5.5). Mean scores are reported in Table 11. Note that the sample size for "Fail" is very small. Students who passed for the subject Dutch selected 22.5 correct word choices compared to only 17.5 for those who failed the subject, but these differences were not statistically significant (*t*(21) = -1.357, *p* = .189).

Table 11

Descriptives Maze Posttest and School Grades, grouped by Pass or Fail

	Fail (grade < 5.5)	*Pass (grade > 8.0)*
N Valid	6	16
Missing	1	0
Mean Maze Posttest	**17.50**	**22.56**
Std. Deviation	10.02	7.34
Skewness	-.350	.115
Minimum	3	10
Maximum	29	37

3.3 Sensitivity to growth

Finally, in order to determine the sensitivity to growth, mean scores on the *Maze Pretest* and the *Maze Posttest* were compared. Means for both measures are reported in Table 12.

Table 12

Descriptives Maze Prestest and Maze Posttest

	Maze Pretest	*Maze Posttest*
N Valid	87	78
Missing	9	18
Mean	**15.839**	**21.09**
Std. Deviation	7.045	8.135
Skewness	.053	-.123
Minimum	1.0	3.0
Maximum	31.0	41.0

A paired samples t-test was carried out to compare the mean scores on the *Maze Pretest* and *Maze Posttest*. The mean difference appeared to be significant ($t(72) = -9.197$, $p = <. 00$).

Discussion

The goal of this study was to examine to what extent CBM *Maze* scores serve as indicators of general reading proficiency for students enrolled in Dutch vocational secondary education at PrO-schools.

The research questions addressed in this research study were:

(1) What is the alternate-form reliability of CBM Maze scores as indicators of general reading proficiency for students enrolled in vocational secondary education?

(2) What is the validity of CBM Maze scores as indicators of general reading proficiency for students enrolled in vocational secondary education?

(3) Are CBM Maze scores sensitive to growth?

4.1 Reliability Maze scores

To examine the reliability of the *Maze* scores used in this study, correlations were calculated between all consecutive *Maze* tasks. All adjacent correlations were significant ($p < .001$). The highest correlations were found between Tasks 1 & 2 (*Maze Pretest*; $r = .89$) and Tasks 18 & 19 (*Maze Posttest*; $r = .86$). The lowest correlations were found between Tasks 5 & 6 ($r = .65$, $p < .001$) and Tasks 6 & 7 ($r = .61$, $p < .001$). The correlations between Tasks 1 & 2 and Tasks 18 & 19 represent a strong relationship between the tasks. This is perhaps not surprising, because both pairs of tasks were administered on the same day. The rest of the Mazes were administered one week apart. In addition, tasks 3 through 12 were all administered in counterbalanced order. Thus, each set of correlations represent different pairs of Maze passages. Despite this time lag and differences in passages, correlations still tended to be in the .61 to .77 range.

In summary, the positive and moderately to strong correlations for the *Maze* tasks provide support for the reliability of the tasks, indicating that they consistently reflect general reading proficiency over repeated applications (*Hypothesis A*). However, correlations did not increase gradually as students completed more *Maze* tasks (*Hypothesis B*).

4.2 Validity Maze scores

With regard to the validity of the *Maze* scores as indicators of general reading proficiency, several analyses were carried out. In the first analysis, the relationship between the scores on the *Maze Pretest* and the scores on the criterion reading test for reading comprehension (*Nieuwsbegrip*) was examined. Analyses were conducted within levels (AA and A). A strong correlation was expected, based on the assumption that the *Maze* scores were indicators of general reading proficiency (*Hypothesis C*). Although the correlations were significant, the strength of the relations was weak (range from $r = .35$ for the Level AA participants to $r = .42$ for the Level A participants). There are several potential explanations for the weak correlations. First, it is possible that the *Maze* is not a good indicator of general reading proficiency. However, before this conclusion is drawn, it is important to consider other factors potentially influencing the outcomes. For example, the sample sizes were small, as was the range of performance levels, due to the fact that correlations were conducted within level (AA and A). Second, there were little data available on the validity and reliability of the *Nieuwsbegrip* test, and as described in the method, the test was not really intended to reflect broad reading comprehension.

In the second analysis, differences in mean *Maze* scores between students in Levels AA and A were examined (*Hypothesis D*). The hypothesis was if *Maze* scores were valid indicators of general reading proficiency, higher mean scores should be seen for better readers (Level A) than for poorer readers (Level AA). Although the mean scores revealed a difference of approximately 1 correct word choice between the two groups, the difference was not significant. A plausible reason for this could be the fact that the Level AA and A designations were not valid themselves. After all, the decision to designate a student to a certain *Nieuwsbegrip* level was largely based on the teacher's judgment and not on prior *Nieuwsbegrip* test scores.

In the third analysis, the relation between the *Maze* scores and the year in school was examined. The hypothesis was that older students would achieve higher scores on the *Maze* than the younger students, because the older students should generally be better readers than the younger students (*Hypothesis E*). A significant effect was observed with respect to the year in school. Follow up analyses (LSD) revealed significant mean differences between years 1 and 4, years 2 and 3 and years 2 and 4. These results support the validity of the *Maze* scores.

As a final examination of the validity of the *Maze* scores, the relation between the *Maze* scores and the school grades for the subject of Dutch was examined. A strong relation was expected between the *Maze Posttest* scores and the school grades in Dutch, based on the

assumption that school grades in Dutch reflect a general language and reading proficiency (*Hypothesis F*). Results revealed only one significant relation (school year 3) (r = .34, p < .05). Although the correlation was significant, the strength of the relationship was weak (r < .50).

Because the large majority of students had a grade of around 7, the variability in the criterion variable (school grades in Dutch) was quite small. Thus, a second analysis was conducted in which mean score differences on the *Maze Posttest* were compared for students who had *Passed* the subject Dutch (grade > 8.0) or who had *Failed* the subject (grade < .5.5). Students who passed for the subject of Dutch selected 22.5 correct word choices compared to 17.5 correct word choices for those who failed the subject, but these differences were not statistically significant. The sample size for the "*Fail*" group of students was extremely small (N = 6).

In summary, the results regarding the validity of the *Maze* scores reflect different outcomes. First, the relation between the *Maze* scores and the criterion test *Nieuwsbegrip* were significant but weak. Second, the relation between the *Maze* scores and the variable year in school were significant. Finally, the relation between the *Maze* scores and the school grades in Dutch were non-significant. Although there may be several competing explanations, the conclusion must be drawn that the present data provide only minimal support for the validity of the *Maze* scores.

4.3 Sensitivity to growth

The final research question in this study regarded the degree to which the CBM *Maze* scores were sensitive to growth. The mean difference between the *Maze Pretest* scores and the *Maze Posttest* scores was expected to be significant, based on the assumption that the general reading proficiency of the students would improve over time (*Hypothesis G*). The time gap between both measurements was three months. In accordance, the mean difference was significant; means increased from 15.84 correct word choices to 21.09 correct word choices. These results suggest that the *Maze* scores were sensitive to change over time.

4.4 Limitations and Implications for Future Research

The goal of this study was to examine whether CBM *Maze* scores are indicators of general reading proficiency for students enrolled in Dutch vocational secondary education at PrO-schools. The results provide support for the reliability of the *Maze* scores, but only provide

minimal support for the validity of the *Maze* scores. Certain limitations of the study may have contributed to the somewhat limited outcomes. These limitations will be discussed below.

First, the use of the reading test *Nieuwsbegrip* as a criterion variable must be questioned. The original plan was to use a *Nieuwsbegrip* test designed to be a general indicator of reading proficiency, but the test used by the school was actually a diagnostic test. It would be good in future research to use a criterion test with known validity and reliability.

Second, the final research sample consisted of 96 students across four school years. With regard to the validity of the *Maze* scores, the analyses required the examination of differences in mean *Maze* scores between groups (Level AA and Level A *Nieuwsbegrip* participants) and differences in mean grades for the participants across the four school years. These analyses led to the use of (sometimes extremely) small sample sizes. For instance, the differences in mean scores on the *Maze Posttest* were compared for students who passed the subject Dutch or who had failed the subject. The sample size for "Fail" was very small (N = 6) and therefore the non-significant outcome of the analysis was not surprising. Larger sample sizes in the future are desirable.

Third, some analyses included data with large standard deviations. For example, the comparison of the mean scores on the *Maze Pretest* for students in the AA and A *Nieuwsbegrip* level included large standard deviations for both groups. An experiment that yields data with large standard deviations is said to have low precision. Given that fact, the non-significant differences found were to be expected.

Finally, the research sample was also limited in the sense that only students enrolled in PrO-schools were examined. The comparison of scores between a variety of educational levels is necessary to further establish the reliability and validity of the *Maze* scores as indicators of general reading proficiency.

4.5 Conclusion

The main question of this study was: *Are the CBM Maze scores indicators of general reading proficiency for students enrolled in Dutch vocational secondary education at PrO-schools?*

Although the study results suggest the *Maze* scores are reliable to some extent, the present data provide only minimal support for the validity of the *Maze* scores. Consequently, the broad implementation of such a monitoring tool will have to be postponed until more research on the technical adequacy of the *Maze* scores has been carried out. Further research in this field is justified for an important reason. Recall that approximately 25% of the Dutch primary school children do not achieve the final AVI reading level by the end of primary

school and that many of these reading difficulties often persist into secondary school (Inspectie van het Onderwijs, 2007). Most of these children are found in preparatory middle-level applied education (Vmbo) and in vocational training programs (practical education) (Gille et al. 2010). Consequently, these students are in need of intensive reading interventions and in turn, the teachers of these students are in need of an easy tool to measure and evaluate the effectiveness of their interventions on students' learning.

Previous research in the U.S. provided support for the reliability and validity of *Maze* scores as indicators of general reading proficiency (Fuchs & Fuchs, 1992; Shin et al., 2000; Wiley & Deno, 2005; Tichá et al., 2009; Wayman et al., 2007). Building on those positive results regarding the technical adequacy of *Maze* measures, means that more research on the use of the CBM *Maze* task in the Dutch school system is needed. Moreover, research supports the importance of good reading skills as a prerequisite for school success and future participation in society (Vernooy, 2006). Therefore, to ensure that the students who 'failed' to learn to read properly in primary school are not overlooked in secondary school, further research on the reliability and validity of easy-to-use and relatively inexpensive monitoring tools such as the CBM *Maze* tasks is necessary and justified.

References

Biancarosa, G. & Snow, C. E. (2004). *Reading next – A vision for action and research in Middle and High School Literacy.* Retrieved January 18, 2013 from http://www.scobre.com/download/readingNext.pdf.

Cain, K. (2010). *Reading development and difficulties.* Oxford, United Kingdom: Wiley-Blackwell.

Commission on Reading of the National Council of Teachers of English. (2004). *On Reading, Learning to Read, and Effective Reading Instruction: An Overview of What We Know and How We Know It.* Retrieved March 13, 2013, from http://www.ncte.org/positions/statements/onreading.

Deno, S. L., & Mirkin, P. K. (1977). *Data-based program modification: A manual.* Reston, VA: Council for Exceptional Children.

Deno, S. L. (1985). Curriculum-based measurement: The emerging alternative. *Exceptional Children, 52,* 219-232.

Deno, S. L. (2003). Developments in Curriculum-Based Measurement. *The Journal of Special Education, 37,* 184-192.

Espin, C., McMaster, K., Rose, S. and Wayman, M.M. (2012). *The Influence of Curriculum-Based Measurement on Education.* Minneapolis, M.N: University of Minnesota Press.

Espin, C., Wallace, T., Lembke, E., Campbell, H. & Long, J. D. (2010). Creating a Progress-Monitoring System in Reading for Middle-School Students: Tracking Progress Toward Meeting High-Stakes Standards. *Learning Disabilities Research & Practice, 25(2),* 60-75.

European Commission. (2007). *Key Competences for Lifelong Learning. European Reference Framework.* Retrieved March 2, 2013, from *http://ec.europa.eu/dgs/education_culture/publ/pdf/ll-learning/keycomp_en.pdf.*

Fuchs, L. S. & Fuchs, D. (1992). Identifying a Measure for Monitoring Student Reading Progress. *School Psychology Review, 21,* 45-58.

Fuchs, L. S., Fuchs, D. & Maxwell, L. (2002). Treatment validity as a unifying construct for identifying learning disabilities. *Learning Disability Quarterly, 25,* 33-45.

Gunning, T. G. (1996). Creating reading instruction for all children (2nd ed.). Needham Heights, MA: Allyn & Bacon.

Hacquebord, H. I., Linthorst, T. R., Stellingwerf, B. P. & de Zeeuw, M. (2004). *Voortgezet taalvaardig. Een onderzoek naar tekstbegrip en naar taalproblemen en taalbehoeften van brugklasleerlingen in het voortgezet onderwijs in het schooljaar 2002-2003.* Groningen: ETOC.

Inspectie van het Onderwijs (2005). *De staat van het onderwijs. Onderwijsverslag 2004/2005*. Den Haag: Inspectie van het Onderwijs.

Inspectie van het Onderwijs (2007). *Basisvaardigheden taal in het voorgezet onderwijs. Resultaten van een inspectieonderzoek naar taalvaardigheden in de onderbouw van het vmbo en praktijkonderwijs. Inspectierapport 2007-32.* Den Haag: Inspectie van het onderwijs.

Inspectie van het onderwijs (2011a). *Basisvaardigheden in taal in het voortgezet onderwijs 2010. Resultaten van een inspectieonderzoek naar niveau basisvaardigheden taal en activiteiten scholen in vmbo leerjaar 1 en leerjaar 3 en 4 havo en vwo*. Utrecht: Inspectie van het Onderwijs.

Jackson, M. D., & McClelland, J. L. (1975). Sensory and cognitive determinants of reading speed. *Journal of Verbal Learning and Verbal Behavior, 14,* 565-574.

Kamil, Michael L. (2003). *Adolescents and Literacy. Reading for the 21st Century.* Retrieved February 24, 2013 from http://www.all4ed.org/files/archive/publications/AdolescentsAndLiteracy.pdf.

Mercer, C. D. & Pullen, P. C. (2009). *Students with learning disabilities.* Upper Saddle River, NJ: Pearson Merril.

Rijksoverheid. (2012). *Kwaliteit voortgezet onderwijs.* Retrieved February 13, 2013 from http://www.rijksoverheid.nl/onderwerpen/voortgezet-onderwijs/kwaliteit-voortgezet-onderwijs.

National Center on Student Progress Monitoring. (2010). *What is progress monitoring?* Retrieved April 5, 2013, from http://www.studentprogress.org.

Nicholson, T. (1997). Closing the gap on reading failure: Social background, phonemic awareness, and learning to read. In B. Blachman (Ed.), *Foundations of reading acquisition and dyslexia: Implications for early intervention* (p. 381-408). Mahwah, NJ: Erlbaum.

Nuffic. (2006). Nederlandse organisatie voor internationale samenwerking in het hoger onderwijs. *"Cijfers ontcijferd".* Notitie aangaande het omzetten van (examen)cijfers: een onderzoek naar de verdeling van cijfers en *grades* in het onderwijs van Nederland, de

Verenigde Staten, Canada en het Verenigd Koninkrijk. Retrieved March 16, 2013 from http://www.nuffic.nl/diplomawaardering/diplomawaardering/cijfersystemen.

Organization for Economic Co-Operation and Development. (1996). *The Knowledge-based Economy.* Retrieved March 3, 2013, http://www.oecd.org/science/sci-tech/1913021.pdf.

Perfetti, C. A. (1977). Language comprehension and fast decoding: Some psycholinguistic prerequisites for skilled reading comprehension. In J. T. Guthrie (Ed.), *Cognition, curriculum and comprehension*. (p. 20-41). Newark, DE: International Reading Association.

PISA. (2009). *Results: Volume II, Overcoming Social Background: Equity in Learning Opportunities and Outcomes*. Retrieved March 8, 2013 from http://www.oecd.org/pisa/pisaproducts/48852584.pdf.

Shin, J., Deno, S. L. & Espin, C. A. (2000). Technical adequacy of the maze task for curriculum-based measurement of Reading Growth. *The Journal of Special Education, 34*, 164-172.

Snow, C. E., Burns, M. S. & Griffin, P. (1998). *Preventing reading difficulties in young children*. Washington, DC: National Academy Press.

Stichting Cito. (2012). *Welk boek kan uw kind al lezen? Een uitleg over leesniveaus van kinderen*. Retrieved April 2, 2013 from http://www.cito.nl/nl/Onderwijs/Primair%20onderwijs/cito_volgsysteem_po/alle_producten/f721733fc9ff48718ff41c825e1a5f19.aspx?cachekey=cf13c0e6-81f1-422c-b92a-898ae289a5a3.

Stichting Lezen en Sectorinstituut Openbare Bibliotheken. (2013). Leesopvoeding Ouders Blijft Invloedrijk – Ook in de Puberteit. Retrieved June 27, 2013 from http://www.leesmonitor.nu/.

Stroop, J. R. (1935). Studies of interference in serial verbal reactions. *Journal of Experimental Psychology, 18,* 643-622.

Tichá, R., Espin, C. A. & Wayman, M. M. (2009). Reading Progress Monitoring for Seconday-School Students: Reliability, Validity, and Sensitivity to Growth of Reading-Aloud and Maze-Selection Measures. *Learning Disabilities Research & Practice, 24*, 132-142.

Tolar, T. D., Barth, A. E., Francis, D. J., Fletcher, J. M. Stuebing, K. K. & Vaughn, S. (2012). Psychometric Properties of Maze Tasks in Middle School Students. *Assessment for Effective Intervention, 37(3),* 131-146.

Van den Broek, P., Young, M., Tzeng, Y. & Linderholm, T. (1999). The landscape model of reading: Inferences and the online construction of memory representation. In H. van

Oostendorp & S. R. Goldman (Eds.). *The construction of mental representations during reading* (p. 71-98). Mahwah, NJ: Lawrence Erlbaum.

Vellutino, F. R., Tunmer, W. E., Jaccard, J. J., & Chen, R. (2007). Components of reading ability: Multivariate evidence for a convergent skill model of reading development, *Scientific Studies of Reading, 11,* 3-32.

Vernooy, K. (2006). Leesonderwijs basisonderwijs behoeft dringend verbetering. In Jansen et. Al (Red), *De nieuwe schoolstrijd*. Amsterdam, Uitgeverij Boom.

Vernooy, K. (2009). *Lezen stopt nooit! Van een stagnerende naar een doorgaande leesontwikkeling voor risicolezers.* Retrieved February 13, 2013, from http://www.surfsharekit.nl:8080/get/smpid:3305/DS2.

Wayman, M. M., Wallace, T. W., Wiley, H. I. & Espin, C. A. (2007). Literature Synthesis on Curriculum-Based Measurement in Reading. *Journal of Special Education, 42*, 85-120.

Wiley, H. I., & Deno, S. L. (2005). Oral Reading and maze measures as predictors of success for English learners on a state standards assessment. *Remedial and Special Education, 26*, 207-214.

Willms, D. J. & Murray, S. T. (2007). *Gaining and losing literacy skills over the lifecourse.* Statistics Canada. Retrieved February 26, 2013 from http://publications.gc.ca/collections/collection_2007/statcan/89-552-M/89-552-MIE2007016.pdf.