

# A new test for measuring text comprehension skills of at-risk adolescents: preliminary results of a validation study

**Author(s)**

Steensel, Roel van; van Gelderen, A.J.S.; Oostdam, R.J.; Trapman, Mirjam

**Publication date**

2009

**Published in**

Articles of Applied Linguistics

[Link to publication](#)

**Citation for published version (APA):**

Steensel, R. V., van Gelderen, A. J. S., Oostdam, R. J., & Trapman, M. (2009). A new test for measuring text comprehension skills of at-risk adolescents: preliminary results of a validation study. In A. Backus, M. Keijzer, I. Vedder, & B. Weltens (Eds.), *Articles of Applied Linguistics* Eburon.

**General rights**

It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

**Disclaimer/Complaints regulations**

If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please contact the library: <https://www.amsterdamuas.com/library/contact/questions>, or send a letter to: University Library (Library of the University of Amsterdam and Amsterdam University of Applied Sciences), Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.

## **A new test for measuring text comprehension skills of at-risk adolescents; preliminary results of a validation study**

Roel van Steensel, Amos van Gelderen, Ron Oostdam & Mirjam Trapman  
Universiteit van Amsterdam

Correspondentie-adres: Universiteit van Amsterdam, SCO-Kohnstammstituut,  
Postbus 94208, 1090 GE Amsterdam, e-mail: R.C.M.vanSteensel@uva.nl

### **1. Introduction**

This paper presents the first results of a validation study of a newly devised test for monitoring the text comprehension skills of at-risk students in the lower tracks of prevocational education. We will start the paper by describing the context within which the test was developed. Its construction was part of the ‘SALSA’<sup>1</sup> project, an extensive study which combines quantitative and qualitative research methods to identify predictors of at-risk adolescents’ literacy development. Subsequently, we will present the design of the test: we will discuss the way the reading texts were selected, and the way the test items were constructed. After explaining how we applied the concept of validity, we will present the research method of the validation study. In the final part of the paper, we will present the outcomes, draw preliminary conclusions, and describe further steps in the validation process.

### **2. The SALSA project**

Even though large groups of students in secondary education are at risk of lags in reading and writing development (OECD, 2001; Dagevos et al., 2003), studies into adolescent literacy and its determinants are relatively scarce (Curtis, 2002). Against this background, the ‘SALSA’ project – a collaboration between the universities of Amsterdam and Utrecht – was set up. SALSA is also part of an international, comparative study: similar projects were started in Toronto (Canada) and Geneva (Switzerland).

Starting from the bio-ecological model of cognitive development (Bronfenbrenner & Morris, 1998), the SALSA project aims to identify those individual attributes (i.e., knowledge, skills, attitudes), and features of both the educational and out-of-school context that further or impede the literacy development of at-risk students in the lower tracks of prevocational education (i.e., *vmbo-basisberoepsgerichte leerweg* and *vmbo-kaderberoepsgerichte leerweg*).

In the first stage of the project, a multiple case study was set up, in which a group of 63 students is followed in the course of three school years (2007/’08, 2008/’09, 2009/’10), starting in year 1. The students were selected on the basis of low scores on the CITO End of Primary School Test<sup>2</sup>, and on the basis of ethnicity and language background: in year 1, the sample consisted of 32 native Dutch, monolingual students and 31 nonnative, bi-/multilingual students. Students were identified as nonnative if both their parents were born in a ‘non-Western’ country (cf. Dagevos & Gijsbers, 2007). Students were identified as bi-/multilingual if they spoke

---

<sup>1</sup> Studie naar Achtergronden van Lees- en Schrijfontwikkeling bij Adolescenten.

<sup>2</sup> As an indicator for being at risk of an educational lag.

another language with their parent(s) at least as much as the Dutch language. Information on these variables was gathered via a student questionnaire.

Data on individual variables (cognitive and language skills, language attitudes), educational variables (e.g., frequency and nature of reading and writing instruction, self-regulated learning), and socio-cultural variables (e.g., students' out-of-school literacy experiences) are gathered by means of multiple techniques (tests, questionnaires, interviews, classroom observations, internet logs), and will be related to actual literacy development. In order to measure literacy performance, a combined reading and writing test was developed. In the context of this test, 'reading' was defined as text comprehension, and 'writing' as text production. This paper deals with the development of the reading part of the test.

### 3. Design of the SALSA reading test

The SALSA reading test was developed in 2007 by a team of researchers from the three parallel projects in Canada, Switzerland and the Netherlands (see section 2). A test version was piloted between November 2007 and February 2008. On the basis of the results of this pilot, the final test was compiled. The test consists of nine reading tasks. Each task is based on one or two texts about which comprehension questions have to be answered.

#### *Text selection*

Research has shown that different rhetorical modes or text types place different demands upon readers (cf. Harris et al., 1998; Lin et al., 2000). In order to prevent test scores from being too dependent on the ability to process one specific kind of text, an equal coverage of different text types was sought. Two distinctions were made in this respect: a distinction between continuous and discontinuous texts, and a distinction between narrative, argumentative, expository, and instructive texts (cf. Adams & Wu, 2000; OECD, 2003). Continuous texts are typically composed of sentences, that are organized into larger units, i.e., paragraphs, sections, chapters etc. Discontinuous texts – or 'documents' as Mosenthal & Kirsch (1998) call them – are texts organized in row/column formats, and include a.o. lists, schedules, tables and graphs. It has to be added that, within the context of the SALSA test, texts were also categorized as discontinuous if they *contained* lists, schedules etc. in addition to continuous text. Narrations or stories describe a (fictional or non-fictional) sequence of events around certain persons or objects. Argumentative texts are understood here as persuasive texts, in which a writer provides arguments to support a proposition. Expository texts give an explanation of a certain mental construct, the elements into which the construct can be analyzed, and the interrelations between those elements. Instructive texts provide directions, and include, for example, procedures, rules and regulations.

Students can come across texts in different kinds of media. Following a line of reasoning comparable to the one applied to text types, the aim was to strive for a balanced coverage of texts from different media types. Four media types were distinguished: (school) books, newspapers and magazines, official documents, and the internet. In all cases, the intention was to select texts that were relatively close to the target group (to prevent test results from being biased by lack in task motivation), and that were 'context neutral', i.e., texts that could be assumed not to have different cultural loadings in the three countries.

In Table 1 the distribution of the reading tasks over the text and media types is presented.

*Table 1: Distribution of texts over text and media types; discontinuous texts are indicated by a \**

	Narrative	Argumentative	Expository	Instructive
(School) books	NAR1 (‘Koning Kostuum’)		EXP1 (‘Vietnam’)	
Newspapers and magazines	NAR2 (‘De plank’)	ARG1 (‘Mobieltjesverbod’) ARG2 (‘Biobrandstof’)	EXP2 (‘Aromajockey’) EXP3* (‘Newsletter’)	
Official documents				INS1* (‘House rules’)
Internet				INS2* (‘Camping website’)

Even though all text and media types are represented, some cells remain empty.<sup>3</sup> This is because some combinations of text and media types are highly improbable (e.g., narrative official documents), or because some combinations were relatively hard to find. Two texts are in fact combinations of text types: EXP1, an informative text from a school book, also contains a narrative part. INS2, a web page about a camping site, contains, besides instructions and directions, also descriptive, expository information.

All texts are (parts of) authentic texts. In most cases the texts were not manipulated, in some cases small changes were made, for instance in order to adapt a text to the contexts of the three countries.

#### *Item construction*

Text comprehension implies that a reader makes a (valid) mental representation of the elements within a text and the connections between these elements (Kintsch, 1998; Land et al., 2008). In order to achieve this, he or she enters in a process where various types of information processing activities simultaneously interact (Royer & Cunningham, 1981; Samuels & Kamil, 1984; Garner, 1987). For this process to be successful, different types of skills and knowledge are assumed: general language abilities (oral language skills, extended discourse skills), lower-order text processing skills (decoding), as well as higher-order or ‘pure’ text comprehension skills (Francis et al., 2006). The aim of the SALSAs reading test is to assess the extent to which students have and apply the latter skills.

With respect to text comprehension processes a distinction can be made between the micro- and the macro-level (cf. Kintsch & Van Dijk, 1978). The former refers to the comprehension of individual propositions and their relations, the latter to the comprehension of larger parts or the text as a whole. As readers need to operate on both levels to fully understand a text, the aim of the SALSAs test was to assess both micro- and macro-level comprehension skills. Comparable to a typology suggested by Rosenshine (1980), the following set of skills was distinguished, where (i) and (ii) refer to micro-level processes, and (iii) to macro-level processes:

- i. the ability to locate or retrieve details in/from the text;
- ii. the ability to make inferences about shorter passages in the text (e.g., identifying causal relationships between propositions);

<sup>3</sup> That there was only one internet task was due to the fact that this medium was not part of the core of the international test: it was only used in the Dutch context. The decision not to include internet tasks in the Canadian and Swiss tests was a result of the observation that the computer infrastructure in Canadian and Swiss schools is, as yet, too limited.

- iii. the ability to reflect on larger passages or the text as a whole (e.g., articulating the main idea of the text or the intention of the writer).

A total of 65 items (including both open and closed question formats) were constructed: 25 items aimed to measure the first type of skill, 21 aimed to measure the second type of skill, and 19 aimed to measure the third type.

#### 4. Design of the validity study

Central to the process of (test) validation is the notion of construct validity. According to Messick, construct validity “is based on an integration of any evidence that bears on the interpretation or meaning of the test scores” (1989: 7). Although ‘any evidence’ can contribute to the understanding of test scores, this contribution becomes stronger if the evidence supports theoretically founded hypotheses with respect to the latent ability assumed to be measured (i.e., the construct). Messick argues that there are two major threats to construct validity. The first is ‘construct underrepresentation’, the situation that the test is too narrow and fails to include important elements of the construct. The second is ‘construct-irrelevant variance’, the situation that the test is disproportionately hard or easy for some students due to variables irrelevant to the measurement of the construct. Messick presents several methods of establishing construct validity, two of which will be discussed here. The first involves the test’s ‘internal structure’: this method deals with the question whether theoretically expected patterns of relationships among item scores can be empirically confirmed, and relies on an analysis of the tests’ internal consistency and factor structure. The second involves the test’s ‘external structure’: this method refers to the empirical investigation of theoretically assumed relationships between test scores and other measures, e.g., the relationship between the test and other tests that aim to measure the same underlying ability (also known as concurrent validity), or measures of skills proven to be related to the ability.

In the context of this paper three aspects of construct validity were analyzed:

- i. *construct-(ir)relevance*. For this purpose, it was investigated whether or not the difficulty level of the test was primarily determined by the difficulty level of the texts to be read (as should be the case in a test of text comprehension);<sup>4</sup>
- ii. *internal structure*. The aim of the test was to measure one overall ability, namely text comprehension, which is assumed to comprise three separate, but interrelated (sub)constructs, namely ‘retrieval of details’, ‘inferences about shorter passages’, and ‘reflections about larger passages or the text as a whole’. It was expected, therefore, that the internal consistency of the test is high, and that the results of factor analysis support the assumed structure of constructs;
- iii. *external structure*. For the process of text comprehension to be successful, the presence of certain precursor skills are assumed (see also section 3). On the basis of previous research into determinants of adolescent literacy development (cf. Van Gelderen et al., 2004), the relationships between the reading test scores and scores on four measures of such precursors skills were investigated: vocabulary, grammar knowledge, metacognitive knowledge (i.e., knowledge about

---

<sup>4</sup> There might also be unwanted difficulties, such as difficulties in the phrasing of test items.

text characteristics, and reading and writing strategies), and working memory.

The test was administered to 200 students in the period April-June 2008. The sample consisted of the 63 students participating in the main study (see section 2) and their classmates. On the basis of the test results the three aspects of construct validity described above were examined.

In order to investigate whether the difficulty level of the test corresponds with the difficulty level of the selected texts, two types of data were compared: the so-called 'readability scores' of each text, and the difficulty level of the accompanying task, as indicated by actual student scores. If both types of data show the same pattern (i.e., provide the same rank order of tasks), this provides support for the 'test difficulty = text difficulty' assumption.

Readability scores provide an estimate of the difficulty readers have in processing a text. One widely used readability index is the Flesch Reading Ease (Flesch, 1948), which is based on the formula  $206.835 - (1.015 * msl) - (84.6 * mwl)$ , where *msl* is mean sentence length (the number of words divided by the number of sentences), *mwl* is mean word length (the number of syllables divided by the number of words), and the constants are language-specific coefficients (in this case based on English). The result is a score between 0 and 100. Douma (1960; in Wagenaar et al., 1987) adapted the index for the Dutch language, resulting in the formula  $206.84 - (0.93 * msl) - (77 * mwl)$ . Flesch and Douma provide the same rules of thumb for deciding on the difficulty level of a text: 0-30=very difficult, 30-50=difficult, 50-60=fairly difficult, 60-70=standard, 70-80=fairly easy, 80-90=easy, and 90-100=very easy. Because the SALSA test is in Dutch, Douma's formula was used. Syllables, words and sentences were counted following the guidelines given by Flesch (1948). It has to be added that readability indexes such as the Flesch/Douma are not all encompassing. For one, they do not take into account other important aspects of text difficulty, such as degree of discourse cohesion, the amount of inferencing required, or the complexity of ideas (cf. Bruce et al., 1980).

In order to investigate the test's internal structure, the reliability coefficient was computed and confirmatory factor analysis was conducted, using Structural Equation Modeling (cf. Dunn et al., 1993). Factor analysis was conducted in two steps. First of all, a one-factor model was tested, or, more specifically, a three-factor model (with each factor representing one of the subskills on which the construction of test items was based) in which perfect correlations were assumed between the factors. In a second step, a three-factor model was tested in which the correlations between the factors were entered as free parameters, based on the assumption that text comprehension involves three distinct, but interrelated subskills. Subsequently, difference scores were computed to see whether the three-factor model lead to a significantly better fit than the one-factor model. Because continuous variables are preferred in factor analysis (Comrey, 1978; Kim & Mueller, 1978), sum scores were computed for subsets of items. These subsets were based on the distinction between the three comprehension skills: two sets of items per skill were selected randomly. On the basis of these item sets, six sum scores were computed (two per skill), and these sum scores were entered as observed variables in the model. Analyses were conducted in EQS.

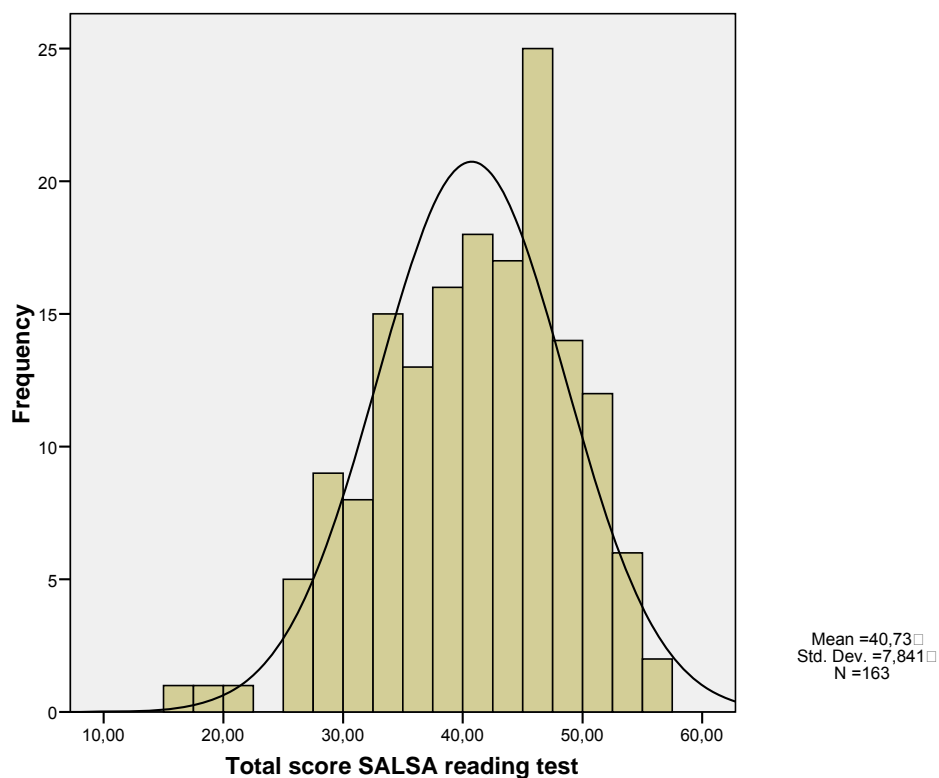
In order to investigate the test's external structure, correlations were computed between the reading test scores and the scores on a vocabulary test, a grammar knowledge test, a metacognitive knowledge test, and a working memory test. These

tests had been developed in a previous study (Van Gelderen et al., 2004; 2007) and were adapted for the current project (see also Trapman et al., this volume).

## 5. Results

Before analyzing the three aspects of construct validity, the distribution of scores was examined. Figure 1 shows that the scores more or less follow the curve of the normal distribution, even though there is a small peak around the score of 45. No evidence for threshold or ceiling effects was found, however: none of the students had the theoretically minimum score of 15.33 (guess chance), no one scored the theoretical maximum of 65.

Figure 1: Distribution of scores on the SALSA reading test



### *Construct (ir)relevance*

Table 2 presents the outcomes of the Flesch/Douma readability analyses of the different texts in the reading test (column 2), the rank order of the tasks based on these readability analyses (column 3), the mean score per task as a percentage of its theoretical maximum (column 5), and the rank order of the tasks based on these percentages (column 4). In both rank orders 1 indicates the easiest, and 9 the most difficult task.

Table 2: Difficulty level of the tasks according to the results of readability analyses and actual student scores (mean of a task as a percentage of its theoretical maximum)

Text	Flesch/Douma Reading Ease	Rank order based on Reading Ease	Rank order based on mean/maximum*100%	Mean/maximum*100%
NAR2	89.7 (easy)	1	1	79%
NAR1	83.8 (easy)	2	3	65%
EXP2	70.8 (fairly easy)	3	5	60%
INS2	67.9 (standard)	4	9	51%
EXP1	62.2 (standard)	5	5	60%
ARG1	57.2 (fairly difficult)	6	4	64%
ARG2	56.3 (fairly difficult)	7	7	57%
EXP3	49.7 (difficult)	8	8	52%
INS1	48.3 (difficult)	9	2	68%

A comparison of the rank orders on the basis of the readability analyses and the ‘means as a percentage of the maximum’ scores indicates that there is, overall, a relatively high correspondence between the two: four texts have exactly the same rank in both rank orders and three texts differ two ranks at the most. There are two exceptions, however. According to the Flesch/Douma Reading Ease, INS1 can be seen as a difficult text, while the means indicate that this task is among the easiest. A hypothetical explanation for this observation is that the difficulty level of the test is compensated for by its format. The text, which consists of a set of house rules in a youth hostel, can be characterized as a discontinuous text, consisting of relatively short passages, each dealing with one specific rule. It can be argued that this makes locating relevant information more straightforward, particularly because some of the test items dealt with specific rules.

Additionally, according to the reading ease index, INS2 can be characterized as a text with a standard difficulty level, while the means show that the task based on this text is the most difficult one. Part of the difficulty in this task seems to be caused by the influence of one item: this item was very difficult ( $p$ -value = .29) and its corrected item-total correlation was negative (-.03). Deletion of this item lead to a somewhat higher means as a percentage of the maximum score (56%); nevertheless, this task remains one of the more difficult ones. A possible explanation is that the difficulty level of the text was, once again, influenced by its format, although now in a negative way. INS2 is a website, which consists of eight subpages. The links to these subpages are given in a central menu; on the basis of the names of these links, students have to make inferences about the contents of these subpages in order to decide where to find the right answer. It may be that this activity increased the difficulty level of the task.

### *Internal structure*

Reliability analysis yielded a Cronbach’s alpha of .80, which can be interpreted as an indication of a high internal consistency (De Heus et al., 1995). There were four items that contributed negatively to the reliability score. Reliability analysis without those items resulted in a slightly higher alpha (.82).

Outcomes of a confirmatory factor analysis of the one-factor model showed that this model fits the data ( $\chi^2=10.025$ ,  $df=9$ ,  $p=.348$ ; all fit indices are above .95; RMSEA=.027, with a 90% confidence interval of .000-.094). Analysis of the three-factor model resulted in a better fit ( $\chi^2=6.147$ ,  $df=6$ ,  $p=.407$ ; all fit indices are above .95; RMSEA=.012, with a 90% confidence interval of .000-.103). Difference scores, however, showed that this improvement was not statistically significant ( $\chi^2=3.878$ ,  $df=3$ ,  $p>.05$ ).



### *External structure*

In Table 3 the correlations are presented between the reading test scores and the scores on the tests measuring vocabulary, grammar knowledge, metacognitive knowledge, and working memory.

*Table 3: Correlations between the scores on the SALSAs reading test, and the scores on measures for vocabulary, grammar knowledge, metacognitive knowledge, and working memory*

	Vocabulary	Grammar knowledge	Metacognitive knowledge	Working memory
SALSAs reading test	.59*	.66*	.41*	.43*

\*  $p < .001$

The table shows that, as expected, all measures – and particularly grammar knowledge and vocabulary – have substantial correlations with the reading test.

## **6. Discussion**

The aim of this study was to analyze the construct validity of a newly devised test for measuring the text comprehension skills of at-risk adolescents. On the basis of the framework provided by Messick (1989) three aspects of construct validity were investigated: construct (ir)relevance, the test's internal structure, and the test's external structure. Analyses of these three aspects largely support the construct validity of the test.

With respect to the first aspect, a comparison was made between the readability scores of the selected texts and the difficulty level of the accompanying tasks as indicated by students' mean task scores. The results showed that both types of data resulted in more or less the same rank order of tasks, which supports the assumption that the difficulty level of the test was primarily governed by the difficulty level of the texts. There were two exceptions: one task proved to be more difficult than was expected on the basis of readability estimates, while another task was easier. In both cases, these differences can be explained by the format of the texts: for one text, its discontinuous nature seemed to have increased the difficulty level of the task, while for the other text its discontinuous nature seemed to have made the task less difficult. It can be argued that, since discontinuity is a fundamental aspect of both texts, these differences do not indicate construct-irrelevant variance, but are a legitimate outcome of the comprehension process.

With respect to the test's internal structure, a sufficiently high reliability coefficient was found. The results of confirmatory factor analysis supported a one-factor solution; testing an additional three-factor model did not improve the fit, however. It can be concluded, then, that the assumption that the test measures one underlying ability can be confirmed, but that the data do not support the presence of the three theoretically defined subskills. This seems to be in line with previous research by Meijer & Van Gelderen (2001). They analyzed data from the PISA test of reading literacy, which is partly based on a distinction in skills similar to the one in the SALSAs test. They too did not find empirical evidence for this distinction.

With respect to the test's external structure, substantial relationships were established between the scores on the reading test and measures of skills known to be related to text comprehension (i.e., vocabulary, grammar knowledge, metacognitive knowledge, and working memory).

The analyses presented here were part of the first stage of a validation study of the SALSA reading test. In a next step, other aspects of validity will be examined. Two partial studies are currently being set up. In the first study, the reading test will be administered to a sample of students in addition to another reading test developed for this age group. The correlations between the scores on both tests will be examined in order to analyse the test's concurrent validity. Additionally, this study will investigate whether the test is able to differentiate between groups of students known to differ in text comprehension skills (e.g., native versus nonnative students, high SES versus low SES students). In the second study the content validity will be judged by a sample of 'field experts': in semi-structured interviews, teachers will be asked how they rate the relevance and difficulty level of the texts and tasks.

## References

- Adams, R. & M. Wu (2000), *PISA 2000 Technical Report*, Paris: OECD.
- Bronfenbrenner, U. & P.A. Morris (1998), The ecology of developmental processes, in: R.M. Lerner, (red.), *Handbook of Child Psychology (5th edition, volume 1)*: 993-1028, New York: Wiley.
- Bruce, B., A. Rubin & K. Starr (1981), *Why readability formulas fail: Reading Education Report No. 28*. Champaign, IL: University of Illinois at Urbana-Champaign.
- Curtis, M.E. (2002), *Adolescent Reading: A synthesis of research*, Online available: <http://www.nifl.gov/partnershipforreading/adolescent/default.html>
- Dagevos, J., M. Gijsbers & C. van Praag (red.) (2003), *Rapportage minderheden. Onderwijs, arbeid en sociale integratie*, Den Haag: SCP.
- De Heus, P., R. van der Leeden & B. Gazendam (1995), *Toegepaste data-analyse. Technieken voor niet-experimenteel onderzoek in de sociale wetenschappen*, Maarssen: Elsevier Gezondheidszorg.
- Dunn, G., B. Everitt & A. Pickles (1993), *Modelling covariances and latent variables using EQS*, London: Chapman & Hall.
- Flesch, R. (1948), A new readability yardstick, *Journal of Applied Psychology* 32 (3): 221-233.
- Francis, D.J., C.E. Snow, D. August, C.D. Carlson, J. Miller & A. Iglesias, (2006), Measures of reading comprehension. A latent variable analysis of the diagnostic assessment of reading comprehension, *Scientific Studies of Reading* 10 (3): 301-322.
- Garner, R. (1987), *Metacognition and reading comprehension*, Norwood, NJ: Ablex.
- Harris, J.L., W.A. Rogers & C.D. Qualls (1998), Written language comprehension in younger and older adults, *Journal of Speech, Language & Hearing Research* 41 (3): 603-617.
- Kim, J.O. & C.W. Mueller (1978), *Factor analysis. Statistical methods and practical issues*, Beverly Hills: Sage.
- Kintsch, W. (1998), *Comprehension. A paradigm for cognition*, Cambridge: Cambridge University Press.
- Kintsch, W. & T.A. van Dijk (1978), Toward a model of text comprehension and production, *Psychological Review* 85 (5): 363-394.
- Land, J., T. Sanders & H. van den Bergh (2008), Effectieve tekststructuur voor het vmbo. Een corpus-analytisch en experimenteel onderzoek naar tekstbegrip en tekstwaardering van vmbo-leerlingen voor studieteksten, *Pedagogische Studiën* 85 (2): 76-94.
- Lin, L.M. (2000), Metacomprehension knowledge and comprehension of expository and narrative texts among younger and older adults, *Educational Gerontology*, 26: 737-749.
- Messick, S. (1989), Meaning and values in test validation. The science and ethics of assessment, *Educational Researcher* 18 (2): 5-11.
- Meijer, J. & A. van Gelderen (2002), *Lezen voor het leven. Een empirische vergelijking van een nationale en een internationale leesvaardigheidsspeiling*, Amsterdam: SCO-Kohnstamm Instituut.
- Mosenthal, P.B. & I.S. Kirsch (1998), A new measure for assessing document complexity. The PMOSE/IKIRSCH document readability formula, *Journal of Adolescent & Adult Literacy* 41 (8): 638-657.
- OECD (2001), *Knowledge and Skills for Life. First results from PISA 2000*, Paris: OECD.
- OECD (2003), *The PISA 2003 Assessment Framework. Mathematics, Reading, Science and Problem Solving Knowledge and Skills*, Paris: OECD.

- Rosenshine, B.V. (1980), Skill hierarchies in reading comprehension, in: R.J. Spiro, B.C. Bruce & W.F. Brewer (red.), *Theoretical issues in reading comprehension. Perspectives from cognitive psychology, linguistics, artificial intelligence, and education*: 535-554, Hillsdale, NJ: Lawrence Erlbaum.
- Royer, J.M. & D.J. Cunningham (1981), On the theory and measurement of reading comprehension, *Contemporary Educational Psychology* 6: 187-216.
- Samuels, S.J. & M.L. Kamil (1984), Models of the reading process, in P.D. Pearson (red.), *Handbook of reading research*: 185-224, New York: Longman.
- Trapman, M., A. van Gelderen, J. Hulstijn & R. van Steensel (this volume), De relatie tussen talige en cognitieve vaardigheden en de leesvaardigheid van risicoleerlingen in het vmbo.
- Van Gelderen, A., R. Schoonen, K. de Glopper, J. Hulstijn, A. Simis, P. Snellings & M. Stevenson (2004), Linguistic knowledge, processing speed, and metacognitive knowledge in first- and second-language reading comprehension: a componential analysis, *Journal of Educational Psychology* 96 (1): 19-30.
- Van Gelderen, A., R. Schoonen, R. Stoel, K. de Glopper & J. Hulstijn (2007), Development of adolescent reading comprehension in L1 and L2: a longitudinal analysis of constituent components., *Journal of Educational Psychology*, 99 (3): 477-491.
- Wagenaar, W.A., R. Schreuder & G.J. Wijnhuizen (1987), Readability of instructional text, written for the general public, *Applied Cognitive Psychology* 1: 155-167.