

Written narrations by 8- to 10-year-old Turkish pupils in Flemish primary education: A follow-up of seven text features

Lieve Verheyden and Kris Van den Branden

Katholieke Universiteit Leuven, Begium

Gert Rijlaarsdam

University of Amsterdam, the Netherlands

Huib Van den Bergh

Utrecht University, the Netherlands

Sven De Maeyer

University of Antwerp, Belgium

This semi-longitudinal study examined the development of narrative writing quality of young Turkish second language learners in mainstream Dutch-only education, and the impact of student-level and classroom-level predictors of narrative writing quality, using hierarchical linear modelling. Writing samples of 106 third graders and 111 fourth graders of seven Flemish primary schools were collected at the beginning and at the end of the school year. Measures included one holistic primary trait judgement, and six objective indices of text quality. Student-level predictors included age, SES and home language, while the classroom-level predictor focused on the home language pattern of the classroom. There was a significant mean growth for each index in each grade, but effect sizes differed from quite large for content and word level indices over moderate for sentence level indices to small for the text level index. Home language (Turkish) had a significant negative effect on all but one variables, particularly in Grade 4, while the negative effect of low SES was much more limited. A supplementary negative effect was found for homogeneity of classroom population. Implications of the study highlight the importance of student and classroom characteristics in writing achievement as well as the need to consider the poor performance of Turkish children.

During the past 20 years, the Flemish government has issued several policy measures and raised extensive funding to further the educational success of learners-at-risk in primary

education. Measures such as the current 'Equity in Education Policy' (EEP) are primarily targeted at children who are raised in another language than Dutch (the medium of instruction in Flemish education), and children of low socioeconomic status. Although the measures have raised teachers' awareness of the 'EEP-target children's' educational needs, recent national and international studies show that the literacy development of the Flemish student population is still determined by the pupils' socioeconomic background, their ethnic origin and/or home language background. For reading skills, there is ample research that documents this observation (Janssen et al., 2008; Pirls Vlaanderen, 2006); for writing skills, however, empirical evidence is lacking. At present, there is hardly any evidence available that enables researchers and practitioners to establish the quality of EEP-target pupils' written texts. As a result, curriculum developers, syllabus developers and teachers lack a firm empirical basis for the design of a writing education approach that is suited to the target pupils' learning needs. This study aims to shed light on this matter, as a first step in further optimising the educational measures that can be taken by the Flemish government to promote EEP-target children's literacy development, and ultimately their chances of school success and success in society in general.

Young writers-at-risk

From the age of 8 onwards, children in Flemish primary education, who have only just begun to read and write, are expected to produce functional, and increasingly extensive, messages for communicative purposes: words need to be combined in sentences, which in turn need to shape texts that meet growing demands of communicative effectiveness and content- and form-related text features. Although most 8- to 10-year-old children are ready to take up this challenge (Kress, 1994), meeting the combination of demands remains quite difficult at that young age. Several grounds for explanation may be at play here.

First, writing research shows that young writers pay primary attention to the exceedingly demanding transcription processes of individual words (orthography, handwriting) (Berninger et al., 1992; McCutchen, 2000; Medwell & Wray, 2008). As a result, little mental space remains for higher-order processes that are crucial for fluently constructing sentences and putting together coherent text. Secondly, young children's approach to writing assignments, described by Bereiter and Scardamalia (1987) as 'knowledge telling', based on on-line and local planning processes (Berninger & Swanson, 1994), does not yet reflect 'content' (knowledge) that has been transformed in the process of creatively revising and reshaping text in line with their own writing goals. Thirdly, 8- to 10-year-old children often lack the full competence to 'imagine' the absent reader, taking into account his/her potential needs: as a result, these children write in a 'writer-based' rather than a 'reader-based' way (Flower, 1979).

For young second language learners the task becomes even more complicated. For one, a child's oral language skills and reading skills in a particular language are known to have a direct impact on the written output that is produced in the same language (Reese, Garnier, Gallimore & Goldberg, 2000). For pupils of Flemish primary education who were raised in another language than Dutch, and whose Dutch language proficiency has been shown to be significantly lower than their Dutch-speaking peers (Ramaut, Roppe, Verhelst & Heymans, 2007), producing Dutch written output may turn out to be a daunting challenge. Alamargot and Chanquoy's (2001) model of *translating* – the second

subprocess in Hayes and Flower's (1980) writing process model – may be helpful to understand this challenge. The authors distinguish four steps in the translating phase:

1. Elaboration: digging up content and reshaping it;
2. Linearisation: transforming the hierarchically arranged content into a linear preverbal semantic-syntactic structure (subject–predicate);
3. Formulation: matching the semantic-syntactic structure with grammatical and lexical structures;
4. Execution: putting the sentence on paper.

The transition from phase 2 to 3 is highly complicated for children with limited target language proficiency: they might lack the grammatical and lexical skills to turn a preverbal message into an adequate verbal construction. This tension is likely to come to the surface of these young children's texts in different ways. For instance, they may avoid treating certain aspects of the message because they cannot find the right words. Or, looking for alternative ways to get parts of their message across, they may lose the general overview of what they have written so far, and as a result, produce incoherent text (Weigle, 2002). The attention they have to devote to trying to express complex ideas may be at the cost of accuracy (or vice versa), in line with Skehan's (2007) limited attention hypothesis. Further, the language they ultimately put on paper will reflect their current state of interlanguage (Ellis, 1985), including sentences that do not yet meet the standards of native speakers. Recent studies of young second language learners confirmed that they produced texts of a significantly lower quality than mother tongue speakers (Howard, Christian & Genesee, 2004).

In addition, a great proportion of the non-Dutch-speaking pupils in Flemish education belong to socioeconomically disadvantaged families. Numerous studies illustrate the strong impact of family support of early literacy development on the development of reading and writing skills (Ravid & Tolchinsky, 2002), irrespective of the language in which this support is provided (Reese et al., 2000). Wells (1985) shows strong correlations between the social status of families and the available sociocultural capital, including the importance attached to literacy. In line with this research, international research of writing skills shows that the writing skills of children of lowly educated parents significantly lag behind (Moats, Foorman & Taylor, 2006).

Finally, a high proportion of socioeconomically disadvantaged non-native speakers of Dutch are enrolled in schools in which Dutch-speaking peers do not constitute the majority of the student population because of the uneven division of the EEP-target children across Flemish schools (Desmedt & Nicaise, 2006). Various studies (Braun, Forges & Wolmainck, 1997; Krom, Verhelst & Veldhuijzen, 2004) have corroborated the impact of school/class population on the quality of children's written output, because of a concentration of low-SES pupils and/or second language learners. In Wallonia non-indigenous (low-SES) pupils perform significantly better for most text features in classes with a majority of indigenous pupils than in classes with a majority of non-indigenous pupils. With these results in mind, Braun et al. wonder whether 'l'égalité des chances ne passeraient-elles pas par une égalité de l'hétérogénéité des classes sur l'ensemble du territoire?' [equal opportunities, don't they go via an equality of (heterogeneous) class composition over the territory?] (p. 113).

In Flemish education today, many pupils of Turkish, Moroccan and other ethnic origins face the above-mentioned challenges. At present, it remains unclear whether these target pupils should be treated as a homogeneous group when it comes to their literacy development. Verhoeven (1997) and Smits and Aarnoutse (1997) have shown significant

differences between the performances and development of pupils of different ethnic origins, and wonder whether the pupils' cultural orientation, their language or a combination of both may explain these differences. In Flanders, Verhaeghe, Knipprath and Mertens (2007) reveal that even when controlling for *socioeconomic status*, the Turkish pupils have a 'relatively big deficit' for 'reading comprehension' as compared with a 'relatively small deficit' for other nonindigenous pupils. On the other hand, Braun et al. (1997) do not reveal systematic differences between Turkish and Moroccan pupils' performances.

Parameters for describing text quality

In this study we aim to describe differences in the quality of texts written by young children on the basis of a well-balanced selection of text features. However, describing the text quality of children's written output (and of its growth in quality) is a complex endeavour: any text is a unique and complicated fabric of various *knots* at different linguistic levels. In order to assess text quality in a valid and reliable way, then, decisions need to be taken as to which knots should be focused upon, and how each of these knots can be described in the most adequate way.

Different studies have made different, and often unique, selections of text features (e.g. Colpin, Heymans & Rymenans, 2005; McMaster & Espin, 2007; National Writing Project, 2008). Nevertheless, though made operational in various ways, a number of broad aspects tend to recur across these studies:

- communicative effectiveness;
- content;
- organisation;
- language use;
- fluency/length.

To select text features that together may describe the quality of the texts collected in our study, we started from the premises that (1) a text is a multifaceted product, different aspects of which may be influenced in unique ways by various background variables such as the pupil's home language and socioeconomic status; and (2) a combination of qualitative and quantitative measurements may yield the most reliable and valid assessments (Tindal & Parker, 1991). Taking into account the specific nature of the written story accounts we worked with, we decided to work with '6+1 traits of writing' (National Writing Project, 2008). To 'Communicative Effectiveness' (Com.Eff.), a qualitative holistic measurement, we added six objective quantitative features which we could use to validate the more subjective Com.Eff.-construct and to compensate for the potentially low inter-rater-reliability of its holistic measurement (Rijlaarsdam & Wesdorp, 1988; Van den Bergh, De Glopper & Schoonen, 1988). The six measures were: 'Content: correctness and completeness' (Content), 'Lexical richness' (Lex.Rich.), 'Complexity of T-unit' (Compl.TU), 'Accuracy of T-unit' (Accu.TU), 'Accuracy of spelling' (Spelling) and 'Referential cohesion' (Ref.Coh.).

As can be inferred, the six features that were selected can be related to different linguistic levels (Nelson & Van Meter, 2007), that is the word level, the sentence level and the overall text level. Methodologically, the selection of these six features involves a combination of production-independent measures (e.g. Compl.TU), probability measures (e.g. Accu.TU) and production-dependent measures (Lex.Rich.), allowing us to put the importance of the absolute length of the text in perspective (Jewell & Malecki, 2005).

Research questions

The main research questions of this study were, first of all, informed by the current lack of empirical evidence with regard to the writing development of young children who belong to families of low socioeconomic status and who were raised in another language than the main medium of instruction. In Flanders, these children are designated the status of target students of the EEP of the Flemish government (EEP-target children). The main research questions of this study, then, can be formulated as follows:

1. To what degree does the average performance of seven text features in a written narration task by 8- to 10-year-old EEP-target children improve from the beginning to the end of the school year?
2. To what extent can the EEP-children's average performance (described in research question 1) be predicted by the EEP-pupils' *home language* and/or *socioeconomic status*, or by *class composition in terms of home language use*?

Method

A semi-longitudinal study was carried out, involving two measurements in the third and fourth grade of seven primary schools.

Participants

Data collection took place at the beginning (September) and at the end (June) of the school year 2006–2007. We collected written output in the third and fourth grades of seven primary schools in which more than 60% EEP-students were enrolled (Van Gorp & Verheyden, 2006). The schools displayed a wide diversity in terms of the ethnic origin of their pupil population: in some schools/classes, more than 80% of the pupils were of Turkish origin (home language Turkish, sometimes in combination with Dutch); in this case we will speak of 'monolingual' classes. Other schools/classes had a 'multilingual' population: these classes included more than five different home languages (sometimes in combination with Dutch), none of which being dominant. A third group included the 'bi- or trilingual' schools/classes, that included two or maximally three dominant subgroups: a group of Turkish origin, a group of Flemish origin and a group of Moroccan-Arabic origin.

In four schools we collected texts from all pupils of the third and fourth grades; for reasons of practicability, we collected texts from only 10 pupils in the three remaining schools. These 10 pupils were selected on the basis of their scores on standardised Dutch reading comprehension tests and of their performance on the pre-test of this study: 6 of the 10 pupils had a relatively low score (within their class group) for reading and writing; the other 4 had a relatively high score. Table 1 provides an overview of the number of pupils per grade in each school.

On the basis of a written survey administered from the pupils and their parents (Reynders, Van Heddegem, Nicaise & Van Damme, 2004) the following background information about the pupils was collected:

1. *Home language*: 'Dutch' (D) was attributed as a feature to pupils who only speak Dutch at home; 'Turkish (+Dutch)' (Tu[D]) to pupils who speak Turkish at home with at least one key interlocutor (mother, father, brothers/sisters); 'Other (+Dutch)'

Table 1. School characteristics: linguistic diversity, number of pupils (numbers).

School	Linguistic diversity (based on home language)	Grade 3	Grade 4	Total
School A	Monolingual Turkish (MONOL)	17	27	44
School B	Monolingual Turkish (MONOL)	10	10	20
School C	Trilingual: Turkish, 'Moroccan' ^a , Dutch ^b (BITRIL)	17	14	31
School D	Trilingual: Dutch, Turkish, 'Moroccan' (BITRIL)	24	17	41
School E	Bilingual: 'Moroccan', Turkish (BITRIL)	10	10	20
School F	Multilingual	18	23	41
School G	Multilingual	10	10	20
Total		106	111	217

^a'Moroccan' stands for Moroccan Arabic or for a Berber language.

^bThe order in which the different instances of ethnic origin are mentioned denotes their relative proportions in the class: from high proportion to lower proportions.

Table 2. Demographic data per grade (percentages).

Variable	Grade 3	Grade 4
Socioeconomic status		
High	29	28
Low	66	69
Missing	05	03
Home language		
Dutch	20	18
Turkish (+Dutch)	54	57
Other (+Dutch)	26	25

(Oth[D]) to pupils who speak another language than Dutch or Turkish with at least one key interlocutor;

2. *Socioeconomic status of the family (SES)*: this attribute is based on the highest diploma obtained by the mother. Following the regulations of the Flemish EEP, 'low' was attributed to families with a mother who has not obtained a diploma higher than lower secondary education; 'high' was attributed to families with a mother who has obtained a diploma in higher secondary or higher education.

Table 2 provides an overview of the results of this survey.

Combining both background variables revealed that most of the pupils who spoke another language than Dutch belonged to a family of low socioeconomic status (76%). Among the children who spoke Dutch as their mother tongue, high and low SES were more evenly distributed (55% vs 45%).

Writing assignment: written narrations

We focused our study on narrative texts, a genre that is familiar to 8- to-10-year-old children (Heesters, 2000). We invited the children to retell a given story, presented in a nonverbal comic mode (Eaton, Collis & Lewis, 1999). By presenting the children the contents of a story, the potential impact of children's foreknowledge and imaginative skills on the quality of the written output is neutralised (Van den Bergh et al., 1988). In addition, chances are that this input helps the children to focus their attention on the 'formulation' and 'execution' phases (Alamargot & Chanquoy, 2001). Finally, since the

storyline that needs to be produced is fixed in advance, the children's working memory is not taxed excessively.

In the pre-test and post-test, the children were invited to retell (in writing) the story represented in the comic strip to a reader who had not seen the comic or been informed in any other way about the storyline. We used the comic strip 'Ice cream' (Appendix A), which was pre-tested with children of the same age. In order to increase comparability of text quality (eg. Lex.Rich., Ref.Coh., Content) the same comic was selected for both test administrations (9 months apart).

The test administration was standardised and all pupils finished their assignment within the previewed span of the writing class (50' *minus* instructions).

Describing text quality

The seven text features were made operational in the following way.

Communicative effectiveness (Com.Eff.) refers to the quality of the message at content level. It was holistically scored on a 7-item-scale, using anchor texts (Rijlaarsdam & Wesdorp, 1988). For reasons of practicability, only 244 of the 434 texts were given a score for 'communicative effectiveness', that is the pre- and post-tests of 7 to 10 pupils in each class. For schools A, C, D and F we adopted the same selection procedure as described in 'Participants'.

Content: correctness and completeness (Content) refers to the extent to which the full contents of the story are correctly described to the reader. This variable was made operational by using a template containing 12 propositions, with a maximum score of 1 to 3 (Cragg & Nation, 2007). The internal consistency of the items was moderate (Cronbach's $\alpha = .61$ for 157 pre- and post-tests).

Lexical richness (Lex.Rich.) refers to the number of different lemmas in the text (Vermeer, 2000). Lemma was defined as the headword in a dictionary (Woordenlijst Nederlandse Taal, 2005). Since we did not correct for text length, the correlations between 'lexical richness' and 'number of words' were very high across the four test administrations ($r = .86$; $p < .001$).

For young writers, we can still assume that longer and more complex sentences reflect a growth of proficiency (Hudson, 2009; Lanauze & Snow, 1989; Saddler & Graham, 2007). Following Hunt (1965) we operationalised *Complexity of T-unit (Compl.TU)* using the average number of words per T-unit.

As far as *Accuracy of the T-unit (Accu.TU)* is concerned (Foster & Wigglesworth, 2007; Wolfe-Quintero, Inagaki & Kim, 1998), we restricted ourselves to syntactic errors such as redundant words or incorrect word order, and we chose to introduce *probability* since the length of the texts showed considerable variance. Therefore, this variable was calculated as the probability that the next syntactic unit (T-unit) would be correct.

We assessed *Accuracy of spelling (Spelling)* in much the same way as 'Accuracy of the T-unit'. We ticked off which of the first 10 content words (nouns and adjectives, adverbs and infinitives) and which of the first 10 conjugated verbs were written in a way that did not deviate from the official spelling norm. We calculated the probability of accurate production.

Referential cohesion (Ref.Coh.) can add to text quality by interweaving separate units (Wigglesworth, 1997). In this study we calculated the probability of referential cohesion, that is we evaluated (1) whether the references to the two main characters in the story were

unambiguous, and (2) whether the references were textually adequate. Only for the cases in which both conditions were fulfilled was a positive score assigned.

Analysis

Scoring procedure. A trained graduate assistant transcribed the stories verbatim. All transcriptions were checked and corrected by the first author. The holistic primary trait evaluation was realised by juries of three experts (Schoonen, 1991), randomly selected from a final-year teacher education student group (K.H.Leuven, academic year 2007–2008). The members of the juries were asked to evaluate each of a random set of texts holistically on a 7-point Likert scale, focusing on one trait, viz. ‘communicative effectiveness’. The raters were helped by three anchor texts that illustrated scores 2, 4 and 6 (Rijlaarsdam & Wesdorp, 1988). The inter-jury reliability is good: $\rho = .86$. The scoring of the objective variables was described in a scoring guide. Inter-rater reliability checks were made until scoring agreement was high: Pearson’s product-moment correlation coefficients ranged between .88 (Ref.Coh.) and .98 (Lex.Rich.; Spelling). For the six objective variables we examined whether the results of the 244 texts that also received a primary trait score for Com.Eff., differed from the results of the other texts ($N = 190$). For none of the variables was an effect for the primary trait subset found.

Method of analysis. Considering the research questions and the multi-stage sampling design, we opted for an exploration of the data by means of multivariate multi-level models for two times two occasions. Multi-level analyses account for the nested structure of students within classrooms. In this study we worked with a three-stage sample: level 1: ‘text’; level 2: ‘pupil’; level 3: ‘class’. For each dependent variable a multi-level model was constructed. Estimations were simultaneously given for the fixed effects (mean score) and for the random effects (variance between scores) for two occasions (Oc) per grade (total number is 4). The simultaneous estimations made it possible to test for significant differences between the random and fixed effects, respectively (Van den Bergh & Kuhlemeier, 1997).

The multi-level analyses per variable were performed stepwise (Singer & Willett, 2003). First we looked for the ‘best fit empty model’, that is the model with the lowest possible number of parameters qualitatively fitting the data. In order to be able to compare effects between the seven variables we then standardised the results for each dependent variable applying the following formula: $(\text{score} - M_{\text{Oc1}}) / SD_{\text{Oc1}}$. Finally we introduced the explanatory variables: first *Home Language* (Dutch = reference category) was entered, then *SES* (SES^+ = reference category). They were coupled with each of the occasions, e.g. Turkish (Dutch)*Oc1 or SES^- *Oc1. In a third step we examined whether there were any significant interaction effects between *SES* and *Home Language* at one or more occasions. Finally *Class Composition* was entered (reference category = ‘multilingual’). Nonsignificant effects were removed.

Results

Seven text features in written narratives of high-risk children

Figure 1 presents the mean scores for the seven variables at four occasions. The variables show significant progress between the beginning and the end of each year, and nonsignificant differences between Oc2 and Oc3. There is a considerable difference as far

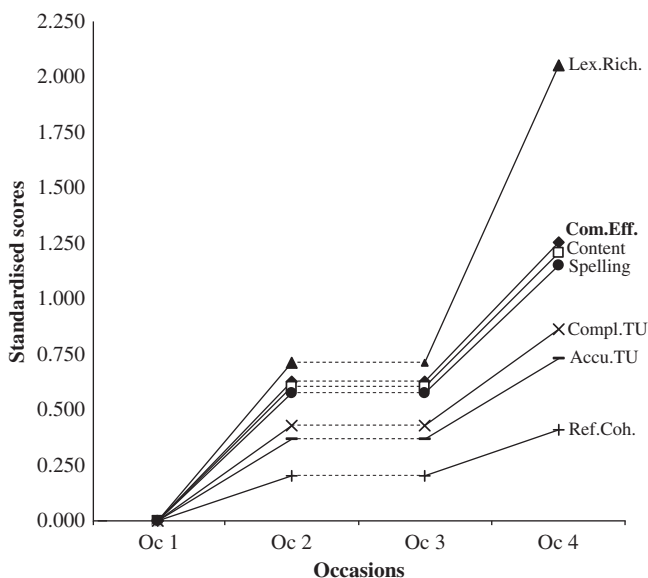


Figure 1. Estimated fixed effects^a for each occasion for Lexical Richness (Lex.Rich.), Communicative Effectiveness (Com.Eff.), Content: completeness and correctness (Content), Accuracy (Spelling), Complexity T-unit (Compl.TU), Accuracy T-unit (Accu.TU) and Referential Cohesion (Ref.Coh.) in third and fourth grade. ^aNonsignificant differences between occasions 2 and 3 are visually equalised.

as effect sizes are concerned. Four patterns can be discerned. Taking Com.Eff. as a starting point (effect size of 0.65 *SD* per year), we notice two quite similarly progressing variables, viz. Content and Spelling. Lex.Rich. belongs to the second pattern: its effect sizes are larger than those of Com.Eff., particularly in Grade 4. Because of its dependence on production, the substantial increase of ‘lexical richness’ also points to an increase of the total number of words written. Both sentence-level variables make up the third pattern: they show a smaller but still considerable effect size, viz. 0.4 *SD* per year. Finally, there is Ref.Coh.: effect sizes are small: only 0.2 *SD* per year.

The different effect sizes support our choice to consider all seven features separately in investigating the predictive power of the explanatory variables below.

Seven text features in written narratives of Flemish and Turkish target children in classrooms with different (home) language patterns

To the ‘best fit empty model’ for Com.Eff. the explanatory variables were entered. Table 3 shows the fixed effects of three different models for Com. Eff. When *Home language* is entered individually, Tu(D) significantly predicts the results at four occasions. Adding *SES* does not change the impact of Tu(D). Since the effects can be added up, SES^- Tu(D)-pupils averagely score 0.7 *SD* less than SES^+ non-Tu(D)-pupils, which means a gap of one school year. We did not observe any interaction effect of *SES* and *Home language*. As far as *Class composition* is concerned, no predictive value was noticed.

Table 4 displays the results of the six objective text features at every occasion as far as the predictive power of *SES*, *Home language* and *Class composition* are concerned. Only significant effects are reported ($p < .1$).

Table 3. HLM results for Communicative Effectiveness: comparison of the impact of the independent variables ‘home language’, ‘SES’ and ‘classroom composition’ in three models.

Fixed effect	Home language		Home language+SES		Home language+SES+classroom composition	
	Coefficient	SE	Coefficient	SE	Coefficient	SE
Oc 1	0.22	0.17	0.42	0.19	0.42	0.19
Oc 2	0.66	0.17	1.03	0.19	1.03	0.19
Oc 3	0.64	0.17	1.05	0.19	1.05	0.19
Oc 4	1.45	0.17	1.66	0.19	1.66	0.19
T(D)	-0.38*	0.15	-0.37*	0.15	-0.37*	0.15
O(D)	NS		NS		NS	
SES ⁻	NS		-0.30*	0.15	-0.30*	0.15
BITRIL					NS	
MONOL					NS	

* $p < .05$.
 Oc1-4, Occasions 1 to 4; T(D), home language Turkish (+Dutch); O(N), home language Other (+Dutch); SES⁻, low socioeconomic status; BITRIL, bi-/trilingual classrooms; MONOL, monolingal Turkish classrooms.
 NS = not significant.

Table 4. HLM results for Content: Completeness and Correctness (Content), Lexical Richness (Lex.Rich.), Complexity of T-unit (Compl.TU), Accuracy of T-unit (Accu.TU), Accuracy Spelling (Spelling) and Referential Cohesion (Ref.Coh.): impact of ‘Home Language’, ‘SES’ and ‘Class Composition’.

Fixed effect	Content	Lex.Rich.	Compl.TU	Accu.TU	Spelling	Ref.Coh.
Oc 1	0.01	0.00	0.37	0.27	0.01	0.9
Oc 2	0.67	0.77	0.69	0.74	0.56	0.57
Oc 3	0.80	1.02	1.48	0.52	0.85	1.28
Oc 4	1.46	2.07	1.80	0.99	1.41	0.95
Tu(D)*Oc 1				-0.44***		-0.75***
Tu(D)*Oc 2				-0.44***		
Tu(D)*Oc 3	-0.55***		-0.65**	-0.44***		-0.85***
Tu(D)*Oc 4	-0.55***		-0.65**	-0.44***		-0.32*
Oth(D)*Oc 1				-0.36(*)		-0.48*
Oth(D)*Oc 2				-0.36(*)		
Oth(D)*Oc 3			-0.82**	-0.36(*)		-0.59**
Oth(D)*Oc 4			-0.82**	-0.36(*)		
SES ⁻ *Oc 1						
SES ⁻ *Oc 2						
SES ⁻ *Oc 3					-0.40*	
SES ⁻ *Oc 4					-0.40*	
BITRIL			-0.44**			-0.48*
MONOL			-0.57**	-0.30(*)		-0.67*

*** $p < .001$; ** $p < .01$; * $p < .05$; (*) $p < .1$
 Oc1-4, Occasions 1-4; Tu(D), home language Turkish (+Dutch); Oth(D), home language Other (+Dutch);
 SES⁻, low socioeconomic status; BITRIL, ‘bi-/trilingual’ classrooms; MONOL, ‘monolingal’ Turkish classrooms.

No significant effect is observed for Lex.Rich.: the variance between the scores is not influenced by *Home language* or *SES*, or by *Class composition*.

Two other text features are predicted by student-level variables only: *Home language* is a significant predictor of Content: on average, fourth grade Tu(D)-pupils’ scores are more than 0.5SD lower than non-Tu(D)-fourth graders. Additionally, *SES* predicts Spelling in the fourth grade (0.4SD). There was no significant effect of *Home language* on Spelling.

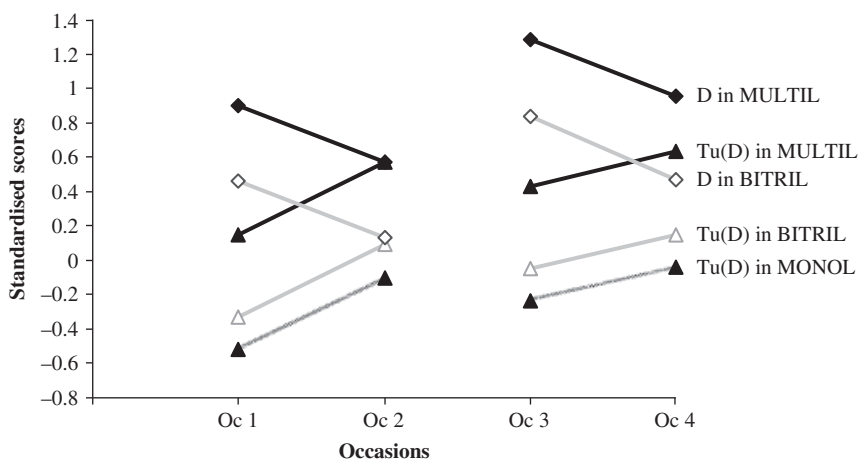


Figure 2. Estimated fixed effects for each occasion in third and fourth grade for 'Referential Cohesion' for five subgroups: Tu(D) in MULTIL: Turkish (Dutch) in 'multilingual' classrooms; Tu(D) in BITRIL, Turkish (Dutch) in 'bi-/trilingual' classrooms; Tu(D) in MONOL, Turkish (Dutch) in 'monolingual' Turkish classrooms; D in MULTIL, Dutch in 'multilingual' classrooms; D in BITRIL, Dutch in 'bi-/trilingual' classrooms.

With respect to the sentence variables (Accu.TU and Compl.TU), there are significant effects of *Home language* and of *Class composition*. Compared with the texts of pupils in 'multilingual' classes, the texts of pupils of 'mono- or bi-/trilingual' classes show significantly less accurate and less complex (only fourth grade) sentences. Moreover, as far as Compl.TU is concerned, Tu(D)- or Oth(D)-pupils in fourth grade write significantly less complex sentences than D-pupils. Since effects can be added up, the difference between Tu(D)-children in 'monolingual' classes and D-children in 'multilingual' classes is 1.22 *SD*, which indicates a performance gap of two school years and more.

As far as the last variable is concerned, Ref.Coh., we also notice an affect of *Class composition* and of *Home language*, but the latter shows a quite different pattern than the one observed with Accu.TU/Compl.TU: the negative impact of *Home language* at the beginning of the school year (Oc 1 and Oc 3) is reduced at the end of the year (Oc 2 and Oc 4). While mean scores show little progress over time, Figure 2 illustrates the amount of variance that lies hidden behind those mean scores. As is the case with Accu.TU and Compl.TU, the gap between Tu(D) scores of Ref.Coh. in different classrooms is significant.

Discussion

Predictors of text features

The results of this research into the impact of pupil background variables and a class composition variable on the scores for seven text features reveal that these features all have their own story to tell (Hamp-Lyons, 2003; Nelson & Van Meter, 2007; Weigle, 2002), only a small part of which they appear to have in common. As for the latter, for all of the features, the variance between children's texts is very large (Howard et al., 2004; Moats et al., 2006; Van de Gein, 2005); in addition, the children's performance with regard to the seven text features is significantly different at each occasion: the fourth

grade pupils' scores are systematically and significantly higher than the third grade pupils' (Howard et al., 2004). On the basis of the data we collected, we cannot determine to what extent these differences are caused by education or merely come as a result of the children's natural cognitive growth, or whether a combination of both factors is at play.

Despite these common trends, the differences in the growth patterns of the different variables are considerable. Some of this variance can be explained by the independent variables: the seven text features appear to be differentially influenced by the variables *SES*, *Home language* and *Class composition*, but a large amount of variance remains unexplained. Particularly, the differences between the large effect sizes of the content-related variables and the more limited effect sizes of the form-related variables are striking. The variable Spelling assumes an interesting, and possibly revealing, position in this respect: the effect size of Spelling is comparable to the content-related variables. We may wonder whether this can be attributed to (a) the nature of this variable; (b) young writers' strong focus on processes of 'transcription' (Medwell & Wray, 2008); or (c) an effect of education, caused by the considerable amount of time and attention teachers devote to spelling (Bonset & Hoogeveen, 2007).

In contrast with many other studies on the development of academic language skills (e.g. Janssen et al., 2008), the variable *SES* – as determined by the highest diploma of the mother – only weakly predicted the children's scores. *SES* only has a significant impact on Spelling (in the fourth grade) and – together with *Home language* – on the holistic scores for Com.Eff. in both grades (National Assessment of Educational Progress, 2002). The effect of *SES* on Spelling in the fourth grade might be explained by the negative impact of the summer holidays (Cooper, Nye, Charlton, Lindsay & Greathouse, 1996). The lack of effect of *Home language* on Spelling that we find in our study is corroborated by Smits and Aarnoutse (1997). As for the effect of *SES* on Com.Eff., we can assume that the relatively limited access to written material that children of socioeconomically weaker families have, may be at play here (Aarnoutse & Verhoeven, 1999; Wells, 1985).

Except for Spelling and Lex.Rich., *Home language* significantly and dominantly predicts the performance of the selected text features (Cameron & Besser, 2004): for some, *Home language* as a background variable of individual pupils is decisive (e.g. for Com.Eff. and Content); for other variables, this effect is complemented by the indirect effect of *Class composition* in terms of home language (e.g. Accu.TU and Ref.Coh.): *Class composition* does not overrule the same variable at individual level.

The impact of *Home language* on the more formal features, such as Accu.TU, Compl.TU and Ref.Coh. may be explained by the relatively limited level of Dutch language proficiency of many L2-L2-pupils: they may have difficulties trying to match the preverbal semantic-syntactic structure of their meaning intentions with adequate grammatical and lexical items in Dutch. However Lex.Rich. does not show this pattern: 'text length', which highly correlates with this production-dependent variable, might be impervious to the impact of *Home language* or *SES* (Verhoeven, 1992). Alternatively, the effect of *Home language* on Lex.Rich. may be intermingled with the effect of *Gender*, which was not part of this study. Olinghouse (2007), among others, has shown that girls significantly outperform boys for production-dependent variables such as 'total number of words'.

Another unexpected result of this study is the progress that non-native speakers of Dutch make for Ref.Coh. For this variable, both the Oth(D) and Tu(D)-pupils of the third and fourth grades are able to close much of the gap with their Dutch-speaking peers. We want to suggest two possible explanations: the first explanation is related to the nature of the feature itself: writers who produce more elaborate stories, including many details,

face a complex task at the level of Ref.Coh. So, for Dutch-speaking pupils, the contextually adequate reference to the main characters as one of the many details to pay attention to in their multilayered stories, may give rise to a higher error rate. The way we operationalised Ref.Coh. in this study, however, does not allow us to further investigate this hypothesis. Another possible explanation draws upon the combination of Turkish pupils' linguistic deficit with regard to the Dutch reference system (Nistov, 2001), and the ease with which incorrect references can be corrected (e.g. 'hij' [= *he*] instead of 'zij' [= *she*]). Teachers or pupils themselves may easily notice this kind of error, and correct it in a relatively straightforward way. This hypothesis is supported by the negative effect of the summer holidays that we found for Turkish pupils in particular.

Contrary to Ref.Coh., the Tu(D)-pupils, especially in 'monolingual' classes, did not succeed in closing the gap in the same way for the two variables at sentence level (Accu.TU and Compl.TU). Several explanations may be raised here, the analyses not allowing us to decide which may be dominant: (1) Since the sentence is a phenomenon typically associated with *written* language (Kress, 1994), young children may experience severe cognitive difficulties in the first phase of writing texts; (2) this may be especially the case for pupils with a relatively limited level of Dutch language proficiency, and/or (3) for pupils who are poorly familiar with written language. (4) In classes in which Dutch is not the lingua franca for colloquial communication, second language learners of Dutch may be offered limited opportunities to further develop their interlanguage at the same speed as the other peers; and finally (5) education may make a difference. For instance, we may wonder whether the EEP-target students receive enough high-quality support and guidance from their teachers in formulating sentences, e.g. through *writing aloud* (Myhill, 2008), or get enough practice opportunities to compose relatively complex sentences (Van de Gein, 1991). Again, further research is needed to thoroughly investigate the explanatory power of these hypotheses.

Finally, we must comment on the differences between the third and the fourth grades: in the fourth grade there are more variables for which the gap between the Dutch-speaking and non-Dutch-speaking becomes significant than in the third grade. This primarily applies to the Turkish pupils, in line with the results of other studies (Pirls Vlaanderen, 2006). At the moment, there is no consensus on what exactly explains the gap between the Turkish and the Dutch-speaking pupils. A number of arguments, however, tend to be raised by various authors. Together, they boil down to a pattern referred to with the terms 'economic exclusion' and 'cultural marginality', which are attributed to the Turkish migrant communities across Europe (Aarts, 1994; Akinci, Jisa & Kern, 2001).

At second glance, the growing gap with the Dutch-speaking children of the fourth grade does not only apply to the Turkish children. For an increasing number of variables, the texts of Oth(D) and low-SES pupils display a similar pattern. We suspect that this indicates that, when it comes to writing, a fundamental change may take place between the third and the fourth grades. After two years of exclusive attention to the spelling of individual words (Grades 1 and 2), the third grade for the first time invites the children to produce written output that moves on to the sentence and text level: pupils are challenged to try and coordinate the many subskills and processes that together constitute the multiplex skill of 'communicative writing'. In view of the fact that the quantity and quality of writing education at primary school level is relatively poor (Bonset & Hoogeveen, 2007; Troia & Graham, 2003), children have to explore this new world mainly by themselves: what is a text composed of? How do I find out what I am supposed to tell? How do I turn thoughts into sentences? What should I pay primary attention to?

Beginning writers are very uncertain and, in looking for answers to many questions, may be very sensitive to any supportive input they receive. As such, we should not be surprised that the extreme variance between the children's texts is hard to explain: all children are actively searching, irrespective of their mother tongue or socioeconomic status, and they may come up with quite different, temporarily satisfactory answers (Nelson & Van Meter, 2007). Building up greater cognitive maturity and a range of experiences with writing assignments, the pupils of the fourth grade who can rely on different literary sources (including confrontations with written materials at home), may become more confident and knowledgeable about what constitutes writing at a quicker pace than the pupils who solely rely on the few hours of writing education that they are offered at primary school (Moats et al., 2006). This may explain that Tu(D)- and low-SES pupils do make progress, while, at the same time, the gap with their high-SES, Dutch-speaking peers widens.

In the Flemish education system, the third and fourth grades of primary school may actually constitute a crucial stage for EEP-target students: either they catch the train (having received the right educational impulses to really launch them into communicative writing processes), or they miss it and come to a standstill. This may have far-reaching consequences. For one, in Flemish education, low-SES, non-Dutch-speaking pupils are strongly overrepresented in the first year of the vocational strand of secondary education; they are usually assigned to this strand at the age of 12 after repeating a grade in primary school several times. As a result, an early system of streaming (Hirtt, Nicaise & De Zutter, 2007; Scheys, 2008) is institutionalised in the second and third grades of primary education, threatening the maximal learning opportunities of many EEP-target students (Nicaise, 2008).

Limitations of this study and recommendations for further research

Clearly, our study of the impact of background variables on the quality of young children's texts as measured with seven text features is limited in a number of aspects.

First, we are aware of the fact that our study was restricted to one particular genre and one particular task. Retelling a story on the basis of a comic does not tell us anything about the pupils' capacities to 'translate' (Hayes & Flower, 1980) while also having to retrieve 'content' and organise its transmittal. Moreover, retelling could be classified as a relatively closed task, which may be to the advantage of the weaker writers, rather than the stronger ones (Schoonen, 1991; Weigle, 2002). This could imply that the gap between the strong texts and the weaker texts could even have been wider than we have documented. Furthermore, we have worked with only one comic strip, enhancing the risk of task-specific effects, for the production of certain text features (e.g. Lex.Rich.) may be dependent on the contents of the story that the pupils are confronted with, and the complexity of its storyline. In addition, the production of certain text features may be genre-dependent: further research is needed to establish to what extent our findings can be generalised to other genres.

Secondly, our sample of informants was limited in a number of ways. The pupils of the third and the fourth grades were two separate cohorts; as a result, we cannot present any true longitudinal analyses of children's progress. Furthermore, the number of Dutch-speaking pupils was rather limited in proportional terms (nearly 20%). In the same vein, the number of high-SES pupils was relatively small (28,5%), and the variance of the mothers' diplomas restricted. Owing to the small number of different classes that were involved in the study, our conclusions with regard to the impact of *class composition* must remain tentative. We are also aware that we cannot generalise our findings in the

Turkish classes to the full range of ‘monolingual’ classes (e.g. monolingually Moroccan classes). On the other hand, the strong significant effects of variables relating to the pupils’ individual background and to class composition strengthen our intentions to further pursue this line of research.

This study showed that young pupils’ background variables and class composition variables have a strong impact on the extent to which they produce a range of text features while performing writing tasks. To further explore this issue in future studies, a greater number of mother-tongue speakers (of the target language) and a greater number of classes of different compositions at the level of home language should be involved. Some of the variables we used (like Ref.Coh. or Lex.Rich.) should be operationalised in a way that allows for more fine-grained analyses, based on the hypotheses we raised in the Discussion section. In addition, it would be extremely useful to gather data relating to different genres.

The differences between the third and fourth grades that we found in our study need to be further explored in follow-up research. Especially the way in which second language learners succeed in building up the cognitive and linguistic maturity to produce higher-quality texts can aid curriculum and syllabus developers, as well as teachers, to optimise writing education to all children in this particularly interesting phase of their scholarly development.

Finally, the underachievement of the Turkish children begs for further research. Future studies might focus on documenting the negative effect of the summer holidays on different text features, the relationship between these children’s oral narrations and written narrations, and the impact of education on their writing development.

Conclusions

This study complements earlier studies in (1) opting for a combination of six analytic variables and one primary-trait assessment; (2) applying multi-level-analyses to the data; and (3) focusing on the written texts of young, low-SES, non-indigenous pupils in Flemish primary education. The analyses show that young pupils’ *Home language* and *SES*, together with *class composition* have a strong impact on the quality of the written texts pupils produce, as measured with seven text features. However, the pace at which they grow and the extent to which they are affected by the independent variables are not the same for the different features. That implies that the process of producing written output of an increasingly higher quality cannot be described in terms of a uniform, linear route that all children follow in the same way. The results of this study also clearly indicate that Turkish children in homogeneously Turkish classes show significantly lower scores for the various text variables we selected. Further research is needed to corroborate the findings of this study, and to gain further insight into the written texts of young, low-SES, nonindigenous pupils in order for teachers to be able to provide higher-quality support during writing education.

Acknowledgements

This study was part of a broader research project into the quality of instruction in schools with a high proportion of EEP-target students, supported by the Flemish Ministry of Education and the K. V. Leuven (2005–2009). The authors are grateful to the anonymous reviewers for their useful comments and suggestions.

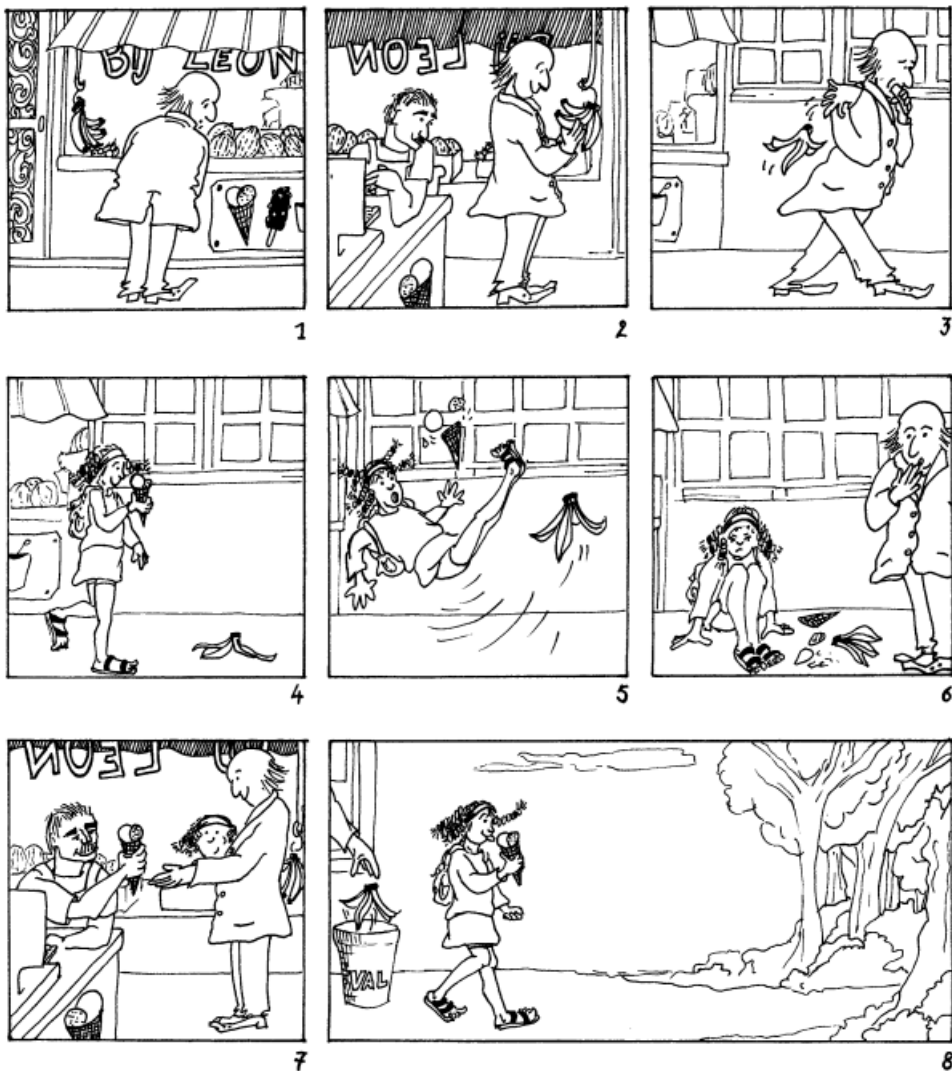
References

- Aarnoutse, C. & Verhoeven, L. (1999). *Tussendoelen Beginnende Geletterdheid. Een leerlijn voor groep 1 t/m 3*. [Intermediate goals emergent literacy for primary schools pupils aged 4 to 7.] Nijmegen: Expertisecentrum Nederlands.
- Aarts, R. (1994). *Functionele geletterdheid van Turkse kinderen in Turkije en in Nederland*. [Functional literacy of Turkish children in Turkey and in the Netherlands.] De Lier: Academisch Boeken Centrum.
- Akinci, M., Jisa, H. & Kern, S. (2001). Influence of L1 Turkish on L2 French narratives. In L. Verhoeven & S. Strömquist (Eds.), *Narrative development in a Multilingual Context*. (pp. 189–208). Amsterdam: John Benjamins Publishing Company.
- Alamargot, D. & Chanquoy, L. (2001). *Through the Models of Writing*. Dordrecht: Kluwer Academic Publishers.
- Bereiter, C. & Scardamalia, M. (1987). *The Psychology of Written Composition*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Berninger, V. & Swanson, H. (1994). Modifying Hayes and Flower's model of skilled writing to explain beginning and developing writing. *Advances in Cognition and Educational Practice*, 2, 57–81.
- Berninger, V., Yates, C., Cartwright, A., Rutberg, J., Remy, E. & Abbott, R. (1992). Lower-level developmental skills in beginning writing. *Reading and Writing*, 4, 257–280.
- Bonset, H. & Hooegeven, M. (2007). *Schrijven in het basisonderwijs. Een inventarisatie van empirisch onderzoek in het perspectief van leerplanontwikkeling*. [Writing in primary education. An inventory of empirical studies informing curriculum development.] Enschede: Stichting Leeplanontwikkeling (SLO).
- Braun, A., Forges, G. & Wolmainck, P. (1997). *Ecrire en français au primaire. Quelles performances pour les enfants issus de l'immigration*. Brussel: De Boeck & Larcier.
- Cameron, L. & Besser, S. (2004). *Writing in English as an Additional Language at Key Stage 2 (Rep. No. 586)*. Leeds: University of Leeds.
- Colpin, M., Heymans, R. & Rymenans, R. (2005). *Ontwikkeling van een instrument voor periodiek peilingsonderzoek schrijven (PPON-S). Ongepubliceerd eindrapport bij OBPWO 01.06*. [Developing an instrument for the nationwide assessment of writing skills at the end of primary education: A pilot study.] Leuven/Antwerpen: K.U.Leuven/U.A.
- Cooper, H., Nye, B., Charlton, K., Lindsay, J. & Greathouse, S. (1996). The effects of summer vacation on achievement test scores: A narrative and meta-analytic review. *Review of Educational Research*, 66, 227–268.
- Cragg, L. & Nation, K. (2007). Exploring written narrative in children with poor reading comprehension. *Educational Psychology*, 26, 55–72.
- Desmedt, E. & Nicaise, I. (2006). Etnische segregatie in het onderwijs: horen, zien en zwijgen? In S. Sierens, M. Van Houtte, P. Loobuyck, K. Delrue & K. Pelleriaux (Eds.), *Onderwijs onderweg in de immigratiesamenleving*. [Education in an immigration society.] (pp. 91–108) Gent: Academia Press.
- Eaton, J., Collis, G. & Lewis, V. (1999). Evaluative explanations in children's narratives of a video sequence without dialogue. *Journal of Child Language*, 26, 699–720.
- Ellis, S. (1985). *Understanding second language acquisition*. Oxford: Oxford University Press.
- Flower, L. (1979). Writer-based prose: A cognitive basis for problems in writing. *College English*, 41, 19–37.
- Foster, P. & Wigglesworth, G. (2007). Capturing accuracy in second language performance: Exploring the reliability of global and local measures. In *Complexity, accuracy and fluency in second language use, learning and teaching. Book of abstracts*. (pp. 6–7). Brussels: VUB.
- Hamp-Lyons, L. (2003). Writing teachers as assessors of writing. In B. Kroll (Ed.), *Exploring the dynamics of second language writing*. (pp. 162–190). Cambridge: Cambridge University Press.
- Hayes, J. & Flower, L. (1980). Identifying the organization of writing processes. In L. Gregg & E. Steinberg (Eds.), *Cognitive processes in writing: An interdisciplinary approach*. (pp. 3–30). Hillsdale, NJ: Erlbaum.
- Heesters, K. (2000). *Een wereld vol verhalen. Ontwikkeling van verhaalstructuur bij 9- tot 13-jarige eerste- en tweedetaalsprekers van het Nederlands*. [A world full of stories. The development of story coherence by young first and second language learners of Dutch.] Groningen: Rijksuniversiteit Groningen.
- Hirtt, N., Nicaise, I. & De Zutter, D. (2007). *De school van de ongelijkheid*. [Inequality at school.] Berchem: Epo.
- Howard, E., Christian, D. & Genesee, F. (2004). *The development of bilingualism and biliteracy from Grade 3 to 5: A summary of findings from the CAL/CREDE study of two-way immersion education (Rep. No. 13)*. Santa Cruz: Centre for Research on Education, Diversity and Excellence.
- Hudson, R. (2009). Measuring maturity. In R. Beard, D. Myhill, J. Riley & M. Nystrand (Eds.), *The Sage handbook of writing development*. (pp. 349–362). London: Sage.
- Hunt, K. (1965). *Grammatical structures written at three grade levels. (Rep. No. 3)*. Champaign, IL: NCTE.

- Janssen, R., Rymenans, R., Van Den Branden, K., Verhelst, M., Tuerlinckx, F., Van den Noortgate, W. et al. (2008). *Peiling lezen en luisteren (Nederlands) in het basisonderwijs. Eindrapport*. [Assessing Dutch reading and listening skills in primary education.] Leuven: Katholieke Universiteit Leuven.
- Jewell, J. & Malecki, C. (2005). The utility of CBM written language indices: An investigation of production-dependent, production-independent, and accurate-production scores. *School Psychology Review*, 34, 27–44.
- Kress, G. (1994). *Learning to write*. London: Routledge.
- Krom, R., Verhelst, N. & Veldhuijzen, N. (2004). *Balans van het schrijfvaardigheidsonderwijs op de basisschool. Uitkomsten van de peilingen in 1999: halverwege en einde basisonderwijs en speciaal onderwijs (Rep. No. 28)*. [National assessment of writing education in primary school (halfway and at the end). Results of 1999.] Arnhem: Citogroep.
- Lanauze, M. & Snow, C. (1989). The relation between first- and second-language writing skills: Evidence from Puerto Rican elementary school children in bilingual programs. *Linguistics and Education*, 1, 323–339.
- McCutchen, D. (2000). Knowledge, processing, and working memory: Implications for a theory of writing. *Educational Psychologist*, 35, 13–23.
- McMaster, K. & Espin, C. (2007). Technical features of curriculum-based measurement in writing. *The Journal of Special Education*, 41, 68–84.
- Medwell, J. & Wray, D. (2008). Handwriting - a forgotten language skill. *Language and Education*, 22, 34–47.
- Moats, L., Foorman, B. & Taylor, P. (2006). How quality of writing instruction impacts high-risk fourth graders' writing. *Reading and Writing*, 19, 363–391.
- Myhill, D. (2008). *Writing Aloud: An Exploration of Oral Rehearsal in the Early Year's Classroom*. Paper presented at Writing Research Across Borders, University of California, Santa Barbara. February 26, 2008.
- National Assessment of Educational Progress & National Center for Educational Statistics (2002). *Writing: The Nation's Report Card, 2002*.
- National Writing Project (2008). *Local site research initiative report. Cohort III. 2005–2006*. Berkeley, CA: National Writing Project, University of California.
- Nelson, N. & Van Meter, A. (2007). Measuring written language ability in narrative samples. *Reading and Writing Quarterly*, 23, 287–309.
- Nicaise, I. (2008). *Kansenongelijkheid in het Vlaamse onderwijs*. [Inequality in Flemish education.] Paper Presented at 'Denkdagen', Centre for Language and Education, K.U.Leuven, 26 May 2008.
- Nistov, I. (2001). Reference continuation in L2 narratives of Turkish adolescents in Norway. In L. Verhoeven & S. Strömquist (Eds.), *Narrative development in a multilingual context*. (pp. 51–86). Amsterdam: John Benjamins Publishing Company.
- Olinghouse, N. (2007). Student- and instruction-level predictors of narrative writing in third-grade students. *Reading and Writing*, 21, 3–24.
- Pirls Vlaanderen (2006). *Resultaten voor Vlaanderen*. [Results for Flanders.] Retrieved 15 January 2009 from <http://ppw.kuleuven.be/pirls/resultaten/Vlaanderen.html>
- Ramaut, G., Roppe, S., Verhelst, M. & Heymans, R. (2007). *Salto. Screeningsinstrument Aanvang Lager Onderwijs Taalvaardigheid*. [Salto. Instrument for the assessment of language skills at the beginning of primary education.] Leuven: Centrum voor Taal en Onderwijs.
- Ravid, D. & Tolchinsky, L. (2002). Developing linguistic literacy: A comprehensive model. *Journal of Child Language*, 29, 417–447.
- Reese, L., Garnier, H., Gallimore, R. & Goldberg, C. (2000). Longitudinal analysis of the antecedents of emergent Spanish literacy and middle-school English reading achievement of Spanish-speaking students. *American Educational Research Journal*, 37, 633–662.
- Reynders, T., Van Heddegem, I., Nicaise, I. & Van Damme, J. (2004). *Longitudoonaal onderzoek in het basisonderwijs. Oudervragenlijst schooljaar 2002–2003 (Rep. No. 22)*. [Longitudinal study in primary education. Parent Questionnaire 2002–2003.] Leuven: Steunpunt LOA.
- Rijlaarsdam, G. & Wesdorp, H. (1988). Het beoordelen van taalvaardigheid in het onderwijs. Nederlandse onderzoeksresultaten op het gebied van het meten van de schrijf-, spreek- en discussie-, lees- en luistervaardigheid. [The assessment of language skills in education. Results of Dutch studies concerning writing, speaking, discussing, reading and listening skills.] In C. Jansen & M. Steehouder (Eds.), *Taalbeheersing in artikelen*. (pp. 276–323). Leiden: Martinus Nijhoff.
- Saddler, B. & Graham, S. (2007). The relationship between writing knowledge and writing performance among more and less skilled writers. *Reading & Writing Quarterly*, 23, 231–247.
- Scheys, M. (2008). *Beginsituatie van leerlingen in het eerste leerjaar B van het secundair onderwijs (Rep. No. OBPWO 06.00)*. [An analysis of the pupils' starting position at the beginning of vocational training.] Brussel: Vlaamse Overheid, Departement Onderwijs en Vorming.

- Schoonen, R. (1991). *De evaluatie van schrijfvaardigheidsmetingen. Een empirische studie naar betrouwbaarheid, validiteit en bruikbaarheid van schrijfvaardigheidsmetingen in de achtste groep van het basisonderwijs. Ongepubliceerd academisch proefschrift.* [Assessing writing evaluation. An empirical study.] Amsterdam: Universiteit van Amsterdam.
- Singer, J. & Willett, J. (2003). *Applied longitudinal data analysis. Modeling change and event occurrence.* Oxford: Oxford University Press.
- Skehan, P. (2007). Complexity, accuracy, fluency, and lexis in task-based performance: A meta-analysis of the Ealing research. In *Complexity, accuracy and fluency in second language use, learning and teaching. Book of abstracts.* (pp. 21–22). Brussels: VUB.
- Smits, D. & Aarnoutse, C. (1997). Een longitudinaal onderzoek naar verschillen in taal- en leerprestaties van autochtone en allochtone kinderen. [A longitudinal study of differences between language and learning results of Dutch and migrant children.] *Nederlands tijdschrift voor opvoeding, vorming en onderwijs*, 13, 33–52.
- Tindal, G. & Parker, R. (1991). Identifying measures for evaluating written expression. *Learning Disabilities Research & Practice*, 6, 211–218.
- Troia, G. & Graham, S. (2003). Effective writing instruction across grades: What every educational consultant should know. *Journal of Educational and Psychological Consultation*, 14, 75–89.
- Van de Gein, J. (1991). Het benul van de zin. [The sense of sentences.] *Toegepaste taalwetenschap in artikelen*, 40, 102–111.
- Van de Gein, J. (2005). *Balans van taalkwaliteit in schrijfwerk uit het primair onderwijs (Rep. No. 29).* [Quality of writing products in primary education.] Arnhem: Citogroep.
- Van den Bergh, H., de Gloppe, K. & Schoonen, R. (1988). Directe metingen van schrijfvaardigheid: validiteit en taakeffecten. [Direct assessment of writing tasks: validity and effects of tasks.] In F. van Eemeren & R. Grootendorst (Eds.), *Taalbeheersing in Ontwikkeling.* (pp. 370–383). Dordrecht: Foris Publications.
- Van den Bergh, H. & Kuhlemeier, H. (1997). Multiniveau modellen voor de analyse van leerwinst vergeleken. [A comparison of different multilevel models for the analysis of learning gains.] *Tijdschrift voor onderwijsresearch*, 22, 54–75.
- Van Gorp, K. & Verheyden, L. (2006). Effecten van omgaan met meertaligheid op school- en klasniveau op de taalvaardigheid Nederlands van anderstalige leerlingen in het Vlaamse onderwijs. [The impact of different methods to deal with multilingualism in Flemish education on non-native speakers' Dutch language skills.] In T. Koole, J. Nortier & B. Tahiti (Eds.), *Artikelen van de Vijfde sociolinguïstische conferentie.* (pp. 187–198). Delft: Eburon.
- Verhaeghe, J.P., Knipprath, H. & Mertens, G. (2007) *Begrijpend lezen. Vierde leerjaar. Schoolfeedbackrapport n.a.v. deelname aan het PIRLS 2006-onderzoek.* [Reading comprehension in the 4th grade.] Centrum voor Onderwijseffectiviteit.
- Verhoeven, L. (1992). Mondelinge en schriftelijke vaardigheid in het Nederlands als eerste en tweede taal. [Oral and written Dutch first and second language skills.] *Toegepaste taalwetenschap in artikelen*, 43, 66–77.
- Verhoeven, L. (1997). Acquisition of literacy by immigrant children. In C. Pontecorvo (Ed.), *Writing development: An interdisciplinary view.* (pp. 219–240). Amsterdam: John Benjamins.
- Vermeer, A. (2000). Coming to grips with lexical richness in spontaneous speech data. *Language Testing*, 17, 65–83.
- Weigle, S.C. (2002). *Assessing writing.* Cambridge: Cambridge University Press.
- Wells, G. (1985). *Language development in the preschool years.* Cambridge: Cambridge University Press.
- Wigglesworth, G. (1997). Children's individual approaches to the organization of narrative. *Journal of Child Language*, 24, 279–309.
- Wolfe-Quintero, K., Inagaki, S. & Kim, H.Y. (1998). *Second language development in writing: measures of fluency, accuracy & complexity.* Honolulu: University of Hawaii.
- Woordenlijst Nederlandse Taal (2005). Tiel: Lannoo Uitgeverij.

Appendix



© Centre for Language and Education, K.U.Leuven.

Received 29 March 2009; revised version received 29 October 2009.

Address for correspondence: Lieve Verheyden and Kris Van den Branden, Centre for Language and Education, K.U.Leuven, Blijde-Inkomststraat 7–9, 3000 Leuven, Belgium.
E-mail: lieve.verheyden@arts.kuleuven.be, kris.vandenbranden@arts.kuleuven.be