Development of Early Numeracy in 5- to 7-Year-Old Children: A Comparison Between Flanders and The Netherlands

Joke Torbeyns, Wim van den Noortgate, Pol Ghesquière, Lieven Verschaffel, Bernadette A.M. van de Rijt & Johannes E.H. van Luit


To link to this article: http://dx.doi.org/10.1076/edre.8.3.249.3855

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.tandfonline.com/page/terms-and-conditions

This article may be used for research, teaching, and private study purposes. Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae, and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or
howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.
Development of Early Numeracy in 5- to 7-Year-Old Children: A Comparison Between Flanders and The Netherlands

Joke Torbeyns¹, Wim van den Noortgate¹, Pol Ghesquière¹, Lieven Verschaffel¹, Bernadette A.M. van de Rijt², and Johannes E.H. van Luit²
¹University of Leuven, Belgium, and ²Utrecht University, The Netherlands

ABSTRACT

This article describes a research project about the development of early numeracy in 5- to 7-year-old Flemish children compared to their Dutch peers. A group of Flemish children was tested 3 times by means of a slightly adapted Flemish version of the Utrecht Early Numeracy Test, and their results were compared to the development of those of a comparable group of Dutch children. The first and the second assessment took place, respectively, at the middle and at the end of the last year of preprimary education, and the third assessment at the middle of the first year of primary education. The results indicate that the Early Numeracy Test is a moderately reliable and construct valid instrument to assess the general early numeracy. At the time of the first assessment Flemish children scored worse on the test than their Dutch peers. However, they showed a faster progression between the first and the third assessment, resulting in a catch up with their Dutch peers at the third assessment. The results are interpreted in terms of differences in the goals, the content, and the organization of preprimary and primary school mathematics education in both countries.

INTRODUCTION

Empirical studies of the development of mathematical competence (see, among others, Baroody, 1987; Fuson, 1992; Geary, 1994; Ginsburg, 1977; TAL-team, 1999; Van de Rijt, 1996), all seem to come to the same conclusion: The acquisition of mathematical knowledge and skills can be viewed as a
developmental process that starts long before the beginning of formal mathematics education in the primary school. Even at preschool age, children are interested in numbers and number facts, they have acquired some intuitive knowledge about the different meanings of numbers and the way numbers can be used, they can recite the counting sequence up to a certain number, and are able to count, group, separate, and compare small sets of concrete objects, and so forth. These “informal” mathematical concepts and skills (referred to as “early mathematical competence” or “early numeracy” in the literature) develop gradually, based on children’s frequent and varied experiences with numbers and quantitative situations both in in-school and out-of-school contexts. In preschool, teachers stimulate the development of these intuitive mathematical notions and informal counting skills through spontaneously occurring learning activities as well as through more or less intentionally and systematically organized educational activities.

Of course, the development of these (informal) mathematical knowledge and skills differs from child to child. For instance, some 3-year-old children have acquired already certain counting skills that most (normally developing) 5-year-olds do not master yet. In contrast, some 5-year-olds do not yet possess the mathematical knowledge and skills that most 3-year-old children have already acquired.

This article focuses on the development of early numeracy in 5- to 7-year-old children. We start with a brief discussion of the definition of the concept of early numeracy as well as the factors that seem to influence its development. Next, we briefly compare the structure of the Flemish and the Dutch educational system in general, and the organization of the learning and teaching of mathematics education in both countries in particular. This is followed by a description of the design of our comparative study, and of its results. Finally, we summarize the major findings and we raise some questions for further research.

DEFINITION OF EARLY NUMERACY

As mentioned earlier, empirical studies of the development of mathematical competence all seem to come to the same general conclusion: Children do possess quite some informal mathematical knowledge and skills before the beginning of formal mathematics education. In spite of this general agreement about the existence of “emerging numeracy,” there is disagreement about the specific structure of this informal mathematical knowledge and these skills,
about the way they develop, and about their impact on the learning and teaching of formal mathematics. In this respect, three major views can be distinguished (for more details, see Torbeyns, 1999).

A first view, which has been heavily influenced by the work of Piaget (1952), focuses on the “logical-mathematical foundations” of the development of number concepts and early mathematical skills. According to Piaget (1952), the concept of number is conceived as the synthesis of the logical operations of number conservation, classification, and seriation. A child who has acquired these logical operations, has true understanding of what numbers are, and is ready to start to do mathematical operations on them in a meaningful way. The development of these logical operations is stimulated through a child’s interaction with concrete materials in his or her environment (Piaget, 1965). In contrast, learning the number words and learning how to count correctly, smoothly, and efficiently are not of vital importance for the development of the concept of number. Or stated in Piaget’s (1952) terms: Counting has no operational value.

In the second view, the pivotal role of Piaget’s logical operations is questioned, and at the same time the importance of children’s “counting skills” as the basis for the number concept is stressed. According to adherents of this view, the development of the number concept can be largely explained in terms of the development of counting skills: Frequent and varied experiences with counting are considered necessary for the development of the concept of number, and of the concepts of addition and subtraction. The view on early mathematical development of the Freudenthal Institute in The Netherlands (TAL-team, 1999), as well as the work of developmental psychologists like Baroody (1987), Frank (1989), and Gelman and Gallistel (1978) can be considered as typical examples of this second view.

A third approach integrates the ideas of the first and the second approach into one synthetic model. In this synthetic approach, both the Piagetian operations and the counting skills are considered crucial for the development of early numeracy. According to Van de Rijt (1996), early numeracy consists of eight close-knit aspects listed below, some of which are reminiscent of the first and others of the second approach. Further theoretical evidence for the distinction between these eight components can be found in the work of Brainerd (1979), Briars and Siegler (1984), Fuson (1988), Sophian (1992), and Wynn (1990).

Van de Rijt (1996) used her synthetic model consisting of eight components as the theoretical framework for the construction of the Early Numeracy Test (ENT; Van Luit, Van de Rijt, & Pennings, 1994). The ENT consists of two
forms: ENT Form A and ENT Form B. Each form contains eight subscales, consisting of five items each (total number of items = 40). The subscales correspond to the eight components of the theoretical framework proposed by Van de Rijt (1996):

1. Concepts of comparison. This subscale focuses on the child’s ability to compare objects with regard to quantitative or qualitative properties. It measures whether a child understands the concepts that are used often in tasks of comparison (e.g., higher, most, . . .).

2. Classification. Classification refers to the ability to group objects in a class or subclass on the basis of one or more criteria. This subscale measures whether the child is able to classify objects on the basis of similarities.

3. One-to-one correspondence. This subscale measures the child’s ability to put a one-to-one relation between different objects. The child is asked to compare numbers of objects (presented simultaneously) through the application of the one-to-one relation.

4. Seriation. Seriation can be defined as the ordering of objects according to one or more criteria. This subscale measures the child’s ability to deal with discrete and ordered entities in a passive as well as in an active way. The child has to recognize whether a row of objects or numbers is presented in correct order (passive) and is asked to put a number of in-ordered objects in correct order (active).

5. The use of number words. The use of number words refers to the following counting skills: Producing the number words up to 20 in forward and in backward order, counting on (i.e., counting, starting with a number other than 1), and using the cardinal and ordinal numbers.

6. Structured counting, which includes both the skill of point counting (i.e., pointing to the individual items in a set while counting them, without missing any of the items and without counting any of them more than once) and the skill of skip counting (i.e., counting every nth number in a series of numbers). To start with, this subscale measures whether the child masters the skill of point counting. The child is asked to count a number of concrete materials. It is allowed that the child points to these materials while counting them. Furthermore, this subscale measures whether the child masters the skill of skip counting. The child is asked to determine the cardinal number of a structured group of objects. These objects are ordered in exactly the same way as the points on a die (skip counting, according to the structure of a die).
7. Resultative counting. This subscale measures whether the child understands the cardinal principle, and whether he or she can apply this principle accurately. The child is asked to determine the cardinal number of structured as well as unstructured entities. The child is not allowed to point to the objects while counting them.

8. General understanding of numbers. This aspect integrates all the other components in the theoretical model of Van de Rijt. It refers to the application of numeracy in real daily life situations, represented in drawings.

An example of an item from each of the eight subscales is given in the Appendix.

Each item is scored dichotomously on correctness (maximum score on the test = 40). The sum score on the test is transformed into an ability score (linear transformation), with values ranging between 0 and 100. This ability score indicates a child’s level of mastery of early numeracy: The higher the ability score of the child, the better he or she masters the early numerical skills.

The ENT is constructed for the assessment of early numerical skills in children aged 4–7 years. The test is administered individually and takes about 20–30 min. During assessment, the test administrator is supposed to write down some qualitative observations about the strategies the child uses to solve the test items.

During the stage of the test construction, all items out of a pool of 120 items were administered to 823 Dutch boys and girls in the 4- to 7-year aged groups (anchor items design) in the months of January–February 1993 (Van de Rijt, 1996; Van de Rijt, Van Luit & Pennings, 1999). This resulted in a definitive pool of 80 items. These data were used to construct norm tables for the ENT as well as to assess the psychometric quality of this test. Taking practical, didactical and psychometric conditions into account, the 80 items out of the definitive item bank were divided equally among two scales, which resulted in the construction of ENT Form A and ENT Form B.

Once the psychometric quality was established, Van de Rijt (1996) started in January 1994 a longitudinal study of the development of early numeracy in Dutch children aged 5–7 years. Two-hundred-ninety-nine children, at the first time of measurement all pupils of group 2 at a school for elementary education in The Netherlands, participated in this study. Based on the results of the norming study as well as the results of the longitudinal study, Van de Rijt (1996) and Van de Rijt et al. (1999) concluded that:

1. the ENT is a reliable test;
2. the ENT shows good construct validity;
3. the eight components of the model develop multilinear (i.e., the eight components can be ordered according to their time of development; see Werner, 1957).

In January 1997, we started a longitudinal study of the development of early numeracy in Flemish children aged 4–7 years as measured by means of (a slightly adapted version of) the ENT. The goal of this study was twofold:

1. the construction of norm tables for the ENT for Flanders;
2. the study of the development of early numeracy in Flemish children, in comparison with their Dutch peers as assessed in the study of Van De Rijt (1996).

MATHEMATICS EDUCATION IN FLANDERS AND THE NETHERLANDS

The stimulation of the acquisition of mathematical knowledge and skills is considered as an important educational task throughout the world. However, the aims, the content, and the organization of mathematics education at school, as well as the instructional materials and tools teachers use to stimulate the mathematical development of children, differ from country to country. This is also the case for Flanders and The Netherlands. Hereafter we point to some differences between both countries with regard to the organization and the content of (mathematics) education for children up to the age of 12 (i.e., preprimary and primary education; see Janssen-Vos & Laevers, 1996; Laevers, Janssens, & Laurijssen, 1999).

A first difference concerns the organization of preprimary education. In Flanders, preschool is open for children from 2.5 to 6 years of age, and consists of three grades. The first grade of preprimary education offers instruction to children aged between 2.5 and 3. The second grade of preprimary education offers instruction to 4-year-olds, and in the third grade, instruction is offered to children aged 5. In The Netherlands, preprimary education begins only at an age of 4, and consists of only two grades (called, respectively, group 1 and 2).

Another major difference between the structure of the Flemish and the Dutch educational system concerns the relation between preprimary and primary education. In Flanders, a clear distinction is made between preschool, which offers instruction to children aged 2.5 to 6, and the six grades for
primary education, wherein children aged 6–12 receive instruction. As a result of this clear distinction between preprimary and primary education, the transition from the former to the latter is generally considered as quite a big step for Flemish children. In some cases it implies the physical movement of children from one school to another, and for most of them it requires a rather quick adaptation from a play-like, informal, unsystematic, “holistic” instructional environment to a more formal, more systematic, and curricular kind of instruction. In contrast, preprimary education forms an integrated part of elementary education in The Netherlands: Dutch schools for elementary education offer instruction to children aged 4–12. Group 1 and 2 offer instruction to 4- and 5-year-olds, respectively; children in the 6–12-years age range receive instruction in groups 3–8. Consequently, the transition from preprimary education to primary education seems less drastic in The Netherlands.

This latter organizational difference is reflected in differences in several aspects of the educational practice and culture between both countries. More specifically, it led not only to clear differences in the aims and scope of the Flemish and the Dutch system of teacher training, but also to the development of different curricular goals, time schedules, and instructional tools in preprimary and primary school mathematics education in Flanders and The Netherlands, which all, in their turn, result in different classroom practices. First, with respect to teacher training, the clear separation in Flanders between preprimary and primary education has led to the development of two quite separated forms of teacher training with clearly different goals, practices, and cultures. In The Netherlands, on the other hand, the integration of preprimary and primary education has resulted in one integrated form of teacher training for both age groups. Arguably, this difference in the organization of teacher training will influence, in its turn, the pedagogical content knowledge and beliefs of the Flemish and Dutch teachers, as well as their actual teaching behavior. For instance, it is plausible that in a more integrated system like the Dutch one, teachers working in groups 1 and 2 will be more informed and concerned about the development of and the teaching to older children and, therefore, will be more inclined to start already with resolute and purposive early (mathematics) instruction to these young children, than their Flemish colleagues, who have been enculturated in a setting that is dominated by a “worship” of the genuine developmental characteristics and a great concern for the special educational needs of preprimary school children. A second difference that follows directly from the abovementioned organizational
difference, concerns the standards and goals that Flemish and Dutch teachers have to take into account during preprimary and primary education, as well as the time scheduled to reach these goals. In Flanders, there exists a clear difference between, on the one hand, the so-called developmental goals that preschool teachers have to pursue during preprimary education, and, on the other hand, the standards that elementary school teachers have to take into account during primary education (Ontwikkelingsdoelen en eindtermen, 1998), as well as the amount of time offered to them to reach these goals. Flemish preschool teachers are supposed to primarily observe and stimulate the development of “the child as a whole”: The development of preschool children’s emotional, behavioral, and social competencies is considered at least as important as their cognitive development. Taking into account the large differences in the rate of development of the emotional, behavioral, social, and cognitive competencies of children in the 2.5–6-age range, and, as a result of the latter, of their special educational needs, preschool teachers have a lot of freedom to decide how much time they spend on the development of these distinct competencies in the classroom. In contrast, Flemish elementary school teachers are expected to focus more on the cognitive development of children. The curricular goals for primary education specify which (mathematical) competencies children have to possess at the age of 12; the amount of time available to reach these goals, is also scheduled rather strictly per grade. In The Netherlands, the development of children’s cognitive competencies is considered more from group 1 onwards. This is reflected in both the curricular goals, which specify already starting at group 1 which domain-specific (mathematical) competencies children should possess, and how much time is available in each group for stimulating the development of each of these competencies. A third consequence of the difference in the educational system for Flemish and Dutch children up to age 12, concerns the development of instructional materials and tools for stimulating the cognitive development of preelementary school children. In line with the integration of preprimary and primary education in The Netherlands, current Dutch instructional methods for mathematics education typically start already in group 1. In other words, already for children aged 4 and 5, Dutch teachers have domain-specific instructional programs, materials, and tools at their disposal to organize the learning and teaching of early mathematical concepts and counting skills in quite an intentional, systematic, and planned way. In Flanders, mathematics textbook series that start already in preprimary education are extremely rare. Before the start of formal mathematics education (in the first
grade of a school for primary education), the development of mathematical knowledge and skills of Flemish children is stimulated in a more implicit and informal way, embedded in daily activities, interactions, and play (see Vossen, 1999). In Flemish schools for preprimary education, working around centers of interest and the use of the impression–expression cycle, are used as pivotal guiding principles of the daily class routine (see Laevers et al., 1999).

A final difference between the approach to mathematics education between Flanders and The Netherlands refers to the general view on the goals of mathematics education, the content, the teaching methods, and so forth, in the primary school in both countries. Until the early 1990s, the Flemish curriculum was dominated by the ‘New Math’ approach to mathematics education, which was characterized by the central and foundation role played by set theory, the early utilization of formal representational tools (e.g., Venn-diagrams), the focus on conceptualizing mathematics (at the expense of applying it in the real world), and the direction toward the abstract (Verschaffel & De Corte, 1996). The Dutch curriculum and instructional practice, on the contrary, were heavily influenced by the so-called ‘Realistic approach to mathematics education,’ with its focus on linking mathematics to real life, its cultivation of children’s free productions and constructions, and its plea for a gradual progression towards higher levels of abstraction and formalization (Verschaffel & De Corte, 1996). Whereas these differences have become more and more blurred in recent years, there are still differences in emphasis between both countries that are reminiscent of these different histories, implying, for instance, that from the start of grade 1 (i.e., group 3 in The Netherlands) on, mathematics education in Flanders moves somewhat more quickly and more eagerly into the direction of abstraction and formalization, and leaves somewhat less room for children’s own inventions and free productions than in The Netherlands (Feys, 1997).

**METHOD**

**Participants**

Subjects were 207 children from 12 schools for preprimary and primary education in Flanders. These schools were selected in such a way that they were as representative as possible for the Flemish population in terms of factors like school size, educational network, SES and ethnic background of the parents, and so forth (Torbeyns, 1999). However, for practical reasons,
only schools located in the province of Vlaams-Brabant and one adjacent province (Limburg) were included in the study. At the first time of measurement, all participating Flemish children were in the third grade of preprimary education. For more information about the (Dutch) children who participated in the study that was executed in The Netherlands, we refer to Van de Rijt (1996). Table 1 gives a description of the sex and the age (expressed in months of age) of the Flemish and Dutch children at the first time of measurement, that is in the middle of the last year of preprimary education.

Table 1 shows that the Dutch children were on average 4 months older than the Flemish children at the first time of measurement. This – unexpected – age difference between the group of Flemish and Dutch children was a direct result of the way children are divided into grades in both countries. In Flanders, children are divided into grades primarily according to their year of birth. For instance, all Flemish children born in 1992 move from the last grade of preprimary education to the first grade of primary education in September 1998 (except some children who are considered as not yet ready to start formal education, and stay in a school for preprimary education for an additional year). In The Netherlands, children are put into a grade according to their actual age at the beginning of the school year. Consequently, not all Dutch children born in the year 1992 move from group 2 to group 3 in the month of September 1998: Children born after September 30, 1992, normally move to group 3 only in September 1999.

### Materials: The Early Numeracy Test (ENT)

The early numerical skills of all participating children were measured by means of the ENT. However, before starting the collection of the research data in Flanders, we slightly adapted this instrument to the Flemish situation in two ways. First, we constructed a third scale of the ENT (which was necessary for the calibration of the items of the ENT by means of the Item Response Theory,
and thus for the realization of the first goal of our study): ENT Form C. ENT Form C consists of 40 items. Half of these items are items of ENT Form A; the other half are items of ENT Form B. Second, the instructions of the ENT were translated in Flemish, which means that some Dutch words and expressions that were quite unfamiliar to Flemish children, were replaced by their more common Flemish synonyms. We will refer to this Flemish version of the ENT by means of the abbreviation ENT(-F).

**Design and Procedure**
The early numerical skills of all Flemish children were measured at three different moments in time, in the same way as was done in the study of Van de Rijt (1996). Children were tested for the first time in the middle of the third grade of preprimary education (January 1997). The second assessment was at the end of the third grade of preprimary education (June 1997). The third assessment took place in the middle of the first grade of primary education (January 1998). All children were tested individually each time by means of 1 of the 3 forms of the ENT(-F) (Form A, Form B, Form C). The three forms were divided equally among the participants, taking into account their sex as well as the class group to which they belonged. Each child received the same scale of the ENT(-F) at each time of measurement. For a detailed description of the similar procedure used in the Dutch study, see Van de Rijt (1996).

**RESULTS**

**Psychometric Characteristics of the ENT and the ENT(-F)**

*Reliability*

Table 2 shows the values of the reliability coefficients of the three forms of the ENT and the ENT(-F). We used Cronbach’s $\alpha$ coefficient to estimate the values of the reliability coefficients.

As shown in Table 2, Form B of the original ENT has a considerable greater N than the other forms of the ENT and the ENT(-F). This greater N of ENT Form B can be explained by the fact that the Dutch population consisted of two subpopulations (Van de Rijt, 1996). The reliability of a test can be considered as being satisfying if the reliability coefficient has a value of at least .90 (Storms, 1998). Using this criterion, the ENT Form A is reliable, whereas Form B and Form C of the ENT can only be considered as moderately
reliable. These reliability data are somewhat lower than those obtained in Van de Rijt et al.’s first study (Van de Rijt & Van Luit, 1999), wherein they obtained a reliability coefficient with a value of .94 for ENT Form A, and a value of .95 for ENT Form B. None of the three reliability coefficients of the ENT(-F) reaches the .90 criterion. Therefore, the three forms of the ENT(-F) can be considered as only moderately reliable.

Construct Validity
The construct validity of the ENT and the ENT(-F) has been tested by means of the following three complementary methods:

1. Testing the appropriateness of the One Parametric Logistic Model (OPLM; Verhelst, Glas, & Verstralen, 1995) as an adequate psychometric model to describe children’s responses to the items of the ENT and the ENT(-F).
2. Computing the correlation between children’s scores on the eight subscales of the ENT and the ENT(-F) (matrix of correlations).
3. Executing a factor analysis (principal components analysis) on the ENT and the ENT(-F).

Generally speaking, the results of each of these analyses suggest that the ENT and the ENT(-F) have a good construct validity:¹ Both the ENT and the ENT(-F) seem to measure one general construct, which can be interpreted as a child’s level of mastery of general early numeracy (for more details see Torbeyns, 1999).

Table 2. Reliability Coefficients of the ENT and the ENT(-F) (Cronbach’s $\alpha$ Coefficient).

<table>
<thead>
<tr>
<th>Form</th>
<th>ENT</th>
<th>n</th>
<th>ENT-F</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form A</td>
<td>.90</td>
<td>33</td>
<td>.85</td>
<td>71</td>
</tr>
<tr>
<td>Form B</td>
<td>.84</td>
<td>232</td>
<td>.83</td>
<td>71</td>
</tr>
<tr>
<td>Form C</td>
<td>.82</td>
<td>34</td>
<td>.86</td>
<td>65</td>
</tr>
</tbody>
</table>

¹Because of the fact that no comparable tests (as far as content and psychometric information are concerned) for testing early numeracy are available in Flanders nor in The Netherlands, it was not possible to acquire any further information concerning the construct validity of the instrument. However, the information available (i.e., the judgement of the experts, the results of the correlation analysis, the results of the factor analysis, as well as the results of the analyses by means of the Item Response Theory) provides sufficient evidence for the assumption that the ENT and the ENT(-F) do measure early numeracy.
The Development of Early Numeracy

Regarding the second goal of our study, namely the comparison of the development of early numeracy in Flemish and Dutch children, two multilevel analyses were executed. In both analyses we used a hierarchical linear model with three levels: The scores (level 1) are “nested” within children (level 2), who are nested within schools (level 3). Indeed, it can be assumed that the scores of children from the same school are more similar than the scores of children from different schools, as a result of selection effects or the influence of, among other things, the educational approach and the approach to mathematics education used in that particular school. Unlike in traditional statistical analyses, such as ordinary linear regression analysis, a multilevel analysis takes this dependence in the data due to the multilevel structure into account.

In the first analysis, we compared the scores of the Flemish children on the ENT(-F) and the scores of the Dutch children on the ENT as a function of their age (in months). In the second analysis, we compared the scores of the Flemish and the Dutch children at the three times of measurement. We estimated the value of the parameters by means of the so-called “iterative generalized least squares” (IGLS) method of the computer program MLwiN (Goldstein et al., 1998).

**Analysis 1**

In the first analysis, we investigated the development of early numeracy in Flemish and Dutch children by exploring the relation between age and score on the ENT(-F) across the different times of measurement. For that purpose, we used the following regression equation on the first level, in which the scores of the children on the ENT(-F) are predicted by means of their age:

\[
\hat{\text{ENT}}_{ijk} = \beta_{0jk} + \beta_{1jk} \times \text{age}_{ijk} + \epsilon_{ijk}
\]

with i, j and k indicating respectively measurement i, child j and school k, and \( \epsilon_{ijk} \) a random residual.

Because we assumed that the mean performance as well as the evolution over age would possibly vary over children (second level) and schools (third level), we used indices j and k for both regression coefficients. Next, because the intercept must be interpreted as the expected performance at age 0 and is thus meaningless, we diminished the variable age with the mean age of the children (i.e., 74.68 months). This means that the value of the intercept in the regression model corresponds with the predicted score on the ENT(-F) at an
average age. Furthermore, we assumed that both the overall performance and the evolution over age would possibly differ in Flanders and The Netherlands. Finally, we corrected the results for a possible influence of the form of the test. Therefore, the regression coefficients can be described using the following equations:

\[
\begin{align*}
\beta_{0jk} &= \gamma_{00} + \gamma_{01} \text{(country)}_k + \gamma_{02} \text{(form2)}_{jk} + \gamma_{03} \text{(form3)}_{jk} + \nu_0 + u_{0jk} \\
\beta_{ijk} &= \gamma_{10} + \gamma_{11} \text{(country)}_k + \nu_1 + u_{1jk}
\end{align*}
\]

and

\[
\begin{align*}
\beta_{0jk} &= \gamma_{00} + \gamma_{01} \text{(country)}_k + \gamma_{02} \text{(form2)}_{jk} + \gamma_{03} \text{(form3)}_{jk} + \nu_0 + u_{0jk} + \\
\gamma_{10} \text{(age)}_{ijk} + \gamma_{11} \text{(country \times age)}_{ijk} + \nu_{1k} \text{(age)}_{ijk} + u_{1j} \text{(age)}_{ijk} + \epsilon_{ijk}.
\end{align*}
\]

with \((\text{form2})_{jk}\) equalling 1 if for pupil \(j\) from school \(k\) ENT(-F) Form B was used, 0 otherwise, \((\text{form3})_{jk}\) equalling 1 if for pupil \(j\) from school \(k\) ENT(-F) Form C was used, 0 otherwise, \((\text{country})_k\) equalling 1 if school \(k\) belongs to The Netherlands, 0 if school \(k\) belongs to Flanders, \(\nu_0\) and \(\nu_1\) random residuals at the school level, and \(u_0\) and \(u_1\) random residuals at the pupil level.

Substituting (2) in (1) gives the following equation:

\[
\begin{align*}
\text{ENT}_{ijk} &= \gamma_{00} + \gamma_{01} \text{(country)}_k + \gamma_{02} \text{(form2)}_{jk} + \gamma_{03} \text{(form3)}_{jk} + \nu_0 + u_{0jk} + \\
&\quad \gamma_{10} \text{(age)}_{ijk} + \gamma_{11} \text{(country \times age)}_{ijk} + \nu_{1k} \text{(age)}_{ijk} + u_{1j} \text{(age)}_{ijk} + \epsilon_{ijk}.
\end{align*}
\]

This results in the following country-specific regression equations for ENT(-F) Form A:

\[
\begin{align*}
\text{predicted ENT(-F)} &= \gamma_{00} + \gamma_{10} \times \text{age} \\
\text{predicted ENT} &= (\gamma_{00} + \gamma_{01}) + (\gamma_{10} + \gamma_{11}) \times \text{age}
\end{align*}
\]

The estimated values of the parameters and the corresponding standard errors of estimation are summarized in Table 3. The ratio of the estimated value of a parameter to its standard error follows an almost standard-normal distribution. Consequently, we can say that if the estimated (absolute) value of a parameter is at least 1.96 times its standard deviation, the value of this parameter is statistically significant at a level of 5%.

On the basis of the results of this first analysis (Table 3), it can be concluded that at an age of 74.68 months (the mean age) the predicted score of the Flemish children on Form A of the ENT(-F) is 27.24. At that age, the Dutch children score 0.43 points less on the ENT (predicted score) than their Flemish peers on the ENT(-F). However, this is not a significant difference. Although scores on ENT(-F) Form B are somewhat higher than on both other forms, the
The difference is not significant. The predicted score of the Flemish children on the ENT(-F) increases each month with 0.94 points. In contrast, the predicted score of the Dutch children on the ENT increases with only 0.58 points each month. The relation between the age of the child and his or her score on the ENT(-F)

| Table 3. Development of Early Numerical Competence: Results of the Multilevel Analyses. |
|---------------------------------|-----------------|-----------------|-----------------|-----------------|-----------------|
| Fixed parameter                | Analysis 1      | Estimate        | SE              | Analysis 2      | Estimate        | SE              |
| Intercept ($\gamma_{00}$)      | 27.24           | 0.7734          | 22.7            | 1.151           |                 |                 |
| Regression coefficient for age ($\gamma_{10}$) | 0.9417 | 0.04335 | 0.2969 | 0.113 |                 |                 |
| Regression coefficient for age x country ($\gamma_{11}$) | -0.429 | 0.9673 | 3.393 | 1.306 |                 |                 |
| Regression coefficient for form2 ($\gamma_{02}$) | 0.8667 | 0.7153 | 0.5256 | 0.6639 |                 |                 |
| Regression coefficient for form3 ($\gamma_{03}$) | 0.03924 | 0.7561 | 0.0257 | 0.6826 |                 |                 |
| Regression coefficient for time2 ($\gamma_{20}$) | - | - | 1.94 | 0.6527 |                 |                 |
| Regression coefficient for time3 ($\gamma_{30}$) | - | - | 8.545 | 1.271 |                 |                 |
| Regression coefficient for time2 x country ($\gamma_{21}$) | - | - | -0.8115 | 0.7092 |                 |                 |
| Regression coefficient for time3 x country ($\gamma_{31}$) | - | - | -3.779 | 1.439 |                 |                 |
| Regression coefficient for time2 x age ($\gamma_{22}$) | - | - | -0.05185 | 0.04348 |                 |                 |
| Regression coefficient for time3 x age ($\gamma_{32}$) | - | - | -0.1017 | 0.04669 |                 |                 |

<table>
<thead>
<tr>
<th>Level Random parameter</th>
<th>Analysis 1</th>
<th>Estimate</th>
<th>SE</th>
<th>Analysis 2</th>
<th>Estimate</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>3 Intercept variance between schools ($\sigma^2 v0$)</td>
<td>3.045</td>
<td>1.381</td>
<td>3.812</td>
<td>1.581</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Covariance between schools ($\sigma v0v1$)</td>
<td>-0.1257</td>
<td>0.07224</td>
<td>-0.1293</td>
<td>0.07561</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Slope variance between schools ($\sigma^2 v1$)</td>
<td>0.01363</td>
<td>0.006118</td>
<td>0.01271</td>
<td>0.005753</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Intercept variance within schools ($\sigma^2 u0$)</td>
<td>26.35</td>
<td>1.912</td>
<td>23.03</td>
<td>1.684</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Covariance within schools ($\sigma u0u1$)</td>
<td>-0.3877</td>
<td>0.08603</td>
<td>-0.5612</td>
<td>0.08208</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2 Slope variance within schools ($\sigma^2 u1$)</td>
<td>0.003073</td>
<td>0.00978</td>
<td>0.006399</td>
<td>0.008799</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1 Variance within pupils ($\sigma^2 e$)</td>
<td>9.097</td>
<td>0.5363</td>
<td>8.695</td>
<td>0.4899</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Description of the observed test scores (maximum score = 40), standard deviation of scores (SD), and range in scores of the Flemish and the Dutch children at the three times of measurement:

- First time of measurement: Flemish children: Mean test score = 20.46, SD = 6.52, Minimum score = 6, Maximum score = 35; Dutch children: Mean test score = 25.43, SD = 6.52, Minimum score = 6, Maximum score = 39.
- Second time of measurement: Flemish children: Mean test score = 24.08, SD = 6.82, Minimum score = 5, Maximum score = 38; Dutch children: Mean test score = 28.04, SD = 6.33, Minimum score = 7, Maximum score = 39.
- Third time of measurement: Flemish children: Mean test score = 32.29, SD = 4.64, Minimum score = 19, Maximum score = 40; Dutch children: Mean test score = 33.29, SD = 4.92, Minimum score = 12, Maximum score = 40.
thus is not the same in both countries: The scores of the Flemish children on the ENT(-F) increase faster than the scores of the Dutch children. This interaction effect is highly significant. Figure 1 shows the influence of the variable age on the score on the ENT(-F), as well as the interaction between the variables age and country.

Finally, the results reported in Table 3 indicate that, even if differences in age and country are taken into account, the overall performance differs to a large extent across schools, across pupils from the same school, and across measurements from the same pupils. The negative covariance estimates at both the school and pupil level indicate that the performance of schools or pupils with a better general performance tends to improve less.

**Analysis 2**

In the second multilevel analysis, we compared the performance of the Flemish and Dutch children at the three different times of measurement: at the middle and at the end of the last year of preprimary education, and at the middle of the first year of primary education. Therefore, in this second analysis, dummy variables referring to the time of measurement were added to our regression model. Moreover, taking into account the difference in age between the Flemish and the Dutch children at the three measurement times – due to the different admittance policies in both countries, as explained above – age was included as a predictor in the second analysis, allowing a correction of
the results for the influence of the age on children’s scores at these distinct times of measurement.

So, in the second analysis, we extended the general regression equation that was used in the first analysis by modeling the time of measurement, using a dummy variable time2 (equaling 1 if the score is obtained at the second measurement occasion, 0 otherwise) and time3 (equaling 1 if the score is obtained at the third measurement occasion, 0 otherwise), as well as the interactions of the time of measurement with the variables country and age. This resulted in the following country-specific regression equations:

\[
\text{predicted ENT}(-F) = \gamma_{00} + \gamma_{02}(\text{form2})_{jk} + \gamma_{03}(\text{form3})_{jk} + \gamma_{10}(\text{age})_{ijk} \\
+ \gamma_{20}(\text{time2})_{ijk} + \gamma_{22}(\text{age} \times \text{time2})_{ijk} \\
+ \gamma_{30}(\text{time3})_{ijk} + \gamma_{32}(\text{age} \times \text{time3})_{ijk}
\]

and

\[
\text{predicted ENT} = (\gamma_{00} + \gamma_{01}) + \gamma_{02}(\text{form2})_{jk} + \gamma_{03}(\text{form3})_{jk} \\
+ (\gamma_{10} + \gamma_{11})(\text{age})_{ijk} + (\gamma_{20} + \gamma_{21})(\text{time2})_{ijk} \\
+ \gamma_{22}(\text{age} \times \text{time2})_{ijk} + (\gamma_{30} + \gamma_{31})(\text{time3})_{ijk} \\
+ \gamma_{32}(\text{age} \times \text{time3})_{ijk}
\]

for Flemish and Dutch children, respectively.

The estimated values of the parameters and the standard error of estimation are summarized in Table 3. As mentioned earlier, the estimated value of a parameter is statistically significant at a level of 5% if the absolute value of this parameter estimate is 1.96 times its standard deviation.

As shown in Table 3, the score on the ENT(-F) is related to the time of measurement: The score of the Flemish and the Dutch children on the ENT(-F), respectively the ENT, increases from the first to the second time of measurement as well as from the second to the third time of measurement. The predicted score on the ENT(-F) of Flemish children aged 74.68 months (the average age of all children – both Flemish and Dutch – involved in the analysis) is 22.70 at the first time of measurement, 24.64 at the second assessment, and 31.25 at the third time of measurement. Their Dutch peers obtain a (predicted) score of 26.10 on the ENT at the first assessment, 27.22 at the second time of measurement, and 30.86 at the third assessment. The evolution in score on the ENT(-F) is thus quite dissimilar for both countries:
There is a statistically significant interaction between the variables time and country, in the sense that the (predicted) score of the Flemish children on the ENT(-F) increases more from the first to the second and – especially – from the first to the third time of measurement than the (predicted) score of their Dutch peers on the ENT (Table 3). The initial performance of the Flemish children is worse than that of their Dutch peers, but due to their faster progression, in the end the former perform even slightly better than the latter. The evolution in the Flemish and Dutch children’s score on the ENT(-F) from the first to the second as well as from the second to the third time of measurement, is summarized in Figure 2.

On the basis of the results shown in Table 3 and in Figure 2, we can conclude that there is a considerable difference in the progression of the performance of Flemish and Dutch children on a test measuring their early numerical skills during the last year of preprimary education and the first months of primary education. On the one hand, by the middle of their last year of preprimary education, Dutch children score on average higher than their Flemish peers on this test. But, on the other hand, between that moment and the administration of the same test 1 year later, Flemish children have completely eliminated their arrears and even passed over their Dutch peers.

Furthermore, the results reported in Table 3 indicate that, in comparison with the first analysis, the effect of the variable age on the scores is smaller. However, age does still have a significant positive effect, which is the same in
both countries. This means that amongst children who received the same amount of instruction, older children perform better. However, this influence of the age seems to be smaller at the third measurement time, that is in the 1st year of primary education.

On the basis of the results shown in Table 3, we finally conclude that the form of the ENT(-F) does not influence the score of the children: None of the two coefficients of the dummies indicating the form of the ENT(-F) is statistically significant at the level of 5%. Furthermore, the results of additional analyses in which we tested the statistical significance of the two coefficients of the dummies together, indicated that the score of the children on the ENT(-F) did not differ significantly in function of the form offered ($\chi^2(2) = 0.80, p = .33$). This means that the three forms of the ENT(-F) can be considered equivalent. This is in line with the results of the first analysis.

**CONCLUSIONS**

In this article, we discussed the design and the results of our comparative study of the development of early numeracy in 5- to 7-year-old Flemish and Dutch children. The most important results of this study can be summarized as follows.

With respect to the psychometric qualities of the test, it can be concluded that both the ENT and the ENT(-F), are test instruments of moderate reliability. Moreover, they give a reasonably accurate description of a child’s mastery of general early numeracy.

With respect to the development of early numeracy in Flemish children compared to their Dutch peers, the present study yielded two remarkable results. First, at the first time of measurement that is in the middle of the year before the start of formal mathematics education – the Flemish children scored worse on the ENT(-F) than their Dutch peers on the ENT. Second, between that moment and the middle of their first year of formal mathematics education, the performance of the Flemish children on the ENT(-F) increased faster than the performance of their Dutch peers on the ENT. Importantly, neither of these results can be explained in terms of the age difference between Flemish and Dutch children at the different measurement times, since we corrected for the influence of the variable age. The difference in the rate of progression in performance between both countries, is most obvious between the second and the third time of measurement, that is between the end of the
last year of preprimary education and the middle of the first year of primary education. As a result of the latter, the initial difference in performance between the Flemish and Dutch children had completely disappeared by the middle of the first year of formal instruction.

The finding that at the first time of measurement the mastery of the early numerical skills was worse among children in Flanders than in The Netherlands is somewhat surprising, taking into account that at that moment most Flemish children had already been in a school for preprimary education for 2½ years, whereas most Dutch children had attended (pre)school for only 1½ year. On the other hand, referring to the abovementioned differences in the teaching and learning of early mathematical skills before the start of formal mathematical education in both countries, it could be argued that the higher score of the Dutch pupils at the moment of the first measurement could be due to the more purposive, more focused, and more systematic instructional attention to these early arithmetic skills in the first years of the Dutch elementary school compared to the way early mathematics education is conceived and handled in preprimary education in Flanders.

How can we explain the other major finding, namely that especially during the first half of the first grade of primary education, Flemish children progressed so well – in comparison to their Dutch peers – that by the time of the third administration of the ENT(-F) they had already overtaken the Dutch children? Opponents of the Dutch realistic approach to mathematics education for the primary school may argue that this result reveals the shortcomings of the too open and pupil-centered nature of the realistic approach to mathematics education compared to the more structured and more teacher-centered Flemish mathematics curriculum for the primary school (Feys, 1997). However, this is only one possible interpretation. Another candidate explanation is that the faster progression in performance of the Flemish pupils at the beginning of formal mathematics education in grade 1 is a delayed long-term effect of the well thought-out and balanced Flemish system of preprimary education (see below), rather than as (merely) the result of the quality of the Flemish mathematics curriculum for primary education (Laevers, personal communication). Finally, school effectiveness researchers may explain the greater progress of the Flemish children from the start of formal (mathematics) education in grade 1 in terms of cross-cultural differences with respect to

---

2Actually, in a recent offensive review of the Dutch realistic approach to mathematics education, Feys and Van Biervliet (2000) use the results of the present study in this way.
certain general characteristics of effective schools (e.g., ambitious goals, time-on-task, etc.) rather than in terms of differences with respect to the specific curricular domain of mathematics (Mortimore, 1998). Evidently, much more research is needed, first of all, to further document and unravel in greater detail the possible differences in the development of early numeracy between Flemish and Dutch children aged 5–7, and secondly, to relate these differences to different aspects of the educational and instructional environment at the different stages of their development in both countries.

More generally, the results of the present study can also be related to those of several recent international comparative studies, especially the Third International Mathematics and Science Study (TIMSS) (Beaton et al., 1996), showing that from the upper grades of the primary school on, Flemish children are ranked higher on a mathematics achievement test than their Dutch peers. In a recent study, Luyten (2000) found that Flemish pupils in the first grade of the secondary school still significantly overscored their Dutch peers when using test materials and procedures that were especially developed to optimally fit the (unique) characteristics and accents of the Dutch (realistic) mathematics curriculum. However, we are not aware of strong empirical evidence of the superiority of Flemish children on their Dutch peers before the age of 10. The present study can be considered as a first attempt into that direction. Its results seem to indicate that the difference does not originate before the start of formal mathematics instruction at the age of 6 – to the contrary: before that age Dutch children perform better than Flemish children – but does show up immediately after the start of formal mathematics education. However, this hypothesis requires more research, not only to document in a more fine-grained and more systematic way the observed cross-cultural differences in early numeracy between Flemish and Dutch children between the age of 5 and 7, but also to relate these cross-cultural differences in development to certain characteristics of education and schooling in general, and of mathematics education, in particular, in both countries. In this respect, we refer to a provocative recent report entitled The Early Years written by two British educationalists, Mills and Mills (1998), containing a comparative analysis of the objectives and organization of preprimary school in the UK, which obtained a disappointingly low rank in several recent international comparative studies such as the latest TIMSS, on the one hand, and three countries that have emerged as being particularly successful in recent international comparative studies, namely Hungary, German Switzerland, and Flemish Belgium, on the other hand. According to these authors, a common
characteristic of these three countries is that they have a preelementary school system that does not aim at immediate success by introducing 4–5-year-olds untimely into the formal skills of arithmetic, reading, and spelling, but that aims at preparing these children for effective formal learning by slowly realizing and consolidating the necessary preconditions for the formal learning and the rapidly escalating whole-class interactive teaching they will experience at primary school, such as attention, listening and memory skills, appropriate group behavior, and understanding of and confidence with tangible (mathematical) concepts, before moving on to the introduction of formal (mathematical) concepts and skills. Whereas this report provides some insights into the qualities of the Flemish system for preprimary school education, it remains unclear to what extent this contrastive analysis applies also to the comparison between Flanders and The Netherlands.

ACKNOWLEDGEMENT

The authors thank Prof. Dr. F. Laevers and the anonymous reviewers for their valuable comments on an earlier draft of this article.

REFERENCES


Example of an Item from Each of the Eight Subscales

Subscale 1: Concepts of comparison

You can see some indians here. Point out the indian who has less feathers than this indian with swift and arrow. (Test administrator points out the indian in the square up-left).

Subscale 2: Classification

You can see some people here. Point out all those persons that are wearing a hand-bag, but don’t have glasses.
Subscale 3: One-to-one correspondence

You see three buses here. (Test administrator points out the buses in the square up-left) **Point out the square that has a many dots as the number of buses.**

Subscale 4: Seriation

(Test administrator gives the child a worksheet and a pencil) **You can see some dogs here. Each dog goes to get a stick. A big dog goes to get a big stick and a small dog goes to get a small stick. Can you draw lines from the dogs to the sticks they are going to get?**
Subscale 5: The use of number words

*Count on from nine until fifteen: Six, seven, eight, . . .*

Subscale 6: Structured counting

(Test administrator puts 12 small blocks in a row at the table with a small space between the different blocks) *Count these blocks.* (The child is allowed to point to the blocks while counting them)

Subscale 7: Resultative counting

(Test administrator puts 19 small blocks at the table in an in-ordered way with a small space between the different blocks) *How many blocks are on the table?* (The child is not allowed to point to the blocks while counting them)

Subscale 8: General understanding of numbers

(Test administrator points out the picture with eight chickens) *A farmer has eight chickens. He buys two chickens more.* (Test administrator points out the picture with two chickens) *How many chickens does the farmer possess now? Point out the square with the right answer.* (Test administrator points out the lowest row of pictures)