ELSEVIER

Teaching and Teacher Education 21 (2005) 49–63

TEACHING AND TEACHER EDUCATION

www.elsevier.com/locate/tate

The development of mathematical competence in Flemish preservice elementary school teachers

Lieven Verschaffel[a,*], Steven Janssens[b], Rianne Janssen[c]

[a]Center for Instructional Psychology and Technology, University of Leuven, Vesaliusstraat 2, B-3000, Belgium
[b]Center for Research on Teaching and Training, University of Leuven, Belgium
[c]Research Group on Cognition and Individual Differences and Center for Educational Effectiveness and Evaluation, University of Leuven, Belgium

Abstract

A large-scale longitudinal study was conducted in which the elementary mathematical knowledge and skills of a large group of Flemish preservice elementary school teachers from 15 different institutes was assessed by means of a paper-and-pencil test that was administered both at the beginning and at the end of their 3-year training. The 30-items test covered the new standards for mathematics in the elementary school curriculum in Flanders. The test was divided in six subsets differing in terms of the curricular subdomain and of the cognitive operations being addressed by the item. The results confirmed the frequently heard concern that at the beginning of their training preservice elementary school teachers have rather weak mathematical competencies. At the end of their 3-year training, the overall test performance had become substantially better, although there were still reasons to be seriously concerned about the readiness of some student teachers to teach mathematics to elementary school children. A number of more specific comparisons helped to identify the relative role of two different factors—selection and instruction—in student teachers' gain from pretest to posttest. Besides documenting the development of elementary mathematical competence in preservice elementary school teachers, the study also resulted in an instrument for (self-) assessment of elementary school mathematical competence. This instrument is now being used in many institutes for elementary school teacher training in Flanders.

1. Theoretical background and research questions

According to Shulman's (Shulman, 1986; Shulman & Grossman, 1988) well-known theoretical model of domains of teachers' professional knowledge, teachers draw from several domains of knowledge as they plan and implement instruction. Three domains received a primary focus in his own research program: knowledge of subject matter, pedagogical content knowledge, and knowledge of learners' cognitions (see also: Berliner et al., 1988; Brown & Borko, 1992; Cooney, 1994; De Corte, Greer, & Verschaffel, 1996; Fennema & Loef

---

[*] Corresponding author. Tel.: +32 16 32 57 18; fax: +32 16 32 62 74.
E-mail address: lieven.verschaffel@ped.kuleuven.ac.be (L. Verschaffel).

0742-051X/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.tate.2004.11.005

Franke, 1992; Noddings, 1992). Applied to the domain of mathematics education, these three domains of knowledge can be described as follows.

First, subject-matter knowledge includes mastery of the key facts, concepts, principles and explanatory frameworks, procedures and problem-solving techniques and strategies within the given domain of instruction. Crucial in this respect is also the level of teachers' understanding of the domain. According to Ma (1999), a profound understanding of mathematics has three related meanings: deep, vast, and thorough. A "deep" understanding is one that connects mathematics with ideas of greater conceptual power. "Vast" refers to connecting topics of similar conceptual power. And "thoroughness" is the capacity to weave all parts of the subject into a coherent whole.

The second category of teachers' knowledge can be defined as "knowledge of subject matter for teaching" (Shulman, 1986, p. 9). It consists of an understanding of how to represent specific subject matter appropriately to the diverse abilities and interests of learners. It includes several subsystems such as knowledge of mathematics lesson scripts and mathematics teaching routines, knowledge about the kinds of problem types, graphical representations, etc. that are best suited to introduce particular mathematical notions and skills to pupils, and knowledge of instructional materials (textbooks, manipulatives, software, tests, etc.) available for teaching various mathematical topics.

Third, there is teachers' knowledge of how students think and learn with respect to mathematics. This third component consists of the teachers' knowledge of the mathematical concepts and procedures that students bring to the learning of a topic, the misconceptions and buggy procedures that they may have developed, and the stages of understanding and skill that they are likely to pass through in the course of gaining mastery of it.

As argued by many scholars, these three categories of knowledge should be considered parts of a larger, integrated functioning system in which each part is difficult to isolate from the other. Moreover, they are closely linked to other components of teachers' professional knowledge base, such as their domain-related beliefs about and attitudes toward their discipline (Fennema & Loef Franke, 1992; Thompson, 1992).

While agreeing fully with the well-established conclusion that in order to become a mathematics teacher student teachers need to develop both extensive subject-matter background as well as knowledge about pedagogical content and about students, the present study focuses on the first of these three categories of teachers' professional knowledge by assessing the mathematical competence of a large group of student teachers both at the beginning and at the end of their teacher training.

Our interest in the development of future teachers' mathematical content knowledge was raised by the apparent tension between the widely accepted claim that skillful teaching of a specific topic in mathematics requires that teachers themselves master that specific topic, on the one hand, and the numerous anecdotal observations and systematic research findings documenting that this basic requirement is not fulfilled among many preservice and in-service elementary school teachers, on the other hand.

There is empirical evidence to support the alarming statement about (future) teachers' insufficient mathematical competence. For instance, in the domain of multiplicative structures, Graeber and Tirosh (1988) studied preservice elementary school teachers' knowledge and skills with respect to multiplication and division involving decimals larger and smaller than 1. They found that a considerable number of preservice teachers made the same errors and shared the same misconceptions as observed in 10- to 12-year-olds. Similar disconcerting results have been reported by Post, Harel, Behr, and Lesh (1988). In another study that focused on the connectedness rather than the correctness of prospective teachers' knowledge of division, Simon (1993) found that their knowledge base was weak with respect to several types of connections, such as the conceptual underpinnings of the familiar algorithm of division, the relationship between partitive and quotative division and between division and subtraction, and the connection between symbolic division and real-world situations to which it is applicable.

Other studies have focused on the impact of teachers' mathematical knowledge on students' learning processes and outcomes. A number of large-scale correlational studies in which global measures of teachers' mathematical knowledge were used (such as the number of college mathematics courses completed or scores on standardized tests) did not show a strong relationship with students' learning processes and outcomes (Fennema & Loef Franke, 1992; Romberg & Carpenter, 1986). Small-scale studies in which teachers' content knowledge of particular topics (rather than a global measure of their mathematical knowledge) was related to the quality of their instructional actions on these topics, showed mixed results. While some of these studies failed to demonstrate that the lack of mathematical content knowledge leads to ineffective mathematics instruction (see, e.g., Leinhardt, Putnam, Stein, & Baxter, 1991; Schwartz & Riedsel, 1994), others found clear positive relations. As an example of the latter, we refer to a case study of one elementary school teacher by Fennema and Loef Franke (1992) showing how differences in his knowledge and understanding of two distinct mathematical topics—elementary addition and subtraction versus fractions—accounted for remarkable differences in the quality of teaching. Taken as a whole, the majority of the results are in favor of the old adage, which is also included in the recent publication *Adding it up* (Kilpatrick, Swafford, & Findell, 2001), namely that you cannot properly teach what you don't know yourself.

Based on the above research findings about the role of content knowledge in effective mathematics teaching, Brown and Borko (1992, p. 220) made a plea for a strong preparation in one's content area prior to teaching: "Without adequate content knowledge, student teachers spend much of their limited planning time learning content, rather than planning how to present the content to facilitate students' understanding. Student teachers with strong content preparation are more likely to be flexible in their teaching and responsive to students' needs, and to provide conceptual explanations, instead of purely procedural ones. They also tend to place greater emphasis on the organization and connectedness of knowledge within the discipline and less on the provision of content information. Student teachers without adequate content knowledge are likely to lack confidence in their ability to teach well."

Against this background, many institutes for the training of elementary school teachers all over the world spend a substantial part of the preservice training of their student teachers on the improvement of their mathematical knowledge and skills, although the content and the form of this training differs greatly from country to country. For instance, substantial differences exist in the amount of time that is spent on the teaching and learning of the elementary school mathematics and in the extent to which the teaching and learning of mathematics is integrated with the development of the student teachers' pedagogical content knowledge and of their knowledge of students' thinking and learning in mathematics (Borko & Putnam, 1996; Noddings, 1992).

The primary goal of the present study was to assess the mathematical competence of preservice elementary school teachers in Flanders at the time they started the teacher education program and to assess how this mathematical competence developed by the time they were finishing the program and were getting a qualification to teach mathematics at the elementary school level. A second goal of the study was to design a diagnostic instrument for (self-) evaluation of one's mastery of elementary school mathematics, which could become a helpful tool in the mathematics (education) part of the teacher-training program for future elementary school teachers in Flanders.

Before presenting the design and the findings of our study, some background information is given about the content and the organization of teacher education in Flanders, and of the courses in mathematics education in particular.

2. Becoming an elementary school mathematics teacher in Flanders: some background information

In Flanders elementary school teachers are trained in departments for teacher education, which belong to larger non-university institutes for higher education, and which provide a 3-year training to their students, most of whom have just finished secondary school when entering the institute. Depending on the age group for which they want to teach, these student teachers are streamed into three quite different sections, one of which is the section that prepares for teaching at the elementary school level. As Flemish elementary school teachers typically have to teach all subject matters, preservice teachers are trained in all curricular subjects, including mathematics. The theoretical part of their 3-year training consists of a mixture of general courses in education and psychology, on the one hand, and courses in specific subject-matter domains like mathematics, language, etc., on the other hand. This theoretical part is complemented with a practical part, consisting mainly of visits to schools, watching and discussing lessons given by expert teachers, and, last but not least, learning to teach themselves. The first year of training is primarily theoretical. The proportion of hours spent on theory decreases during the 3 years of the training, while gradually more time becomes available for practice. In the third year half of the time is devoted to practice.

As far as mathematics education is concerned, the three major components of professional domain-specific knowledge discussed above (i.e., mathematical competence, pedagogical content knowledge, and knowledge of students) are typically addressed in one course that is spread over the 3 years of training (with more time devoted to it in the first and the second year than in the third year). However, there are substantial differences between the institutes in terms of the relative proportion of instruction time that is devoted to each of these three components, the level of integration of these components, and, what exactly is being taught and learnt during this course (Verschaffel, 1999).

Unlike in many other countries, in Flanders there is no entrance exam or any other form of selection at the beginning of programs for higher education, including the training of elementary school teachers. Anyone who finished secondary school successfully can enter this teacher-training program. As a consequence, many students drop out during the first year of training or do not succeed in their exams. Only about 60% of those who started the training move up to the second year.

For more information about the goals, the content and the organization of elementary school teacher education in Flanders, we refer to the website of the Ministry of Education (http://www.ond.vlaanderen.be/english/educationinFlanders.pdf).

The low level of mathematical content knowledge and skills of students who want to become an elementary school teacher is increasingly being considered a major issue of concern among policy makers, curriculum developers, and teacher trainers involved in the training of future elementary school teachers in Flanders. This growing concern was the major reason to set up this study, in which the vast majority of Flemish institutes participated voluntarily and actively.

3. Method

3.1. Participants

Participants were 1475 preservice teachers who started the academic year 1997–1998 training to become an elementary school teacher in Flanders. These student teachers belonged to 15 institutes for teacher training. Given that the total number of institutes in Flanders was 18 and that the three institutes that did not participate in the study were among the smaller ones, it can be concluded that this study involved almost the whole population of students who started a training to become an elementary school teacher in Flanders that year. A paper-and-pencil mathematics test was administered to these 1475 student teachers during the first two weeks of the academic year.

At the end of the academic year 1999–2000 a parallel version of this pretest was administered to the third-year students of 11 of the 15 institutes for teacher training that participated in the first part of the study. The posttest was administered to only 11 of the 15 institutes because it was not possible to find an appropriate moment between the end of the theoretical courses (including mathematics) and the end of the school year to bring together all their third-year students to take the test in 4 of the 15 institutes. Because the reason for the dropout of these four schools was merely practical, there is no reason to assume that this dropout distorted the results either in a positive or negative way.

The total number of third-year students who completed the posttest was 534. Clearly, this number is much smaller than the 1475 first-year student teachers who completed the pretest. The great difference is partly caused by the above mentioned dropout of four institutes. More importantly, however, is the fact that 40–50% of the student teachers that started the training in 1997–1998 did not belong any more to the 1999–2000 population of last-year student teachers, either because they had failed in the first or in the second year of their training, or because they had spontaneously stopped their study for another reason. A third reason for the fact that the initial set of 1 475 participants was reduced to 534 is that a number of third-year student teachers were absent at the moment of posttesting, either because of illness or because they were involved in other curricular activities (e.g., an international exchange program).

For 368 of the 534 student teachers that were tested at the end of their third year of training, we were able to compare the first-year and third-year results at the individual level. The major reason why this individual comparison could not be done for all 534 third-year student teachers, was that both during the pretest and the posttest participants were strongly encouraged, but not forced, to identify themselves by writing their name on the top of the test sheet. While the majority of the participants did it, a substantial number did not. Another reason why we could not link the posttest results of all 534 third-year students with their results in the first year, was that some of the third-year students did not belong to the population of first-year student teachers being tested in the year 1997–1998, either because they were absent at the on the day of the pretest or because they had doubled the first or the second year of their training.

3.2. Materials

The starting point for the construction of the mathematical competence pretest and posttest were the new standards for elementary education that have become operational in the Flemish part of Belgium since September 1998 (Ministerie van de Vlaamse Gemeenschap, 1997). These standards cover all domains of the curriculum, including mathematics, and state the competencies that children should master at the end of elementary school.

Characteristic of these new standards for mathematics is that they are in line with the developments on the international scene as reflected in curriculum reform documents published over the past years in many countries such as the *Curriculum and Evaluation Standards for School Mathematics* (National Council of Teachers of Mathematics, 1989) in the United States and the *Proeve van een Nationaal Programma voor het Reken/Wiskundeonderwijs op de Basisschool* in the Netherlands (Treffers, De Moor, & Feys, 1989). Generally speaking, the Flemish standards are well in accordance with the new objectives and contents of mathematics education at the elementary school level as stated in the above mentioned documents, although they may go somewhat less far in emphasizing the constructive and realistic view on mathematics education and in de-emphasizing the place of some mechanistic and structuralistic elements of mathematics education than the above mentioned curriculum reform documents from the US and the Netherlands do (Verschaffel & De Corte, 1996).

The Flemish curriculum standards for mathematics education are officially classified into different categories. Starting from this classification, we decided to divide the standards into six subdomains that were formed by combining a content and a cognitive dimension. The content dimension divided the mathematical content into three categories: (a) numbers and arithmetic, (b) measurement, and (c) geometry. Because more than half of the standards refer to the content domain of number and arithmetic, we decided to combine the two other content domains (measurement and geometry) into one other domain. The cognitive dimension distinguished among three categories: (a) declarative knowledge, (b) procedural knowledge, and (c) strategies and problem solving skills. This resulted in a classification schema consisting of six subdomains.

In a next stage, test items were constructed for each of these six domains. A provisional version consisting of 30 items—five items per subdomain—was sent to the teacher trainers of all participating institutes together with a questionnaire, which consisted of a specific and a general part. In the specific part of the questionnaire the following questions were asked with respect to each of the 30 items: (a) Is this a valid item for measuring the corresponding standard of the elementary school mathematics curriculum? (b) What percentage of beginning student teachers do you expect to answer this item correctly? (c) What percentage of student teachers that have reached the end of their teacher training do you expect to answer this item correctly? (d) Do you have suggestions for improving the formulation and/or presentation of the item or its scoring criteria? The general part of the questionnaire involved some general questions about the test as a whole, such as the following: (a) Do you think that the level of difficulty of this test is appropriate for assessing the mathematical competence of (preservice) elementary school teachers? (b) Are there important curricular standards for elementary school mathematics that were not tested but that should be represented in the test to increase its content validity? (c) Do you agree with the length of the test? (d) How long will it take for beginning student teachers to complete the whole test?

Taking into account the teacher trainers' responses on the questionnaire, a second version of the mathematics test was constructed, consisting also of five items for each of the six subdomains giving a total of 30 items. Table 1 presents one item from each of the six subtests, together with the scoring criterion, while Enclosure 1 provides a brief description of all 30 test items. As can be seen, none of the items required mathematical knowledge or skills beyond the content of the mathematics curriculum for the elementary school in Flanders. Nevertheless, the test contained several items that demanded a good understanding of these elementary school mathematical notions and/or the application of problem-solving strategies for using these mathematical notions in context problems.

In a further stage of the project, we also constructed a parallel version of the first mathematics test. This parallel test—to be used as posttest—contained problems that were isomorphic to the problems from the pretest, but that were different in terms of superficial task characteristics (i.e., the concrete numbers used, the names of the persons and objects in the word problems, etc.). To control for the parallelism of both tests, a draft of the posttest was sent to teacher trainers from the 11 institutes that participated in the second part of the study with the request to assess the equality of the parallel items from both tests in terms of (a) the standard they aimed to test, (b) their level of difficulty, (c) their scoring criteria. The teacher trainers were also asked to make suggestions to improve the parallelism, if necessary. Based on the reactions of the teacher trainers, improvements were made leading to a posttest that mirrored the pretest, except for superficial task characteristics.

3.3. Procedure

Shortly before the start of the academic year 1997–1998 copies of the pretest were sent to the 15 participating institutes, together with specific instructions on how the test had to be administered to the student teachers and how the completed forms had to be returned to the researchers.

The pretest was administered in all institutes during the first two weeks of the academic year 1997–1998. The administration of the test took two time slots (= 100 min). This maximum duration was based on the teacher trainers' responses to the questionnaire.

At the beginning of the pretest session the teacher trainer introduced the test and motivated the student teachers to do their best. At the same time, it was stressed that the results would not be used for evaluative purposes within the context of their teacher training. As explained in the Method section, while student teachers were strongly invited to write their name on the test sheet (together with their gender and the name of their institute), it was stressed that identifying themselves was not obligatory.

At the end of the pretest session all copies were returned to the researchers who scored all test sheets according to strict criteria, leading to either

Table 1
Examples of items from the six different subdivisions of the mathematics pretest

	Numbers and arithmetic	Measurement and geometry
Declarative knowledge	What number is closest to 1.98? (a) 2.12 (b) 1.9 (c) 1.895 (d) 2.001	Right or wrong? Explain your answer • Every rhomb is a square. • Every square is a rhomb
Procedural knowledge	Solve by means of mental calculation: • 1743–997 = • $48 \times 25 =$	How much time is there between 10.47 a.m. and 1.22 p.m.?
Strategies and problem-solving skills	The baker needs 2.2 liters of cream. He only has little bottles of 1/8 liter cream. How many bottles does he need to open?	To paint a square-shaped ceiling with a side of 6 m, a painter needs 20 liters of paint. How many liters of paint will this painter need to paint a ceiling with the same shape but with a side that is double of 6 m?

1, $\frac{1}{2}$ or 0 points for each of the 30 items. Most items were scored dichotomously on correctness of the answer. For the other items credit was given to partially correct answers. These could either refer to a correct response on a subset of questions or problems that were framed within one item, or to a partially correct response to a simple item (like when solving correctly the first step of a multi-step word problem). As soon as the scoring was finished, an overview of the pretest results was sent to each institute. Each institute received an overview of the results on each item and on the test as a whole, together with an overview of the results for the whole group of 1 475 student teachers to allow a comparison of the results of one's own student teachers with a relative norm.

The organization and administration of the posttest was done in exactly the same way as the pretest.

4. Results

4.1. Results on the pretest

The mean score of the 1 475 first-year student teachers on the 30 items of the mathematics pretest was 17.8; the standard deviation was 4.83. The test scores followed a rather normal distribution with scores ranging between 5.5 and 30.

Table 2 gives the means and standard deviations for the six subtests. These data indicate that the results for the numbers and arithmetic (9.14 on 15) were somewhat better than for measurement and geometry (8.67 on 15), and that the scores for declarative and procedural knowledge (respectively, 6.16 and 6.89 on 10) were considerably better than for strategies and problem solving (4.77 on 10).

Within each subtest, we observed big differences in terms of item difficulty. While some items from the pretest showed a mean score of more than .80, the mean score of other items was less than .30. Among the easiest items were: item 4 (.85), item 6 (.84), item 8 (.78), item 22 (.82), and item 23 (.82), while the most difficult items were: item 5 (.24), item 14 (.32), item 15 (.35), item 28 (.31) and item 29 (.18) (for a description of the items, see Enclosure 1).

An analysis of variance (ANOVA) was performed with the student teacher's gender and their teacher-training institute as independent variables, and their result on the pretest as dependent variable. There was a significant effect of gender ($F[1, 1446] = 57.61$, $p < .0001$) explaining 4% of the variance in the results: Male student teachers (19.45) scored higher than female students (17.18). There was also an effect of institute ($F[14, 1446] = 5.64$, $p < .0001$), which was responsible for 6% of the variance: there was a difference

Table 2
Means (and standard deviation) on the six subtests of the mathematics pretest

	Declarative knowledge	Procedural knowledge	Strategies and problem-solving skills	Total
Numbers and arithmetic	3.24 (1.03)	3.29 (1.11)	2.61 (1.25)	9.15 (2.68)
Measurement and geometry	2.91 (1.02)	3.60 (.94)	2.16 (1.34)	8.68 (4.83)
Total	6.16 (1.68)	6.89 (2.22)	4.77 (4.83)	17.82 (4.83)

of 3.73 points between the institute with the lowest (15.96) and the institute with the highest (19.69) score. Taking into account that the pretest was administered in the very beginning of the academic year, these differences could, of course, not be due to characteristics of the teacher training provided in these distinct institutes.

With respect to the psychometric characteristics of the pretest, internal consistency was determined using Cronbach's α, which gives an estimate of the lower bound of the reliability of a test. The α coefficient was .83, which is marginally below the limit of .85 that is typically requested for that kind of tests (Eggen & Sanders, 1993). Removal of items did not result in a significant increase of the internal consistency of the test. To further evaluate the construct validity of the test, a factor analysis was performed. This factor analysis revealed that the test scores were basically determined by one single factor, which could be defined as "general competence in elementary school mathematics" and which accounted for 52% of the variance. So, the six different subdivisions of test in terms of content and cognitive activity did not measure different psychological constructs.

4.2. *Results on the posttest*

529 of the 534 third-year student teachers from the 11 remaining institutes for teacher training involved in the posttest completed all items of the test, and could, thus, be included in the analysis. A descriptive analysis of the results of these 529 student teachers revealed a mean score on the posttest of 24.09 on 30 and a standard deviation of 3.57. This mean score was 6.27 points higher than the pretest mean score (17.82). The student teachers' score showed a slightly negatively skewed distribution with scores ranging between 10.5 and 30.

Table 3 gives the means and standard deviations for the six subtests of the posttest. As for the pretest, the results for the three subtests on numbers and arithmetic (12.37 on 15) were somewhat better than for the three subtests on measurement and geometry (11.72 on 15). The subtests measuring declarative and procedural knowledge (7.95 and 8.97 on 10, respectively) yielded again a somewhat higher score than the subtest about strategies and problem-solving skills (7.26 on 10).

Within each category, there were big differences in terms of problem difficulty, although these differences were less extreme than during the pretest. Compared to the pretest, the number of items with a mean score higher than .80 had increased from five to 18, while only one item, namely item 26, elicited a mean score lower than .50 (compared to nine pretest items).

As for the pretest data, an ANOVA was performed with the student teachers' gender and their teacher-training institute as independent variables and their score on the posttest as dependent variable. There was again a significant effect of gender ($F[1, 508] = 14.56$, $p < .0001$), explaining 2% of the variance in the results: Male student teachers (25.01) scored still higher than female students (23.95), but the difference had become smaller. We also observed an effect of institute ($F[10, 508] = 9.84$, $p < .0002$), which was responsible for 15% of the variance in the results.There was a difference of 6.88 points between the institute with the lowest mean score (19.90) and the institute with the highest mean

Table 3
Means (and standard deviation) on the six subtests of the mathematics posttest

	Declarative knowledge	Procedural knowledge	Strategies and problem-solving skills	Total
Numbers and arithmetic	4.16 (.76)	4.38 (.69)	3.95 (.93)	12.37 (2.11)
Measurement and geometry	3.79 (.87)	4.59 (.57)	3.33 (1.31)	11.72 (2.21)
Total	7.95 (1.28)	8.79 (1.01)	7.26 (1.90)	24.09 (3.57)

score (26.78). It is not surprising that the variance explained by this latter factor was much greater for the posttest than for the pretest results. At the moment of the pretest the students had just entered their institute and, therefore, the institutional factor could not yet have had a significant impact on the development of their mathematical knowledge and skills as measured by the test.

With respect to the psychometric characteristics of the posttest, the internal consistency was determined using Cronbach's α. This coefficient was .75, which is smaller than the .83 obtained for the pretest. This decrease in internal validity from pretest to posttest is probably due to the considerable decrease in the variance of the results from pretest to posttest. Removal of two items led to a very small increase in the internal consistency of the test. With respect to construct validity, we performed again a factor analysis, which revealed that the posttest scores were also determined by one factor, which could be defined as "general competence in elementary school mathematics" and which accounted for 42% of the variance in posttest results (versus 52% explained variance for the pretest).

4.3. Comparison of the results of the student teachers being tested twice

The general comparison of the mean scores during the pretest and posttest reported in the two previous sections suggests that the 3 years of teacher training had a significant and beneficial impact on the student teachers' competence in elementary mathematics. However, the conclusion is jeopardized by the fact that the results on the posttest are coming from a rather small, and more importantly, a selective subset of the 1 475 student teachers being pretested. The most important reason why only a small subset of students being pretested also did the posttest, was that about 40% of those being pretested failed at the end of the first year of their preservice teacher training and, therefore, either left the institute or subscribed for the first year again next year. So, it could be argued that the difference of 6.27 points between pretest and posttest was mainly, if not exclusively, due to the fact that the posttest was only administered to the better students from the group of 1 475 being pretested. In a first attempt to disentangle the two possible explanations for the gain of 6.27 points between pretest and posttest—namely: the selection effect versus the actual effect of the 3 years of teacher training—we performed a new ANOVA involving only those 368 student teachers of whom we could directly compare their pretest and posttest data. Because only student teachers who reached the end of their teacher training were involved in this analysis, gains in test results from pretest to posttest could no longer be due to the selection factor. This new ANOVA had a split-plot factorial design with as independent variables the between-subject variables gender and institute and the within-subject variable test moment (pretest versus posttest), and as dependent variable the student teachers' total test score.

First, and most importantly, this ANOVA revealed a main effect of test moment ($F[1,346] = 260.73$, $p < .0001$): The mean score of the 368 student teachers increased between pretest and posttest from 19.13 (SD: 4.53) to 24.07 (SD: 3.74). Whereas this increase of 4.90 points is smaller than the pretest/posttest difference of 6.27 points reported earlier, it remains a highly significant increase. Taking into account that the selection effect can no longer account for this difference of 4.90 points, we now can conclude that the 3 years of teacher training had indeed a beneficial effect on the development of the elementary mathematical competencies of student teachers. However, it should be acknowledged that the design of the present study does not permit us to conclude that this increase of 4.90 points on an elementary mathematics test was due to the teacher training program *alone*, especially because of the absence of a control group of post-secondary school students with similar mathematical skills as the first-year student teachers in our study but who did not follow courses in elementary mathematics (education) between the administration of the pretest and the posttest.

Second, there was a significant main effect of gender ($F[1,346] = 21.64$, $p < .0001$) and of institute ($F[10,346] = 4.22$, $p < .0001$). However, more important than these two main effects (which only confirm what we have already been reported about the role of these two factors) are the interaction effects between each of these two factors and the factor test moment. There was a significant interaction effect of gender and test moment ($F[1,346] = 14.86$, $p < .0001$). A closer look at the means revealed that this interaction effect was due to the fact that the observed difference between males and females during pretest (21.54 versus 18.43) had become significantly smaller during posttest (24.70 versus 23.89). There was also a significant interaction between institute and test moment ($F[10,346] = 5.89$, $p < .0001$), revealing remarkable differences in gains from pretest to posttest between institutes. For instance, while the score in one institute increased from 17.33 to 24.41, the mean score in another one increased only by about 1.5 (i.e., from 18.31 to 19.92).

Finally, an analysis of the correlations of the pretest and posttest results of the student teachers of each of the 11 institutes involved in this comparative analysis yielded significant correlations ranging between .36 and .78, except for one institute where a correlation of −.07 was found.

4.4. Comparison of the pretest results of the student teachers who did and who did not participate in the posttest

To allow an appropriate interpretation of the pretest and posttest results reported before, we finally compared for the 11 institutes that were involved in both test administrations the results on the pretest of the first-year students who did not participate in the posttest (mainly because they were not third-year students at the time of the posttest) versus of those who did. So, while the first group (Group 1, $n = 590$) involves (mainly) students who started a training to become an elementary school teacher but who failed in the first year of their training, the latter group (Group 2, $n = 368$) consisted of student teachers who started their training at the same time and in the same institutes as Group 1 student teachers, but who did succeed in completing their teacher training within the regular period of 3 years. The major purpose of this final comparative analysis was to check whether the mean score of the two groups was indeed different in favor of Group 2. This finding would yield additional empirical evidence in favor of our claim that the difference of 6.27 points between the overall results on the pretest and the posttest was at least partly due to a selection effect.

We performed an ANOVA with group (Group 1 versus Group 2), gender, and institute as independent variables and student teachers' total test score on the pretest as dependent variable The ANOVA revealed a significant main effect for each the three factors involved in the analysis, but no interaction effects. The effect group was significant at the 1% level ($F[1,912] = 37.51$, $p < .0001$), explaining 4% of the variance. There was a difference of 2.18 between the pretest scores of both groups (16.95 for Group 1 versus 19.13 for Group 2). This finding is in line with our expectation that student teachers who succeeded at the end of the first year would perform significantly better on an entrance test assessing their elementary mathematical knowledge and skills than their peers who failed at the end of the first year. The main effects of gender ($F[1,912] = 41.56$, $p < .0001$) and institute ($F[10,912] = 3.35$, $p < .0002$) were also significant, confirming the findings about the impact of these two factors on first-year student teachers reported before.

4.5. Comparison of the pretest and posttest results with the predictions of the teacher trainers

As explained in the Method section, the questionnaire that was sent to the teacher trainers of the 15 participating institutes in the stage of the construction of the pretest contained two questions in which the teacher trainers were asked to indicate for each item the percentage of Flemish student teachers who would solve the item correctly at the beginning of their training and at the time of their graduation.

An analysis of these predictions revealed that the actual mean pretest score of 17.82 on 30 (= 59%) was slightly better than the mean score of 51% predicted by the teacher trainers. Anyhow, this finding indicates that Flemish teacher trainers certainly do not underestimate the weakness of the mathematical content knowledge of their incoming students.

The actual overall posttest score of 24.09 (= 80%) was lower than the 90% mean score that was predicted by the teacher trainers. So, whereas the teacher trainers of the participating institutes slightly underestimated the mathematical competencies of their student teachers at the start of their 3-year training, they tended to overestimate somewhat the competencies of their graduates.

5. Conclusion and discussion

Starting from the state-of-the-art in the international research literature on preservice and in-service teachers' insufficient mastery of one the major components of their domain-specific professional competence, namely their mastery of the content to be taught to their students, and its relationship with classroom practice, a longitudinal study was set up in which we assessed the elementary mathematical content knowledge and skills of a large group of Flemish preservice elementary school teachers at the beginning and at the end of their teacher training.

Taking into account the new Flemish standards for elementary school mathematics, a pretest and a parallel posttest were constructed consisting of 30 items divided in six subtests of five items each, representing the major categories of these standards. Although none of the items required mathematical knowledge or skills beyond the content of the mathematics curriculum for the elementary school in Flanders, the test contained several items that demanded a thorough understanding of certain mathematical notions and/or the application of problem-solving strategies for using these mathematical notions in context problems.

The results of the pretest confirmed the frequently heard concerns about the problematic level of mathematical competence of students who want to become an elementary school teacher given the low overall mean score as well as the detailed results for some very difficult items and for some very low performing subjects. The comparison of the actual mean pretest score and the score predicted by the teacher trainers indicated that Flemish teacher trainers certainly do not underestimate the weakness of the mathematical content knowledge of their incoming students. As far as implications for teacher training are concerned, the pretest results support the common practice in Flanders of spending a lot of the instructional time available for the mathematics education of future elementary school teachers—typically up to 40–60% of the available time and sometimes even more (Verschaffel, 1999)—on improving student teachers' mathematical competence prior to or together with the development of their pedagogical content knowledge and of their knowledge of student cognitions.

Although the posttest results were considerably better than those for the pretest, the overall mean score was still not more than 24.09 on a total of 30, and the detailed results revealed that even at the very end of their teacher training there were still remarkable low scoring items and participants. The latter finding raises doubts about the readiness of all Flemish student teachers who have come at the end of their training to teach mathematics to

6- to 11-year-old children in line with the new standards for mathematics, which put a strong emphasis on understanding and application. In this respect, it is interesting to recall that whereas the teacher trainers of the participating institutes seemed to underestimate slightly the mathematical competencies of their student teachers at the beginning of their 3-year training, they tended to overestimate somewhat the competencies of their graduates.

The additional analysis of the test results involving only those students who participated in both tests revealed that the difference of 6.27 points between pre-test and posttest was the result of the interplay of two factors, namely (a) the severe selection criteria at the end of the first year of the teacher training and (b) the impact of the 3 years of teacher training. The comparison of the gain from pretest to posttest of all participants being tested either during pretest or during posttest (6.27) with the gain of those participants who were tested during pretest and posttest (4.94) suggested that the second of these two explanatory factors, namely the 3 years of teacher training, was the most important one. Further evidence for the fact that the selection factor also contributed to the difference of 6.27 points, was provided by the substantial difference in the total score on the pretest between the group of student teachers who had failed at the end of first year of training and the group of student teachers who were allowed to continue their training and could therefore be posttested two years later (16.95 versus 19.13, respectively). Unfortunately, the design of the present study does not allow a more fine-tuned analysis of the relative contribution of the selection factor and the instruction factor, and even leaves open the possibility that other factors besides these two contributed to the observed gain in test scores between pretest and posttest.

The substantial differences in test score gain from pretest to posttest between the 11 institutes for teaching training that participated in the study suggest that these institutes were not equally successful in developing the elementary mathematical competencies of their student teachers. Although the questionnaires that were sent to the teacher trainers at the time of the pretest and the posttest yielded valuable information about certain characteristics of the content and organization of mathematics education in the participants institutes, its nature did not allow a systematic, quantitative analysis of the relation between features of the teacher training program in the participating institutes, on the one hand, and the mathematical learning gains obtained by their graduates, on the other hand. Additional research is needed to further unravel what characteristics of the teacher-training program in general, and of the specific mathematics (education) courses in particular, are decisive for the development of the elementary mathematical competence of preservice teachers.

Beside documenting the development of mathematical content knowledge and skills of preservice elementary school teachers in Flanders, the present study also resulted in two parallel versions of an instrument that is useful for the (self-) assessment of student teachers' mastery of the mathematical content they will have to teach after their graduation. Although the psychometric qualities of the test are not totally convincing, the test as a whole proved to be a valuable instrument. Actually, in response to one of the items in the questionnaire, most of the teacher trainers declared that they would continue to use the test, or possibly a shortened one-hour version of it, in the coming years either to assess the entrance level and the progress of mathematical content knowledge of their students, or to assist their student teachers in the self-assessment of (the development of) that level. We are well aware that in order to become a valuable tool for such continuous (self-) assessment, some adaptations of the existing testing materials and some additional elements are required. First, the 30-items test, which covers the whole elementary school curriculum, should be complemented with additional diagnostic test items allowing a more fine-grained and a more process-oriented analysis of the strengths and the weaknesses in student teachers' mathematical content knowledge. Second, rather than just providing scoring criteria and rules for computing (sub) totals, more informative forms of feedback for teacher trainers and/or student teachers should be provided, including hints towards appropriate teaching and/or learning materials. Third, to allow a regular follow-up of the development of student teachers' mathematical competence, there is a need for an (electronic) item bank from which both general and specific calibrated tests can be assembled on the spot, rather than of merely two parallel versions of a paper-and-pencil test. Currently, we are planning, in close collaboration with teacher trainers from the institutes who participated in the present longitudinal investigation, a follow-up study aimed at the development of such a computer-based instrument for continuous (self-) assessment of the mathematical content knowledge and skills of preservice teachers.

Acknowledgement

The authors want to thank the teacher trainers of all institutions for teacher training involved in the study for their help in the data collection as well as Catherine de Rijdt and Bart Bunkens for their assistance in the data collection and data analysis.

Enclosure 1: Short description of the 30 test items[1]

Number and arithmetic: Declarative

1. What digit represents the ten thousands, and what digit represents the hundredths in the number 23 654.917
2. Which two numerical sentences produce the same result: $2+4\times6+2$ $(2+4)\times6+2$ $2+(4-6)+2$ $2+4\times(6+2)$
3. What number is closest to 1.98? 2.12 1.9 1.895 2.001
4. Order these fractions from smallest to largest: 3/2 1/4 1/8 3/4 1/2 3/8
5. Define and illustrate the smallest common multiple of two numbers.

Number and arithmetic: Procedural

6. Pete gets 4/5 of a bottle of wine of .75 cl and Ann gets 5/6 of a similar bottle of wine of .75 cl. Who gets most?
7. Last year's school party resulted in a profit of 25.000 BEF. This year the profit was 31.000 BEF. With what % has the profit increased since last year?
8. Solve by means of mental calculation: $1743-997=$ $48\times25=$
9. Compute by means of written computation: $979.3\times2.75=$ $14004:16=$
10. Yvonne computed on her pocket calculator 835–4.345–7.795. When writing the result of this computation in her book, she forgot to copy the decimal point and wrote: 822286. Where should she put the decimal point?

Number and arithmetic: Strategies and problem-solving skills

11. At the fair John and Mary spent together 780 BEF. John spent 120 BEF more than Mary. How much did each of them spend?
12. The baker needs 2.2 liters of cream. He only has little bottles of 1/8 liter cream. How many bottles does he need to open?
13. [A copy of a book page is given.] About how many words are there in a book of 153 pages? About 3000 About 30 000 About 300 000 About 3 000 000?
14. [A drawing of a couple of domino stones is given.] A domino game consists of stones divided in two parts each part showing either 0, 1, 2, 3, 4, 5 or 6 dots. How many different domino stones are there in the game if all possible combinations are available?
15. [A picture of a paperclip together with a scale is given.] This picture of a paperclip is first reduced by 40% and afterwards the decreased figure is again enlarged by 40%. Indicate the correct response: the final picture is larger than/smaller than/equal to the original picture?

---

[1] The items given in this enclosure are the pretest items. The posttest involved parallel items of these 30 items.

Measurement and geometry: declarative

16. [A picture of a road map with a scale of 1/200 000 is given.] What does the fraction 1/200 000 mean? What is the length of a road of 1 km on the map?
17. Right or wrong? Explain your answer. Every rhomb is a square. Every square is a rhomb.
18. Complete the empty box with one of the following measures: mm, cm, dm, mm^2, cm^2, dm^2, mm^3, cm^3, dm^3: the surface of this page is about 6.2 …
19. [A time × distance diagram representing the course of the running competition between three runners is given.] Write a brief report (in maximum 3–4 sentences) about the course and the outcome of this running competition.
20. One hectare is … m^2

Measurement and geometry: procedural

21. What's the volume in cm^3 of a bottle of drugs of 200 cl?
22. How much time is there between 10.47 a.m. and 1.22 p.m.
23. [Two beam-shaped boxes are given together with the necessary measures.] What's the volume of box A and box B?
24. [A picture of two adjacent pieces of land—one in the form of a triangle, one in the form of a parallelogram—are given together with the necessary measures.] What's the area of these two pieces of land?
25. [A still life consisting of a three objects—a box, a candle, a puppet—is given together with four pictures taken from four different perspectives.] What picture corresponds with which camera perspective?

Measurement and geometry: strategies and problem-solving skills

26. What is the largest rectangular figure with a perimeter of 12 cm that you can draw?
27. To paint a square-shaped ceiling with a side of 6 m, a painter needs 20 liters of paint. How many liters of paint will this painter need to paint a ceiling with the same shape but with a side that is double of 6 m?
28. Two cyclists make a trip of 60 km. The first one rides the first half of the distance at a speed of 40 km per hour and the second part at a speed of 20 km per hour. The second cyclist rides the first half as well as the second half at a speed of 30 km per hour. Who will arrive first?
29. [Two different views on a 20 × 25 m swimming pool with a gradually changing depth from 1 meter until 3 meter are given.] What is the volume of water in this swimming pool when it is completely filled?
30. [One three-dimensional picture of a cube with a black line dividing the four side walls of the cube is given, together with four two-dimensional construction maps of the cube on which the black line is partially drawn and partly missing.] Complete the black line segments in each of the four construction maps.

References

Berliner, D. C., Stein, P., Sabers, D., Clarridge, P., Cushing, K., & Pinnegar, S. (1988). Implications of research on pedagogical expertise and experience for mathematics teaching. In Grouws, D. A., Cooney, T. J., & Jones, D. (Eds.), *Perspectives on research on effective mathematics teaching* (pp. 67–95). Reston, VA: National Council of Teachers of Mathematics.

Borko, H., & Putnam, R. (1996). Learning to teach. In Berliner, D. C., & Calfee, R. C. (Eds.), *Handbook of educational psychology* (pp. 673–708). New York: MacMillan.

Brown, C. A., & Borko, H. (1992). Becoming a mathematics teacher. In Grouws, D. A. (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 209–242). New York: MacMillan.

Cooney, T. J. (1994). Research and teacher education. In search of common ground. *Journal for Research in Mathematics Education*, *25*, 608–636.

De Corte, E., Greer, B., & Verschaffel, L. (1996). Mathematics teaching and learning. In Berliner, D. C., & Calfee, R. C. (Eds.), *Handbook of educational psychology* (pp. 491–549). New York: MacMillan.

Eggen, T., & Sanders, P. (1993). *Psychometrie in de praktijk. [The practice of psychometrics]*. Arnhem, The Netherlands: CITO.

Fennema, E., & Loef Franke, M. (1992). Teachers' knowledge and its impact. In Grouws, D. A. (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 147–164). New York: MacMillan.

Graeber, A., & Tirosh, D. (1988). Multiplication and division involving decimals: Preservice elementary teachers' performance and beliefs. *Journal of Mathematical Behavior*, *7*, 263–280.

Kilpatrick, J., Swafford, J., & Findell, B. (Eds.). (2001). *Adding it up*. Washington, DC: National Academy Press.

Leinhardt, G., Putnam, R. T., Stein, M. T., & Baxter, J. (1991). Where subject knowledge matters. In Brophy, J. E. (Ed.), *Advances in research on teaching: Teachers' subject matter knowledge and classroom instruction*, Vol. 2 (pp. 87–113). Greenwich, CT: JAI Press.

Ma, L. (1999). *Knowing and teaching elementary mathematics: teachers' understanding of fundamental mathematics in China and the United States*. Mahwah, NJ: Erlbaum.

National Council of Teachers of Mathematics (1989). *Curriculum and evaluation standards for school mathematics*. Reston, VA: National Council for Teachers of Mathematics.

Ministerie van de Vlaamse Gemeenschap (1997). *Basisonderwijs. Ontwikkelingsdoelen en eindtermen. [Developmental goals and standards for elementary education.]*. Brussel: Ministerie van de Vlaamse Gemeenschap, Departement Onderwijs.

Noddings, N. (1992). Professionalisation and mathematics teaching. In Grouws, D. A. (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 97–208). New York: MacMillan.

Post, T. R., Harel, G., Behr, M., & Lesh, R. (1988). Intermediate teachers' knowledge of rational number concepts. In Fennema, E., Carpenter, T., & Lamon, S. J. (Eds.), *Integrating research on teaching and learning mathematics* (pp. 92–131). Madison, WI: Wisconsin Center for Education Research, University of Wisconsin.

Romberg, T., & Carpenter, T. (1986). Research on teaching and learning mathematics: two disciplines of scientific inquiry. In Wittrock, M. C. (Ed.), *The third handbook of research on teaching* (pp. 850–873). New York: MacMillan.

Schwartz, J. E., & Riedsel, C. A. (1994). The relationship between teachers' knowledge and beliefs and the teaching of elementary mathematics. Paper presented at the annual meeting of the American Association of College for Teacher Education, February 16–19, Chicago, IL.

Shulman, L. S. (1986). Paradigms and research programs in the study of teaching: A contemporary perspective. In Wittrock, M. C. (Ed.), *Handbook of research on teaching*, (3rd ed.) (pp. 3–36). New York: MacMillan.

Shulman, L. S., & Grossman, P. L. (1988). *Knowledge growth in teaching. A final report to the Spencer Foundation*. Stanford, CA: Stanford University.

Simon, M. A. (1993). Prospective elementary teachers' knowledge of division. *Journal for Research in Mathematics Education*, *24*, 233–254.

Thompson, A. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In Grouws, D. A. (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127–146). New York: MacMillan.

Treffers, A., De Moor, E., & Feys, E. (1989). *Proeve van een nationaal programma voor het reken/wiskundeonderwijs op de basisschool. Deel 1. Overzicht einddoelen. [Towards a national curriculum for mathematics education in the elementary school. Part 1. Overview of the goals]*. Tilburg, The Netherlands: Zwijssen.

Verschaffel, L. (1999). Rekenen en wiskunde in de opleiding van leraren basisonderwijs in Vlaanderen. [Mathematics education in elementary teacher training in Flanders.]. In Dolk, M., Janssen, H., Letschert, J., Nienhaus-Van Lint, A., & Oonk, W. (Eds.), *Verhalen van een leven lang ontwerpen* (pp. 31–58). Enschede, The Netherlands: Stichting voor Leerplanontwikkeling.

Verschaffel, L., & De Corte, E. (1996). Number and arithmetic. In Bishop, A. J., Clements, K., Keitel, C., Kilpatrick, J., & Laborde, C. (Eds.), *International handbook of mathematics education*, Vol. 1 (pp. 99–137). Dordrecht, The Netherlands: Kluwer.