ORIGINAL ARTICLE

Kindergartners' performance in length measurement and the effect of picture book reading

Marja Van den Heuvel-Panhuizen · Iliada Elia

Abstract This paper addresses: firstly, kindergartners' performance in length measurement, the components of their performance and its growth over time; secondly, the possibility to develop kindergartners' performance in length measurement by reading to them from picture books. To answer the research questions, an experiment with a pretest–posttest experimental control group design was carried out involving nine experimental classes and nine control classes. The children in the experimental group participated in a 3-month picture book program that, among other things, spotlighted the measurement of length situated in meaningful contexts. Before and after the intervention, the children's performance in length measurement was assessed in both groups. The responses of 308 kindergartners (4- to 6-year-olds) from two kindergarten years (K1 and K2) were analyzed. Analysis of the pretest data showed that the measurement tasks included in the test were not easy to solve. However, the children belonging to K2 did better than the younger children belonging to K1. Within children's performance, three components could be identified: holistic visual recognition, ordering and unitizing. Finally, the effect of the intervention was investigated by comparing the performances of the experimental and control group in the pretest and the posttest. We found a weak but significant effect of reading picture books to children on their general measurement performance. However, this effect was only found for K1 children on the component of holistic visual recognition.

Keywords Kindergartners · Picture books · Mathematics · Measurement tasks · Assessment

M. Van den Heuvel-Panhuizen (✉) · I. Elia
Freudenthal Institute for Science and Mathematics Education,
Utrecht University, P.O. Box 85170,
3508 AD Utrecht, The Netherlands
e-mail: m.vandenheuvel@fi.uu.nl

I. Elia
Department of Education, University of Cyprus,
P.O. Box 20537, 1678 Nicosia, Cyprus

1 Introduction

1.1 The role of length measurement in developing measurement understanding

Measurement as a mathematical competence refers to the ability to assign a numerical value to a measurable attribute of an object or an event. The numerical values emerge by identifying how many times a particular unit representing that attribute is 'in' the object or event. This measurement can involve a broad variety of attributes, for example, weight, area, duration, volume and length. In addition, measurement also includes composed attributes such as speed and density.

Of all attributes, length can be considered as the most elementary one. It has a kind of universal character, since several other attributes can be converted into length. Think, for example, of the length of a spring that indicates the weight of an object and a measuring strip that indicates the water volume in a water heater. In a way, the elementary nature of length is supported by Curry's and Outhred's (2005) finding of a parallel development in understanding linear measurement and volume measurement based on filling, which allows the 'linear' reading off of volume. Another reason for the elementary nature of length is that it is a very accessible attribute for children. Length is very often connected to questions that emerge in their free play.

Moreover, children have means available to answer these questions about length in a natural way.

Because of the basic character of length measurement, it is no wonder that a good foundation on it is generally considered as a necessary condition for other forms of measurement (e.g., Curry & Outhred, 2005). This central role of length measurement in measurement gives rise to the need to explore and extend our knowledge about the development of young children's understanding of length measurement.

1.2 Development of the understanding of length measurement

Measurement as a mathematical domain, including measurement of length, is incorporated in many kindergarten curricula (e.g., Board of Studies NSW, 2006; NCTM, 2000; Van den Heuvel-Panhuizen & Buys, 2008). Teaching length measurement in kindergarten, however, should not begin with an assumption of a blank slate. Before entering school, children have already developed some knowledge of measurement through playful activities. In kindergarten, the teaching builds on this informal knowledge and offers children meaningful situations in which they can extend their understanding of length measurement. This view of the supportive role of intuitive and informal knowledge when learning mathematics and a recognition of the importance of a meaningful context in establishing mathematical thinking (e.g., Hughes, 1986) are widely accepted in current theories on learning and teaching mathematics (Bransford, Brown, & Cocking, 2000).

Children move through several stages in learning length measurement by the end of the lower primary grades. Early learning begins with the ability to make qualitative comparisons and to order objects by their length. The next major advancement involves the development of the ability to quantify length by assigning a numerical value to it. In the final stage, children learn to use measurement instruments, such as a ruler.

One of the first indications that children are developing a qualitative understanding in measurement is their use of measurement-related words, such as "big" and "small" (Clarke, Cheeseman, McDonough, & Clarke, 2003). Next, children develop the ability to compare the length of two objects by placing objects physically next to each other or by visual comparison (Boulton-Lewis, Wills, & Mutch, 1996; Outhred, Mitchelmore, McPhail, & Gould, 2003). Children are already able to perform this direct comparison of objects at a kindergarten age (Barrett, Jones, Thornton, & Dickson, 2003; Clarke et al., 2003). A more demanding activity involves indirect comparison. For example, asking children to determine whether a table that is positioned in another room can go through a particular doorway (Sarama

& Clements, 2008). To solve this problem, a child might use a rope as a mediator to represent either the table's or the door's dimension.

A further step in children's early development of understanding length measurement is the ordering of objects with respect to their length. Here, ordering is based on repetitive comparison. Research findings show that this ordering ability can be reached by the age of 5 or 6 years. Clarke et al. (2003) reported that 90% of the kindergartners[1] in their study were capable of ordering objects with respect to length. In the study by Outhred et al. (2003), the initial success rate on ordering tasks was below 50%, but after a teaching sequence an increase of more than 30 percentage points was found. This result suggests that children's ability to order lengths can be improved through learning environments that offer experience with length measurement.

The next stage in children's development of measurement is the ability to determine the length of an object by assigning a numerical value to it—an ability not necessary for comparing or ordering objects. The quantification of length is prompted by the need to know "how much longer?" and is established through the use of a unit of measurement, which can be a natural unit such as a footstep or a hand span or a standard unit such as a centimeter or meter.

A further requirement in measuring magnitudes is understanding the concept of unit iteration. This means that children have to realize that they must use units that are of equal size and that these units should be properly aligned without gaps or overlap. These requirements make unit iteration demanding for young children. For example, Barrett et al. (2003) and Clarke et al. (2003) have found that kindergartners were not yet capable of using units to determine the length of an object.

In general, the use of measurement instruments like rulers and measuring tapes is considered the final stage of learning measurement at the end of the lower primary grades (Buys & De Moor, 2008). There is, however, evidence that children can use standard measuring devices, before they understand them fully, or are able to use them accurately (Boulton-Lewis et al., 1996) or have been taught to use them (Nührenbörger, 2001).

---

[1] The term 'kindergartner' in this paper describes a child attending kindergarten. The age of these children is between 4 and 6 years. In the Netherlands, kindergarten involves 2 years of schooling (K1 and K2), in which children are offered some mathematics-related activities, but no formal mathematics instruction is given. In most of the kindergarten classes, the K1 and K2 children are grouped together in one class.

1.3 Teaching length measurement in kindergarten by reading picture books

Since the late 1980s, using picture books—and children's literature in general—has become more and more popular as a way to teach children mathematics (Griffiths & Clyne, 1991). Increasingly, it has been recognized that reading picture books to children provides a learning environment in which they can experience mathematics in a meaningful and informal way.

Research has shown, in general, that the use of picture books in teaching mathematics has positive effects on kindergartners' mathematical understanding (e.g., Casey, Erkut, Ceder, & Mercer-Young, 2008; Hong, 1996). There are also indications that children's literature may be helpful for students' learning of measurement. For example, Castle and Needham (2007) investigated first graders' understanding of measurement and found that the use of children's books stimulated children's length measurement performance. Malinsky and McJunkin (2008) found a positive effect of using children's literature on third graders' understanding of measurement. In the study of Malinsky and McJunkin, children's literature was used as an introduction for measuring objects by non-standard measurement tools such as pencils and straws. These activities were meant to induce discussion about the importance of standard measurements.

However, with respect to kindergartners, we did not find any study investigating whether reading picture books contributes to the understanding of measurement. To gain further insight into this topic, we set up the present study.

1.4 Research questions

In light of the above, we addressed three main research questions. Our *first question* was: what performance do kindergartners show in length measurement? In particular, what is children's general performance (Question 1a), and what are the components of this performance (Question 1b)? Our focus on performance is based on the assumption that what generates kindergartners' performance is their understanding and competence in length measurement. Therefore, performance in length measurement and its components can be considered as an indication of how able the kindergartners are in this mathematical content strand.

The *second question* addressed the growth in performance: how does the performance in length measurement increase over the kindergarten years? In particular, what are the differences between children in the first year of kindergarten and children in the second year of kindergarten with respect to their general length measurement performance (Question 2a) and with respect to the components of their performance (Question 2b)?

The *third research question* was: what is the effect of a picture book reading program on kindergartners' performance in length measurement? In particular, what is the effect of the program on the general performance in length measurement (Question 3a) and what is the effect on the components of this performance (Question 3b)? In addition, what are the effects on the general performance as well as on the components for the first kindergarten year (Question 3c) and for the second year (Question 3d)?

2 Method

To answer the research questions, we collected data about Dutch children's measurement performance in 18 kindergarten classes. In these classes, we also carried out an experiment with an experimental control group pretest–posttest design. The nine classes in the experimental group followed an intervention program that involved reading to children from picture books that address the measurement of length. In the control group, instruction in measurement followed the regular curriculum and picture book reading program.

2.1 Participants

The participating children were from primary schools, which include kindergarten classes. The schools were situated in the province of Utrecht in the Netherlands. We set up a multi-stage sampling procedure. To limit differences in teaching methods, we first excluded schools that had special educational approaches and schools that had first-year kindergartners (K1) and second-year kindergartners (K2) in separate classes. Excluding these schools resulted in a remaining group of about 80 schools subdivided at three urbanization levels (schools in small, medium and large towns). Within each urbanization level, we selected a subset of schools made up of pairs of schools based upon similarity in school size and average socioeconomic status of the school population. As a start, we matched schools into 25 pairs. Of each pair, the schools were randomly assigned to either the experimental group or the control group. Next, these schools were contacted to invite them to participate with one kindergarten class. When a school declined, we searched for another school with characteristics close to the 'missing' school of the pair. When a complete pair of schools decided not to take part in the study, we looked for a new matched pair. This process continued until we had identified nine pairs of schools, which yielded a sample size large enough to satisfy the statistical requirements. The final sample of 18 schools consisted of 6 schools in small towns, 7 schools in medium-sized towns and 5 schools in large towns. After excluding children who did not complete the pretest or

the posttest ($N = 76$), the total number of children involved in the analyses was 308, of which 158 were in the experimental group (61 belonging to K1 and 97 to K2) and 150 were in the control group (48 belonging to K1 and 102 to K2).

The children in the experimental group and the control group were quite comparable with each other with respect to:

- age: $M = 63.8$ months and SD $= 7.4$ months in the experimental group; $M = 64.4$ months and SD $= 7.4$ months in the control group
- gender: girls–boys ratio was .95 in the experimental group and .85 in the control group
- general mathematical ability as assessed by the CITO Ordering Test (see Sect. 2.3.2): $M = 53.57$, SD $= 15.05$ in the experimental group and $M = 51.58$, SD $= 11.19$ in the control group
- home language: proportion of non-Dutch versus home language was 14% in the experimental group and 11% in the control group
- socioeconomic status: proportion of lower socioeconomic status was 11% in the experimental group and 11% in the control group.

2.2 Intervention

The intervention included the use of eight picture books that addressed measurement. The books were read aloud in class (each class contained both K1 and K2 children) over four consecutive weeks (2 books every week).

The books used in the intervention were all trade books of literary quality that had not been purposively written to teach children mathematics; at least it was not explicitly stated, for example on the back cover of the book, that the book was meant for instructional aims. The books were selected with the help of a framework of learning-supportive characteristics of picture books (Van den Heuvel-Panhuizen, Aaten, & Van den Boogaard, 2011). Decisions for selecting books were made on the basis of consensus among the members of the research team. For example, we discussed (1) whether a picture book contained measurement issues that were valuable for children to learn, (2) whether these issues were presented in a meaningful context for the children, that is, in a context which could be recognized from daily life by the children, (3) whether the picture book showed coherence between concepts and connected different appearances and representations of these concepts and, finally, (4) whether it offered opportunities to make children actively involved in mathematical thinking when they were read the book. These criteria were used to select books that offered an environment in which children could think and reason on measurement issues related to length.

One of the picture books included in the intervention program was *De lievelingstrui* [The Favourite Jumper] by Veldkamp and Van der Linden (2001). This book is about a little pig that wants to grow (Fig. 1).

The book offers interesting measurement experiences. The pictures in the book illustrate the use of a measuring strip and the story has the potential to evoke children's active participation in reasoning about whether the pig has grown or not.

Another picture book that was read to the children was *Rosa's reuze zonnebloem* [Rosa's giant sunflower] by Damon (1997). In this book, a girl named Rosa aims to grow a giant sunflower. After some setbacks (e.g., a curious mole accidentally removes the newly sown sunflower seed from the soil), Rosa finally succeeds in growing a very large sunflower. The book depicts the increasing length of the sunflower by the use of a fold-out page (Fig. 2).

Together, the eight picture books addressed a broad range of length measurement issues. The story lines of the books and pictures encompassed, for example, direct and indirect comparison, the increase of length as a function of passing time, the distance to be bridged to reach a certain point and the use of measuring strips. As a result, the books were expected to elicit children's actions of measuring and comparing lengths, and reasoning about change in length.

We developed reading guidelines to ensure similar reading practices in all of the classes in the experimental group. These reading instructions describe in detail, in every page of the book, what the teachers should do, say and ask when reading each book, to help the teachers make full use of the book's potential. For example, for page 12

Fig. 1 Page 3 from *De lievelingstrui* (Veldkamp & Van der Linden, 2001)
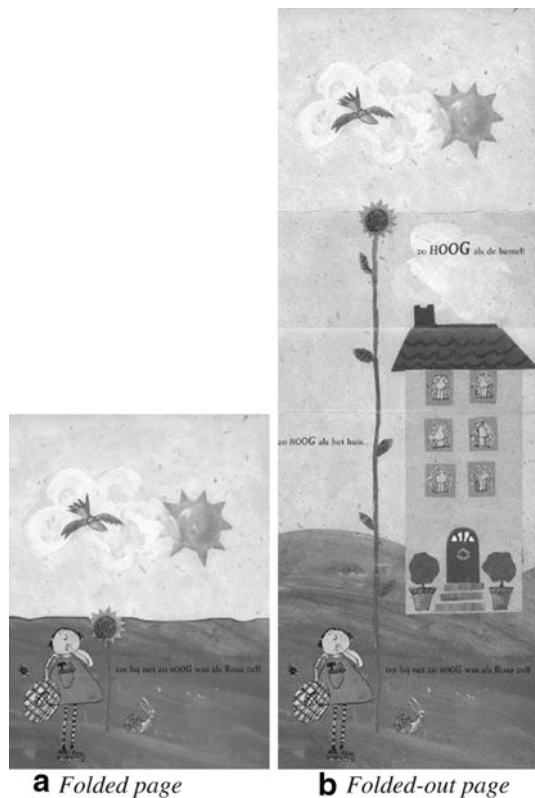
Fig. 2 Page 12 from *Rosa's reuze zonnebloem* (Damon, 1997) [text from *bottom* to *top*: "till the flower is as HIGH as Rosa", "as HIGH as the house", and "as HIGH as the sky!"]

a *Folded page*   **b** *Folded-out page*

from *Rosa's reuze zonnebloem* (see Fig. 2a, b) the following guideline was given:

"Read the text 'till the flower is as HIGH as Rosa'. Ask: 'Is that really true? Invite the children to explain how they can determine this. Then fold out the page and read the text 'as HIGH as the house'. Ask: 'Where about was the top of the flower when it was as high as the house?' The children have to indicate the height of the house on the stem of the flower, which is approximately at the third leaf up from the ground. Then read the text 'as HIGH as the sky!'. Wait for the children's reactions.

(Note. The flower and the house are positioned behind Rosa. Strictly speaking, this means that they appear smaller in the picture than they actually are. Bring this perspective issue only into discussion in case the children mention it.)"

Similarly, other guidelines include directions for posing additional questions, giving children time to respond to the events in the story and the pictures, and repeating the story or parts of it to give the children the opportunity to better grasp the ideas in the story or the logic of the story line.

While carrying out the book reading program, teachers kept logs to document how they read the books. Based on the teachers' logs, we concluded that the book reading was done in agreement with the guidelines.

2.3 Assessment instruments

The children's performance in length measurement was assessed by a collection of items designed by the members of the research team. These items were used for pretesting as well as for posttesting. To have an additional norm-referenced score of the children's mathematical ability in general, we also used a standardized test developed by CITO (Dutch National Institute for Educational Measurement). This test was administered before the intervention. Including this CITO score in the analysis enabled us to investigate whether the intervention works equally well for different levels of mathematical ability.

2.3.1 PICO measurement items

To assess children's ability in length measurement, a number of test items were developed, called "PICO measurement items" (see Appendix). The items have a paper-and-pencil multiple-choice format similar to that of the CITO Ordering Test. Each question and its set of multiple-choice responses are represented visually by drawings. The instruction for each item is read aloud and the children work individually. To answer each question, the children have to underline the particular drawing that represents the correct answer.

The PICO measurement items were designed in such a way that they referred to measurement contexts that the children knew from their experiences in daily life, for example, from the physical world around them, cartoons, and stories and pictures in picture books. Since the measurement issues in these items are all connected to situations with which the children are familiar, the children can easily imagine what the items are about.

Before the PICO measurement items were used for the data collection, the items were tried out in one class not belonging to the research sample. The items were revised where needed. The focus in this revision was on the clearness of the wording and the drawings. Data collection was carried out by trained test administrators in both the experimental group and the control group. The pretesting took place in January and the posttesting in May/June.

A reliability analysis of the collection of PICO measurement items was carried out based on the data collected in the research sample. Analysis of these data showed that 3 out of the 14 test items had a negative item discriminant. Therefore, these three items were excluded from further analyses. The Cronbach's alpha for the remaining items was .40 for the

pretest and .49 for the posttest. The relatively low internal coherence of the items could be explained by the small number of items included in the analysis (11 items), as well as the limited variation in the scores, and also by the fact that the scores, especially in the pretest, were very low. The low reliability value could also mean that the items referred to different measurement components. To figure out the different components and how they were related, we applied a hierarchical similarity analysis (see Sect. 3.1.2).

2.3.2 CITO mathematics test

For assessing kindergartners' general ability in mathematics, CITO has developed the CITO Ordering Test. This is a paper-and-pencil test, which is also available digitally. The test is made up of 42 multiple-choice items. The children have to answer the questions by underlining the picture that shows the correct answer. The total scores on the test are converted into a mathematical ability level ranging from Level A (highest level) to Level E (lowest level).

The test has two versions: one for the children who are in their first year of kindergarten (K1) and one for the children who are in their second year (K2). As a whole, the test comprises the following mathematical topics: shapes, classifying, ordering objects with respect to size, comparing numbers of objects and resultative counting. The K1 version also includes items on color and size, while the K2 version has additional items on number symbols.

The CITO Ordering Test is intended to be administered by the teachers themselves. In the present study this was done in January of that school year. The K1 children took the K1 version and the K2 children the K2 version. The reliability of the test was .85 in K1 and .81 in K2 (Van Kuyk & Kamphuis, 2001).

3 Results

The analyses of the data collected in this study provided us with two types of results: firstly, they informed us about kindergartners' performance in the domain of measurement of length (Questions 1 and 2); secondly, the analyses generated knowledge about whether reading picture books to kindergartners contributed to their performance (Question 3).

3.1 Kindergartners' performance in length measurement

3.1.1 General performance in length measurement: results from pretest

The mean performance for the total sample ($N = 308$) of the 11 PICO measurement items in the pretesting was .34

(SD = .15), which means that the average number of correct items was 3.74. The minimum total score was 0 correct items and the maximum total score was 9 correct items. The older children who were in K2 demonstrated a higher performance ($M = .39$, SD = .14) than the younger children who were in K1 ($M = .25$, SD = .12). The difference between these two age groups was significant [$t(306) = -8.65$, $p < .01$].

Table 1 shows children's success rates per item in the pretest for the whole sample and for the K1 and K2 children separately. Furthermore, this table contains the results of the chi-square tests that were carried out to examine the success differences between the K1 and K2 children.

The easiest items were the *Baby* item (86%) and the *Rope* item (85%). The children were less successful (49%) in the *Door* item. A proportion of 42% of the children could respond correctly to the *Plant* item. The children's success was relatively low in the *Tree* (27%), *Flower* (24%) and *Snail* (20%) items. The *Plants*, *Snake* and *Shawl* items were even more difficult for the children to answer correctly, as the success rates were 16 and 10% for the latter two items, respectively. Most difficult was the *Steps* item as only 4% of the children provided a correct answer.

The K2 children were significantly more successful than the K1 children in 7 out of 11 items including the *Baby*, *Rope*, *Plant*, *Plants*, *Door*, *Snail* and *Shawl* items. This result indicates that the general performance in length measurement increases with age.

3.1.2 Components of length measurement performance

To get more insight into the structure of children's ability in length measurement, we carried out a hierarchical

Table 1 Success percentages per PICO item in the pretest

Item	Success percentage in pretest			Chi-square ($df = 1$)
	K1 + K2	K1	K2	
Baby	86	78	91	9.12*
Rope	85	73	91	16.79*
Door	49	29	60	26.11*
Plant	42	28	49	13.68*
Tree	27	22	30	2.08
Flower	24	18	27	2.67
Snail	20	10	26	10.57*
Plants	16	6	22	13.65*
Snake	10	6	13	2.85
Shawl	10	2	15	13.26*
Steps	4	5	4	.06

* $p < .05$

similarity analysis on the assessment items by using the computer software Classification Hiérarchique, Implicative et Cohésitive (C.H.I.C.) (Gras, Suzuki, Guillet, & Spagnolo, 2008). This analysis identifies hierarchical similarity between groups of variables (Lerman, 1981). In our study, these variables consist of the children's responses to the different measurement items. For instance, the similarity of two distinct items can be determined by the probability that the number of subjects who simultaneously satisfy the two variables, that is, the number of children who answer consistently (i.e., correctly or incorrectly) to the corresponding items, is greater than the random number expected in this situation.

The similarity groups were established in an ascending manner as a function of their strength: the stronger the similarity connections were in the groups, the higher was the level at which they were established. Thus, the similarity groups are represented in a hierarchically constructed similarity diagram, which allows us to study and interpret groups of items in terms of a resemblance of performance characteristics. The similarity diagram in this study revealed a number of components in kindergartners' performance in length measurement.

Figure 3 shows the similarity relations based on the correctness of the children's responses to the items in the pretest in the total sample, including K1 and K2 children. The responses to the *Plant* and the *Plants* items are more similar than to any other pair of items. The similarity relation is situated at the first level of the hierarchical tree. Then, the similarity group consisting of the *Plant*, *Plants* and *Shawl* items, which is formed at the next level, presents a better aggregation than any other pair of items. Next, the similarity group is extended by the responses to the *Flower* and the *Door* items. The *Snake* item and the *Steps* item are linked at the next level. They are more similar than any other extension of the group of items consisting of *Plant*, *Plants*, *Shawl*, *Flower* and *Door*. The next level consists of the group including *Plant*, *Plants*, *Shawl*, *Flower*, *Door*, *Snail* and
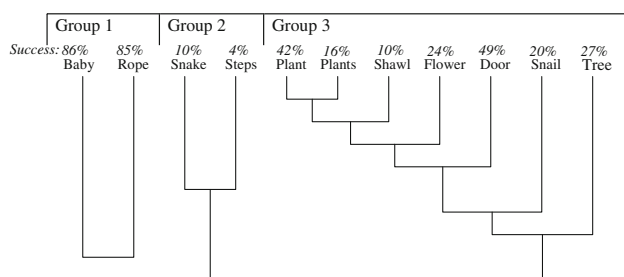
Tree. This group is higher than the level of the pair *Baby* and *Rope*, which in turn is higher than any extension of the pair *Snake* and *Steps*. Thus, in total three groups were identified by the similarity analysis. To enhance the interpretation of the structure found by the similarity analysis, we added the percentage of success to the item names.

Next step was that we had to reason why the items belonging to these groups were solved correctly by the same children. A deliberation among the research team about the determining characteristics of the items resulted in the following interpretations.

The first similarity group involves the responses to the *Baby* and *Rope* items. The remarkable thing about this group is the children's high success rate on the two items compared to the other test items. A distinguishing characteristic that differentiates these two items from the other items is that they strongly trigger the use of *holistic visual recognition*. Thus, in the *Baby* item, the children could have asked themselves: which picture looks like a baby? In the *rope* item they could have asked themselves: which picture has more "rope"? We think this solution approach fits well with children of this young age.

The second similarity group involves the responses to the *Snake* and *Steps* items. What these items have in common is, in one way or another, the partitioning of the length of an object (i.e., a snake or a pathway) into equal-sized units. Therefore, the second similarity group can be considered as reflecting children's understanding of measurement related to *unitizing*.
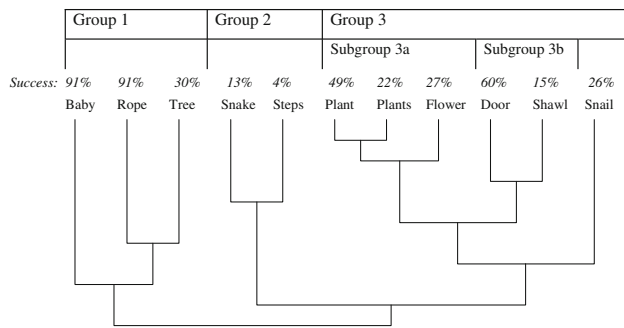
The third similarity group is based on the children's responses to the *Plant*, *Plants*, *Shawl*, *Flower*, *Door*, *Snail* and *Tree* items. All these items require ordering abilities based on the length of objects along a continuum. This is clearer in the *Plant*, *Plants*, S*hawl* and *Flower* items, but applies also to the other items. For example, in the *Tree* item, to understand the relationship between the height of the tree and the height of the girl and to find out which photograph showed the highest tree, the children probably used the order of the photographs. This is because the photographs were ordered according to the depicted height of the girl, starting with the photograph with the girl who looked the tallest. Because the tree is of the same height in all the photographs, the height of the depicted girl determines the height of the tree. So, the taller the depicted girl, the smaller is the tree in reality. Similarly, in the *Snail* item, the possible covered distances are ordered from the shortest to the longest one. This ordering probably helped the children to select the required distance. As for the *Door* item, children probably used ordering as well. Possibly, they imagined a girl's height increasing over the years and projected this height on the measuring strip

Fig. 3 Similarity diagram of the performance of the total sample of kindergartners

next to the door. On the whole, we interpreted the third similarity group to contain items that required *ordering*.

A closer look at the items revealed that a fourth group might have appeared, but it did not show up in the statistical analysis. We identified three other items that shared a common characteristic. The *Shawl*, *Tree* and *Steps* items all include an inverse relation. In the *Shawl* item, as the shawl grows longer the ball of wool grows smaller. In the *Tree* item, the smaller image of the girl relates to a taller tree in real life. In the *Steps* item, the number of steps is larger when the size of the steps is smaller. The fact that this similarity is not reflected in the children's responses can be interpreted to indicate that an inverse relation is too advanced a concept for children of this age to use in their reasoning.

The success rates on the items of each similarity group suggest that the three components of length measurement performance, which correspond to the three similarity groups, did not have the same level of difficulty for the kindergartners. The holistic visual recognition items (similarity group 1) were the easiest ones, whereas the unitizing items (similarity group 2) were the most difficult ones. The items that required ordering abilities (similarity group 3) were of moderate difficulty level.

Carrying out the similarity analysis for the responses of the children in K1 and K2 separately, revealed that the components in the children's performance in length measurement differed only slightly from the components identified in the whole group. The results from the K1 sample are shown in Fig. 4.

In the K1 sample, the *Baby* and *Rope* items, which in the whole sample belonged to the similarity group of responses labeled as holistic visual recognition, were linked with a second similarity group including the *Snake* and *Steps* items, which in the whole sample were identified as requiring unitizing. The linking of these two similarity groups indicates that the youngest kindergartners are not yet able to solve these latter items just by unitizing and may have used holistic visual recognition as well.

Another difference that was found between the results from the similarity analysis in the K1 sample and the whole sample was that, in the K1 sample, the items that comprised the third similarity group in the total sample were distinguished into the similarity groups 3 and 4, which were linked to each other. Comparable to the third similarity group of the diagram of the total sample, all the items in these groups of the K1 diagram require ordering abilities. However, there seems to be two types of responses triggered in the children. The items in group 3, that is, the strongly related *Plant* and *Plants* items, and the *Tree* and *Snail* items, were potentially solved by starting from the four possible answers. Therefore, we labeled group 3 as *recognizing answers*. This group clearly differs from group 4, which could be based on responses that involve *producing answers*. In the *Door*, *Flower* and *Shawl* items, children may first have produced their answer and then looked for the matching answer. For example, in the *Door* item children could have reasoned: I reach just to the doorknob, so the arrow next to it is the answer. In the *Shawl* item, children might have known directly that in the first picture of the ball of wool should be the largest, and consequently have looked for that ball. Finally, the approach of first producing the answer is probably most obvious in the *Flower* item where children could have imagined the height of the next flower followed by looking for the flower of that height in the answer boxes. This *Flower* item clearly contrasts with the *Plant* and *Plants* items, which belong to the third similarity group and which concern finding missing plants within a series of plants with increasing length. To sum up, our hypothetical interpretation for this distinction is that in the items of the fourth similarity group, the strategy of starting with the given answers and checking whether they each fit is more difficult to use and less spontaneous than it was with the third group.

The relative difficulty levels of the items across the three components of length measurement performance by K1 children are similar to the ones referring to the whole group. Specifically, difficult items require unitizing (similarity group 2), items of moderate difficulty involve ordering (similarity groups 3 and 4) and easy items include holistic visual recognition (similarity group 1). Within the third and the fourth similarity groups, though, the *Plants* and the *Shawl* items appear to be more complex for K1 children than the other items that require ordering.

The results from the similarity analysis of the pretest responses in the K2 sample are shown in Fig. 5.

The component structure of the children's length measurement performance in this sample is globally the same as in the K1 sample, but there are also some changes in

Fig. 4 Similarity diagram of K1 children's performance

Fig. 5 Similarity diagram of K2 children's performance

how the items group together. The second group that may represent the unitizing response turned out to be quite stable over the kindergarten years. The differences between the K1 and K2 samples were only found with respect to the two other similarity groups.

In the K2 sample, the first similarity group referring to holistic visual recognition is extended with the *Tree* item. This change could indicate that for the older kindergartners it is not necessary anymore to use the order of the photographs to find the answer, but that they see directly that the last photograph containing the smallest girl shows the tallest tree in reality. Another difference in the K2 sample is that the remaining six items in the K1 sample belonging to the third and the fourth similarity groups, in which the items require ordering, are more strongly linked, thus forming one similarity group. However, they are distinguished into two subgroups. Different from the K1 sample, in the K2 sample the items in these subgroups are not differentiated as requiring recognizing answers versus producing answers. Instead, in this K2 sample with the older kindergartners, there is rather a division in items that include the ordering of *just length* versus items that deal with *length in connection with other physical quantities*. The first category includes the *Plant*, *Plants* and *Flower* items. The second category includes the *Door* and *Shawl* items. The *Door* item involves length and age, and the *Shawl* item length and an informal understanding of volume. The *Snail* item has characteristics of both. It deals with two physical quantities, length and time, and the way it is presented has a strong connection to ordering. This is why there is a similarity between this item and the items belonging to subgroups 3a and 3b.

As in the K1 sample, the difficulty level of the items in the K2 sample varies as a function of the length measurement components the children encounter, with holistic visual recognition items (similarity group 1) as the easiest tasks, the ordering items as the tasks of moderate difficulty (similarity group 3) and the unitizing items as the most complex tasks (similarity group 2).

3.2 Effect of the intervention on length measurement performance

To answer the third research question, we compared the results of the experimental group with those of the control group. We did this for the general performance in length measurement and for its components. Before investigating whether the intervention had an effect on children's measurement performance, we examined whether there were initial differences in measurement performance between the two groups. We found that the experimental group children ($M = .33$, SD $= .14$) and the control group children ($M = .35$, SD $= .15$) demonstrated similar initial performance [$t(306) = 1.24$, $p = .22$].

3.2.1 Effect on general performance

The effect of the intervention program on kindergartners' general performance in length measurement was analyzed by means of a repeated measures univariate analysis of variance (ANOVA) with the factors Condition (control or experimental group), Test Moment (pretest or posttest), Mathematical Ability (levels A, B, C, D or E) and Gender (boy or girl) as independent variables and children's achievement on the PICO test as the dependent variable.

The findings of the analysis showed a significant main effect of Test Moment [$F(1, 280) = 25.71$, $p < .001$, $\eta^2 = .08$] on the general performance of length measurement. A weaker but significant interaction effect was found between Condition and Test Moment [$F(1, 280) = 4.04$, $p < .05$]. An effect size (η^2) of only .01 was found (Cohen, 1988). This finding indicated that the intervention had only a small positive impact. A further analysis revealed no significant triple interactions with Mathematical Ability [$F(4, 280) = 1.71$, $p = .15$, $\eta^2 = .02$] or Gender [$F(1, 280) = .51$, $p = .50$, $\eta^2 = .002$], indicating that all mathematical ability levels and both boys and girls did not differ with respect to their contribution to the Condition and Test Moment interaction effect. The estimated marginal means of the children from the experimental group (pretest: $M = .28$; posttest: $M = .37$) and the control group (pretest: $M = .33$; posttest: $M = .38$) on the pretest and the posttest are illustrated in Fig. 6.

The same analysis was carried out for the K1 and K2 children separately to investigate whether the impact of the intervention on children's results would vary with kindergarten year. The findings in both K1 and K2 revealed that the interaction effect between Condition and Test Moment was not significant [K1: $F(1, 86) = 1.18$, $p = .28$, $\eta^2 = .01$; K2: $F(1, 177) = 1.51$, $p = .22$, $\eta^2 = .01$]. However, in both grades the main effect of Test Moment remained significant [K1: $F(1, 86) = 21.48$, $p < .001$, $\eta^2 = .20$; K2: $F(1, 177) = 10.53$, $p < .01$, $\eta^2 = .06$].
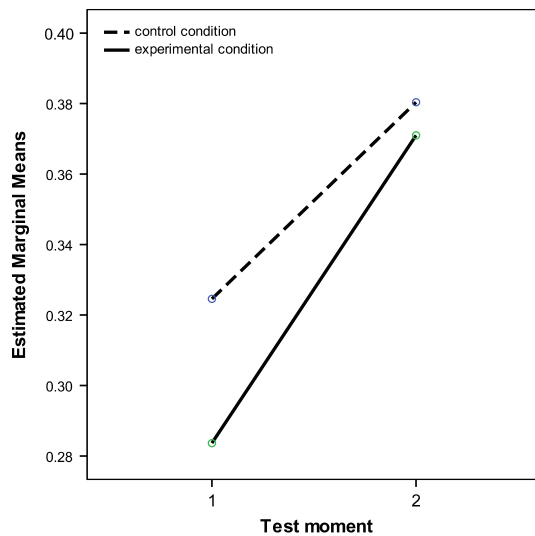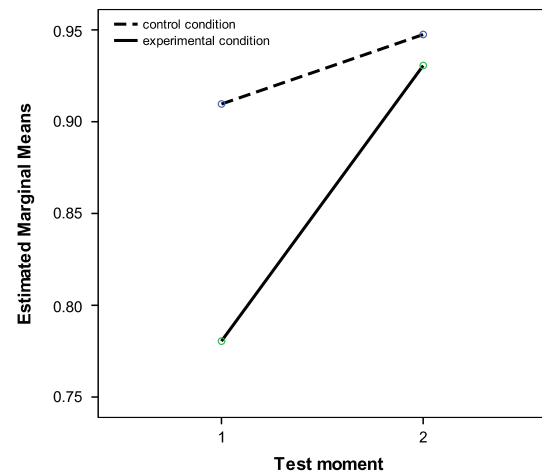
Fig. 6 Estimated marginal mean scores of the experimental and the control group on the general performance in length measurement in the pretest and posttest
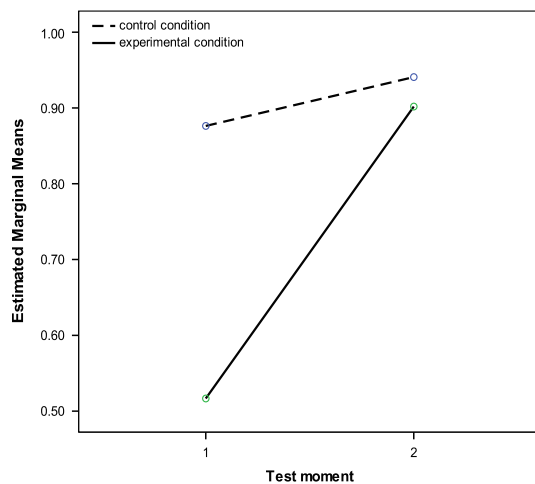

Fig. 7 Estimated marginal mean scores of the experimental and the control group of the total sample on holistic visual recognition of length measurement performance in the pretest and posttest

3.2.2 Effect on the components of the performance

The impact of the intervention program was further examined on the components of the length measurement performance, *holistic visual recognition*, *ordering* and *unitizing*, which were identified previously in the similarity analyses. Repeated measures multivariate analyses of variance (MANOVA) were applied to the data of the total sample, and of the K1 and K2 children, separately—with these three components as dependent variables and Condition, Test Moment, Mathematical Ability and Gender as independent variables.

The analysis of the total sample showed a significant main effect of Test Moment [$F(3, 278) = 10.09$, $p < .001$, $\eta^2 = .1$]. Nevertheless, we did not find a significant interaction effect between Condition and Test Moment [$F(3, 278) = 2.52$, $p = .059$, $\eta^2 = .03$]. Univariate analyses revealed though that this interaction was significant for the first component of the length measurement performance, namely, holistic visual recognition [$F(1, 280) = 5.15$, $p < .05$, $\eta^2 = .02$]. As illustrated in Fig. 7, children of the experimental group made more progress in holistic visual recognition than the children of the control group.

The application of the same analysis on the two kindergarten years, separately, showed similar results for K1 children and rather different results for K2 children. With respect to K1 children, a significant main effect of Test Moment [$F(3, 84) = 13.26$, $p < .001$, $\eta^2 = .32$] was found, but no significant interaction between Condition

and Test Moment was revealed [$F(3, 84) = 2.62$, $p = .056$, $\eta^2 = .09$]. However, again the univariate analyses revealed that this interaction was significant for holistic visual recognition [$F(1, 86) = 7.94$, $p < .01$, $\eta^2 = .09$]. This finding suggests that the K1 children in the experimental group realized considerably more length measurement improvement on the holistic visual recognition ability than did the control group (Fig. 8).

As regards K2 children, the findings showed a significant main effect of Test Moment [$F(3, 175) = 3.96$, $p < .05$, $\eta^2 = .06$], but no significant interaction between Condition and Test Moment [$F(3, 175) = .50$, $p = .69$, $\eta^2 = .01$]. Univariate analyses revealed that this interaction was not significant for any of the three components of length measurement performance [holistic visual recognition: $F(1, 177) = .29$, $p = .59$, $\eta^2 = .002$; unitizing: $F(1, 177) = .08$, $p = .78$, $\eta^2 < .001$; ordering: $F(1, 177) = .98$, $p = .33$, $\eta^2 = .01$], indicating that the intervention program did not result in a significant increase of the older children's general performance in length measurement or of its components.

4 Discussion

The purpose of the study was to explore kindergartners' length measurement ability and to find out whether this ability can be enhanced by reading to them picture books that address measurement issues.

Fig. 8 Estimated marginal mean scores of the experimental and the control group in K1 on the component holistic visual recognition of the length measurement performance in the pretest and posttest

4.1 Children's general performance

The results revealed that kindergartners encountered great difficulty in solving most of the length measurement tasks used in this study (Question 1a). This difficulty can be attributed to the high complexity of these tasks. In particular, some tasks require the mental use of unit of length measurement and unit iteration or complex ordering abilities. Furthermore, although measurement is included in many mathematical curricula, it might be that this mathematical domain is not emphasized in the kindergarten mathematics teaching as much as is suggested by these curricula. Consequently, children's mathematical thinking in length measurement might not be developed sufficiently so that they are able to deal successfully with most of the tasks of this study.

4.2 Components in length measurement performance

In investigating the structure of the kindergartners' length measurement performance (Question 1b), the hierarchical similarity analysis of the children's response performances identified three major components, which we interpreted as *holistic visual recognition*, *ordering* and *unitizing*. Holistic visual recognition does not require reasoning. The children directly "see" the correct answer. Ordering refers to multiple comparisons between the lengths of objects. Unitizing requires the partitioning of the length of objects into equal-sized units. These three length measurement components were found to have different difficulty levels for the children. In

unitizing tasks, the kindergartners encountered the greatest difficulty. The ordering tasks were more easily tackled, while the holistic visual recognition was most easily achieved. This order of relative difficulty among the length measurement components is in line with the development of understanding length measurement in young children that has been suggested by previous studies (e.g., Barrett et al., 2003; Clarke et al., 2003).

Additionally, within the component of ordering, the statistical program identified a number of sub-components. This suggests that the ordering ability of young children can be further analyzed into sub-abilities. The findings of this study showed that the nature of this distinction within ordering can be a result of two task features, that is, the behavior elicited by the type of question and the mathematical content of the item. In particular, among K1 children a distinction was found between producing or selecting answers in ordering tasks. Among K2 children, a division emerged between the children's performance in solving ordering tasks that involve just length and their performance in solving tasks that combine length with other attributes, such as duration or volume.

4.3 Growth in length measurement performance

The study showed a significant increase of the general performance of length measurement over the kindergarten years (Question 2a). This finding in growth was not repeated for all the components of length measurement. We only found a difference between the K1 and K2 children for holistic visual recognition and ordering, and not for unitizing ability (Question 2b). A possible explanation for this deviant finding for unitizing may be that this component belongs to a higher cognitive level than the others. Therefore, substantial attention in teaching might be necessary to enhance this ability at this age.

Although we found a difference in performance of the components of length measurement between the children in K1 and K2, in both age levels the identified components were the same.

4.4 Effect of the intervention program

We found a weak but significant effect of the picture book reading program on children's general length measurement performance (Question 3a). On further consideration, this weak effect is not that surprising. The duration of the program was rather short and the program did not involve any explicit training on measuring

skills; the program was mainly based on incidental learning, which might need a certain amount of time to become effective. Another explanation is that the high rate of children's cognitive growth at this age (Bowman, Donovan, & Burns, 2000) might account for the improvement that also occurred in the children in the control group.

With respect to the components (Question 3b), the intervention effect is exerted mainly on the development of holistic visual recognition among the younger K1 children of the study (Question 3c). Obviously, there is something to gain at this young age, and picture book reading, through its focus on interpretation of pictures and comparison of various lengths, can contribute to children's visual recognition ability. For K2, this was apparently not the case because their performance on holistic visual recognition was already high in the pretest (Question 3d).

It is understandable that no effect was found for ordering. It is an ability included in the CITO Ordering Test that was used in the analysis as a covariate. Furthermore, we should also take into account that the teachers of the children who were in the control group may have taught this ability in their teaching because of its inclusion in the CITO Ordering Test. Therefore, it was probably difficult to find a significant difference for this component between the control and the experimental group.

The lack of an effect for unitizing is probably caused by the fact that—as discussed already—this component requires a higher level of thinking than the other components. Thus, its development probably demands more systematic, intentional and practical experience than that offered by picture book reading only.

4.5 Limitations of the study and suggestions for further research

To place findings in the right perspective, it is necessary to take into account the fact that our study was carried out with a limited number of classes and that the length measurement performance was measured with a small number of items covering a limited number of possible components. A special difficulty was the use of the CITO Ordering Test to measure the general mathematical ability of the children, for which we controlled in the ANOVA and MANOVA analyses. Because a number of items in the CITO test ask children to put objects in order of size, the score of this test may interfere with the score on the PICO measurement items.

Another shortcoming of the study was that the duration of the intervention was rather short and no data were collected about the retention of the effect. The way this intervention was performed is also a point of concern. The teachers were provided with picture book reading guidelines and training for carrying out the intervention, and logs of their reading book sessions were collected. Nevertheless, we could not be completely sure about how the intervention was implemented.

Further research is needed to improve our knowledge of children's development of length measurement ability and the potential contribution of picture book reading to this development. It is necessary to design methods that cancel out the limitations of the present study. Additional research should also pay attention to the development of children's length measurement ability after the kindergarten years. Our understanding might be improved if children from grade 1 and beyond are involved, especially with respect to the difficult component that is related to the unitizing concept. Moreover, a deeper understanding of the component of ordering might be obtained by using a more differentiated collection of items in which the types of questions (recognizing answers vs. producing answer) are systematically varied. A further improvement of our knowledge of children's ability to deal with length measurement could be achieved by using a one-to-one interview in addition to a class-administered paper-and-pencil test.

While the present study has shown that picture book reading might have the potential to contribute to kindergartners' development of length measurement ability, we do not yet know which particular classroom conditions are probably needed to make this happen. Future research should collect more specific information about these conditions. However, this focus on examining the conditions does not imply that the reading sessions should become a systematically built up sequence of instructional activities in length measurement where the picture books are used simply to illustrate what the teacher is teaching. Such instructional intervention is not in alignment with the goals of the research program to use picture books to support children's learning of mathematics. In our research program, the focus is on the power of the picture books themselves. As shown earlier (Van den Heuvel-Panhuizen & Van den Boogaard, 2008), there are good reasons to continue with this focus.
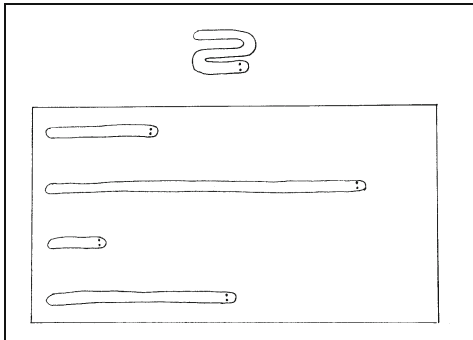
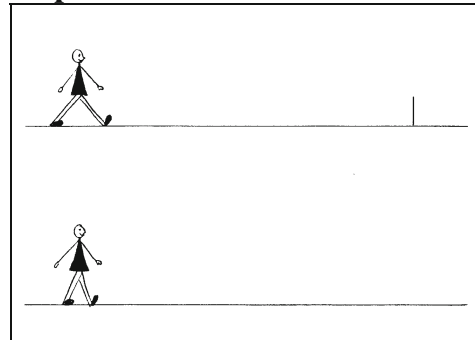**Appendix: PICO measurement items**

Baby
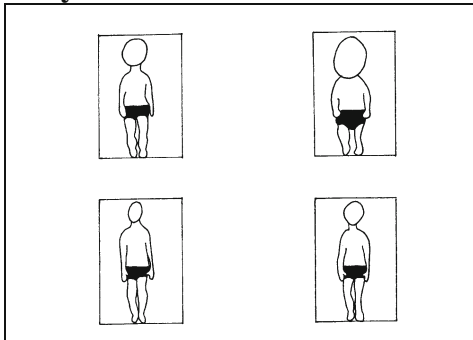
On which picture do you see a baby? Put a line under the baby.
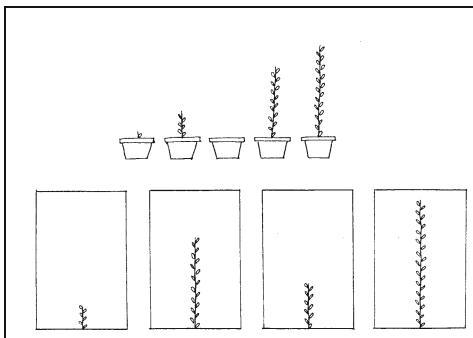
Rope
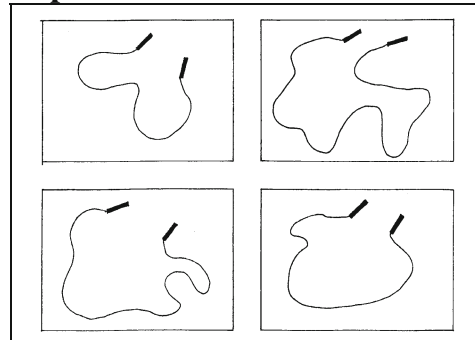
Which skipping rope is the longest? Put a line under the longest rope.

Snake

Which snake is as long as the writhing snake? Put a line under the snake that is just as long.

Steps

The two men take just as many steps. The man on the top takes big steps. He gets up to the small line. The man on the bottom takes small steps. Up to where he will get? Put a line there.

Plant

The plant gets taller and taller. Which plant belongs in the empty flowerpot? Put a line under the plant that should be in between.

Plants
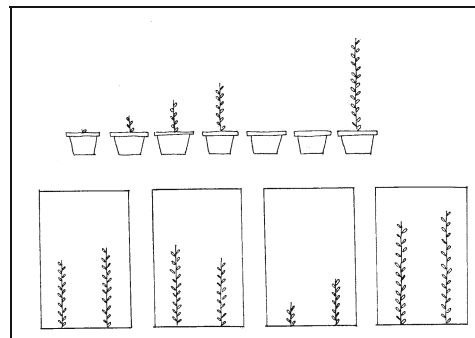
The plant gets taller and taller. Which plants belong in the empty flowerpots? Put a line under the picture with the plants that should be in between.

Tree
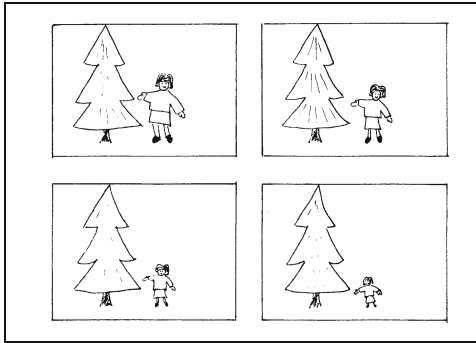
Mommy has taken 4 photos of the girl. On all photos the girl stands at a tree. Which tree is the biggest in reality? Put a line under the tree that is the biggest in reality.

Door
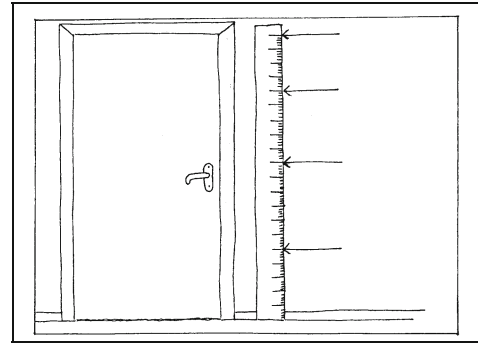
Next to the door is a growth meter. How tall is a five-year old girl? Put a cross through the arrow that shows the height of the girl.

Snail
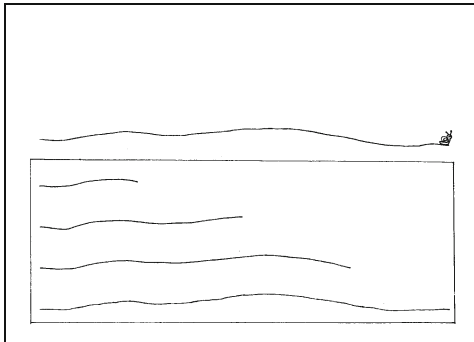
After 1 hour the snail has come this far. Where was the snail after half an hour? Put a line where the snail was after half an hour.

Flower
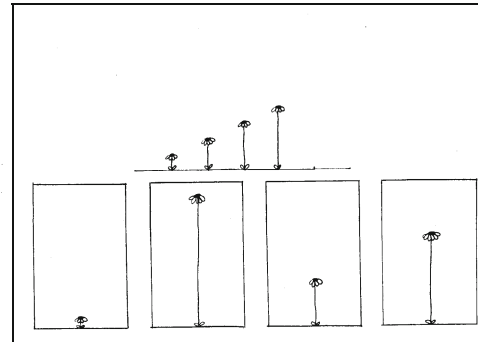
The flower gets taller and taller. Which flower belongs to the empty spot? Put a line under the flower that belongs here.

Shawl
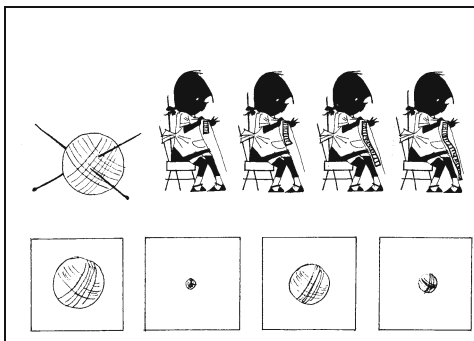
Janneke is going to knit a shawl. She has a ball of wool. When she starts knitting, the ball is getting smaller and smaller. Connect the thread to the right ball.

References

Barrett, J. E., Jones, G., Thornton, C., & Dickson, S. (2003). Understanding children's developing strategies and concepts for length. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement: 2003 Yearbook* (pp. 17–30). Reston, VA: National Council of Teachers of Mathematics.

Board of Studies NSW. (2006). *Mathematics K-6, syllabus 2002*. Sydney: Author.

Boulton-Lewis, G. M., Wills, L. A., & Mutch, S. L. (1996). An analysis of young children's strategies and use of devices for length measurement. *Journal of Mathematical Behavior, 15*, 329–347.

Bowman, B., Donovan, M. S., & Burns, S. (Eds.). (2000). *Eager to learn. Educating our preschoolers*. Washington, DC: National Academy Press.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: brain, mind, experience, and school*. Washington, DC: National Academy Press.

Buys, K., & De Moor, E. (2008). Domain description measurement. In M. Van den Heuvel-Panhuizen & K. Buys (Eds.), *Young children learn measurement and geometry* (pp. 15–36). Rotterdam: Sense Publishers.

Casey, B., Erkut, S., Ceder, I., & Mercer-Young, J. (2008). Use of a storytelling context to improve girls' and boys' geometry skills in kindergarten. *Journal of Applied Developmental Psychology, 29*, 29–48.

Castle, K., & Needham, J. (2007). First graders' understanding of measurement. *Early Childhood Education Journal, 35*, 215–221.

Clarke, D., Cheeseman, J., McDonough, A., & Clarke, B. (2003). Assessing and developing measurement with young children. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement: 2003 Yearbook* (pp. 68–80). Reston, VA: National Council of Teachers of Mathematics.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Curry, M., & Outhred, L. (2005). Conceptual understanding of spatial measurement. In P. Clarkson, A. Downton, D. Gronn, A. McDonough, R. Pierce, & A. Roche (Eds.), *Building connections: Theory, research and practice (Proceedings of the 28th annual conference of the mathematics education research group of Australasia*, Melbourne) (Vol. 1, pp. 265–272). Sydney: MERGA.

Damon, E. (1997). *Rosa's reuze zonnebloem*. Amsterdam: Sjaloom [this is a Dutch version of *Daisy's giant sunflower*. London: Tango Books].

Gras, R., Suzuki, E., Guillet, F., & Spagnolo, F. (Eds.). (2008). *Statistical implicative analysis: Theory and applications*. Heidelberg: Springer.

Griffiths, R., & Clyne, M. (1991). *Books you can count on. Linking mathematics and literature*. Portsmouth, NH: Heinemann.

Hong, H. (1996). Effects of mathematics learning through children's literature on math achievement and dispositional outcomes. *Early Childhood Research Quarterly, 11*, 477–494.

Hughes, M. (1986). *Children and number*. Oxford: Blackwell.

Lerman, I. C. (1981). *Classification et Analyse Ordinale des Données*. Paris: Dunod.

Malinsky, M. A., & McJunkin, M. (2008). Wondrous tales of measurement. *Teaching Children Mathematics, 14*(7), 410–413.

National Council of Teachers of Mathematics (NCTM). (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

Nührenbörger, M. (2001). Insights into children's ruler concepts—grade-2 students' conceptions and knowledge of length measurement and paths of development. In M. van den Heuvel-Panhuizen (Ed.), *Proceedings of the 25th conference of the international group for the psychology in mathematics education* (Vol. 3, pp. 447–454). Utrecht: Freudenthal Institute.

Outhred, L., Mitchelmore, M., McPhail, D., & Gould, P. (2003). Count me into measurement: A program for the early elementary school. In D. H. Clements & G. Bright (Eds.), *Learning and teaching measurement: 2003 Yearbook* (pp. 81–99). Reston, VA: National Council of Teachers of Mathematics.

Sarama, J., & Clements, D. H. (2008). Mathematics in early childhood. In O. N. Saracho & B. Spodek (Eds.), *Contemporary perspectives on mathematics in early childhood education* (pp. 67–94). Charlotte, NC: Information Age Publishing.

Van den Heuvel-Panhuizen, M., Aaten, A. B., & Van den Boogaard, S. (2011). *Experts' view on learning-supportive characteristics of picture books for learning mathematics* (in press).

Van den Heuvel-Panhuizen, M., & Buys, K. (Eds.). (2008). *Young children learn measurement and geometry*. Rotterdam: Sense Publishers.

Van den Heuvel-Panhuizen, M., & Van den Boogaard, S. (2008). Picture books as an impetus for kindergartners' mathematical thinking. *Mathematical Thinking and Learning, 10*(4), 341–373.

Van Kuyk, J., & Kamphuis, F. (2001). *Verantwoording van de toetsen uit de pakketten Ruimte en Tijd, Taal voor Kleuters en Ordenen*. Arnhem, The Netherlands: Citogroep.

Veldkamp, T., & Van der Linden, G. (2001). *De Lievelingstrui*. Rotterdam: Lemniscaat.