Foreword

In a cubic centimeter, there are 1 000 cubic millimeters, in a cubic decimeter 1 000 000, in a cubic meter 1 000 000 000, and so forth. Why on earth is it usually so difficult to teach and to learn such simple facts, and many others of a similar vein? Of course, some questions of this type are more intricate. There is no easy computation showing that a giant ten times as high as a given dwarf weighs about one thousand more. So the nature of the problem is a crucial factor, and the authors of this study are fully aware of that.

This book deals with the illusion of linearity, mainly in the context of enlargement and reduction of figures and solids. An elementary example is when somebody believes that multiplying the side of a square by 2 implies that its area is also multiplied by 2. The book approaches also, to some extent, the context of probabilities. The authors rely essentially on two methods of investigation, namely experiments involving an experimental and a control group of students, and individual interviews on the other. A number of important variables are scrutinized, the most important being:

- drawings made by students themselves versus ready-made drawings, using squared paper or not
- *direct* versus *indirect* measures
- problems stated in the missing-value format, or in the comparison format
- awakening students' consciousness by a preliminary significant question
- degree of authenticity of the situation.

Two of these variables deserve some further explanation.

Direct versus indirect measures. Expressing an area in square meters is an example of a *direct measure,* while relating the area of a surface to the amount of paint required to cover it is an example of an *indirect measure*.

Missing-value or comparison format. As an example, one side of a polygon measures 2 dm and its area is 6 dm^2. What happens to the area in an enlargement operation where the 2 dm become 8 dm? This is a missing-value format. And if the 2 dm were increased by a factor of 4? This is a comparison format. In short, the givens are three measures in the former case, and two measures and a ratio of measures in the latter.

An impressive result of the study is how deep-rooted the illusion of linearity is and how strongly it resists many variations of the teaching and learning parameters. The principal circumstance in which the illusion is substantially weakened is when the students are asked, instead of drawing or computing, to physically cover the enlarged surface (of the problem) by an appropriate paving. And even in that case, the improvement of the students' awareness does not substantially withstand returning to more academically stated questions.

The authors also tried a series of ten classroom one-hour sessions inspired by the principles of realistic mathematics education: meaningful and attractive problems, small group work, whole-class discussions and, as far as mathematical matters are concerned, a variety of representations and symbols (drawings, tables of functions, graphs, formulas). Even such a more concentrated and well-oriented pedagogical action did not yield entirely satisfactory results: "many students did not develop a deeper understanding of (non)-proportionality."

For such persistence of the illusion of linearity, three main causes are identified. One of them pertains to the way proportionality is often taught, namely when some parts of the curriculum pay an almost exclusive attention to proportionality as compared to non-linear relations, when there is an overuse of missing-value problems and an overemphasis on routine solving processes as compared to meaningful analysis of situations. Indeed, proportionality is more than a four-term relation. There are the classical rule of three, tables of proportionality showing more than four terms, straight-line graphs passing through the origin, the constant slope of such graphs identified with the coefficient of proportionality, etc. Of course, these features are better understood when contrasted with non-proportional (non-linear) relations. If the teaching ignores these meaningful facets of linearity and remains confined to the narrow domain of four-term missing-value questions, and if it does not contrast proportionality with non-proportionality, then the students are likely to remain like short-sighted prisoners in an obscure intellectual cell. As was so convincingly explained by Wertheimer (1945), a perspicacious problem solver in a given domain is one who knows the landscape familiarly, i.e. not only its various parts, but the ways to circulate amongst them. Perceiving the very structure as a whole is crucial.

Further, if linearity and non-linearity ought to be regularly confronted, doesn't it mean that the real notion at stake is the one of function with its various modalities? As Klein (1939) wrote a long time ago (the quotation remains surprisingly timely after such a long period):

> We, who are readily called reformers, want to place the concept of function at the centre of teaching. For it is that mathematical concept of the last 200 years which, wherever mathematical thought is needed, plays a central role.

Another cause of the persistent illusion of proportionality can be found, according to the authors, in some shortcomings of the general geometrical knowledge of the students. However, this second cause is akin to the first one. One refers to the teaching of proportionality, the other to an unsuccessful teaching of geometry in general. But what does it mean that, when solving proportionality questions, the students show some gaps in their general geometrical knowledge? It means that their understanding of proportionality lacks some structural links with significant adjoining geometrical questions. Generalizing this comment, one might say that mathematics is not a juxtaposition of items, it is an integrated culture. What is at stake is the mobility of mind. The authors are aware of that. As a remedy, they propose to displace the emphasis from computing correct numerical answers to building appropriate mathematical models. But what is the substance of such models if not those mathematical notions and properties that faithfully express the structure of the situation on hand?

Let us now leave aside the deficiencies of the teaching system. A third cause of the illusion of linearity is of a more intrinsic nature. It relates to the intuitiveness and simplicity of the linear relation. This deserves some comments. Let us assume that a student received a fully appropriate instruction on proportionality and non-proportionality. She or he might still be seduced by the charms of proportionality. This is an effect of what might be called the inertia of concepts. Proportionality is similar, to some extent, to a paradigm in the sense of Kuhn (1962). When you have an intellectual instrument at your disposal, if this instrument properly solved a lot of previous problems, if it appears simpler and more elegant than others, then you stick to it until further notice. What happens here pertains simultaneously to the pleasant and simple nature of the knowledge and the indolent nature of the human mind. The inertia of the concepts is also illustrated by a striking finding of the study. In fact, when students have been duly trained, on a number of examples, to identify the non-linear situations, they show a tendency to overuse a non-linear model. Changing ones mind is not so easy, but once this is done... Le mieux est l'ennemi du bien[1].

Now, what could be done to avoid such seductions, if not developing the habit of doubting, a critical mind, a constant circumspection in front of any problem, the habit of checking everything? To conclude, may I express how I appreciate the honesty of this study. It brings us a most careful survey of a number of real difficulties more than a wealth of solutions. All the more so that one of the findings is that even ten classroom sessions are not enough to bring a persistent change. No panacea is proposed. These questions have to be considered in a long run perspective. In the mean time, some doubts will remain. But after all, as Dante wrote in the Inferno,

Che non men che saper, dubbiar m'aggrada,

which means

As well as knowing, doubting is praiseworthy.

Nicolas Rouche
Professor Emeritus of the Catholic University of Louvain, Belgium

1 The best is the enemy of the good.

Chapter 2

IN SEARCH OF EMPIRICAL EVIDENCE

1. INTRODUCTION

Notwithstanding the numerous illustrations of students' tendency towards the overuse of linearity, systematic observations and analyses of this phenomenon based on empirical research have been absent until recent years. In this chapter, we report a first pair of studies aimed at filling this gap in the research literature (see also De Bock, Verschaffel, & Janssens, 1998). Both studies focused on application problems about the effect of a linear enlargement or reduction of a geometrical figure on the perimeter or area of that figure. As illustrated in the previous chapter, many scholars in the field have mentioned the occurrence of students' overuse of linearity for this particular type of problem situation, most often incidentally in the course of studies with other main foci (e.g. Mogensen, 2004; Outhred & Mitchelmore, 2000; Rogalski, 1982; Simon & Blume, 1994; Tierney, Boyd, & Davis, 1990).

A first study investigated the occurrence and the strength of 12–13-year old students' tendency to overuse linearity in solving word problems about the effect of a linear enlargement or reduction of a geometrical figure on the perimeter or area of that figure. Moreover, the impact of two task variables was investigated: (1) the shape of the geometrical figure involved in the problem situation, and (2) the availability of self-made or ready-made drawings accompanying the problem statement. Due to the extremely low occurrence of correct answers in this study, however, it was hardly possible to determine the impact of these task variables. Therefore, a second study was conducted, which was basically a replication of the first one, but now with an older age group of 15–16-year old students.

2. STUDY 1

2.1 Subjects, materials and procedure

Participants in the first study were 120 12–13-year olds (7th-graders) of a large secondary school in Flanders recruiting its students from a wide range of elementary schools in the region. Therefore, although all students attended the same secondary school at the time of testing, their educational histories with respect to (elementary) mathematics education were not identical but there are common elements. Word problems dealing with direct proportionality constitute an important topic of the mathematics program in the upper grades of the primary school in Flanders (Ministerie van de Vlaamse Gemeenschap, 1997). With the exception of inverse proportionality ($y = k/x$), other types of functional relationships are normally not systematically addressed at this level of schooling. With respect to elementary geometry, Flemish students are expected to master the names and the basic properties of the most familiar geometric figures, including the formulas for calculating their perimeter, area and volume, by the end of elementary school. Their expected experience with scaling and similarity in relation to mensuration involves declarative and procedural knowledge of the rules of the metric system for one-, two-, and three-dimensional measures.

In this first study, the 120 students were divided into three equivalent groups (Groups I, II, and III consisting of 40, 42, and 38 students, respectively). Each group was composed of two intact classes, one of which had six hours of mathematics a week, while the other class had four hours of mathematics a week.

The experiment consisted of two phases. During the first phase all 120 students were administered the same paper-and-pencil test consisting of 12 experimental items and 3 buffer items (Test 1). No clues or special instructions were given. All 12 experimental items involved enlargements of similar plane figures, and belonged to three categories: 4 items about *s*quares (S), 4 about *c*ircles (C), and 4 about irregular figures (I). Within each category of figures (S, C, and I), there were 2 proportional and 2 non-proportional items. Table 2-1 lists examples of one proportional and one non-proportional item for each of the three categories of figures from Test 1.

Table 2-1. Examples of experimental items

Enlargement of a square figure
Proportional item: Farmer Gus needs approximately 4 days to dig a ditch around a square pasture with a side of 100 m. How many days would he need to dig a ditch around a square pasture with a side of 300 m? (Answer: 12 days)
Non-proportional item: Farmer Carl needs approximately 8 hours to fertilise a square piece of land with a side of 200 m. How many hours would he need to fertilise a square piece of land with a side of 600 m? (Answer: 72 hours)
Enlargement of a circular figure
Proportional item: You need approximately 6 hours to sail around a circular island with a diameter of 70 km. How many hours would you need to sail around a circular island with a diameter of 140 km? *(Answer: 12 hours)*
Non-proportional item: You need approximately 400 grams of flower seed to lay out a circular flower bed with a diameter of 10 m. How many grams of flower seed would you need to lay out a circular flower bed with a diameter of 20 m? *(Answer: 1 600 grams)*
Enlargement of an irregular figure
Proportional item: On a map of Belgium in an atlas the distance from Genk to Leuven is approximately 5 cm and the distance from Genk to Ghent approximately 11 cm. On a map in front of the classroom the distance from Genk to Leuven is approximately 20 cm. How long is the distance from Genk to Ghent on this map? *(Answer: 44 cm)*
Non-proportional item: On a map of Belgium in an atlas the distance from Genk to Tongeren is approximately 2 cm and the area of Belgium approximately 250 cm^2. On a map in front of the classroom, the distance from Genk to Tongeren is approximately 6 cm. How large is the area of Belgium on this map? (Answer: 2 250 cm^2)

As illustrated in Table 2-1, the variables 'length' and 'area' were mostly replaced by more concrete, indirect variables that are proportional to them (or are reasonably supposed to be so), with a view to constructing a set of meaningful application problems. We come back to this issue in chapter 3. All other possibly relevant task variables – such as the degree of familiarity with the problem context, the grammatical complexity of the problem formulation, and the nature of given numbers – were controlled as much as possible. For instance, we used only 'simple' natural numbers as scale factors, so that all required computations had a similar technical difficulty.

The response sheets could be used not only to write the answers, but also to make calculations, drawings, or comments.

Two weeks after the first test the three groups of students were confronted with a second test (Test 2), which was a parallel version of Test 1. Once again, the problems were the same in all three groups, but the way in which the test was introduced and presented was different. In Group I, which functioned as the control group, the testing conditions were exactly the same as during the first test, i.e., these students received the problems without any additional clues and in exactly the same format as during Test 1. The students of Group II were explicitly instructed to make a sketch or drawing before answering each problem. This instruction was given at the beginning of the test and was illustrated by means of an example item (which did, of course, not involve similar plane figures). In Group III every problem was accompanied by a relevant ready-made drawing like the one given in Figure 2-1 (belonging to the non-proportional item about squares in Table 2-1).

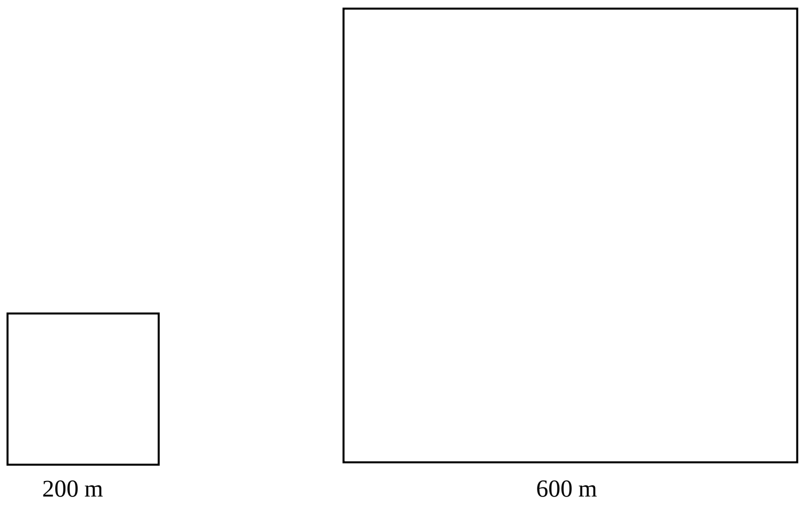

Figure 2-1. Example of a ready-made drawing

2.2 Hypotheses

First, on the basis of what is generally acknowledged in the field, we hypothesized that the predominance of the linear model would be a serious obstacle for the vast majority of the students. Consequently, we predicted that their performance on the 6 proportional items would be very high, while their scores on the 6 non-proportional items would be very low.

Second, we hypothesized that a sketch or drawing would have a beneficial effect on the students' performance, especially for the non-proportional experimental items. This hypothesis is based on the vast amount of theoretical and empirical research on how and why drawings and diagrams are a useful in enhancing people's ability to represent and solve (mathematical) problems (Aprea & Ebner, 1999; De Corte, Greer, & Verschaffel, 1996; Hall, Bailey, & Tillman, 1997; Larkin & Simon, 1987; Pólya, 1945; Reed, 1999; Schoenfeld, 1992; Vlahovic-Stetic, 1999). When students are asked to *make a drawing themselves*, they are stimulated to construct a proper (mental) representation of the essential elements and relations involved in the problem (Pólya, 1945; Schoenfeld, 1992). Especially for the non-proportional items, this representational activity should help students to detect the inappropriateness of a stereotyped linear reasoning, and to determine the nature of the non-linear relationship connecting the known and the unknown elements in this problem representation. Of course, the heuristic of making a drawing or diagram does not guarantee that one will find the solution of a given problem. For instance, a drawing that reflects an incorrect understanding of the problem will be of little help for the problem solution (Van Essen & Hamaker, 1990). When students do not succeed in making a *correct, usable* drawing themselves, it might be more effective if they are *provided with a correct ready-made drawing*. Starting from these hypotheses, we predicted that in Group I the results will be the same for Test 1 and Test 2, while in Group II and Group III, the percentage of correct responses will increase from Test 1 to Test 2. This increase would be essentially due to a decrease of inappropriate solutions based on linear reasoning on the non-proportional items during Test 2. No specific hypothesis was stated with respect to the relative impact of the two experimental manipulations.

Third, we predicted that students' performances would be different for the distinct types of plane figures involved in the study. More specifically, the items about squares (S items) were supposed to be the easiest and those about irregular figures (I items) the most difficult. We also expected that the size of the anticipated effect of drawings (see hypothesis 2) would be affected by the type of figure, in the sense that this drawing effect would be the greatest for the (non-proportional) S items and the lowest for the (non-proportional) I items. The rationale behind these predictions is exemplarily worked out for the non-proportional item about squares given in Table 2-1. To find the answer to this item, the problem solver can choose among three appropriate solution strategies: (1) 'paving' the big square with little ones, (2) calculating and comparing the areas of both squares by means of the formula ‘area = side × side’, and (3) applying the general principle ‘if length × r, then area × r^2’. For the corresponding non-proportional item about circles, the first solution strategy is less obvious and can only provide an approximate answer, while the second strategy is more error-prone (because of the greater complexity of the formula for finding the area of a circle). For the corresponding non-proportional item dealing with irregular figures, a formula for calculating the area is not available and the paving strategy – or a variant, namely approximately transforming the irregular figure into one or more regular ones – does not provide an ‘exact’ answer. Applying the general principle is the only ‘direct’ solution strategy for this item.

2.3 Analysis

All responses on the proportional and the non-proportional items were categorized as ‘correct’ or ‘incorrect’. A response was considered correct when it was the result of a mathematically appropriate reasoning process; therefore, answers that differed from the correct one because of a technical mistake in a correct overall solution process were considered correct too. All other kinds of erroneous answers were scored as incorrect. Because a detailed analysis of a random sample of 300 incorrect answers on non-proportional items revealed that 95% of them resulted from an inappropriate linear reasoning process (see De Bock, Verschaffel, & Janssens, 1996), we decided not to split up the incorrect answers any further in the present analysis.

The hypotheses were tested by means of a ‘3 × 2 × 2 × 3’ analysis of variance (ANOVA) with ‘Group’ (Group I, II, and III), ‘Test’ (Test 1 vs. Test 2), ‘Proportionality’ (proportional vs. non-proportional items), and ‘Figure’ (squares, circles, and irregular figures) as independent variables, and the number of ‘Correct answers’ as the dependent variable. Significant main and interaction effects were further analyzed using a posteriori Tukey tests.

2.4 Results with respect to the hypotheses

Table 2-2 gives an overview of the percentages of correct responses of the three groups of students (I, II, and III) on the proportional and the non-proportional problems involving squares (S), circles (C), and irregular figures (I) in Test 1 and 2.

The results provide a very strong confirmation of the first hypothesis. Indeed, there was a strong difference in the performance on the proportional and non-proportional problems[4]. For the three groups and the two tests together, the percentages of correct responses for all proportional and for all non-proportional items were 92% and 2%, respectively. This shows that most 12–13-year olds were able to solve the proportional items correctly, whereas the non-proportional items were seldom solved correctly.

The results did not support the second hypothesis concerning the beneficial effect of drawings on students' performance. For none of the three groups did we find a significant increase in the students' scores from Test 1 to Test 2 in general, or in their performance on the non-proportional items in particular. For Group I (= the control group) the percentage of correct responses on the non-proportional items even decreased slightly from 2% to 1% between Test 1 and Test 2. For Group II (= the self-made drawing group) this percentage unexpectedly remained the same at 2% during both tests. For Group III (= the ready-made drawing group) the percentage of correct answers on the non-proportional items increased slightly between Test 1 and Test 2, from 2% to 5%, but remained still extremely low during the latter test. In sum, the anticipated beneficial impact of the instruction to make drawings (in Group II) and of the provision of ready-made drawings (in Group III) was too weak to break the predominance of the linear model in the reasoning of these 12–13-year olds.

In line with the third hypothesis, the type of figure had a significant effect on the percentage of correct responses[5]. The scores for the S, the C, and the I items were in the expected direction – the overall percentages of correct answers for these three kinds of problems were 49%, 48%, and 45%, respectively, but only the difference between the S items and the I items and between the C items and the I items was statistically significant. Moreover, the observed effect of the type of figure was found in the proportional items as well as in the non-proportional items.

Table 2-2. Percentage of correct responses of the three groups of 12–13-year olds on the different categories of proportional and non-proportional items in Test 1 and Test 2

Group	Test 1						Test 2					
	Proportional items			Non-proportional items			Proportional items			Non-proportional items		
	S	C	I	S	C	I	S	C	I	S	C	I
I	96	98	89	5	0	1	99	96	85	3	0	0
II	93	95	89	6	1	0	93	95	95	4	2	0
III	91	91	87	4	3	0	93	89	89	8	5	1

4 'Proportionality' main effect: $F(1,117) = 4994.92$, $p < .01$
5 'Type of figure' main effect: $F(2,234) = 12.96$, $p < .01$

2.5 Additional findings

After presenting the quantitative results with respect to the three research hypotheses, we also briefly discuss some qualitative findings based on a systematic and fine-grained analysis of the students' written notes on the response sheets, which may help to explain these quantitative results.

First, the analysis of the notes on the response sheets of Test 1 revealed that only 2% of the students *spontaneously* constructed a sketch or drawing of the non-proportional problems. Apparently, these 12–13-year olds were not inclined to apply the heuristic of making a sketch or a drawing when modelling and solving a verbally stated geometric problem situation.

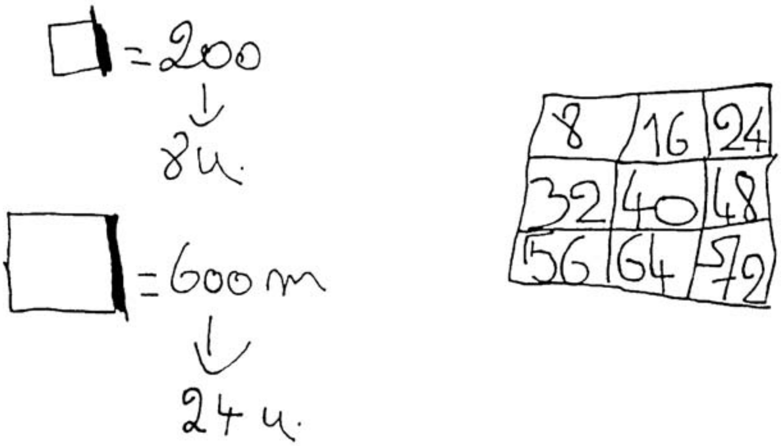

Figure 2-2. Examples of self-made drawings with low (left) and high (right) representational quality

Second, the inspection of the notes on Test 2 in Group II revealed that – in spite of the explicit instruction to do so – the students produced drawings for the non-proportional items in only 46% of the cases. It remains unclear why even the students from Group II who *did* follow the instruction to make a drawing still failed to solve the corresponding non-proportional item. Possibly, the representational quality of these drawings (both in terms of correctness and richness) was mostly too low to really help students in interpreting and solving these non-proportional items correctly. For examples of self-made drawings with a low and a high representational quality, we refer, respectively, to the drawings left and right in Table 2-2, both belonging to the non-proportional item about squares in Table 2-1. On the basis of the available data we cannot determine whether the poor quality of most student drawings was due to their inability to make better drawings or to their unwillingness to make drawings for word problems that seemed trivial to them.

Third, although the notes of the students from Group III on Test 2 provide no *direct* information about the extent to which the ready-made drawings were effectively used by these students, the finding that only 6% of the given drawings for the non-proportional items had been ‘edited’ by these students as part of their problem-solving process, suggests that they generally paid little or no attention to them.

2.6 Conclusion and discussion

The first study convincingly demonstrated the strength and omnipresence of the linear model with respect to problems involving length and area of similar plane figures in 12–13-year old students. However, for two reasons we found it necessary to set up a follow-up study with an older target group.

First, the extremely small number of correct responses on the non-proportional items made us wonder how strong the predominance of linearity would be for students who were older and, therefore, probably mathematically better equipped for overcoming the obstacle of unlimited linear reasoning.

Second, because the predominance of the linear model proved to be so strong for 12–13-year olds, the first study did not yield adequate information about the possible impact of both the kind of figure used in the problem (square, circular or irregular) and of self-made or ready-made drawings on the occurrence of errors based on inappropriate proportional reasoning. For instance, because of the extremely low occurrence of correct answers on the non-proportional items in Study 1, it was impossible to perform a statistical analysis of the relationship between making and/or using drawings, on the one hand, and producing correct answers, on the other hand. A follow-up study with older students – who were expected to suffer less from the predominance of the linear model – should result in a better understanding of the effect of these two additional task variables.

3. STUDY 2

3.1 Subjects, materials and procedure

For this follow-up study, we decided to work with 10th-graders (15–16-year olds). This choice was induced by the Flemish mathematics program for

Grades 9–10 (Ministerie van de Vlaamse Gemeenschap, 2002) which includes a systematic study of plane geometry (including similarities and scaling in relation to mensuration). Furthermore, in Grades 9 and 10 students also learn how to identify and use a diversity of non-proportional functional relationships (e.g. quadratic and general polynomial functions).

To work with a cohort of participants of an intellectually, educationally and socially comparable level to that of the first study, we decided to undertake the second study in the same school as the first one. Participants were the 222 10th-graders (15–16-year olds).

Contrary to the first study, the test was administered only once to all students. Therefore, students were immediately matched in three equivalent groups. For practical reasons, it was not possible to match the students from the three experimental groups on an individual basis; rather, we had to work with intact classes. The selection and placement of the distinct classes in the three experimental groups was based on the following subject variables: (1) the study stream to which the students of a certain class belonged (e.g., 'Ancient languages', 'Modern languages', 'Economics', ...), (2) the number of hours a week spent at mathematics, and (3) the students' results on the previous mathematics examination. Using these available data, three groups were formed in which (1) the different study streams were equally represented, (2) the average number of hours of mathematics per week was the same, and (3) the average result on the previous mathematics examination was similar (for a detailed description of the characteristics of the three experimental groups, see De Bock et al., 1996). The testing conditions for the three experimental groups were the same as in Test 2 of Study 1, i.e. in Group I no special help or instructions were given, in Group II students were explicitly instructed to make a drawing before computing their answer, and in Group III every item was accompanied by a correct ready-made drawing.

The same 12 experimental items from the first study (Table 2-1) were used. The testing procedure, the layout of the response sheets, and the data-analysis procedure were also identical.

3.2 Hypotheses

The predictions of Study 2 were largely the same as for Study 1. However, with regard to the first hypothesis, we assumed that 15–16-year olds would suffer less from the predominance of the linear model than 12–13-year olds. Therefore, we predicted that they would perform better on the experimental items in general and on the non-proportional ones in particular.

The hypotheses regarding the influence of self-generated or ready-made drawings (= hypothesis 2 from Study 1) and of the type of geometrical figure involved (= hypothesis 3 from Study 1) remained unchanged.

3.3 Results with respect to the hypotheses

Table 2-3 lists the percentages of correct responses for the three groups of 15–16-year olds (I, II, and III) for the proportional and non-proportional items involving squares (S), circles (C), and irregular figures (I).

Table 2-3. Percentage of correct responses of the three groups of 15–16-year olds on the different categories of proportional and non-proportional items

Group	Proportional items			Non-proportional items		
	S	C	I	S	C	I
I	91	97	97	26	11	1
II	89	91	97	26	20	5
III	89	93	97	39	21	7

The first hypothesis was confirmed. Once again, there was an extremely strong difference in performance on the proportional and non-proportional items[6]. For all three groups together, the percentages of correct responses on the proportional and non-proportional items were 93% and 17%, respectively. (In Study 1, these percentages were 92% and 2%). A comparison of the results from the two studies revealed that only on the non-proportional items was the difference in the number of correct responses between the 7^{th}- and the 10^{th}-graders significant[7].

The second hypothesis about the positive influence of drawings was again not confirmed. Global percentages of correct responses in Groups I, II, and III showed indeed a positive trend – i.e. 54%, 55%, and 58% correct answers, respectively – but the differences were again too small to produce a significant effect.

As predicted in the third hypothesis and in accordance with the results of Study 1, the type of figure did make a difference[8]. Global percentages of correct responses for S, C, and I items were in the expected direction (60%, 56%, and 51% correct answers, respectively), and all mutual differences between these three problem types were significant. Interestingly, the percentages of correct responses on the non-proportional items were in the expected direction (30%, 17%, and 4% for the S, C, and I items,

6 ‘Proportionality’ main effect: $F(1,219) = 1591.64$, $p < .01$

7 ‘Age’ × ‘Proportionality’ interaction effect: $F(1,340) = 25.33$, $p < .01$

8 ‘Type of figure’ main effect: $F(2,438) = 24.67$, $p < .01$

respectively), whereas the percentages of correct responses on the proportional items were in the opposite direction (90%, 94%, and 97% for the S, C, and I items, respectively).

3.4 Additional findings

Students' use of self-made and ready-made drawings. First, we will describe some additional findings concerning the use of drawings in students' solutions of the non-proportional items. According to hypothesis 2, the instruction to make a drawing and – even more – the provision of a ready-made drawing should help students to overcome the obstacle of unlimited linear reasoning. Therefore, we predicted better results (on the non-proportional items) for Groups II and III than for Group I. But, as said before, the results of Study 2 again yielded no confirmation of this prediction. The qualitative analysis of the response sheets of the students of Group II and III in Study 1 suggested that the instruction to make drawings – and even the given drawings – were often ignored by these students. Therefore, the absence of a significant difference between the three groups in both studies did not allow the conclusion that the *actual* making of a drawing and the *actual* use of a ready-made drawing did not have a positive influence on students' performance. To make that conclusion, we first needed to demonstrate that there was no relationship between *actually* making a drawing or *actually* using a ready-made one, on the one hand, and giving a correct response to a non-proportional item, on the other hand. Because of (1) the extremely low overall score on the non-proportional items and (2) the relatively small number of self-generated and 'edited' drawings in the first study, it was impossible to analyze (statistically) the relationship between students' *actual* use of these drawings and their performance on the non-proportional items. Study 2, however, did allow such an analysis, because of the greater number of (1) self-generated and 'edited' drawings for the non-proportional items and of (2) correct responses on these items. For each of the three groups (Group I, II, and III) a contingency table with the variables 'Drawing' and 'Answer' was constructed (Table 2-4), and the (in)dependence of the two variables was investigated.

Chi-square tests revealed that this null hypothesis must be rejected in all groups[9]. The comparison of the observed and expected frequencies in the various cells for Group I, II, and III shows that the dependence of the two variables is in the expected direction. In all three tables the number of subjects in the cells on the main diagonal (i.e., 'drawing/correct answer' and

[9] $\chi^2(1, N = 444) = 16.18, 25.26$ and 14.82 for Group I, II, and III, respectively

‘no drawing/incorrect answer’) is greater than could be expected if these two variables were mutually independent. To investigate if the application of a drawing indeed *provoked* the apprehension of non-linearity, or if – conversely – this heuristic was used more often by students who already had detected the non-linear nature of the mathematical model underlying the problem, we compared the occurrence of spontaneous drawings for proportional and non-proportional items. This analysis revealed that the incidence of these drawings was not significantly higher for non-proportional than for proportional items, which suggests that students’ insight in the non-linear character of an item generally *came after* the (successful) application of the drawing heuristic.

Table 2-4. Contingency tables for Group I, II, and II (Frequencies and percentages between brackets represent the expected cell frequencies and percentages under the null hypothesis of independence of both variables)

Group I	Spontaneous drawing	No spontaneous drawing	Row totals
Correct answer	17 (7) 4% (2%)	41 (51) 9% (11%)	58 13%
Incorrect answer	40 (50) 9% (11%)	346 (336) 78% (76%)	386 87%
Column totals	57 13%	387 87%	444 100%

Group II	Drawing made	No drawing made	Row totals
Correct answer	65 (45) 15% (10%)	12 (32) 3% (7%)	77 17%
Incorrect answer	196 (216) 44% (49%)	171 (151) 39% (34%)	367 83%
Column totals	261 59%	183 41%	444 100%

Group III	Given drawing edited	Given drawing not edited	Row totals
Correct answer	32 (19) 7% (4%)	66 (79) 15% (18%)	98 22%
Incorrect answer	53 (66) 12% (15%)	293 (280) 66% (63%)	346 78%
Column totals	85 19%	359 81%	444 100%

Students’ solution strategies. Because the 15–16-year olds in Study 2 produced considerably more correct responses on the non-proportional items than the 12–13-year olds from Study 1, it also became possible to get further insight into the kind of strategies underlying these correct answers. For this purpose, all solution processes underlying a correct answer on a non-proportional item were scored in terms of one of the following categories: (1) the ‘paving’ strategy, (2) the strategy of computing and comparing the length or area of both figures, and (3) the strategy of applying the general rule. For instance, the non-proportional item about squares in Table 2-1 can be solved in three different ways:

1. ‘Paving’ the big square with little ones, observing that there are 9 little squares, and then concluding that the farmer will therefore need 9 times 8 hours = 72 hours;
2. Calculating the area of both squares (200 × 200 = 40 000 m^2 and 600 × 600 = 360 000 m^2), determining the result of the division (360 000 : 40 000 = 9), and therefore concluding that the farmer will need 9 times 8 hours = 72 hours; and
3. Immediately applying the general rule ‘side × *r*, thus area × r^2’ (‘if the length is multiplied by 3, then the area (and thus the fertilising time) needs to be multiplied by 9’).

A detailed analysis of the written protocols of all correct solutions revealed that the second solution strategy was – by far – the most frequent one. For instance, for the non-proportional items about squares, this strategy (sometimes used in combination with one of the other methods) was applied for 90% of the correctly solved non-proportional items. ‘Pure’ applications of the first and the third strategy were very rare (7% and 3% of the correct solution strategies, respectively). The low frequency of the ‘paving’ strategy is quite remarkable. Indeed, ‘paving’ is a very easy, intuitive, context-bound method requiring only little sophisticated formal-mathematical knowledge. The rare use of this informal strategy was probably due to students’ well-documented belief that solving a mathematical problem is primarily a matter of finding and executing the correct mathematical formula(s) previously taught in school (Schoenfeld, 1992; Verschaffel, et al., 2000).

Students’ errors on proportional items. With respect to the type of figure, the percentages of correct responses on the different kinds of non-proportional items were in the expected direction (i.e. highest score for the S items and lowest score for the I items), whereas for the proportional items the percentages of correct responses were in the *opposite* direction (i.e. a higher percentage of correct answers for the I items than for the S items). A detailed analysis of students’ incorrect responses on the proportional items showed that they were typically the result of an inappropriate *non-proportional* reasoning process. Apparently, students found it easier to discover the non-proportional nature of a given problem situation when a square figure was involved (and to some extent also when the figure had a circular form), but, as a result of this discovery, they sometimes began to question the correctness of a linear model for problem situations wherein that model *was* appropriate. In this respect, we point out that also in Study 1 some of the rare students who produced correct answers on non-proportional items, erroneously applied non-proportional reasoning for one or more proportional items too.

4. DISCUSSION

The major results of this first pair of studies were as follows. First, the tendency to apply linear reasoning in the solution of non-proportional problems proved to be extremely strong in the age group of 12–13-year olds, and was still very influential among 15–16-year olds. The 12–13-year olds solved 92% of the non-proportional items correctly, whereas they only answered 2% of the non-proportional items correctly; in the age group of 15–16-year olds, the overall percentages of correct responses on the proportional and non-proportional items were, respectively, 93% and 17%. These data, which can be interpreted as strong evidence for students' overuse of linearity, elicited a lot of amazement and unbelief among practitioners to whom we presented the results of these studies. Most of them were aware of the issue, but had not realised that it affected their students' solutions so strongly.

Second, the type of figure played a significant role. Students performed significantly better on the non-proportional items when the enlarged figure involved was regular (a square or a circle), but, as a drawback, they performed worse on the proportional items about these regular figures, because some students started to apply non-proportional reasoning on the proportional items too! This latter finding seems to indicate that the knowledge base underlying students' (rare) correct answers on non-proportional items was typically still quite fuzzy and unstable. Third, we unexpectedly did not find a beneficial effect of the self-made or given drawings, either for the test as a whole or for the non-proportional items in particular. An additional qualitative analysis of the data revealed that the vast majority of correct responses on the non-proportional items were found by applying an appropriate mathematical formula; informal strategies, such as paving a self-made or a given drawing, were chosen far less. It appears that we did not succeed in integrating the drawing activity or the given drawings in students' problem-solving process.

While this first set of investigations documented students' strong tendency to apply linear reasoning in the solution of non-proportional problems, it did not yield an explanation for this tendency. One could argue that students' unbridled use of the linear model was just an artefact of certain elements in our experimental setting. In that case, it would be possible to improve students' performance on the non-proportional items by simply modifying these elements. In the next chapter, we report a series of follow-up studies, each testing one or more particular hypotheses with respect to the influence of the testing setting on students' tendency to overuse linearity.