Do realistic contexts and graphical representations always have a beneficial impact on students' performance? Negative evidence from a study on modelling non-linear geometry problems

D. De Bock [a,b,*], L. Verschaffel [a], D. Janssens [c],
W. Van Dooren [a,1], K. Claes [a]

[a] *Center for Instructional Psychology and Technology (CIP&T), University of Leuven, Vesaliusstraat 2, B–3000 Leuven, Belgium*
[b] *Europese Hogeschool Sint-Aloysius (EHSAL), Brussels, Belgium*
[c] *Department of Mathematics, University of Leuven, B–3000 Leuven, Belgium*

Abstract

The aim of the present study was to investigate the influence of authentic contexts and of self-made graphical representations on students' well-documented tendency to improperly apply the linear model to represent and solve non-proportional word problems about area and volume. A paper-and-pencil test on this kind of geometrical problems was collectively administered in different experimental groups of 13–14- and 15–16-year old students. Problem authenticity was experimentally enhanced for half of the students by prefacing the test by an assembly of well-chosen video fragments telling the story of Gulliver's visit to the world of the Lilliputians and by linking all test items directly to these video fragments. The impact of self-made graphical representations was examined by asking half of the students to draw a reduced copy of the geometrical figure described in the problem before actually solving it. None of the experimental manipulations yielded the expected results. To the contrary, both factors even yielded a negative effect on students' performance. Several explanations for these unexpected results are discussed.
© 2003 Elsevier Science Ltd. All rights reserved.

* Corresponding author. Tel: +32-16-32-57-55; fax: +32-16-32-62-74.
   E-mail address: dirk.debock@avl.kuleuven.ac.be (D. De Bock).

1. Theoretical and empirical background

Because of its wide applicability—both in everyday life situations and in scientific contexts—proportional reasoning constitutes a major topic in mathematics education at the elementary and secondary school. But according to several researchers and educators, the attention given to the proportional (or linear) model may have a serious drawback: it may lead to the illusion that this model has a 'universal' applicability and to develop in students a tendency to use the linear model also in situations in which it is not applicable (Freudenthal, 1983; Gagatsis & Kyriakides, 2000). This tendency to see and apply the linear model 'everywhere' is often referred to by names as the 'illusion of linearity', 'linear trap', 'linear obstacle' or 'linear misconception' (De Bock, Verschaffel, & Janssens, 1998). The literature contains several examples of the occurrence of this illusion in different domains of mathematics, such as elementary arithmetic (Säljö, 1991; Verschaffel, Greer, & De Corte, 2000), algebra (Berté, 1993; Gagatsis & Kyriakides, 2000; Matz, 1982), geometry (De Blocq-Docq, 1992; National Council of Teachers of Mathematics, 1989) and probability (Fischbein & Schnarch, 1996; Freudenthal, 1973; Shaughnessy, 1992; Van Dooren, De Bock, Verschaffel, & Janssens, 2001), and in diverse scientific disciplines, such as physics (Anderson, 1983).

Undoubtedly, the most famous and most often quoted example of the occurrence of the linearity illusion can be found in Plato's dialogue *Meno* (Berté, 1993; Rouche, 1992) in which a slave, when asked to draw a square having two times the area of a given square, initially proposes to *double the side* of that square. So, the slave spontaneously applies the idea of linear proportionality (between length and area) and changes his mind only when Socrates helps him in diagnosing and correcting the error in his reasoning by confronting him with a drawing. Recent research by De Bock et al. (1998, 2002), that will be summarised in the next section, provides strong evidence that—more than two millennia later—an alarming number of students still consider the relations between length and area or between length and volume in similar geometrical figures as linear instead of quadratic or cubic, and apply the linear scale factor instead of its square or cube to determine the area or volume of an enlarged or reduced figure.

A second typical example, issued from the field of probabilistic thinking, is students' misbelief that the probability of at least one success in a game of chance is proportional to the number of trials (e.g. the probability of 'at least one six' in n rolls with a fair die equals $n \times 1/6$) (De Bock, Verschaffel, & Janssens, 1999; Van Dooren et al., 2001). This example relates to another historical problem, namely the 'problem of the dice' Chevalier de Méré posed to his friend Pascal (Freudenthal, 1973). By experience, de Méré knew the advantage of betting on the event 'at least *one six* in *4* rolls of one fair die' and he deduced that it must be equally advantageous

to bet on 'at least *one double-six* in *24* rolls of two fair dice'. From 4 to 24 rolls, the number of opportunities to reach a successful outcome increases by factor six, but, correspondingly, the likelihood of reaching a successful outcome with two dice instead of one, divides by six, so the two effects cancel out and the probability remains the same. This proportional reasoning procedure in a binomial probabilistic setting is clearly erroneous—one of the underlying assumptions, namely the proportional increase of probability with the number of trials, leads to probabilities larger than one!—and elicited the following (sharp) comment by Freudenthal:

> He [*de Méré*] applied the mathematics he knew, the kind of mathematics which in my childhood was called the rule of three… Maybe he would have performed better if he had never learned mathematics at all! Then there would have been some chance that he would have applied not the mathematics he had learned but the mathematics that he would have to create himself (Freudenthal, 1973: p. 585).

Until recently, empirical data on the scale and persistence of the phenomenon of unbridled proportional reasoning were practically absent. A first attempt to fill this gap in the research literature was a series of studies by De Bock et al. (1998, 2002) on the occurrence of the illusion of linearity as well as on its resistance to change in students working on word problems about the relationship between the linear measurements and the area of similar plane figures. In these studies, paper-and-pencil tests consisting of the same set of proportional and non-proportional items about length and area of similar plane figures were collectively administered to large groups of 12–13- and 15–16-year old students under different experimental conditions. The items were formulated as traditional word problems and were built around different types of plane figures (regular figures such as squares and circles, as well as irregular figures). The next problem—a non-proportional item about the enlargement of a square—is an example of such an item: "Farmer Carl needs approximately 8 hours to manure a square piece of land with a side of 200 m. How many hours would he need to manure a square piece of land with a side of 600 m?" Of course, when the side of the square piece of land is multiplied by 3, the area has to be multiplied by 9 and thus, also the manuring time which is, by approximation, directly proportional to the area, has to be multiplied by 9 too.

The major results of these studies were as follows. First, the tendency to apply proportional reasoning in the solution of the non-proportional problems proved to be extremely strong in the age group of 12–13-year olds (2–7% correct responses on non-proportional items), and was still very influential among 15–16-year olds (17–22% correct responses). It was shown that almost all incorrect answers resulted from an inappropriate proportional reasoning process. Second, the type of figure involved 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), than when it was irregular (such as a map of a country). Third, neither explicitly instructing the students to make a sketch or drawing of the problem situation before computing their answer, nor providing them with ready-made drawings of the original and the enlarged figure had a beneficial effect on students' perform-

ance. As evidenced by a detailed analysis of the notes on their answer sheets, students often ignored the instruction to make drawings and even the given drawings. However, students who *actually followed* the instruction to make a drawing or who *actually used* the given drawings were more successful in discovering the incorrectness of a stereotypical, linear proportional reasoning and in finding the correct solution for the non-proportional items. Fourth, the provision of metacognitive support (by confronting students with a correct and an incorrect solution strategy for a representative non-proportional item immediately before the start of the actual test), yielded a significant, but still remarkably small effect on students' performance on the non-proportional items. Fifth, the misuse of the proportional model proved to be partially caused by the formulation of the items in a traditional 'missing-value' format (in which three numbers are given and a fourth unknown number has to be determined; see, e.g., the non-proportional item about the enlargement of a square mentioned before). Indeed, students confronted with mathematically equivalent problems, but (re)phrased as 'comparison problems' (such as, e.g., "Farmer Carl manured a square piece of land. Tomorrow, he has to manure a square piece of land *with a side being three times as big*. How much more time would he approximately need to manure this piece of land?"), resisted more easily the trap of proportional reasoning than students confronted with familiar missing-value problems. However, students receiving these comparison problems still solved the non-proportional items erroneously in more than half of the cases. Sixth, each time a certain experimental manipulation made it—at least for some students—easier to discover and resolve the non-proportional nature of a problem, this positive effect was accompanied by the overgeneralisation of the non-proportional reasoning schema towards proportional situations among these students, revealing the fragile and unsteady nature of their emerging non-proportional reasoning schemata.

In sum, the studies by De Bock et al. (1998, 2002) showed that the vast majority of 12–16-year-old students failed on non-proportional word problems about length and area of similar plane figures, because of their alarmingly strong tendency to apply proportional reasoning 'everywhere'. Although several experimental manipulations carried out so far yielded significant positive effects on students' performance on non-proportional items, these effects were disappointingly small and did thus not suffice to make disappear the illusion of linearity. On the contrary, the tendency towards unbridled proportional modelling turned out to be a widespread phenomenon affecting many students from a broad range of ages and in different kinds of experimental settings, and being very resistant to change. The present study focuses on two new experimental manipulations that were expected to result in a considerable increase of students' success rates on representing and solving this type of non-proportional geometry problems.

First, it can be argued that students' weak performance on the non-proportional items in the previous studies is—at least partially—caused by the inauthenticity of the problem context and setting. Over the past decades, several researchers and educators have pointed out the benefits of constructing and organising mathematical activities around rich, attractive and realistic contexts (de Lange, 1987; Freudenthal, 1983; Treffers, 1987). A dual function is assigned to these contexts: (1) they act

upon students' intrinsic motivation and task involvement and thus, enhance the likelihood that they will do a serious effort to complete the problem; and (2) they help them to make a correct representation of the problem and to find a correct solution strategy by eliciting the activation and use of prior contextualised knowledge (real-world experiences, intuitions, models, strategies, …) that is helpful for understanding and solving the problem. According to the before-mentioned authors, realistic contexts cannot only refer to aspects of the 'real' social or physical world; they can also refer to imaginary, fairy-like worlds as long as they are meaningful, familiar and appealing to the students. It is not the amount of realism in the literal sense that is crucial for considering a context as realistic or authentic, but rather the extent to which it succeeds in getting students involved in the problem and engage them in situationally meaningful thinking and interaction. The problems used by De Bock et al. (1998, 2002), were merely constructed for research purposes, and were traditional word problems built around rather 'poor' contexts having no special meaning or attractivity to 12–16-year youngsters nowadays. Therefore, students' performance on the non-proportional items might increase significantly if one would succeed in increasing the authenticity of these problems for the students. Some support for this assertion can be found in the work of scholars of the Freudenthal Institute, who have explored students' difficulties with the influence of linear enlargement on area and volume in realistic contexts (like the instructional units 'With the giant's regards' in Streefland, 1984, or 'Gulliver' in Treffers, 1987). In 'With the giant's regards' for example, the hand of a giant, being four times as long as a human hand, was drawn on the blackboard and pupils of Grade 3 were asked to estimate various lengths, areas and volumes in the giant's world (e.g. How big would the giant's newspaper be? How much would he weigh? …). According to Streefland (1984), these activities proved to be helpful for many students in breaking the 'pattern of linearity' and making them aware of the 'multi-dimensional' impact of increase on area and volume. Treffers (1987) adopted and elaborated these ideas in a design experiment with sixth graders built around the context of 'Gulliver's travels'. In his report, Treffers states: "This question [*How many Lilliputian handkerchiefs make one for Gulliver?*] introduces the influence of linear enlargement on area. The students have no difficulty with the problem." (Treffers, 1987: p. 5). Unfortunately, Treffers does not support this latter statement with empirical data.

Second, there is a vast amount of theoretical and empirical research on how and why drawings and diagrams may enhance people's ability to represent and solve (mathematical) problems. In this respect, we refer the reader to Larkin and Simon's (1987) famous article 'Why a diagram is (sometimes) worth ten thousand words,' but also to several more recent meta-analyses on the role of diagrammatic and pictorial representations in (mathematical) problem solving (Aprea & Ebner, 1999; Hall, Bailey, & Tillman, 1997; Reed, 1999; Vlahovic Stetic, 1999) as well as to the literature on the role of heuristics in skilled (mathematical) problem solving in general (Collins, Brown, & Newman, 1989; De Corte, Greer, & Verschaffel, 1996; Pólya, 1945; Schoenfeld, 1992). According to all of these authors, the heuristic of visualising the problem using a drawing or diagram does, of course, not guarantee that one will find the solution of a given problem. But because it induces a systematic analysis and elabor-

ation of the problem situation, because it enhances a planned solution of the task, and because it can also be used in interpreting and checking one's answer, making a drawing or a diagram is generally considered a very helpful and successful cognitive tool in (mathematical) problem solving. In accordance with the active, constructive, and self-regulated view of effective learning (De Corte, 1996), the positive effect of visualisations is considered to be even stronger when they are made by the learners themselves rather than merely being given to them by the teacher or the researcher, since self-made visualisations stimulate an even more deep-level and mindful approach of the task (Aprea & Ebner, 1999; Dirkes, 1991). Of course, although generating a drawing increases the chance that a problem will be conceptualised correctly, a drawing that reflects an incorrect understanding of the problem will be of little help for the problem solution (Van Essen & Hamaker, 1990). Applying these considerations to the topic of the present investigation, namely the illusion of linearity, it can be hypothesised that giving students the instruction to make a drawing that explicitly shows the topological and geometrical relations among the essential components of the problem before actually solving it, will also result in a significant improvement of their score on non-proportional items. However, we remind the reader that De Bock et al. (1998) found no significant differences between students who solved non-proportional problems without drawings (condition 1), on the one hand, and students who received the instruction to make drawings themselves (condition 2) or were provided with ready-made drawings (condition 3), on the other hand. In retrospect, two elements may have caused this unexpected finding: (1) as reported above, only a very small number of the students in the two drawing conditions effectively made a drawing themselves or effectively used the given drawing; and (2) due to the way in which the drawings and the drawing instructions were presented in that study, the construction of a drawing or the use of a given drawing did not constitute a genuine part of the problem solving process (De Bock et al., 1998). For those reasons, we were reluctant to consider the absence of a significant effect of drawings or drawing instructions in that previous study as convincing evidence against the drawing hypothesis, but rather as an encouragement to operationalise the drawing factor in a more appropriate way in further research.

The present study was executed to empirically test the facilitating power of the two experimental manipulations mentioned before, namely (1) making the problem context and setting more authentic; and (2) making the drawing activity a truly integrative part of the problem solving process.

2. Design

Hundred and fifty-two eighth graders (13–14-year olds) and 161 tenth graders (15–16-year olds) participated in the study. The participants came from a school for general secondary education that is located in a medium-sized Flemish town and that pursues high educational standards. Boys and girls were almost equally represented in the school as well as in our sample. All participants were confronted with a paper-and-pencil test consisting of six experimental items (two proportional

items and four non-proportional items) about the relationships among the lengths, areas and volumes of different types of rectilinear and non-rectilinear figures, together with some buffer items. Students were maximally prevented from making calculation errors by allowing them to use their own scientific calculator while doing the test. To allow a proper evaluation of the two experimental manipulations mentioned before, we applied a 2×2-design with four matched groups at both age levels. The matching of the four groups was done by guaranteeing that: (1) the different study streams provided in the school were equally represented in the four groups; (2) the average number of hours a week students from the four groups spent at mathematics was the same; and (3) students' average result on last year's final mathematics examination did not significantly differ between the four groups (Claes, 2000). Four different versions of the test were administered: problems were administered in or out of an authentic context and setting (A+ and A− condition), and either with or without an integrated drawing instruction (D+ and D− condition), leading to four combinations (A+D+, A+D−, A−D+ and A−D−) that were administered to the four matched groups at each age level.

In the two A+ groups, students' involvement in the problems was experimentally enhanced by preceding the test by a series of well-chosen video fragments of a screen-version (Kenworthy & Sturridge, 1996) of Jonathan Swift's world-famous novel 'Gulliver's travels', first published in 1726. All fragments were selected from the story of Gulliver's visit to the isle of the Lilliputians, a world in which all lengths are 12 times as small as in our (and Gulliver's) world, and were assembled to tell in 10 min the whole plot in a visually attractive way. To direct students' attention to the relevant features while watching these video fragments, they were instructed to look carefully at the ratio's between the Lilliputian and Gulliver's world and it was announced that they would have to solve a test consisting of mathematical problems linked to the video fragments afterwards. Because the linear scale factor was imposed by the context, all items involved reductions by factor 12, which was clearly announced in the following way at the beginning of the test.

> In the world of the Lilliputians, all lengths are 12 times as small as in our world, the world of Gulliver. A tree, being 12 m high in our world, would thus be only 1 m high in the world of the Lilliputians. A road, being 12 km long in our world, would be only 1 km long in the Lilliputian world. And, of course, the Lilliputians themselves are also 12 times as short as Gulliver.

All problems were stated in a so-called comparison format, in which one measure (a length, an area or volume) of an object in Gulliver's world was given, and students had to determine a corresponding measure in the Lilliputian world (which arithmetically leads to divisions by 12, 12^2, or 12^3 for, respectively, one, two, or three dimensional quantities). In the two opposite A− groups, the experimental items were equivalent from a mathematical point of view, but were presented in the form of a series of non-related traditional school problems without any attempt to embed the mathematical task in a contextually meaningful and attractive setting. To bring some

variation in the test, three different reducing factors were used (11, 12 and 13). Table 1 gives an overview of the test items used in the A+ and A− groups.

In the D+ groups, students were provided with a drawing of the (geometrical) object introduced in the problem and they were firstly asked to complete the drawing by making, next to the given original, a reduced copy using the given scale factor (which was, as mentioned before, always 12 in the A+ groups, and 11, 12 or 13 in the A− groups). To assure that the students would perform the drawing task as a genuine part of their solution process, the drawing instruction was given before all necessary numerical data were presented. Fig. 1 exemplifies how this drawing task was presented on students' response sheets for the first non-proportional item in the A+ and A− groups given in Table 1.

In all four experimental conditions, students could not only write the answer on their response sheets, but also to make calculations, sketches or comments. When finished with their test, students were asked to complete a short questionnaire consisting of several statements, which had to be rated on a five-point scale (from full disagreement to full agreement). Beside six general statements (G1, G2, G3, G4, G5

Table 1
Experimental items used in the A+ and A− groups

Test items in A+ groups	Test items in A− groups
Proportional items	
Gulliver's walking-stick is 96 cm high. What's the height of a similar Lilliputian walking-stick? (*Answer: 8 cm*)	The length of a line segment *A* is 13 times as large as the length of a line segment *B*. If the length of line segment *A* is 78 cm, how long is line segment *B*? (*Answer: 6 cm*)
Gulliver's waistband is 108 cm long. What's the length of a similar Lilliputian waistband? (*Answer: 9 cm*)	The length of a line segment *C* is 11 times as large as the length of a line segment *D*. If the length of line segment *C* is 132 cm, how long is line segment *D*? (*Answer: 12 cm*)
Non-proportional items	
Gulliver's handkerchief has an area of 1296 cm^2. What's the area of a similar Lilliputian handkerchief? (*Answer: 9 cm^2*)	The side of square *Q* is 12 times as large as the side of square *R*. If the area of square *Q* is 1440 cm^2, what's the area of square *R*? (*Answer: 10 cm^2*)
The sole of Gulliver's shoe has an area of 288 cm^2. What's the area of the sole of a similar Lilliputian shoe? (*Answer: 2 cm^2*)	The diameter of a circle *E* is 11 times as large as the diameter of a circle *F*. If the area of circle *E* is 242 cm^2, what's the area of circle *F*? (*Answer: 2 cm^2*)
In Gulliver's world, a cheese cube has a volume of 172 800 mm^3. What's the volume of a Lilliputian cheese cube? (*Answer: 100 mm^3*)	The side of a cube *I* is 13 times as large as the side of a cube *J*. If the volume of cube *I* is 2197 cm^3, what's the volume of cube *J*? (*Answer: 1 cm^3*)
In Gulliver's world, a wineglass has a volume of 1728 mm^3. What's the volume of a similar wineglass in the Lilliputian world? (*Answer: 1 mm^3*)	The diameter of sphere *M* is 12 times as large as the diameter of a sphere *N*. If the volume of sphere *M* is 172 800 mm^3, what's the volume of sphere *N*? (*Answer: 100 mm^3*)

A+ groups

Below the handkerchief of Gulliver is drawn. Draw beside a similar handkerchief of a Lilliputian.
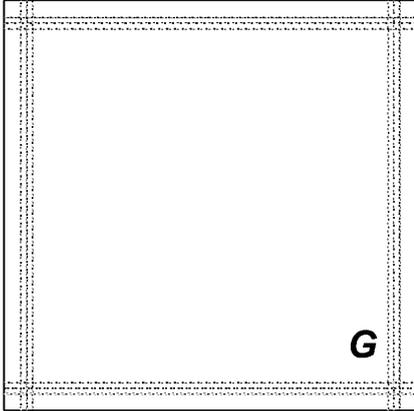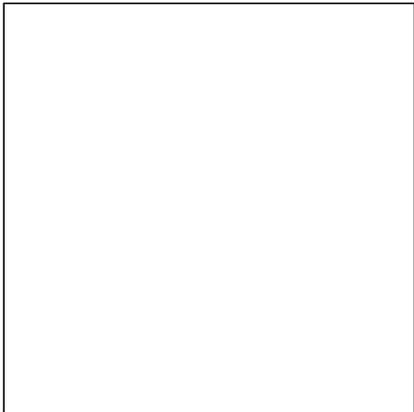

Gulliver's handkerchief   *Handkerchief of a Lilliputian*

A+ groups

The side of square Q is 12 times as large as the side of square R. Below square Q is drawn. Draw

beside square R.

Square Q   *Square R*

Fig. 1.   Example of presentation of drawing activity in A+ and A− groups.

and G6) submitted to all participants, there were two video related statements (V1 and V2, of course, only submitted in the A+ groups), three drawing related statements (D1, D2 and D3, for the D+ groups) and one video and drawing related statement (VD1, for the A+D+ group). Table 2 gives an overview of these statements. The main purpose of the questionnaire was to provide some information on how the test and the experimental setting were perceived and appreciated by the students.

3. Hypotheses

First, in line with the results of De Bock et al. (1998, 2002), we predicted that students' performance on the proportional items would be much higher than their scores on the non-proportional items (=hypothesis 1), and, second, that the 15–16-year olds would perform better on the test in general and on the non-proportional items in particular than the 13–14-year olds (=hypothesis 2).

Third, based on the claims of the advocates of authentic or realistic mathematics education (cf. supra), we hypothesised that plunging students into a rich and attractive context—in our case: the story of Gulliver's visit to the Lilliputians, coming alive in a fantastic video faithfully respecting the original 1:12 ratio for lengths—would help them to engage themselves in meaningful and thoughtful mathematical thinking and problem solving, especially on the difficult non-proportional items. Accordingly, we predicted for both age groups higher scores for the A+ groups than for the A− groups, due to a positive effect of the authenticity factor on students' willingness and capacity to model and solve the non-proportional items.

Fourth, we anticipated a better performance, especially on the non-proportional items, for the students who had to make a drawing of a figure similar to a given

Table 2
Questionnaire used in the different experimental conditions

General statements
(G1) I think I shall have a good score on this test.
(G2) I liked to work on this test.
(G3) The problems are similar to the problems we solve in the classroom.
(G4) I did my best to work on this test as much as possible.
(G5) I considered this as an easy test.
(G6) It was an instructive experience to complete this test.
Video statements
(V1) While solving the arithmetical problems, I often recalled the video fragments on Gulliver.
(V2) The video fragments helped me to find the answer.
Drawing statements
(D1) While solving the arithmetical problems, I made use of the drawing above.
(D2) The drawings helped me to find the answer.
(D3) I usually make a drawing before solving a geometrical problem.
Video and drawing statement
(VD1) The video fragments helped me to make the drawing.

---

original before the start of their solution process. This hypothesis is in line with numerous investigations reporting positive effects of drawings and diagrams, and especially of self-made visualisations, on students' mathematical problem solving (cf. supra). Accordingly, we predicted for both age groups higher scores for the D+ groups than for the D− groups, due to the anticipated positive effect of making a drawing on students' capacity to represent and solve the non-proportional items.

4. Analysis

In view of the quantitative data analysis, all responses on the proportional and the non-proportional items were categorised as 'correct' or 'incorrect'. A response was considered as correct when it was the result of a mathematically appropriate reasoning process; so, answers that differed from the correct answer because of a purely technical mistake in a correct overall solution process were considered as correct too. All other kinds of erroneous answers were scored as incorrect. Because incorrect answers were mostly due to inappropriate proportional reasoning,[2] we decided not to split up the incorrect answers any further in the present analysis.

The hypotheses were tested by means of a $2 \times 2 \times 2 \times 2$ analysis of variance with as independent variables the within subject variable 'proportionality' (proportional vs. non-proportional items) and the between subject variables 'age' (13–14- vs. 15–16-year olds), 'authenticity' (A+ vs. A− groups) and 'drawing' (D+ vs. D− groups), and the number of 'correct answers' as the dependent variable. To obtain the most pure and valid comparison between the A+ and A− groups, we decided to restrict the before-mentioned quantitative analysis to only four of the six experimental items (the two proportional items and the two non-proportional items about rectilinear figures, and thus not the two non-proportional items about non-rectilinear figures).[3] However, as clearly evidenced by Claes (2000), the results of this restricted analysis are representative for the whole data set.

In addition to this quantitative analysis, aiming at testing the four before-mentioned hypotheses, we performed a qualitative analysis of students' solution strategies and their schematic drawings. Students' answers to the different statements of the questionnaire were analysed for each statement separately.

---

[2] In our previous studies (De Bock et al., 1998) we analysed in detail a random sample of 300 incorrect answers on non-proportional items. This analysis revealed that 95% of them indeed resulted from an inappropriate proportional reasoning process.

[3] In the A− groups, the non-proportional items about non-rectilinear figures handled about circles and spheres, geometrical forms students typically encounter in school geometry, while in the A+ groups, these forms were replaced by 'real', but different non-rectilinear forms issued from Gulliver's world (a sole of a shoe and a wineglass). For the non-proportional items about rectilinear figures, the same geometrical forms (squares and cubes) could be used in both experimental conditions.

5. Quantitative results

The results provided a very strong confirmation of the first hypothesis. Indeed, the analysis of variance revealed a strong main effect of the task variable proportionality, $F(1,302) = 430.64$, $p < 0.01$. The overall percentages of correct responses for all proportional and for all non-proportional items were 95 and 33%, respectively.

The second hypothesis was confirmed too: the factor age had a significant main effect, $F(1,302) = 16.63$, $p < 0.01$: the 15–16-year olds performed better than the 13–14-year olds on all experimental items. Percentages of correct answers were 69 and 59%, respectively. Furthermore, the predicted age × proportionality interaction was found, $F(1,302) = 13.81$, $p < 0.01$: while the 15–16-year olds answered nearly twice as much non-proportional items correctly than the 13–14-year olds (43 and 23% correct responses, respectively), the performance of the two age groups on the proportional items was exactly the same (95% correct responses in both age groups).

Third, there was a main effect of the authenticity variable on students' performance, but the effect was in the opposite direction: students who watched the video fragments and who received the video-related items performed significantly *worse* than the students from the other groups, $F(1,302) = 5.91$, $p < 0.05$ (61% correct responses for the A+ groups vs. 67% for the A− groups). We also found an interaction effect between authenticity and proportionality, $F(1,302) = 12.43$, $p < 0.01$. While the percentages of correct responses on the proportional items were in the expected direction (97 and 93% respectively, for the A+ and A− groups), the percentage of correct responses on the non-proportional items in the A+ groups (25%) was much smaller than in the A− groups (41%).

Fourth, the results did not support the drawing hypothesis. On the contrary, as for the authenticity factor, students who had to make a drawing performed even significantly *worse* than students from the non-drawing groups, $F(1,302) = 19.52$, $p < 0.01$ (59% correct responses in the D+ groups vs. 70% for the D− groups). Furthermore, the analysis of variance revealed a drawing × proportionality interaction effect, $F(1,302) = 10.38$, $p < 0.01$: the negative effect of making a drawing was much greater for the non-proportional items (23 and 44%, respectively correct responses for the D+ and D− groups) than for the proportional items (94 and 96%, respectively correct responses for the D+ and D− groups).

Because of the very surprising nature of the results with respect to the two main hypotheses, we decided to do a replication study. Although the matching of the students in the four groups had been done very carefully and systematically taking into account all available possibly relevant information about the students (Claes, 2000), and although the administration of the test happened as planned, one cannot exclude that the results might have been accidentally affected by some unattended initial differences between the four experimental groups. For practical reasons, we restricted the replication study to only one of the two experimental manipulations. We choose the drawing factor because its negative results had surprised us most.

Hundred and sixty-five students from the same age groups, but from a different school and region participated in this replication study. Using all available possibly relevant information (the study stream to which the students belonged, the number

of hours a week spent at mathematics, and students' examination results for mathematics), the students were individually matched in a drawing and a non-drawing group, defined in the same way as in the original study. The same paper-and-pencil tests were administered, and we absolutely guaranteed that all participants could work as long as they wanted to complete the task. The results of that replication study were exactly the same, in the sense that the students who were instructed to make drawings again performed significantly worse on the test in general, $F(1,159) = 9.27$, $p < 0.01$ (60 vs. 73% correct responses in the drawing and non-drawing condition, respectively) and on the non-proportional items in particular, $F(1,159) = 22.95$, $p < 0.01$ (40 vs. 62% correct responses in the drawing and non-drawing condition, respectively) (for a detailed description of this replication study, see Claes, 2000).

6. Additional findings

As mentioned before, we also collected qualitative data about the nature of students' solution strategies and their drawings. Besides, we administered a questionnaire consisting of a number of statements about how the students had perceived and appreciated the test. In this section we briefly report the major results of these additional analyses, restricting ourselves to those findings that are of direct help to understand the surprising results with respect to the authenticity and the drawing hypothesis. For a more detailed and systematic overview of the findings from these additional analyses, we refer the reader to Claes (2000).

6.1. Students' solution strategies

A detailed analysis of the written protocols of all students' correct solution strategies for the non-proportional items was undertaken. Because only the non-proportional items handling about rectilinear figures were included in the quantitative analysis, we also restrict the present discussion to these items. In the studies reported by De Bock et al. (1998), the solution processes underlying a correct answer on a non-proportional item were categorised using the following classification schema: (1) 'paving': finding the area of a plane figure by paving it with small, similar figures; (2) 'formula': finding and applying an appropriate mathematical formula for the area; (3) 'general principle': applying the general rule 'length $\times r$, thus area $\times r^2$'. For the present study, we had to adapt this former classification in two different ways. First, because the present study included also items dealing with solids, the paving category should also include its three-dimensional variant: finding the volume of a solid by filling it up with small, similar solids. Second, we added a fourth and fifth category: (4) 'mixed' strategy for the responses found by means of a combination of at least two different strategies; and (5) 'remainder' to include the correct solutions for which the solution strategy could not be determined.

Table 3 gives an overview of solution strategies students used to correctly solve

Table 3
Strategies underlying students' correct responses on the non-proportional items about rectilinear figures

Group (n)	Formula		Principle		Mixed		Paving		Remainder	
	Number	%	Number	%	Number	%	Number	%	Number	%
A+ (75)	8	11	61	81	2	3	4	5	0	0
A− (131)	87	66	26	20	14	11	1	1	3	2
D+ (72)	30	42	34	47	7	10	0	0	1	1
D− (134)	65	49	53	40	9	7	5	4	2	1
Total (206)	95	46	87	42	16	8	5	2	3	1

n=number of correct responses on the non-proportional items about rectilinear figures. *Note*: A+=authentic setting groups; A−=no authentic setting groups; D+=drawing groups; D−=no drawing groups.

a non-proportional item about a rectilinear figure in the different experimental conditions.

All together, 46% of the correct responses on non-proportional items about rectilinear figures were found by applying the formula strategy. Besides, the analysis revealed a remarkable difference between the A− and A+ groups: in the A− groups much more correct solutions were found by applying an appropriate mathematical formula than in the A+ groups. Apparently, the traditional context and setting more often elicited this school-like approach, which not only prevented the students from falling into the linearity trap but also brought them to the correct solution. The general principle strategy—length $\times r$, thus area $\times r^2$—was the second most popular strategy, accounting for 42% of the correct responses on non-proportional items about rectilinear figures. In the A+ groups, the little use of the formula strategy was compensated by students' application of the general principle. The three other strategies were chosen far less (for the mixed, paving and remainder category, we found percentages of 8, 2 and 1%, respectively). Although in line with students' well-documented resistance to informal, context-bound mathematical strategies (Schoenfeld, 1992; De Bock et al., 1998; Verschaffel & De Corte, 1997), the very rare use of the intuitive paving strategy is surprising. In the D+ groups, not a single student (visibly) found the correct solution to a non-proportional item about a rectilinear (!) figure by paving. Remarkably, the drawing instruction did absolutely not elicit this informal but efficient graphical strategy by the students.

6.2. Analysis of students' drawing activity

Because students' actual drawing activity could also shed more light on the quantitative results of the test, we scrutinised their notes on the presence of drawings of the non-proportional problems, and, in the case a drawing was actually made, we verified its correctness. For a proper interpretation of the unexpected results with respect to the hypothesised effect of the drawing factor on students' performance on the non-proportional items, it was very important to know if the students from both

groups (D+ and D−) had actually made drawings for those items and if they had done it correctly. The criteria for scoring *if* a drawing was made were the same in the D+ groups and the D− groups: each graphical representation referring to the geometrical shape described in the item was scored as a drawing. Mere visualisations or schematisations of the numerical data, such as arrow diagrams, tables or graphs, were thus not considered as drawings. The scoring of the *correctness* of the drawings was done somewhat differently in the D+ and D− conditions. We remind that in the D+ groups, students were explicitly instructed to draw a reduced copy of the geometrical shape described in the item, starting from an original provided on their response sheet and using a given reducing factor (Fig. 1). To evaluate the correctness of this drawing, we tolerated a 20% error margin: if the linear dimensions of the students' drawing of the geometrical figure differed 20% or less from its correct linear dimensions, the drawing was scored as a correct one (e.g., if the student had to draw a square with a side of 1 cm, his drawing of that square was considered as a correct one if the side measured between 0.8 and 1.2 cm). In the D− groups, who received no (partial) drawings nor instructions to make a drawing, we additionally required that both the original geometrical form and its reduced copy were drawn. Table 4 gives an overview of the number of correct and incorrect drawing in the D+ and D− conditions.

The instruction to make a drawing in the D+ groups was followed in 94% (584/624) of the cases. This percentage is remarkably high, compared with the results of De Bock et al. (1998), who reported only 46% of drawings in groups that were *recommended* to make a drawing *after* the numerical data of the problem were presented. So, from that point of view, the change in the operationalisation of the drawing factor had been very successful. Moreover, 83% (485/584) of these drawings in the D+ groups were correct ones, suggesting that students' numerous errors on the non-proportional items typically did not originate in the phase of the problem representation. In the D− groups, only a minority of the students made visibly effective use of the drawing heuristic: spontaneously made drawings were found in only 10% (64/616) of the cases and, in addition, only 9% (6/64) of these drawings could be scored as correct.

The results on students' drawings made it clear that the significant lower score of the D+ groups on the non-proportional items could not be attributed to the fact that

Table 4
Number and percentage of drawings in the D+ and D− conditions

Group (*n*)	Correct drawing		Incorrect drawing		No drawing	
	Number	%	Number	%	Number	%
D+ (624)	485	78	99	16	40	6
D− (616)	6	1	58	9	552	90

n=number of non-proportional items that was administered. *Note*: D+=drawing groups; D−=no drawing groups.

the students in these groups had neglected the drawing instruction nor to the fact that they had made incorrect drawings. On the contrary, the very high percentage of actual and correct drawings followed by incorrect responses to the non-proportional items in the D+ condition suggests that making a (correct) drawing rather lured them into the proportionality trap instead of protecting them for it—a conclusion that is hard to draw in the light of our current knowledge on the role of heuristics in general, and of visualisations in particular, in skilled (mathematical) problem solving. We come back to this issue in the discussion section.

6.3. Results with respect to the questionnaire

Students' degree of agreement with the different statements in the questionnaire (Table 2) was first analysed at a descriptive level by looking at the frequency distributions and by determining students' mean rating for the different statements in the four different experimental conditions and for all four conditions together. For calculating the mean rating, we assigned numbers to the values of the scale in the following way: 1=full disagreement; 2=disagreement; 3=no opinion; 4=agreement; and 5=full agreement. An overview of the mean ratings in the different experimental conditions and in the total group is given in Table 5.

Because a comparison between the A+ and A− groups could reveal important information for both the validation of our experimental design and for the interpretation of the results with respect to the authenticity hypothesis, possible differences in students' rating of the general statements (G1, G2, G3, G4, G5 and G6) in both groups were statistically tested by means of a χ^2-test. We limit the discussion of the results to the most important findings (for a detailed description, see Claes, 2000).

First, with respect to the six general statements, no clear conclusions could be drawn from students' rating of the statements G1, G2 and G5 in the total group, but the χ^2-test revealed some interesting differences between the A+ and A− groups. The A+ groups reported higher confidence that their test scores would be good, $\chi^2(2, N = 303) = 6.91$, $p < 0.05$ (statement G1), liked the test more, $\chi^2(2, N = 302) = 11.02$, $p < 0.01$ (statement G2) and considered this test to be easier, $\chi^2(2, N = 295) = 9.09$, $p < 0.05$ (statement G5) than the A− groups. We consider

Table 5
Mean rating of the questionnaire statements in the different experimental groups and for the total group

Group	G1	G2	G3	G4	G5	G6	V1	V2	D1	D2	D3	VD1
A+D+	3.24	3.28	1.84	3.93	3.45	3.07	2.34	2.04	3.00	2.86	2.90	2.16
A+D−	3.50	3.61	1.81	4.12	3.55	3.03	2.62	2.01				
A−D+	3.11	3.15	2.11	4.06	3.17	3.19			3.84	3.71	3.34	
A−D−	3.10	3.17	2.30	4.20	3.18	3.16						
Total	3.23	3.30	2.02	4.08	3.33	3.11	2.48	2.03	3.42	3.29	3.12	2.16

Note: A+D+=authentic setting and drawing group; A+D−=authentic setting and no drawing group; A–D+=no authentic setting and drawing group; A−D−=no authentic setting and no drawing group.

the greater task attractiveness in the A+ groups (statement G2) as support for the validity of our design, because a major purpose of the authentic setting consisted in the creation of an attractive and stimulating context for the students, and, therefore, in an increase of their involvement in the task. With respect to statement G3, 74% of the students in the total group (fully) *dis*agreed that the problems in the test were comparable with those they had to solve in the classroom. However, students in the A− groups more often agreed with this statement (18%) than those in the A+ groups (6% of the students), $\chi^2(2, N = 302) = 5.22$, $p < 0.05$. Apparently, the students in the A− groups experienced the presentation and formulation of the problems as more traditional and school-like than those in the A+ groups. With respect to statement G4, 82% of all students declared, they did their best working on the test; there was no significant difference between the A+ and A− groups. With respect to the last general statement G6 (inquiring if students considered the task as an instructive experience), the reactions were very neutral—the frequency distribution was more or less symmetric around 'no opinion' that itself represented 42% of the data—and no significant differences between the A+ and A− groups were found.

Second, with respect to the two video related statements (V1 and V2), the general tendency was rather negative: 61% of the students in the A+ groups claimed not to have recalled the video fragments of Gulliver while solving the problems (statement V1) and 77% asserted that these fragments had not helped them to find the answers on the problems (statement V2).

Third, with respect to the three drawing related statements (D1, D2 and D3), 63% of the students in the D+ groups claimed they had made use of the drawings while solving the problems (statement D1), 55% asserted that the drawings helped them to find the answer (statement D2) and only 44% claimed they usually make a drawing before solving a geometrical problem (statement D3). Apparently, the drawing instruction was perceived as more effective than the presentation of the video.

Finally, with respect to the video and drawing related statement VD1, 72% of the students in the A+D+ group claimed that watching the video did not help them to make the drawings.

7. Discussion

Studies by De Bock et al. (1998, 2002) convincingly demonstrated that the vast majority of students fail to solve problems involving the relationships between the linear measurements and the area and/or volume of similarly enlarged or reduced geometrical figures, even when considerable help is provided (giving metacognitive support, presenting the problem in another format, …). In the present study, we tested the effect of two additional forms of help, namely increasing the authenticity of the problem context and asking students to make a drawing as an integral part of the solution process, on students' performance on this type of geometrical problems. Contrary to our expectations, none of the experimental manipulations yielded the expected result; they both even yielded a significant negative effect on students' performance on the test in general and on the non-proportional items in particular.

With respect to the *absence of a positive authenticity effect* one can point a finger to the way authenticity was operationalised in the present study. Compared to the studies by Streefland (1984) and Treffers (1987) in which students were actively involved in a series of math lessons around an authentic context with several learning tasks (e.g. drawing a giant's footprint, making a newspaper for the giant, …), watching a video and completing a paper-and-pencil mathematics test about this context is a rather weak operationalisation of the authenticity factor. Although the video did have a positive effect on students' task-specific motivation (cf. statement G2), the majority of the students also claimed that the video fragments had not helped them to find the answers on the test items (statement V2). Probably, a more performance-based form of assessment (involving, for instance, subtasks like effectively 'making a Lilliputian handkerchief' or, 'filling a Lilliputian wineglass') (see also Reusser & Stebler, 1997) is needed for helping students to discover the non-linear nature of the problem situations. However, while the way in which the authenticity factor was operationalised in the present study can account for the absence of a positive effect, it cannot explain the observed *negative* effect. To explain this negative effect, several hypothetical explanations can be raised. Although some of these explanations are (partially) sustained by students' reactions on the questionnaire, we acknowledge their speculative nature. First, the way in which the problems were stated and presented in the A− condition was more similar to how these problems are typically formulated and presented in mathematics lessons at school (some support for this explanation can be found in the results with respect to the questionnaire). By adding the non-routine element of 'authentisation' the students might have lost the good sense of what they were tested for—the discourse became unfamiliar, with its rules no longer clear. A second possible explanation can be found in Salomon's (1981) account of the mediating effects of people's perceptions of media characteristics on their willingness to invest mental effort during learning and problem solving, and, consequently, on their performance. According to Salomon (1981), students perceive video as a less difficult medium than written materials, and therefore are inclined to invest less mental effort in working with information transmitted by this easy medium as compared to media that are perceived as difficult. Once again, some support for this explanation can be found in the questionnaire: students in the authentic setting perceived the difficulty level of the test lower and their subjective competence higher than students in the traditional setting (see the results with respect to the statements G5 and G1). For a third possible explanation, we refer the reader to a related finding reported by Boaler (1994), namely that students—and especially girls—were likely to underachieve in realistic mathematical contexts as compared to problems presented in an abstract way. Boaler analysed the performance of 50 students from a school with a traditional approach to mathematics education on two sets of three questions assessing the same mathematical content through different contexts. Interestingly, the girls from that school attained a lower grade on an item built around the context of fashion, than on an isomorphic items that were embedded in an abstract context or in a context that was less appealing for these girls (e.g. football). According to Boaler, these girls' underachievement on the fashion item was caused or influenced by their greater involvement in the problem context, which

led them away from its deeper mathematical structure. Although there are quite some differences between the way in which the authenticity factor was operationalised as well as between the participants' backgrounds in both studies, it is possible that students' emotional involvement in the fantastic world of the Lilliputians may have had a negative instead of a positive influence on their performance, just like the girls in Boaler's study. A fourth and last possible explanation refers to the time factor. Students from the A+ groups, who first watched the video fragments for 10 min, had less time for solving the test than students from the A− groups. Although students from all groups were not forced to give in the test before they were ready, the time factor may have played a subtler negative role in the A+ groups. According to the notion of 'experimental contract' (Greer, 1997), which can be described as a system of implicit norms, rules and expectations that influence how a participant thinks and acts during an experiment, the time provided for a test yields signals about the nature c.q. complexity of the task(s) to be fulfilled. Consequently, the less time the participants are given for completing a task, the more chance they will perceive it as easy. The finding that the students in the authentic setting perceived the test as easier than those of the traditional setting (statement G5) is in line with this assertion.

The *absence of a positive effect* of asking students to make a drawing both in the major study and in its replication, was not in agreement with our drawing hypothesis, but corresponds with a number of other studies reporting either no or only marginal effects of the use of heuristics (like making a diagram or a drawing) on (mathematical) problem solving (De Corte et al., 1996; Schoenfeld, 1992; Van Essen & Hamaker, 1990). Among the reasons that are given in this literature for the weak relationship between heuristics and success in (mathematical) problem solving are the following: (a) the description of the heuristics are not sufficiently detailed to enable students not yet familiar with them to implement and use them efficiently in their problem solving endeavours; (b) heuristics are only efficient when they are inculcated in conjunction with metacognitive executive control skills. Whereas the first of these explanations seems not very helpful to explain our results (the qualitative analysis of the students' drawings revealed they were mostly correct), the second one seems more relevant (see later). But, once again, these explanations cannot account for the strong *negative* effect of making a drawing, which is, as far as we know, never reported before in the research literature on the role of heuristics in general and of visualisation in particular, in (mathematical) problem solving. After having replicated these remarkable results in a follow-up replication study with the drawing factor as the only experimental variable (as reported before), it became clear that the explanation for this remarkable result had to be searched in the precise role played by the drawing heuristic in the students' problem solving processes. Ultimately, a detailed rational analysis of the cognitive processes of students' solving of the non-proportional items in the D+ conditions, yielded a reasonable explanation for this remarkable finding, which needs, empirical validation in further research. When students were asked to make a reduced copy of the given geometrical shape (cf. Fig. 1), they first measure a linear element of that shape (e.g. the side of the given square, the height of the wineglass, …) and then divide that element by the

given linear scale factor to obtain the length of the corresponding dimension of the reduced shape that is needed for drawing this reduced shape. So, what they have in their working memory at the beginning of their actual solution process of a non-linear problem is exactly the *linear scale factor* they just applied in the drawing activity, rather than its square or cube which is needed for the solution of the problem. The activation of the linear scale factor in students' working memory during the drawing task may have enhanced the (erroneous) inclination to re-use that activated information element in the actual solution of the problem through a kind of automatic priming mechanism. Another mechanism that may underlie this tendency is students' more or less deliberate reliance to the experimental contract (Greer, 1997): if the experimenter involves the students in a preparatory activity, such as the making of a drawing, they will normally expect that this activity will be of help for solving the subsequent problem. Most likely, in our case this preparatory activity was perceived by the students as merely a matter of 'dividing by the linear scale factor,' and it was exactly the transfer of this information element to the subsequent solution of the problem that lured them into the linearity trap.

We end this article with two major concluding remarks and recommendations for further research. First, in line with our previous investigations (De Bock et al., 1998, 2002), the present study shows that the illusion of linearity as observed in a variety of students and in a variety of testing conditions, is not an artefact of a particular experimental setting that can be easily overcome by changing this experimental setting. In contrast, it is a recurrent phenomenon that seems to be quite universal and resistant to a variety of forms of support aimed at defeating it. The present study has added two additional factors to the list of candidate forms of help which have proven to be of little or no help addressed in our previous studies: the authenticity of the test setting and the integrative use of drawings. Some of the qualitative findings indicate that the reason why these two new forms of help did not yield the expected positive effect, was that they conflicted with students' implicit norms, expectations and beliefs about doing mathematics (Davis, 1989; Yackel & Cobb, 1996), especially about their appreciation of formal and informal strategies and of drawings as a valuable modelling tool. From that point of view, classroom interventions that are only partial and instantaneous, and that are unable to influence or alter these more fundamental attitudes and beliefs, have little chance of success. Most likely, only a long-term classroom intervention, not only acting upon students' deep conceptual understanding of proportional reasoning in a modelling context, but also taking into account the social, cultural and emotional context for learning, can produce a positive effect in defeating the illusion of linearity.

Second, at a more general level, the results of the present study support the warnings of several authors from the mathematics education community discipline (Freudenthal, 1978; Treffers, 1987) to be very cautious when applying general principles and findings from cognitive and educational psychology to teaching and learning in a particular subject-matter domain like mathematics—with little regard for the specific nature of (the subdomain of) mathematics and mathematics education that is addressed in these studies (Freudenthal, 1991). Research conceived and supervised by cognitive and instructional psychologists together with experts in the domain (e.g.,

mathematics and mathematics education) will contribute to unravel when and how certain kinds of support that are widely accepted as effective and valuable within the community of cognitive and instructional psychologists are indeed helpful to prevent and/or overcome certain learning difficulties in a particular curricular domain like mathematics.

Acknowledgements

References

Anderson, N. H. (1983). Intuitive physics: understanding and learning of physical relations. In T. J. Tighe, & B. E. Shepp (Eds.), *Perception, cognition and development: Interactional analyses* (pp. 231–265). Hillsdale, NJ: Lawrence Erlbaum Associates.

Aprea, C., & Ebner, H. J. G. (1999, August). The impact of active graphical representation on the acquisition and application of knowledge in the context of business education. Paper presented at the 8th European Conference for Research on Learning and Instruction, Göteborg, Sweden.

Berté, A. (1993). *Mathématique dynamique* [*Dynamical mathematics*]. Paris: Nathan.

Boaler, J. (1994). When do girls prefer football to fashion? An analysis of female underachievement in relation to "realistic" mathematic contexts. *British Educational Research Journal*, *20*(5), 551–564.

Claes, K. (2000). De lineariteitsillusie doorbroken? Een empirisch onderzoek naar het effect van de authenticiteit van de context en van tekeningen op het oplossen van niet-lineaire toepassingsproblemen bij leerlingen van het secundair onderwijs [The illusion of linearity defeated? An empirical research on the effect of context authenticity and of drawings on secondary school students' solution of non-linear application problems]. Unpublished master's thesis, Center for Instructional Psychology and Technology, University of Leuven, Leuven.

Collins, A., Brown, J. S., & Newman, S. E. (1989). Cognitive apprenticeship: teaching the crafts of reading, writing and mathematics. In L. B. Resnick (Ed.), *Knowing, learning and instruction. Essays in honor of Robert Glaser* (pp. 453–494). Hillsdale, NJ: Lawrence Erlbaum Associates.

Davis, R. B. (1989). The culture of mathematics and the culture of schools. *Journal of Mathematical Behavior*, *8*, 143–160.

De Block-Docq, C. (1992). Analyse épistémologique comparative de deux enseignements de la géométrie plane vers l'age de douze ans [Epistemological comparative analysis of two approaches to plane geometry on the age of twelve]. Unpublished doctoral dissertation, Faculté des Sciences, Université Catholique de Louvain, Louvain-la-Neuve.

De Bock, D., Verschaffel, L., & Janssens, D. (1998). The predominance of the linear model in secondary school students' solutions of word problems involving length and area of similar plane figures. *Educational Studies in Mathematics*, *35*, 65–83.

De Bock, D., Verschaffel, L., & Janssens, D. (1999). Some reflections on the illusion of linearity. In P. Radelet-De Grave (Ed.), (pp. 153–167). *Proceedings of the Third European Summer University on History and Epistemology in Mathematical Education*, Vol. 1. Belgium: Leuven/Louvain-la-Neuve.

De Bock, D., Verschaffel, L., & Janssens, D. (2002). The effects of different problem presentations and formulations on the illusion of linearity in secondary school students. *Mathematical Thinking and Learning*, *4*(1), 65–89.

De Corte, E. (1996). Instructional psychology: overview. In E. De Corte, & F. E. Weinert (Eds.), *Inter-*

national encyclopedia of developmental and instructional psychology (pp. 33–43). Oxford, UK: Elsevier Science.

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

de Lange, (1987). *Mathematics, insight and meaning. Teaching, learning and testing of mathematics for the life and social sciences*. Utrecht: OW&OC.

Dirkes, M. A. (1991). Draw to understand. *Arithmetic Teacher*, *39*(4), 26–29.

Fischbein, E., & Schnarch, D. (1996). Intuitions and schemata in probabilistic thinking. In L. Puig, & A. Gutiérrez (Eds.), (pp. 353–360). *Proceedings of the 20th Conference of the International Group for the Psychology of Mathematics Education*, Vol. 2. Valencia, Spain: University of Valencia.

Freudenthal, H. (1973). *Mathematics as an educational task*. Dordrecht: Reidel.

Freudenthal, H. (1978). *Weeding and sowing*. Dordrecht: Reidel.

Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht: Reidel.

Freudenthal, H. (1991). *Revisiting mathematics education*. Dordrecht: Kluwer.

Gagatsis, A., & Kyriakides, L. (2000). Teachers' attitudes towards their pupils' mathematical errors. *Educational Research and Evaluation*, *6*(1), 24–58.

Greer, B. (1997). Modelling reality in mathematics classroom: the case of word problems. *Learning and Instruction*, *7*(4), 293–307.

Hall, V. C., Bailey, J., & Tillman, C. (1997). Can student-generated illustrations be worth ten thousand words. *Journal of Educational Psychology*, *89*(4), 677–681.

Kenworthy, D. (Producer), & Sturridge, C. (Director) (1996). Gulliver's travels [Film]. (Available from Twentieth Century Fox Home Entertainment, Inc.).

Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, *12*, 101–138.

Matz, M. (1982). Towards a process model for high school algebra errors. In D. Sleeman, & J. S. Brown (Eds.), *Intelligent tutoring systems* (pp. 25–50). London: Academic Press.

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

Pólya, G. (1945). *How to solve it*. Princeton: Princeton University Press (2nd edition, 1957).

Reed, S. K. (1999). *Word problems: Research and curriculum reform*. Mahwah, NJ: Lawrence Erlbaum Associates.

Reusser, K., & Stebler, R. (1997, August). Realistic modelling through the solving of performance tasks. Paper presented at the 7th European Conference on Learning and Instruction, Athens, Greece.

Rouche, N. (1992). *Le sens de la mesure* [The sense of measurement]. Bruxelles: Didier Hatier.

Säljö, R. (1991). Learning and mediation: fitting reality into a table. *Learning and Instruction*, *1*, 261–272.

Salomon, G. (1981). *Communication and education, social and psychological interactions*. Beverly Hills, CA: Sage Publications.

Schoenfeld, A. (1992). Learning to think mathematically: problem solving, metacognition, and sense making in mathematics. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 334–370). New York: Macmillan.

Shaughnessy, J. M. (1992). Research in probability and statistics: reflections and directions. In A. D. Grouws (Ed.), *Handbook of research in mathematics teaching and learning* (pp. 465–494). New York: Macmillan.

Streefland, L. (1984). Search for the roots of ratio: some thoughts on the long term learning process (Towards … a theory). Part I: reflections on a teaching experiment. *Educational Studies in Mathematics*, *15*, 327–348.

Treffers, A. (1987). *A model of goal and theory description in mathematics instruction. The Wiskobas project*. Dordrecht: Reidel.

Van Dooren, W., De Bock, D., Verschaffel, L., & Janssens, D. (2001, July). Searching for the illusion of linearity in probabilistic misconceptions: A literature review. Poster presented at the 25th Annual Conference of the International Group for the Psychology of Mathematics Education, Utrecht, The Netherlands.

Van Essen, G., & Hamaker, C. (1990). Using self-generated drawings to solve arithmetic word problems. *Journal of Educational Research*, *83*(6), 301–312.

Verschaffel, L., & De Corte, E. (1997). Teaching realistic modelling in the elementary school. A teaching experiment with fifth graders. *Journal for Research in Mathematics Education*, *28*, 577–601.

Verschaffel, L., Greer, B., & De Corte, E. (2000). *Making sense of word problems*. Lisse, The Netherlands: Swets & Zeitlinger.

Vlahovic Stetic, V. (1999). Word-problem solving as a function of problem type, situational context and drawing. *Studia Psychologica*, *41*(1), 49–62.

Yackel, E., & Cobb, P. (1996). Sociomathematical norms, argumentation, and autonomy in mathematics. *Journal for Research in Mathematics Education*, *27*, 458–477.