Assessing Mathematical Proficiency
MSRI Publications
Volume **53**, 2007

Chapter 7
Aspects of the Art of Assessment Design

JAN DE LANGE

Educational design in general is a largely underestimated and unexplored area of design, and its relationship with educational research can be characterized as somewhat less than satisfying. The design of assessments is often seen as an afterthought. And it shows.

Of course, there is a wealth of publications on assessment, but quite often these focus on psychometric concerns or preparation for high-stakes tests (a very profitable industry). What is lacking is a tight linkage between research findings and the creation of mathematically rich and revealing tasks for productive classroom use. The report "Inside the black box" [Black and Wiliam 1998], which looks in depth at current research, shows clearly that we should not only invest more in classroom assessment in mathematics, but also that the rewards will be high if we do so.

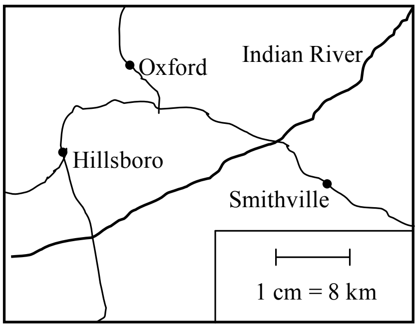
Linking Items with Framework

To provoke some discussion, and to invite the reader to reflect on the items and make a judgment, I start the examples with an item from the Third International Mathematics and Science Study (TIMSS). According to the TIMSS web site http://www.timss.org, TIMSS was "the largest and most ambitious international study of student achievement ever conducted. In 1994–95, it was conducted at five grade levels in more than 40 countries (the third, fourth, seventh, and eighth grades, and the final year of secondary school)." Mathematics coverage on TIMSS was, in essence, international consensus coverage of the traditional curriculum.

The item in question, given to students about 14 years old in the mid-1990s, involved a simplified portion of a regional road map, replicated on the next page. The scale is reinforced by a sentence above the map:

One centimeter on the map represents 8 kilometers on the land.

Distance on a Map



About how far apart are Oxford and Smithville on the land?

A. 4 km B. 16 km C. 25 km D. 35 km

The problem seems simple, actually too simple, for the intended age group. The multiple-choice format makes it even easier for the students to find the correct answer, 35 km. There is nothing wrong with easy questions as such; the surprise lies in the "intended performance category," which is "using complex procedures." The consequence of this classification is that students who have chosen answer C are considered able to use complex mathematical procedures. This seems somewhat far-fetched.

This example shows clearly that we need to make sure that there is a reasonable relationship between "expected performances" (or "competencies" or "learning goals") and assessment items. This may seem straightforward but is at the heart of the design of any coherent and consistent assessment design.

Framework and Competencies

Discussions of mathematical competencies have really evolved during the last decade, not least because of large-scale assessments like TIMSS, the Program for the International Student Assessment, known as PISA (see http://nces.ed.gov/surveys/pisa/AboutPISA.asp), and the U.S. National Assessment of Educational Progress, known as NAEP (see http://nces.ed.gov/nationsreportcard/). Items from TIMSS and PISA are discussed in this chapter. (NAEP is discussed in Chapters 1 and 3.)

The context used in the TIMSS item above is excellent. The task is an authentic problem: estimate distances using a map and a scale. But simply placing a problem in a context can create difficulties as well. This has been discussed in the literature, although not widely. That it is necessary to pay more attention to the connection between context and content is shown by the following example, taken from a popular U.S. textbook series:

One day a sales person drove 300 miles in $x^2 - 4$ hours.
The following day, she drove 325 miles in $x + 2$ hours.

- Write and simplify a ratio comparing the average rate the first day with the average the second day.

This example shows clearly what we would classify as a non-authentic problem in the sense that the mathematics is sound, but is unlikely to be needed in the context in which the problem has been set. To call this "problem solving using algebra," as the textbook suggests, clearly shows that we need clearer understandings of and guidelines for the design of authentic problem solving items.

Figure 1, adapted from [PISA 2003, p. 30], suggests concerns illustrated by the examples. In the center of the figure are the problem as posed, and the desired solution. The ovals at the top of Figure 1 indicate that special attention is given to mathematical situations and contexts, and that problems on the assessment will focus on overarching mathematical concepts. These are the mathematical arenas in which students will be examined. The bottom oval in Figure 1 refers to competencies — the specific mathematical understandings we want to measure. The question is, do the students have these competencies? Especially in classroom assessment, this question is often overlooked: the teacher poses a question, the students give an answer, and the issue considered is whether or

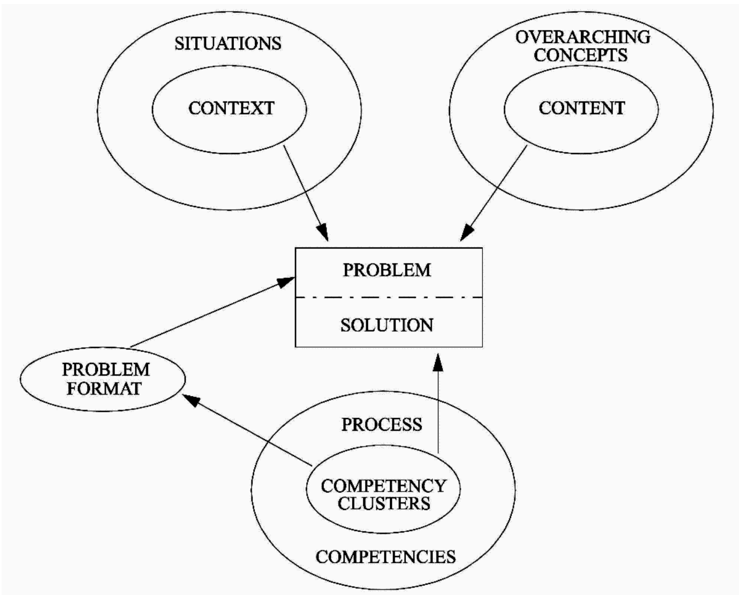


Figure 1. Problems, competencies, context, and content. Adapted from [PISA 2003, p. 30].

not the answer is correct. More often than not both teacher and students fail to reflect on what competencies were used in order to solve the "problem."

Figure 1 by itself does not shed any light on how these relationships function in good task design, it just indicates that these relationships need to function properly. But we need much more research in order to come close to answers that are scientifically based. At this moment many high-quality tasks are the result of creative impulses, and often seem more the result of an art than the result of scientific principles.

As we look at different frameworks we see some correlation between the competencies identified, albeit with slight differences. In an earlier *Framework for Classroom Assessment* [de Lange 1999] the following clusters of competencies were identified:

- *Reproduction*: simple or routine computations, definitions, and (one-step or familiar) problems that need almost no mathematization.
- *Connections*: somewhat more complex problem solving that involves making connections (between different mathematical domains, between the mathematics and the context).
- *Reflection*: mathematical thinking, generalization, abstraction and reflection, and complex mathematical problem solving.

These clusters of competencies are at the heart of the PISA framework. Each of the clusters can be illustrated with examples. (Note: These are not meant as examples of psychometrically valid test items, merely as illustrations of the competencies needed.)

Examples of reproduction

What is the average of 7, 12, 8, 14, 15, 9?

1000 zed are put in a savings account at a bank, at an interest rate of 4%. How many zed will there be in the account after one year?

The savings account problem above is classified as reproduction, because it will not take most students beyond the simple application of a routine procedure. The next savings account problem goes beyond that: it requires the application of a chain of reasoning and a sequence of computational steps that are not characteristic of the reproduction cluster competencies, hence is classified as in the connection cluster.

Examples of connections

1000 zed are put in a savings account at a bank. There are two choices: one can get a rate of 4%, or one can get a 10 zed bonus from the bank and a 3% rate. Which option is better after one year? And after two years?

Mary lives two kilometers from school. Martin five. What can you say about how far Mary and Martin live from each other?

Example of reflection

In Zedland the national defense budget is 30 million zed for 1980. The total budget for that year is 500 million zed. The following year the defense budget is 35 million zed, while the total budget is 605 million zed. Inflation during the period covered is 10%.

(a) You are invited to give a lecture for a pacifist society. You intend to explain that the defense budget decreases over this period. Explain how you would do this.

(b) You are invited to give a lecture for a military academy. You intend to explain that the defense budget increases over this period. Explain how you would do this.

Item Difficulty and Other Considerations

I observed, when discussing the TIMSS example, that items may be "simple," not in the sense that many students succeed in doing the task, but in the sense that students seem to have the mathematical skills required to do so. The story of how the following example stopped being a potential PISA item raises questions about the kinds of tasks that should be included in large-scale assessments.

Heartbeat

For health reasons people should limit their efforts, for instance during sports, in order not to exceed a certain heartbeat frequency.

For years the relationship between a person's recommended maximum heart rate and the person's age was described by the following formula:

$$\textit{Recommended maximum heart rate} = 220 - \textit{age}$$

Recent research showed that this formula should be modified slightly. The new formula is as follows:

$$\textit{Recommended maximum heart rate} = 208 - (0.7 \times \textit{age})$$

Question 1

A newspaper article stated: "A result of using the new formula instead of the old one is that the recommended maximum number of heartbeats per minute for young people decreases slightly and for old people it increases slightly."

From which age onwards does the recommended maximum heart rate increase as a result of the introduction of the new formula? Show your work.

Question 2
The formula *Recommended maximum heart rate* = 208 − (0.7 × *age*) also helps determine when physical training is most effective: this is assumed to be when the heartbeat is at 80% of the recommended maximum heart rate.

Write down a formula for calculating the heart rate for most effective physical training, expressed in terms of age.

This problem seems to meet most standards for a good PISA item. It is authentic; it has real mathematics; there is a good connection between context and content; and the problem is nontrivial. Therefore it was no great surprise that when this item was field tested, it did well as a test of the relevant student knowledge. There was only one small difficulty: fewer than 10% of the students were successful. This was a reason to abandon the problem.

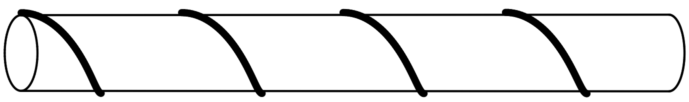
Abandoning such problems is a problem! We should look into why this item is difficult for so many 15-year-old students! Why are we excluding items of this kind, with a rather low success rate, from these kinds of studies? We are probably missing a valuable source of information, especially if we consider the longitudinal concept underlying the whole PISA study.

Knowing How to Think

TIMSS underscores this point — at least in my opinion. Let us look at an item for 18-year-olds from the 1990s. This item was the subject of a *New York Times* article by E. Rothstein, titled "It's not just numbers or advanced science, it's also knowing how to think" (9 Mar 1998, p. D3). The title is not a quote from the PISA framework regarding the clusters, but a very good observation that the journalist made when looking at this item.

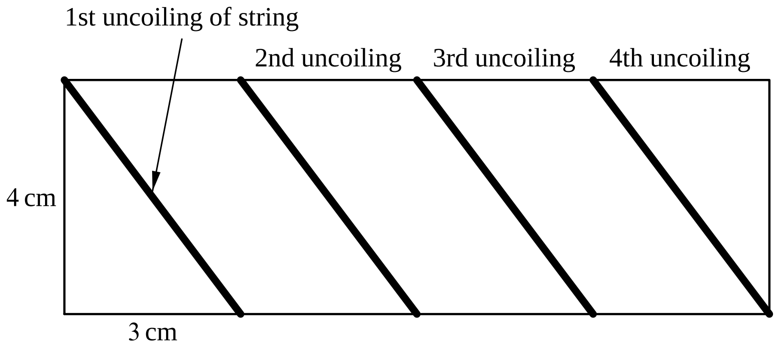
Rod

A string is wound symmetrically around a circular rod. The string goes exactly four times around the rod. The circumference of the rod is 4 cm and its length is 12 cm.



Find the length of the string. Show all your work.

The *Times* article included a solution that has all the essentials in the first sentence: "Imagine that you unwrap the cylinder and then flatten it." If you do that, you see that the cylinder becomes a rectangle that is 4 cm (the circumference of the cylinder) in height. The string unrolls as a straight line, and it first hits the bottom of the rectangle at a distance 1/4 of the length of the cylinder, or 3 cm. The next part of the string starts again at the top, hits the bottom 3 cm further along, and so on.



Once the unwrapping has been visualized, an application of the Pythagorean theorem is all that remains to be done. However imagining the unwrapping is something that caused a major problem for almost all students (about 10% got the problem correct) because "knowing how to think" in this creative, applied sense is not part of many curricula. Despite the low success rate it is very useful to have an item like the Rod Problem in the study, because it can reveal significant problems with the curriculum.

This last example might suggest that we reconsider the aims of mathematics education. *If* we consider that the TIMSS Distance on a Map Problem shows that kids are able to carry out complex mathematical procedures; *if* we accept that the Heartbeat Problem should not be in the item collection because it is too difficult; and *if* we similarly reject the Rod Problem, then there seems to be reason to reflect on mathematics education as a whole, and on assessments in particular. Are our students really so "dumb," or as a German magazine asked its readers in reaction to PISA: "Are German kids *doof*?"

This is an interesting question, worth serious reflection, especially in light of some more or less recent outcomes of educational research about mathematics learning. An alternative is to consider the possibility that poor student performance on such tasks is a result of the lack of curricular opportunities, both in terms of specific mathematical experiences (such as in the case of the rod problem) and because, in general those designing curricula have seriously underestimated what students might be able to do if appropriately challenged and supported.

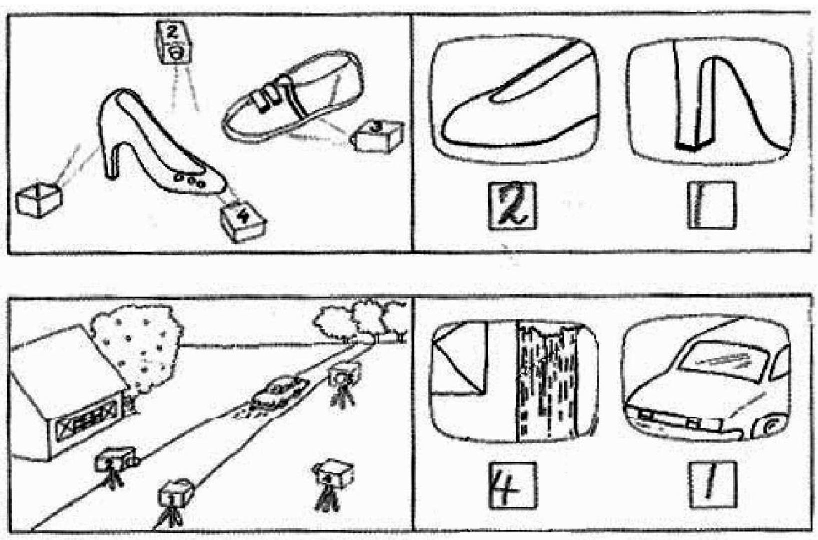
Enhancing Mathematics Learning

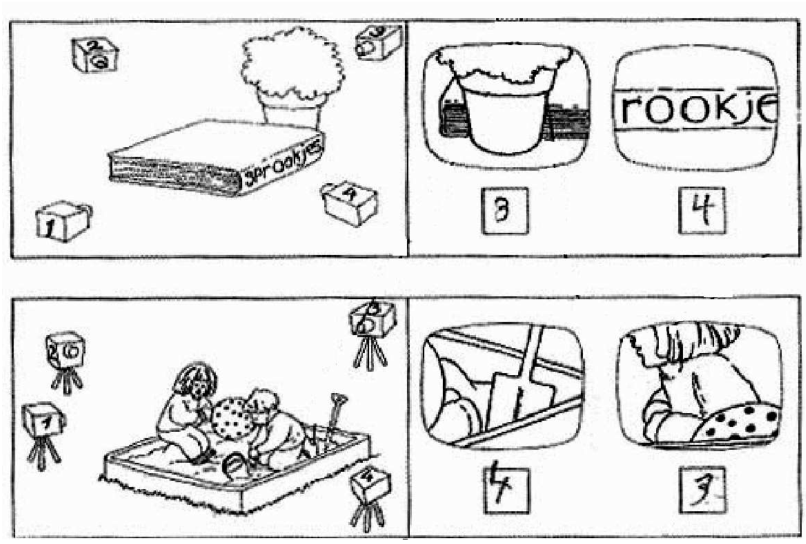
Neuroscientists have entered the arena that was considered the domain of cognitive psychologists, mathematics educators, and others. Barth et al. [2006] stated: "Primitive (innate) number sense can be used as a tool for enhancing early mathematics." This is a point well made (about young children) and once again mentioned by an Organisation for Economic Co-operation and Development report [OECD 2003, p. 25]: "Children can do much better than was expected." (The question here of course is: who expected what? But we will not go into that discussion here.) The German mathematics educator Wittmann expressed similar sentiments, probably shared by many who study young children's learning of mathematics: "Children are much smarter than we tend to think" (quoted in *Der Spiegel*, 12 Jun 2004, p. 190).

If all this is true (and we think it is), this has great implications, not only for curriculum design but for assessment as well. It means that assessment should assess as many learning outcomes as possible, at the edge of what children can solve. We need challenging tasks that give us real and valuable information about students' thinking.

Here are examples, taken from regular schools and regular kids, that not only show that there seems to be much truth in these quotes, but also that good assessment design can deliver us little miracles.

This is a test designed for 5- to 6-year-olds. The images shown here (continued on the next page) contain written responses from a student, including the lines from the camera to the shoes in the upper left rectangle:





The only question the children were asked is: “Who makes the photo?” The teacher explained orally that this means that the kids had to find the camera that made the pictures shown on the right.

Here is a student’s explanation for his response to the first problem (see preceding page):

> Look, you have camera 1, 2, 3 and camera 4. And you see two shoes. Camera 2 makes photo 1. This is because camera 2 points towards the nose of the shoe [points to the two lines he has drawn]. And camera 1 films the heel of the shoe.
>
> Camera 4 films the other side; otherwise you would have seen the three dots. And camera 3 films the other shoe!

And here are his comments on the last of the four problems shown:

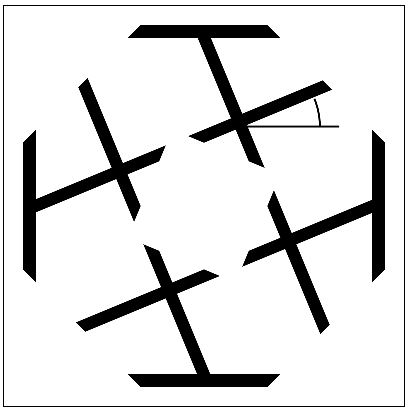
> This is a difficult one!
>
> Camera 3 films the left side of the girl and in the picture you face the other side. Camera 1 films that side, and camera 2 films the back of that girl. Camera 4 shows no lens. This [the lens] sits in front of the camera [points to the front of camera 4] because they are filming towards the playground. And so [puts finger on line of vision] the camera films the shovel.

Yes, this problem does involve complex geometric reasoning. But as we see, many students are capable of doing it, if given the experience.

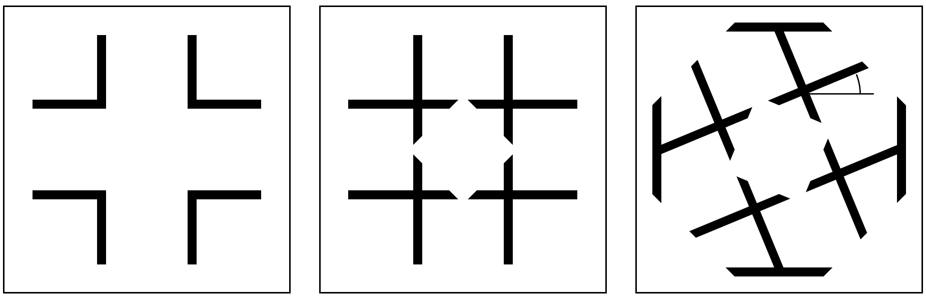
Reasoning Is an Art

Another example that was trialed both at middle school level and with graduate mathematics students will give us some insight in the possibilities of assessment, and how to challenge students to become mathematical reasoners.

The problem starts with a little strip that shows how the following figure was created.

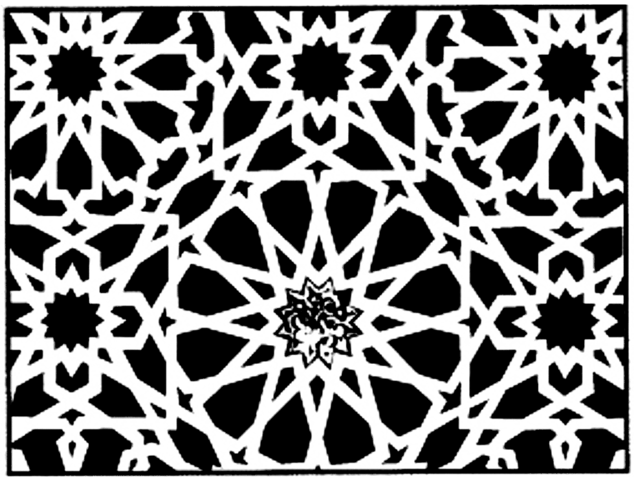


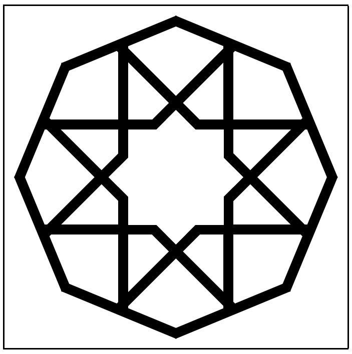
The first figure of the strip is a "cross." The horizontal and vertical segments are extended to form the second figure. The second figure is rotated 22.5 degrees about its center and four new segments are drawn to create the third figure. (Thus the angle marked in the third figure is 22.5 degrees.)



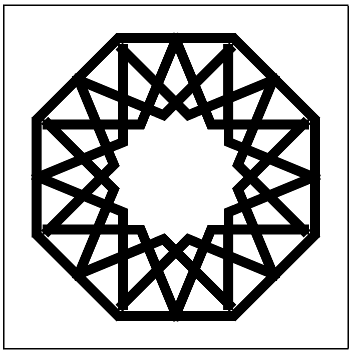
The task: Use transparencies that show the shape to make a Moorish star pattern. An example of such a Moorish star pattern is shown at the top of the next page.

After a while most students are able to make the star. The real question follows: if you know that a key angle is 22.5 degrees, what are the measurements of all the other angles in the star pattern? (See middle picture on the next page.)



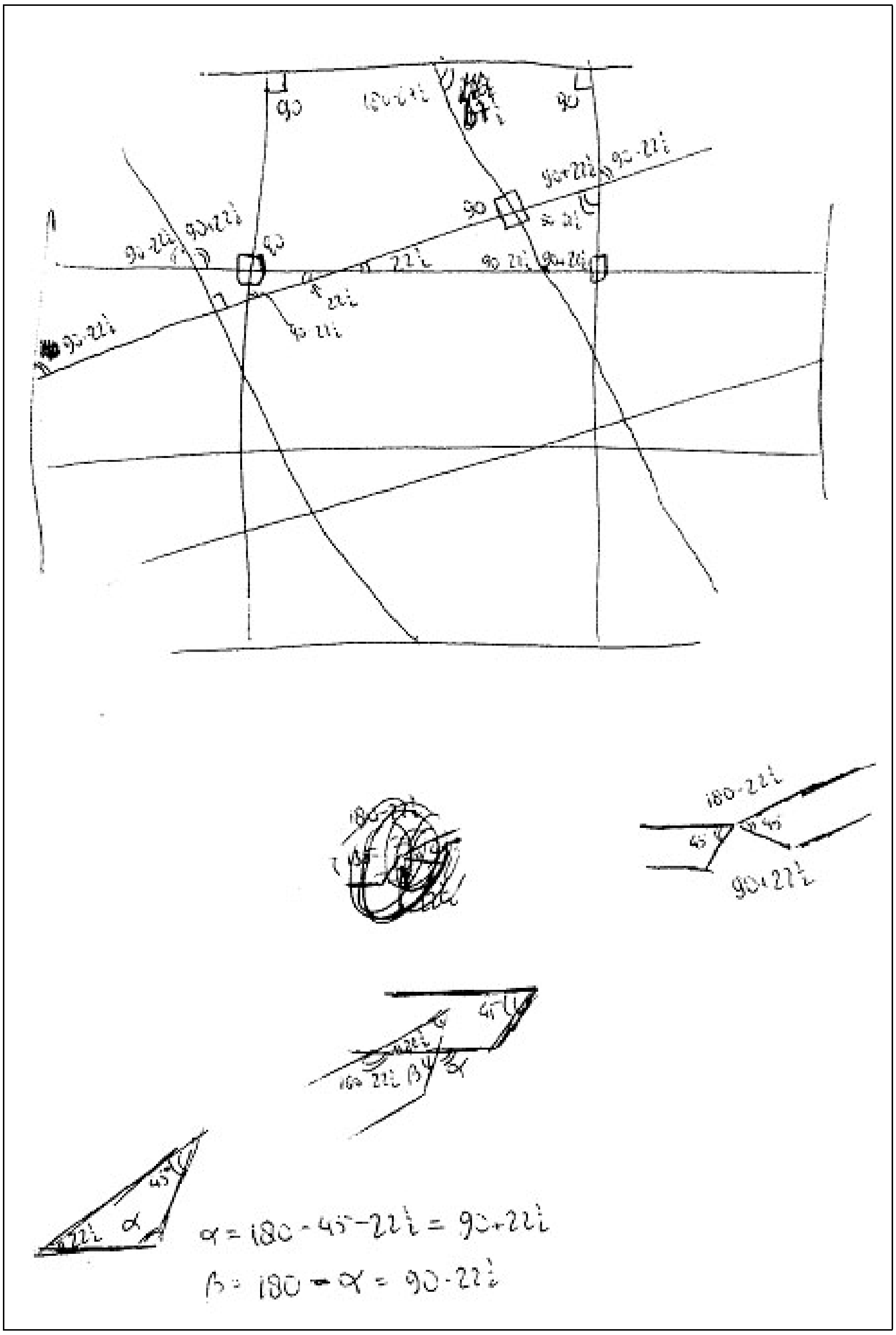


And how did we get the following star with the same transparencies? What do you know about the number of star points, and about all angles?



It goes almost without saying that both the 14-year-olds and 24-year-olds were indeed challenged and surprised. But above all they encountered the beauty of mathematics in a very unexpected way.

Here is a 22-year-old student's work:



In sum: Mathematics meets art; mathematics assessment needs to be an art.

References

[Barth et al. 2006] H. Barth, K. La Mont, J. Lipton, S. Dehaene, N. Kanwisher, and E. S. Spelke, "Non-symbolic arithmetic in adults and young children", *Cognition* **98**:3 (2006), 199–222.

[Black and Wiliam 1998] P. Black and D. Wiliam, *Inside the Black Box: Raising standards through classroom assessment*, London: King's College, 1998.

[de Lange 1999] J. de Lange, *A framework for classroom assessment in mathematics*, Madison, WI: National Center for Improving Student Learning and Achievement in Mathematics and Science, 1999. Available at http:/www.fi.uu.nl/catch.

[OECD 2003] Organisation for Economic Co-operation and Development, *A report on the literacy network and numeracy network deliberations*, Paris: Author, 2003.

[PISA 2003] Programme for International Student Assessment, *The PISA 2003 assessment framework: Mathematics, reading, science and problem solving knowledge and skills*, Paris: Organisation for Economic Co-operation and Development, 2003. Available at http://www.pisa.oecd.org/dataoecd/46/14/33694881.pdf. Retrieved 13 Jan 2007.