Equivalence of Computer-Based and Paper-Based Testing for the topic of Geometry

Rebekka Arntzen

Utrecht University

Student number: 6035604

30 ECTS

September 2018

Supervisor: prof. dr. P.H.M. Drijvers

Second examiner: dr. A. Bakker

This research was carried out at and supported by Cito, Arnhem.

Participating schools:

Twents Carmel College, Wellantcollege Gorinchem VMBO, Burgemeester Harmsma School, Terra VO Winsum, Staring College Herenlaan, Sprengeloo, Reeshof College, Daaf Gelukschool, Scholengemeenschap Panta Rhei, and Hilfertsheem College

Abstract

This study investigated whether paper-based testing (PBT) and computer-based testing (CBT) can be considered equivalent testing modes for geometry in the VMBO-KB level final exams. Parallel tests were designed and administered in both delivery modes to 129 third-year VMBO-KB level students using a research design that controlled for order and version effects. Equivalence was measured by two components: similarity in rank order between scores on the test modes and similar characteristics of the score distributions. The disattenuated correlation between the test scores in both conditions was found to be substantially high. However, for the PBT, the mean was higher, the variance lower and the distribution was more negatively skewed. Moreover, an investigation of item-level differences led to the conclusion that practicing questions decreased the test mode effect, while questions that required more mathematical thinking skills increased the test mode effect. Nevertheless, since the rank order was similar and rescaling eliminated any differences in characteristics of the score distributions, PBT and CBT can be considered equivalent test modes for VMBO-KB final exams.

Keywords: computer-based testing, geometry assessment

Despite the countless changes that our society has gone through during the last 2000 years, the ancient tradition of practicing geometry with a ruler and compass has survived the test of time. Learning how to hold and use a compass is still an important part of high school mathematics curricula. However, in the last decade the digitalization of learning methods has gradually gained traction in schools in the form of methods such as digital textbooks and electronic learning environments. A recent and major step in the digitalization of education is the use of computerized testing. Will the use of a tangible ruler and compass survive this development as well?

Computerized testing has become increasingly popular due to the many opportunities it offers. In contrast to paper-based testing (henceforth PBT), computer-based testing (henceforth CBT) facilitates testing independent of place and time (Bugbee Jr & Bernt, 1990). Besides that, CBT saves time for the teacher due to the automatic creation and scoring of the test (Inouye & Bunderson, 1986). Moreover, with automatic scoring the objectivity and consistency of the scoring is increased (Drijvers, 2018). Furthermore, the test can be adapted to different levels while students complete the test in order to measure the level of the student more adequately (Inouye & Bunderson, 1986). There are many potential uses of CBT and it is therefore to be expected that CBT will be the dominant mode for delivering tests in the future (Wang, Jiao, Young, Brooks, & Olson, 2008).

The transfer from PBT to CBT for the subject of geometry can only take place when the quality of digital geometry tests is guaranteed. The quality of a CBT can be established by the test developer in two ways. The validity requirements for the development and use of any test can be fulfilled (APA, 1985) or the CBT can be demonstrated to be equivalent to the PBT variant (APA, 1986). In the latter option, which is the focus of this study, the validity of the PBT passes on to the CBT and thus the validity of the CBT is assured.

A recent meta-analysis indicates that PBT and CBT can be considered equivalent testing modes for the topic of mathematics (Wang et al., 2008). However, no research in this field has been conducted for geometry tests specifically. Moreover, most of the current research is performed for multiple choice tests only, while geometry tests require more extensive computer skills that may hinder a valid transfer from PBT to CBT (Bennett et al., 2008). The present paper attempts to fill the gap in the literature on the equivalence of computerized geometry tests.

Theoretical background

Computerized testing has existed for many decades. The first testing computer was an IBM machine designed in 1935 with the purpose of objectively scoring millions of American tests every year (Khoshsimaa & Toroujenib, 2017). In the course of the 20th century, a technological revolution has occurred and currently computerized testing is omnipresent in our society.

With the transfer from PBT to CBT, it was noticed that CBT and PBT "are not the same thing even if their items are verbatim copies of one another" (Bugbee & Bernt, 1990, p. 97). Consequently, many studies were performed on whether different test modes could be considered equivalent. Equivalence can be defined as the comparability of test tasks represented in different test conditions (Choi, Kim, & Boo, 2003). If test conditions are established as equivalent for a specific test, the same underlying construct is measured and different formats of the test can replace each other (Neuman & Baydoun, 1998).

Differences in scores between the test modes PBT and CBT can be attributed to several causes. The causes can be divided into two main categories: technology issues and the characteristics of the test taker. Technology issues can affect the compatibility of PBT and CBT scores through issues such as font size, screen resolution, or lack of reliable tools (Gould, Alfaro, Finn, Haupt, & Minuto, 1987). Participant factors could also induce differences in scores. Characteristics that have been researched include sex, ethnicity, familiarity with technology, socioeconomic status, and content ability (Hensley, 2015).

Test mode effect

An extensive amount of research has been conducted to investigate the size and sign of the test mode effect between PBT and CBT. While some studies found that CBT leads to lower scores than PBT (Bennett et al., 2008; Lee, Moreno, & Sympson, 1986) and some studies found higher scores for CBT compared to PBT (Clariana, & Wallace, 2002; Pomplun, Frey, & Becker, 2002), other studies do not report significant differences between the scores of computer-based and paper-based tests (Bodmann & Robinson, 2004; Neuman & Baydoun, 1998).

Even meta-analyses do not provide a consensus on the direction in which the test scores are slanted. A number of meta-analyses have concluded that achievement tests produce similar scores. However, Griffin, McGaw and Care (2014) report several issues on the conclusions of these meta-analyses:

> This conclusion [of equivalence between PBT and CBT], however, is best viewed as preliminary, because the summarized effects have come largely from: analyses of distribution differences with little consideration of rank-order differences; multiple-choice measures; unrepresentative samples; non-random assignment to modes; unpublished studies and a few investigators without accounting for violations of independence (p. 184).

The researchers also note that for “analyses more sensitive to rank order and constructed-response items, the conclusion that scores are generally interchangeable across modes has not been supported” (Griffin et al., 2014, p. 184).

Mathematics tests on computer

Threllfall, Pool, Homer, and Swinnerton (2007) performed a research on translating PBT mathematics tests into computer versions, with the focus on the validity of specific item types. Year 6 students (age 11) and year 9 students (age 14) were administered test items both on paper and on computer, by use of a research design that removed any sample and practice effects. The researchers examined item types that revealed large differences in test scores between PBT and CBT more closely. The researchers found that a “difference in medium sometimes leads to better performance, and sometimes to better assessment, but not necessarily at the same time” (p. 346). They therefore argued that validity of question types in different test modes must be researched on a question to question basis.

As previously discussed, most articles on equivalence of tests conducted research on rather simple test formats such as multiple choice. A study on mathematics tests (Bennett et al., 2008) concluded that it is harder to maintain a constant level of difficulty for constructed-response questions than for multiple choice questions. The reasons for this shift are not clear. It could be that the need for computer skills has increased or that the nature of what is being measured has changed. Another reason may be that it has become impossible for students to answer in a way different from the proposed answer format. For example, in a text box it is impossible to answer diagrammatically, while this of course is possible in a paper-and-pencil test.

Geometry tests. Both reasons for the possible difficulty shift also apply for geometry tests. Firstly, the interaction with the computer is rather complicated. Students need to use digital geometry tools and provide answers that consist of the construction of lines and circles. Therefore, the transfer of this type of questions requires more advanced digital tools. Secondly, these digital tools may puzzle students and hinder them from answering in their desired manner. Drijvers (2018) has drawn attention to this issue. He notes that digital tools can prevent the students from "express[ing] themselves mathematically and to show what they can, compared to the paper-and-pencil environment, in which they can sketch and scratch whatever they want." (p. 8)

Geometry in computer-based tests has been studied before. Kreis, Dording, Keller, Porro, and Jadoul (2011) investigated the use of the dynamic geometry software GeoGebra in a multimedia interactive platform by 9-year-old children. 30 children followed a traditional paper-and-pencil geometry course while another 29 children worked additionally with GeoGebra. The researchers found that the group used to the technology performed at least as well in tests through GeoGebra as they did in a test using paper-and-pencil. Hence, this study supports a possible transfer to computer-based geometry education.

Equivalence of tests

No research has been conducted yet on the equivalence of PBT and CBT for geometry assessment specifically. Van de Vijver and Harsveld indicate that equivalence must always be demonstrated and never assumed (1994).

The explication of equivalence of two test-modes is provided by the American Psychological Association in the Guidelines for computer-based tests and interpretations:

> The scores from different test-modes can be considered equivalent when (a) the rank orders of scores of individuals tested in alternative modes closely approximate each other, and (b) the means, dispersions, and shapes of the score distributions are approximately the same, or have been made approximately the same by resealing the scores from the computer mode (APA, 1986, pp. 13-14).

The first component of equivalence, the similarity in rank order, is the validation of the underlying construct that is measured in the tests. If a highly performing student in geometry receives the highest score on the PBT, it is expected that this student is also amongst the top students of the CBT. If similarity in rank order between the test-modes is not established, other factors such as computer ability may have influenced the scores and therefore PBT cannot be transferred to CBT in that particular format.

The second part of equivalence, the mean score, variance and skewness of the distributions of the student scores are the same, is less restrictive than the first part. If students' scores are on average three points higher on the CBT than on the PBT, the CBT version of the test can be considered easier. When three points are subtracted from the students' scores of the CBT, the scores of PBT and CBT can be *made* the same and are therefore equivalent. Hence, differences in mean, variance and skewness do not necessarily hinder the transfer between test-modes, but should be taken into consideration as well.

Research context

The high school levels VMBO-BB and VMBO-KB are the only two high school levels in the Netherlands for which schools can choose the delivery mode for the final exams: paper or computer. Currently, 92% of the VMBO-KB students take their exams on computer (College voor Toetsing en Examens, 2017). These digital exams are administered by use of a digital assessment player, which includes multiple geometry tools such as a ruler and protractor. Students can use these tools in order to answer geometry questions. However, testing geometry construction is not possible in the digital mathematics exams yet, since no suitable drawing tool is available. In this research, it is investigated whether and how testing geometry construction would be possible in the VMBO-KB final exams, including the design of a possible drawing tool. Therefore, this study is supported by and contributes to Cito, the assessment institute that creates national examinations including the VMBO-KB final mathematics exams.

For this study, the following research question and sub questions are specified:

Can paper-based testing and computer-based testing be considered equivalent testing modes for VMBO-KB final exams for the topic of geometry?

1. Do the rank orders of scores of students tested in PBT and CBT closely approximate each other?
2. Are the characteristics of the score distributions – means, dispersions and shapes – of PBT and CBT approximately the same, or can be *made* approximately the same?
3. Do differences in PBT and CBT scores vary between item types?

Methods

Participants

The study was conducted in the Netherlands. The participants were high-school students primarily of ages 14-15. All students were recruited from the same school type and class, namely 3 VMBO-KB. The four-year VMBO level can be translated as "preparatory secondary vocational education", and prepares students for further education that will lead to a practical profession such as a mechanic, gardener, or nurse. The suffix KB specifies the level further: in this level, students follow a combination of practical courses and theoretical courses. In the fourth year of study, the final exams are administered. Since the fourth-year students had to take final exams during the time of this study, third-year students were selected.

Teachers of 3 VMBO-KB level classes were recruited to participate in the study on a voluntary basis. This led to responses from teachers located throughout the Netherlands. The participating schools are displayed in Figure 1. Some schools participated with more than one class: in total, thirteen classes from eight schools were involved, with a total of 129 students.

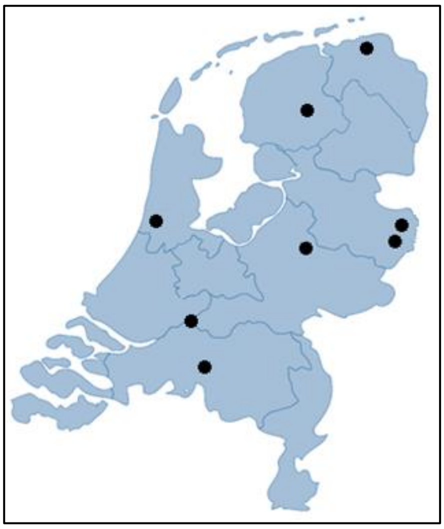


Figure 1. Participating schools

Instruments

Geometry tests were used as measurement tools. For the purpose of comparing PBT and CBT, tests were developed for the two conditions. In order to control for possible differences in skill level of students, each student was required to complete tests in both conditions. Furthermore, since administering the same test twice is not preferable, two parallel versions A and B of the test were developed. The order, subjects, score points, and difficulty of the items were similar. Hence, in total four tests were used: PBT-A, PBT-B, CBT-A and CBT-B.

Design process. First, the syllabus and old exams of the VMBO-KB level were investigated on the topic of geometry. Thereafter, a number of test questions with different activity types were created by the researcher. This led to an initial test design. The researcher received feedback on the initial test design by two experienced test experts of Cito, who both had more than 10 years of experience in creating national final mathematics exams. The test experts were closely involved during the design process and provided feedback multiple times. In pilot rounds, the tests were administered to students to improve the quality of the tests further.

Pilot rounds. Four small-scale pilot rounds were held in a separate room in a high school. In this room, one 3 VMBO-KB level student at a time voluntarily participated in the research. After a short instruction, the student was administered the test. In the first pilot round, it became clear that the students needed more content instruction before they could answer the test questions. Therefore, both an instruction and corresponding practice exercises set were developed. For the digital test, the instruction and practice exercises also contained an explanation on how the computer program operated.

The instruction, practice exercises and tests were tested on suitability of level, duration and clarity. Besides that, the digital program was still in developmental stage and was therefore critically reviewed at the end of each pilot.

After the small-scale pilot rounds, a classroom pilot round was performed. In this pilot class, a PBT test was administered. The pilot led to the purchase of geometry tools by the researcher since most of the students did not bring the necessary tools. Subsequently, some minor changes were made to the test items. At that moment, the design process was finished.

Test design. The eventual tests consisted of seven open questions that all required another activity type. In this way, most of the exam subjects within geometry construction were covered. The test activities per item are displayed in Table 1. The tools necessary for completing the test were a ruler, a set square with integrated protractor of 180 degrees, a compass, and a 360 degree protractor. For the CBT, these tools were available in digital form. In Appendix A and B, the paper-based and computer-based tests are included respectively. The answer sheets for all tests can be found in Appendix C.

Technical support. The computer-based tests were administered in Facet, which is the digital assessment player used for the VMBO-KB level final exams. The students opened the test on their computer by logging in with personal login codes. Facet displayed the questions of the test individually on a screen. This was in contrast to the paper-based test, in

Table 1

Question	Activity/goal
1	To measure a direction in degrees
2	To draw lines of sight
3	To draw an angle of a given size
4	To draw a circle of a given size
5	To draw a direction with a given size and a length of a given size
6	To draw the contours of a given object
7	To measure a length

which multiple questions were displayed on one sheet of paper, as can be seen in the comparison of PBT and CBT in Figure 2. Another difference between the CBT and PBT was that the CBT was in color and the PBT in black-and-white. The students could make use of digital geometry tools and a drawing tool for answering questions.

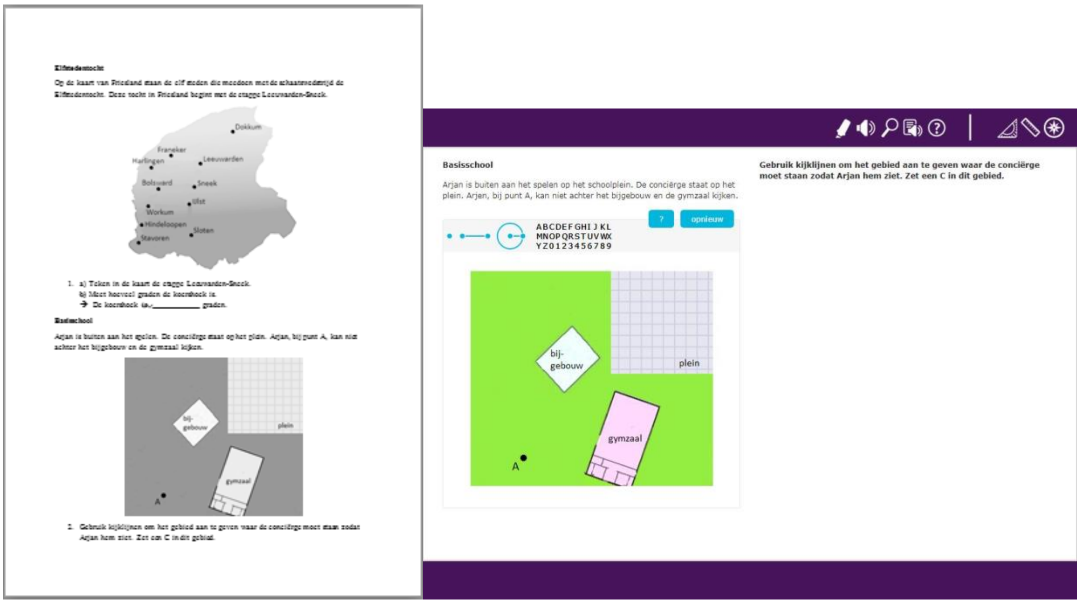


Figure 2. A test page in PBT (left) and a test screen in CBT (right)

Geometry tools. The digital geometry tools used were built-in features of Facet that have been used for a number of years in the digital final exams. The digital geometry tools consisted of a set square with integrated protractor of 180 degrees, a ruler, and a 360 degree protractor, see Figure 3. The tools could be activated by clicking on the pictogram of the tool and this tool then appeared in the middle of the screen. The students moved the tools by dragging the tool over the screen to the desired place. All tools displayed red dots: the set square and 360 degree protractor had one red dot on top and the ruler displayed two red dots on the ends of the ruler. These dots could be used by clicking and moving which resulted in a turning movement or enlargement of the tool.

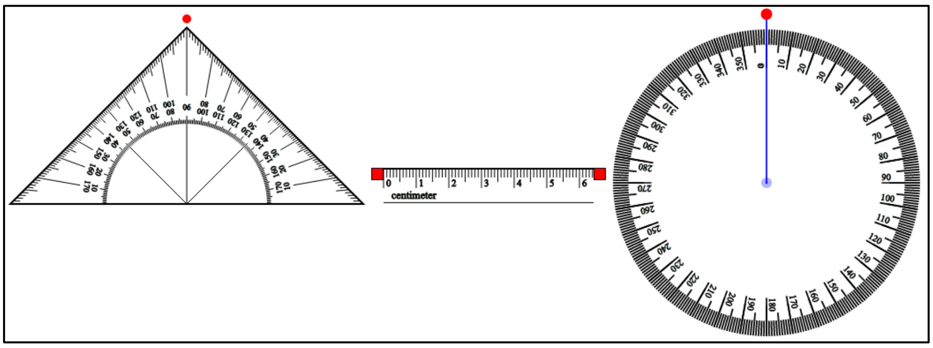


Figure 3. The geometry tools

Drawing tool. Since Facet did not facilitate a drawing tool, a drawing tool that suited this research was sought. The first potential program for this drawing tool was GeoGebra, which is dynamic geometry construction software. However, in the aforementioned pilot rounds it was found that GeoGebra in Facet resulted in technical issues and therefore did not suit this research. Thereafter, a potential drawing tool was programmed by ICT experts of Cito. This drawing tool was also tested in the subsequent pilot rounds. After each pilot round, the tool was adjusted by the programmers to the desired format. The eventual drawing tool, displayed in Figure 4, met all requirements and was used in this research.

Figure 4. The drawing tool

The drawing tool consisted of a top bar with the potential objects (points, lines, circles, letters and ciphers) and a drawing area. Drawing was equivalent to moving the objects from the top bar into the drawing area. Line segments could be lengthened and twisted by movable points at the end of each line. Circles could be enlarged by a movable point on the outer shape. In addition to the moving objects, a help button and reset button were also present in the top bar. The help button, represented by a question mark, explained how to use the drawing tool after clicking it. The complete drawing could be removed by clicking on the reset button. It should be noted that no “previous” or “delete” buttons were available. For that reason, it was told to the students that they could move away useless objects to the side of the drawing area, which would be neglected by the grading. Furthermore, the background of the drawing area of the tool differed for each question. For example, the background could be empty, a grid of squares, or a ground plan. Circles and angles with given sizes could be drawn with all of the geometry tools that were available.

Procedure

All examinees of the study were administered both a CBT and a PBT. To eliminate order effects, a counter-balanced repeated measure design was used. At the student level, this was impractical, and therefore half of the classes first completed the PBT and then the CBT and the other half vice versa. Kim and Huynh (2007) also employed this design with alternate versions of the test for PBT and CBT to examine the comparability of student scores. The researchers commented that the version effect could not be distinguished from the mode effect and that therefore “caution should be exercised when interpreting the results” of their study (p. 26). Taking this into consideration, this study was designed to control for both the effects of the test order and different versions.

Each participating class was assigned to a research group. The research groups differed by delivery mode and whether they were administered test version A or B, that can be found in Table 2. It can be seen that group 1 and group 4 made the same test, but in different order, similar as group 2 and group 3. Besides that, group 1 and 2 had the same test mode order, but differed in versions of the test, similar as group 3 and 4. Therefore, both the test order and version effect were controlled for in this design. Furthermore, the time span between the first and the second test moment was at least two weeks, with the exception of one class that had one week in between. The minimum of two weeks between the tests was chosen to minimize the effects of practice and fatigue (Hosseini & Toroujeni, 2017).

Table 2

Group	First test	Second test
1	PBT-A	CBT-B
2	PBT-B	CBT-A
3	CBT-A	PBT-B
4	CBT-B	PBT-A

Lesson procedure. The tests were part of 40 minutes lessons on geometry with the researcher as a guest teacher. Every class took two of these lessons, one lesson completely with paper and pencil and one lesson on computer. Both lessons consisted of three activities: a classical instruction, practice exercises, and the test.

The first activity, the classical instruction, was performed by the researcher. In this instruction, the concepts of angle and course were explained to the students in order to acquire the necessary knowledge. In the digital version of the lesson the digital program used in the CBT was also explained. In Appendix D, the classical instruction is included for both the paper-based and computer-based lesson.

The second activity included the students working on the practice exercises provided by the researcher. At this moment, the students were allowed to ask questions to each other and to the researcher. The content of the exercises was on angles and course. In order to get used to the digital environment, some extra practice exercises were included that were not correlated in terms of content with the exercises of the tests. In Appendix E, the practice exercises can be found. Screenshots of the exercises are also included for the computer-based lesson.

The final activity of the lesson was the geometry test. The class was silent for the duration of the test and the students were required to work individually. However, the tests did not count as a grade for the students.

Grading

The researcher graded all of the student work. First, the researcher globally examined the test answers and prepared a list of the possible mistakes and how to assign points. Thereafter, the researcher used this list as a reference with all revising work. The digital tests could be revised on the computer after the researcher was assigned the role of reviser in Facet.

A portion of the tests were revised by a second grader, who worked for more than three years as a mathematics teacher. This grader was randomly assigned to score all tests for one specific group and time moment and a complementary group. The test groups that were selected randomly for the second revision were group 1 test 1, which was PBT-A, and group 4 test 1, which was CBT-B. This resulted in 57 tests scored by the second reviser, or 22% of the tests corrected in total. There was substantial agreement between the two test raters, determined by the values of $\kappa = .749$ and $p < .0005$ (Landis & Koch, 1977).

Data collection

The data were collected within a time span of two months. During the data collection, some issues hindered proper execution of the lesson procedure. Three classes suffered from time-consuming technical issues in the computer-based lesson and as a consequence ran out of time before the test was finished. Therefore, the results of these tests were not used for analysis. Another class was not motivated to do the tests both times and the researcher therefore did not use the results of this class as well. Consequently, the results of nine classes were analyzed.

Analysis

Reliability. In order to determine the internal consistency of the test, the Cronbach's alpha statistical test was performed. Although it is necessary to calculate the reliability measure because it is needed for other analyses, Cronbach's alpha is expected to be low for the following reason. Since the students did not get the opportunity to study the subjects beforehand, they were solely dependent on their prior knowledge about geometric topics. This prior knowledge is likely to differ between students from different classes or schools. A good student from one school may receive full points on the question about "lines of sight", while he or she may not score any points on the question about "views". If a good student from another school does exactly the contrary, the correlation between the questions equals 0. This issue of differing prior knowledge may systematically decrease the value of Cronbach's alpha.

Cronbach's alpha should be performed for a sample that has at least a solid number of correctly administered tests, such as more than 100. This was possible for one test version only: PBT version B (N=113). It should be noted that this includes tests that were not used further in the analysis, since for example the computer-based lesson was not executed properly. The low Cronbach's alpha of the test, $\alpha = .374$, implied that the internal consistency of the paper-based test version B was unacceptable (Bland & Altman, 1997). Since the internal consistency did not reach an acceptable level by removing items, the researcher decided to perform the analysis on the original test. The issue of the low Cronbach's alpha value is examined further in the discussion section.

Multivariate analysis of variance. A multivariate analysis of variance (MANOVA) was performed with test scores as the independent variable, one within-subject factor (test mode), one between-subject factor (group) and the interaction term. The interaction term cancels out possible effects between test modes and groups such as when one group has practiced more with mathematics on a computer. For this research, the main effect of the within-subject factor mode is important, since this determines whether the test mode affects the test score.

In addition, the correlation between the scores in both conditions is of interest. Classical test theory states that the true score of a test is latent because the observed test score is the true score plus error. Since this study requires the measurement of the correlation in the true scores, hence the two latent variables, it is necessary to calculate the disattenuated correlation. If we have two observed test scores, x and y, the "disattenuated correlation is the raw correlation between x and y (r_{xy}) divided by the square root of the product of the reliability of x (r_{xx}) and the reliability of y (r_{yy})" (Murphy & Davidshover, 1998, p. 130). For this case, this can be calculated by the equation:

$$r_{att,s_p s_d} = \frac{r_{s_p s_d}}{\sqrt{r_{s_p s_p}}\sqrt{r_{s_d s_d}}}$$

with s_p = observed score on paper-and-pencil test

s_d = observed score on digital test

If the correlation is almost as high as the reliability of the tests, and the disattenuated correlation coefficient is therefore high, the latent correlation is almost perfect and the same underlying construct is measured in both tests.

Results

Rank order

To answer the first sub research question on possible differences in rank order between the two test modes, the disattenuated correlation between the student scores was calculated. The formula for the disattenuated correlation was already provided in the Methods section. Since it was only possible to calculate the Cronbach's alpha value of one version of the test, it is necessary to assume that the reliability for each version of the test is equivalent (0.375). Furthermore, it was determined that the observed correlation between the test scores is 0.285. Therefore, the disattenuated correlation is calculated with the following result:

$$r_{att,s_p s_d} = \frac{0.285}{0.375} = 0.76$$

Therefore, the true correlation between the tests is strong.

Characteristics of the score distributions.

To answer the second research question, the characteristics of the score distributions for both test modes are investigated. The characteristics that will be compared are the means, dispersions and shapes of the distributions.

Means. In Table 3, descriptive statistics are provided. For each group, test score means and standard deviations for paper and computer can be found. It can be seen that all groups score higher on paper than on computer and that the average test score mean for PBT (14.59) is circa two score points higher than the average test score mean for CBT (12.69).

Table 3

Descriptive statistics

Group	*PBT Mean (Sd.)*	*CBT Mean (Sd.)*	*N*
1	14.91 (3.95)	11.59 (3.83)	32
2	12.76 (3.58)	11.82 (3.71)	34
3	14.95 (2.88)	13.76 (3.67)	38
4	15.76 (2.18)	13.56 (3.07)	25
Total	14.52 (3.40)	12.67 (3.71)	129

A MANOVA was conducted to test the null hypothesis that there was no difference in scores between the paper-based and computer-based test. A statistically significant MANOVA effect was obtained, with Pillai's Trace = .178, $p < .000$. The interaction term turned out to be insignificant. Students in the PBT condition scored significantly higher than in the CBT condition, with a 95% confidence interval of the difference between means from 0.7 to 3.1 score points on a scale of 21 score points.

Dispersions. In order to compare the dispersions of the score distributions, a comparison between variances of both test modes scores is used. In Figure 3 and 4 frequencies of tests scores are given for PBT and CBT respectively. From these figures, it becomes clear that CBT has more low scores than PBT. This indicates that the variance of the CBT is higher than the variance of PBT. From the reported variances from the residual covariance matrix provided after the MANOVA was run, the variance of PBT equals 10.62 and the variance of CBT equals 13.09. This arouses the conjecture again that the variances are not equal to each other, which is confirmed by the Bartlett's Test of Sphericity, that is statistically significant ($\chi^2(2) = 11.87$, p=.003). Hence, the variances of the score distributions are not equal and should be rescaled when a transfer from PBT to CBT takes place.

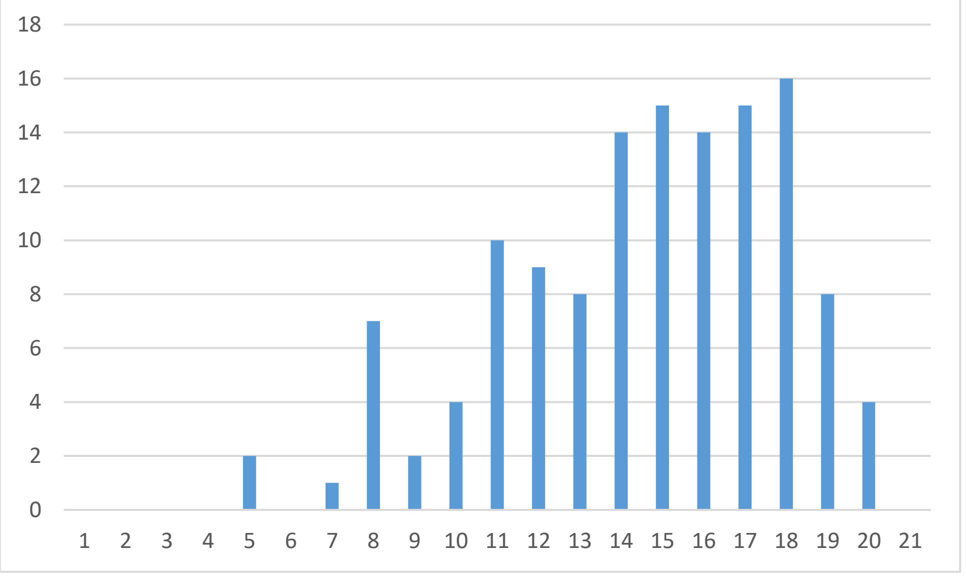


Figure 3. Frequency distribution for PBT scores

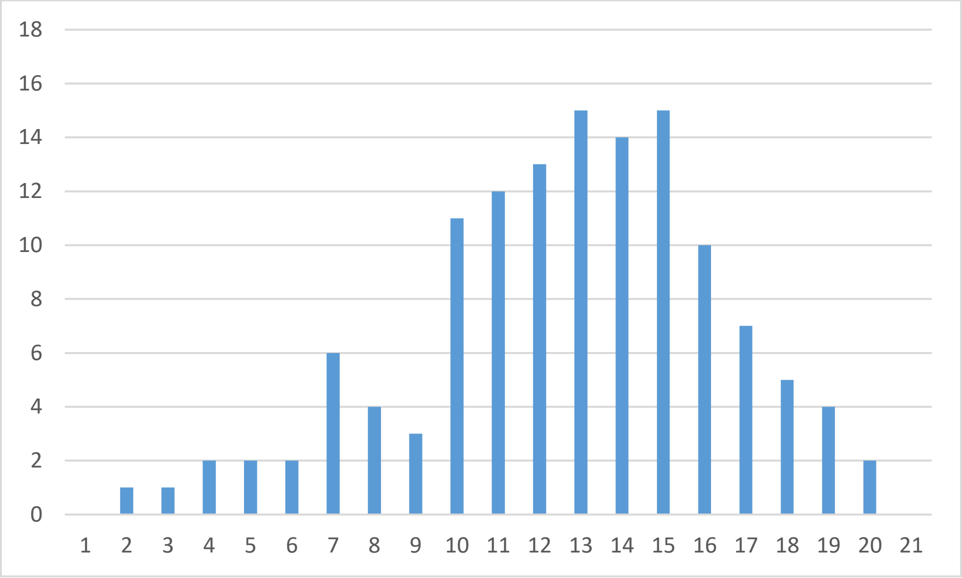


Figure 4. Frequency distribution for CBT scores

Shapes. The shapes of the distribution are defined by the skewness of the distribution. Since we have rather small group sizes, this characteristic is only investigated by eyeballing and not by statistical tests. Figures 3 and 4 show a difference in skewness. The PBT looks quite negatively skewed, while CBT shows a negatively skewed to normal distribution. Hence, the shapes of the distribution should be rescaled when equivalence is required.

Differences on item-level

Since the interaction term in the MANOVA was insignificant, the effect of which test group was administered which specific test is ignorable. Therefore, the different test groups were aggregated and mean item level scores of PBT and CBT were calculated. Since the items differ by maximum number of score points, the means were converted into relative percentages of correct answers; these are provided in Table 4. In this table, it is evident that the item types behave differently when transferred to CBT. Some item types, such as item 1, 2, and 5, have relatively small differences in percentages correct over the different test modes, while item 3 has a relative large difference. This stresses the importance of not only considering the equivalence of the complete test in different test modes, but also to focus on the specific item types (Threllfall et al., 2007).

Table 4.

Item	*PBT*	*CBT*	*Difference (PBT-CBT)*
1	71%	68%	3%
2	63%	61%	2%
3	59%	37%	22%
4	68%	59%	10%
5	69%	67%	2%
6	81%	70%	11%
7	83%	74%	9%

To investigate whether the differences are statistically significant or not, a dependent *t*-test was performed to compare the scores on the different items of the tests for the same participants. The results of this procedure are displayed in Table 5. The scores on questions 1, 2 and 5 can be considered to be similar on the PBT and CBT ($p>.05$). On the other hand, the scores of questions 3, 4, 6 and 7 were found lower on the CBT ($p<.05$).

Table 5

Item number	*Activity/goal*	*Paired difference in mean scores (PBT-CBT)*	*Sd.*	*t*	*Sig.*
1	To measure a direction in degrees	0.07	1.06	0.75	0.457
2	To draw lines of sight	0.05	1.54	0.34	0.731
3	To draw an angle of a given size	0.87*	1.74	5.67	0.000
4	To draw a circle of a given size	0.29*	1.62	2.01	0.047
5	To draw a direction with a given size and a length of a given size	0.07	1.66	0.48	0.634
6	To draw the contours of a given object	0.33*	1.17	3.24	0.002
7	To measure a length	0.17*	0.95	2.03	0.044

* significant on 95% confidence level

Both questions 1 and 5 can be considered “the same” on the PBT and CBT ($p > .05$). These questions required rather complicated operations with the geometry tools and drawing tool. Question 5 was the most complex because a direction with a given length had to be drawn; this question required the use of the 360 degree protractor, ruler, and drawing tool. Since both questions 1 and 5 have the same score points on PBT and CBT, students were indifferent to paper and computer for measuring and constructing with the 360 degree protractor and ruler. This indicates that the students were not hindered by technological difficulties of the computer program. Moreover, questions 1 and 5 measure the activities that were practiced in the practice assignments, which may indicate that practicing questions on computer enables a reduction or even elimination of the test mode effect.

However, the drawing of an angle was also practiced in the practice exercises, but question 3 demonstrates the largest difference between PBT and CBT, namely 0.87 ($p < .000$). Since drawing an angle with a given size was also practiced, it was expected that the students would not be hindered by ignorance of the tools in this question either. Therefore, another systematics may explain the large difference in scores. The primary difference between question 3 and questions 2 and 5 was that question 3 required mathematical thinking. The students had to investigate where and how the angle should be drawn, which was not practiced prior to the test. The lower scores on the CBT may indicate that when students need to reason mathematically, PBT is easier for the students. This may also explain the significantly lower scores for CBT in question 4, because mathematical reasoning was required in this question as well.

For questions 6 and 7, it is unclear what the significant difference in PBT and CBT score can be attributed to. For question 6, a common mistake was that students drew the complete figure instead of only the lowest part. Drawing the complete figure costs a significant amount of time on the computer, which may have lowered motivation and resulted in mistakes. However, it is not possible to explain this difference with certainty. For question 7, a possible reason for the lower CBT scores is that some students began to measure from the red dot on the end of the digital ruler instead of from the zero. This may have caused some difference, although other questions show that working with the ruler was not a problem. However, the difference between PBT and CBT is barely significant and small for this question.

Conclusion

In this study, the possible transfer from paper-based to computer-based geometry examination was researched for the VMBO-KB level. For this purpose, it was investigated whether paper-based testing (PBT) and computer-based testing (CBT) can be considered equivalent testing modes for geometry tests for this specific student population. Therefore, parallel tests were designed and administered in both delivery modes to 3 VMBO-KB school students. The internal consistency could only be determined for one out of four versions of the test, and this internal consistency turned out to be unacceptable (7 items, $\alpha = .374$). The low internal consistency can likely be attributed to the different prior knowledge of the students, which may have systematically decreased Cronbach's alpha.

Equivalence of test-modes consists of two aspects: a similar rank order between scores on the test modes and similar characteristics of the score distribution (APA, 1986). The rank order was measured by the disattenuated correlation, the correlation corrected for error variance, which was 0.76. Therefore, the latent correlation between the test modes was substantially high and the same underlying construct was measured in both tests.

The characteristics means, dispersions, and shapes of the score distributions were investigated. Scores of the PBT were significantly higher than scores of the CBT, with a 95% confidence interval of the difference between means from 0.7 to 3.1 score points on a scale of 21 score points. In order to compare the variances of the score distributions, Bartlett's Test of

Sphericity was run. The test appeared statistically significant ($\chi^2(2) = 11.87$, p=.003), which implied a higher variance of the CBT distribution. For the determination of similar skewness of the distributions, no statistical test was performed. However, a comparison between bar charts of the frequency distributions indicated a difference in skewness, with PBT more negatively skewed than CBT.

An investigation of the differences on item-level led to several notions. Firstly, for some questions the test-mode effect was eliminated. This was the amongst others the case for questions that were practiced during the practice exercises, which may indicate that practicing questions on computer enables a reduction or even elimination of the test mode effect. Secondly, the test-mode effect seemed to be larger for questions that required more mathematical thinking. Lastly, for some questions it was not clear where the difference in score points could be attributed to – which stresses the importance of comparing PBT and CBT on item level as well.

Since the rank order was similar and rescaling eliminates any differences in characteristics of the score distributions, PBT and CBT can be considered equivalent test modes for VMBO-KB final exams.

Discussion

The results of this study are in line with a recent meta-analysis on mathematics tests, in which no statistically significant difference between CBT and PBT for mathematics was found (Wang et al., 2008). Although other studies found that for more complex question types the test modes will likely differ (Bennett et al., 2008; Griffin et al., 2014), this study shows that even for geometry construction questions, PBT and CBT are interchangeable. It should be noted, however, that this result holds for the VMBO-KB level specifically: the level of geometry questions is rather low compared to other high school levels and therefore the result cannot be generalized directly.

Although the rank order was found to be similar for PBT and CBT in this study, the characteristics of the score distributions were all different. A possible explanation for the CBT mean to be statistically lower than the PBT mean, is the short time span for the students in order to get familiar with the computer program. Students were only provided with 20 minutes time to both listen to the instruction and work on the practice exercises. Therefore, possibly the lower 2 score points on the CBT can be removed or lessened when more practice time is provided.

Besides a significant difference in mean test score, the variances of the scores turned out to be unequal. The variance of the test score distribution of CBT was statistically higher than variance of PBT test scores. This can possibly be explained by the decrease in test scores of students who find it harder to work on computer, while students who were indifferent between the test modes received the same test score. In that way, the dispersion of CBT scores increases. This systematics may also explain the more negatively skewed distribution of the CBT. However, all these differences might disappear when the students are more used to computer geometry.

Limitations

The major limitation of this study is the low internal consistency of the test instruments. However, the researcher does not claim that the complete test measures one underlying concept such as “geometric ability”. The students were only allowed to take *parallel tests* in different conditions. If a parallel question was asked on paper or on the computer, it does not matter whether the correlation with other questions is high or low. Only the *difference* between the total test scores in the different conditions is important for this research.

However, the size of this contribution to the low Cronbach’s alpha value cannot be distinguished from other factors that lower the internal consistency of the tests. Hence, the argument that these tests are reliable is only supported by the qualitative argument that the tests were verified by two experienced test experts. Since the reliability cannot be supported by a quantitative argument, the results of this study should be interpreted with care.

Another issue regarding this research is the fact that students did not receive a grade for the tests that counted as a study result. Therefore, the students could likely have been less motivated to do the tests as they would have been if they had received a grade for the tests. Although the tests of students that had clear motivational issues were not included in the analysis, it was not possible to eliminate the motivation effect completely in the study design.

Although the students completed the test in proper conditions, purposeful or accidental cheating may also have influenced the test scores. In the computer rooms where the CBT was conducted, shelves between the computers were placed to avoid cheating for most of the classes. However, these were not available for some classes. It is also possible that in the PBT students observed each other working with a compass and then realized that that was the tool they needed for answering a particular question.

Further research

Since this study indicates that mathematical reasoning is not easily transferred from PBT to CBT, it may be promising for future research to examine this issue further. For example, it can be researched how students' mathematical reasoning can be stimulated on the computer. In addition, research on this topic may be expanded to other high school levels, since this may open the way for digital testing in those levels of education as well.

Lastly, this research has been the only research conducted comparing paper-based and computer-based testing in the Netherlands on such a scale. However, because computerized testing will be the dominant testing mode in the future (Wang et al., 2008), this research has revealed how important it is that such research is conducted and how insightful this area of research can be for the field of science education.

Acknowledgements

I would like to express my gratitude to prof. dr. Drijvers, for his patient guidance and valuable suggestions. I would also like to thank Jos Remijn for the pleasant internship and useful discussions. Besides that, advice given by Marieke van Onna has been a great help in the statistical analyses. I also wish to acknowledge the help provided by Merijn IJperlaan, who helped me by being the second grader, and Melanie Steentjes, for her constructive feedback on the tests.

Special thanks should be given to Fabian Pothof for facilitating the creation of the drawing tool, and Kevin Schuit, Sander Masselink, Bart Veenhof, and Jaap van der Sluis for programming and testing the drawing tool. Technical support regarding Facet provided by Frederik de Boer was greatly appreciated.

My special thanks are extended to all participating teachers and Cito colleagues for their enthusiasm and help, and my friends and family for their feedback and support.

References

American Psychological Association. (1986). *Guidelines for computer-based tests and interpretations*. Washington, DC: Author.

Bennett, R.E., Braswell, J., Oranje, A., Sandene, B., Kaplan, B., Yan, F. (2008). Does it Matter if I Take My Mathematics Test on Computer? A Second Empirical Study of Mode Effects in NAEP. *Journal of Technology, Learning, and Assessment, 6*(9).

Bland, J. M., & Altman, D. G. (1997). Statistics notes: Cronbach9s alpha. *Bmj*, *314*(7080), 572.

Bodmann, S. M., & Robinson, D. H. (2004). Speed and performance differences among computer-based and paper-pencil tests. *Journal of Educational Computing Research*, *31*(1), 51-60.

Bugbee Jr, A. C., & Bernt, F. M. (1990). Testing by computer: Findings in six years of use 1982-1988. *Journal of Research on Computing in Education*, *23*(1), 87-100.

Choi, I. C., Kim, K. S., & Boo, J. (2003). Comparability of a paper-based language test and a computer-based language test. *Language Testing*, *20*(3), 295-320.

Clariana, R., & Wallace, P. (2002). Paper–based versus computer–based assessment: key factors associated with the test-mode effect. *British Journal of Educational Technology*, *33*(5), 593-602.

College voor Toetsing en Examens. (2017). Kaders en flexibele en digitale centrale examens BB en KB 2018.

Drijvers, P. (2018). Digital assessment of mathematics: opportunities, issues and criteria. *Mesure et évaluation en education*.

Gould, J. D., Alfaro, L., Finn, R., Haupt, B., & Minuto, A. (1987). Reading from CRT displays can be as fast as reading from paper. *Human factors*, *29*(5), 497-517.

Griffin, P., McGaw, B., & Care, E. (2012). Assessment and Teaching of 21st Century Skills. *Springer*

Hensley, K. K. (2015). Examining the effects of paper-based and computer-based modes of assessment on mathematics curriculum-based measurement.

Hosseini, M., & Toroujeni, S. M. H. (2017). Replacing Paper-Based Testing with an Alternative for the Assessment of Iranian Undergraduate Students: Administration Mode Effect on Testing Performance. *International Journal of Language and Linguistics*, *5*(3), 78.

Inouye, D. K., & Bunderson, V. C. (1986). Four generations of computerized test administration. *Machine-Mediated Learning*, *1*(4), 355-371.

Kim, D. H., & Huynh, H. (2007). Comparability of computer and paper-and-pencil versions of algebra and biology assessments. *The Journal of Technology, Learning and Assessment*, *6*(4).

Khoshsima, H., & Toroujeni, S. M. H. (2017). Comparability of Computer-Based Testing and Paper-Based Testing: Testing Mode Effect, Testing Mode Order, Computer Attitudes and Testing Mode Preference. *International Journal of Computer (IJC)*, *24*(1), 80-99.

Kreis, Y., Dording, C., Keller, U., Porro, V., & Jadoul, R. (2011). Dynamic mathematics and computer-assisted testing: GeoGebra inside TAO. *ZDM The International Journal on Mathematics Education*, *43*(3), 1-10.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.

Lee, J. A., Moreno, K. E., & Sympson, J. B. (1986). The effects of mode of test administration on test performance. *Educational and Psychological Measurement*, *46*(2), 467-474.

Murphy, K. R., & Davidshofer, C. O. (1988). Psychological testing. Principles, and Applications, Englewood Cliffs.

Neuman, G., & Baydoun, R. (1998). Computerization of paper-and-pencil tests: When are they equivalent?. *Applied Psychological Measurement*, *22*(1), 71-83.

Pomplun, M., Frey, S., & Becker, D. F. (2002). The score equivalence of paper-and-pencil and computerized versions of a speeded test of reading comprehension. *Educational and psychological measurement*, *62*(2), 337-354.

Threlfall, J., Pool, P., Homer, M., & Swinnerton, B. (2007). Implicit aspects of paper and pencil mathematics assessment that come to light through the use of the computer. *Educational Studies in Mathematics*, *66*(3), 335.

Van de Vijver, F. J., & Harsveld, M. (1994). The incomplete equivalence of the paper-and-pencil and computerized versions of the General Aptitude Test Battery. *Journal of Applied Psychology*, *79*(6), 852.

Wang, S., Jiao, H., Young, M. J., Brooks, T., & Olson, J. (2008). Comparability of computer-based and paper-and-pencil testing in K–12 reading assessments: A meta-analysis of testing mode effects. *Educational and psychological measurement*, *68*(1), 5-24.

Appendix A – Paper-based tests

Test version A

Meetkundetoets versie A

Naam: ______________________

Klas: ______________________

School: ______________________

Datum: ______________________

Benodigdheden:

- Geodriehoek
- Koershoekmeter
- Passer
- Potlood + pen
- Gum

De toets heeft **7 vragen** en je hebt **30 minuten** de tijd. Succes!

Texel

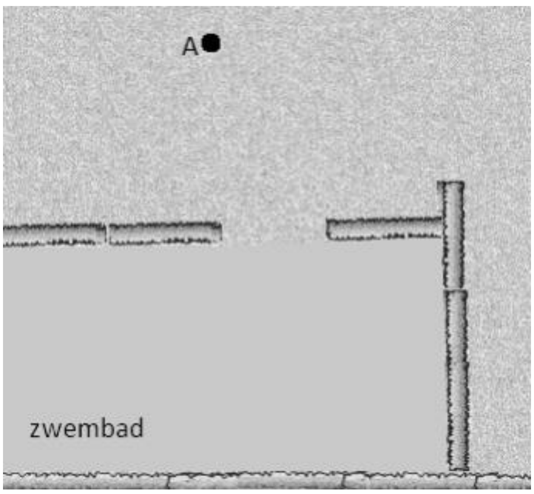
Semna is op vakantie op Texel. Ze rijdt met de auto van Den Burg naar De Koog.

1. a) Teken in de kaart hoe Semna rijdt.
   b) Meet hoeveel graden de koershoek is.
   ➔ Het aantal graden is _______________.

Achter de heg

Anja staat bij punt A. Zij kan niet over de heg kijken. Marit zwemt in het zwembad.

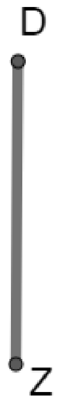
2. Gebruik kijklijnen om het gebied aan te geven waar Marit moet zwemmen zodat Anja haar ziet. Zet een M in dit gebied.



Over de rand

Hiernaast zie je een zijaanzicht van een persoon op een schommel.

Hieronder zie je een schets van de schommel. Het is het zijaanzicht van het touw van de schommel in rust. Punt D is het draaipunt en punt Z het zitvlak. De schommel schommelt vanuit het draaipunt D naar links met een maximale hoek van 70°.



3. Teken het touw in deze maximale positie en geef hierbij het nieuwe punt Z aan. Zorg ervoor dat het touw even lang blijft.

DE TOETS GAAT VERDER OP DE VOLGENDE PAGINA

België

Er staan twee telefoonzendmasten in België, beide met een bereik van 3 cm op de kaart. De telefoonzendmasten staan in Gent en Luik.

4. a) Teken het bereik van de zendmasten op de kaart.
   b) Er zijn steden die door beide zendmasten kunnen worden bereikt. Welke steden kunnen door beide zendmasten worden bereikt?
   → De steden
   ____.

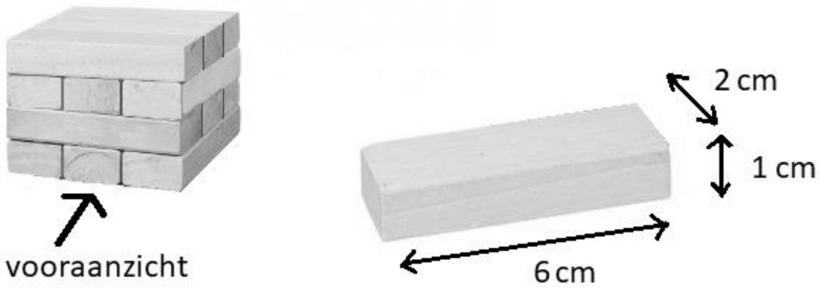
Horloge zoeken

Pieter is zijn horloge verloren ergens in het gras. Hij loopt met een koershoek van 190° op de kaart een afstand van 2 cm. Op die plek vindt hij zijn horloge.

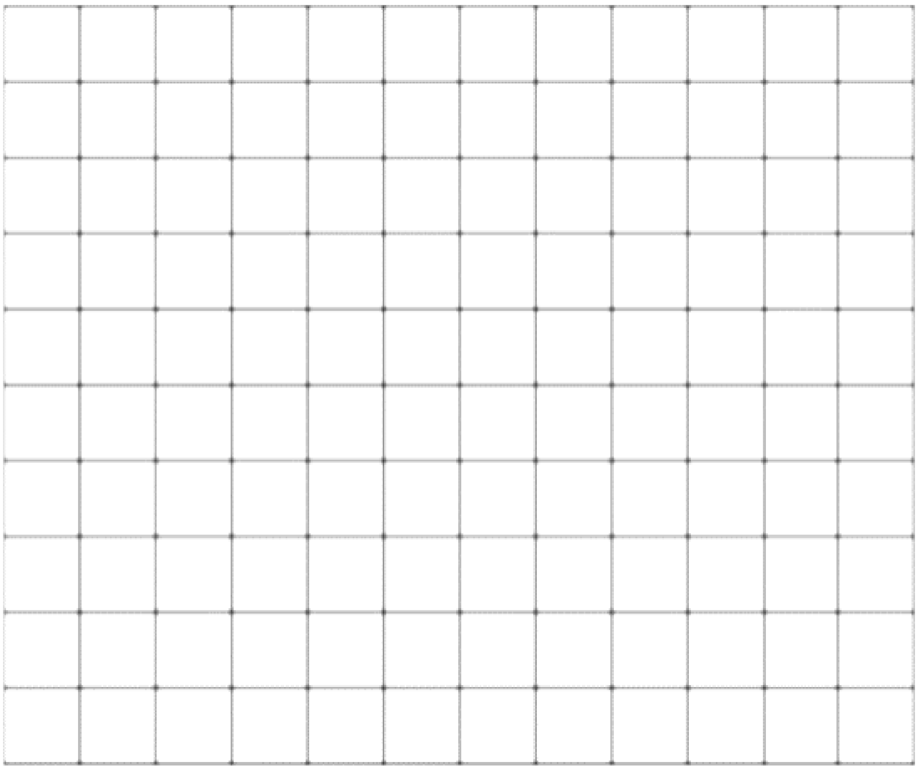
5. a) Teken op de kaart de route die Pieter vanuit punt P loopt.
   b) Zet een kruis op de plek waar hij zijn horloge vindt.

Blokjesstapel

Hieronder zie je een afbeelding van een stapel blokjes. Elke laag bestaat uit drie balkvormige blokjes. Bij het losse blokje zijn de afmetingen gegeven.



6. Teken het vooraanzicht van de onderste laag van de stapel op het ruitjespapier. De ruitjes op het ruitjespapier zijn 1 bij 1 cm.



DE TOETS GAAT VERDER OP DE VOLGENDE PAGINA

Ansichtkaart

Nika meet de hoogte van de Domtoren op een ansichtkaart op.

7. Meet de hoogte van de Domtoren op de kaart. Rond je antwoord af op hele mm.
   ➔ De hoogte is ______________________________.

<u>EINDE TOETS</u>. CONTROLEER OF JE NAAM OP DE VOORKANT STAAT

Test version B

Meetkundetoets versie B

Naam: ______________________

Klas: ______________________

School: ______________________

Datum: ______________________

Benodigdheden:

- Geodriehoek
- Koershoekmeter
- Passer
- Potlood + pen
- Gum

De toets heeft **7 vragen** en je hebt **30 minuten** de tijd. Succes!

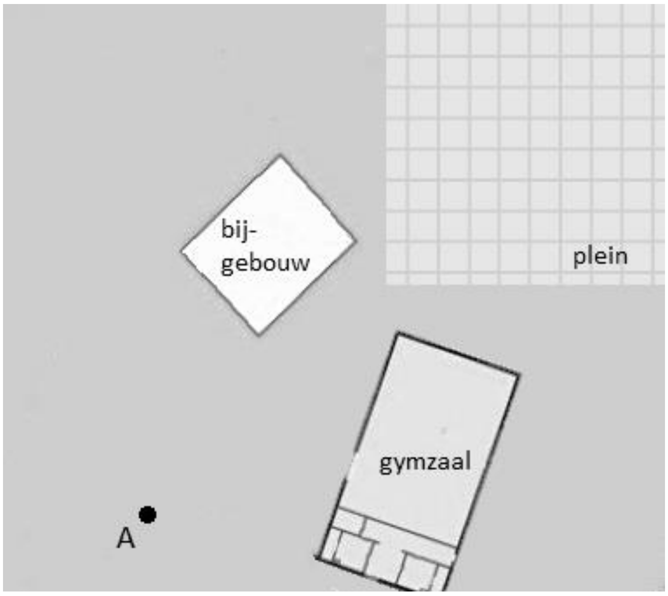
Elfstedentocht

Op de kaart van Friesland staan de elf steden die meedoen met de schaatswedstrijd de Elfstedentocht. Deze tocht in Friesland begint met de etappe Leeuwarden-Sneek.

1. a) Teken in de kaart de etappe Leeuwarden-Sneek.
   b) Meet hoeveel graden de koershoek is.
   ➔ De koershoek is _____________ graden.

Basisschool

Arjan is buiten aan het spelen. De conciërge staat op het plein. Arjan, bij punt A, kan niet achter het bijgebouw en de gymzaal kijken.

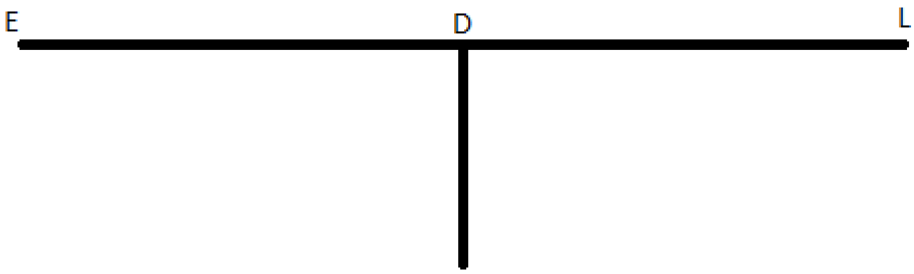


2. Gebruik kijklijnen om het gebied aan te geven waar de conciërge moet staan zodat Arjan hem ziet. Zet een C in dit gebied.

Wip

Twee meisjes, Eva en Lisa, zitten op een wip in een speeltuin.

Hieronder zie je een schets van de zijkant van de wip in rust. Eva is aangegeven met punt *E* en Lisa is aangegeven met punt *L*. De wip beweegt om punt D, het draaipunt, met een maximale hoek van 30°.

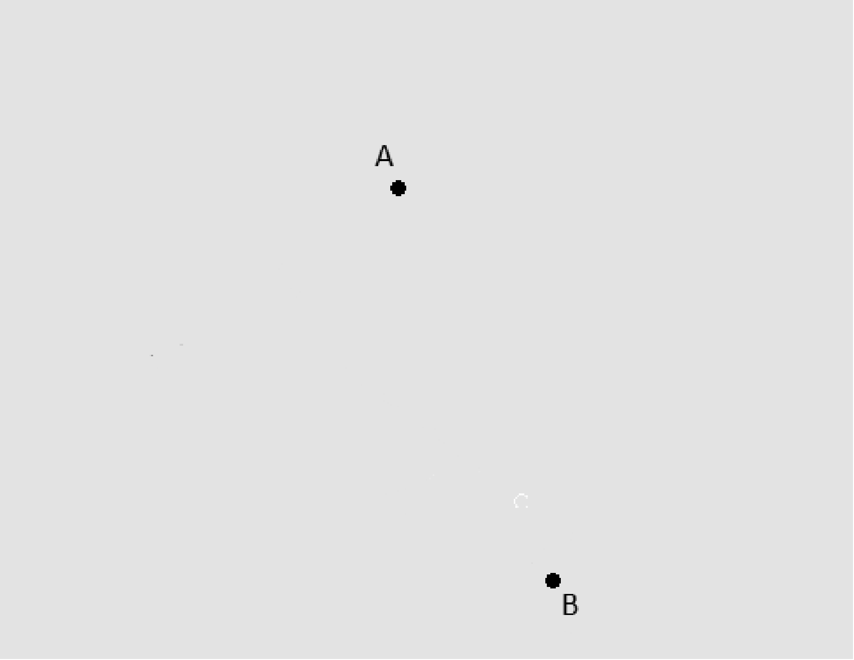


3. Teken de wip in de maximale positie waarbij Eva lager is dan Lisa. Zet de letters *E* en *L* op de goede plaats. Zorg ervoor dat de wip even lang blijft.

DE TOETS GAAT VERDER OP DE VOLGENDE PAGINA

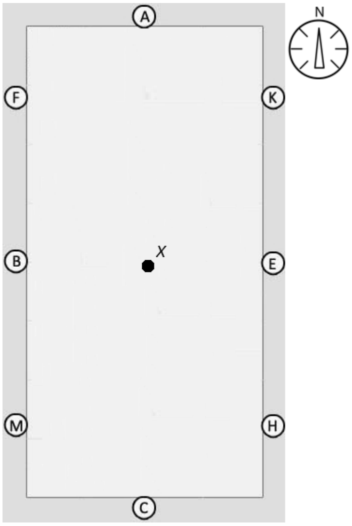
Schatzoeken

Manuel wil een schat opgraven. Hij weet dat de schat op de kaart 3 cm van punt *A* af ligt en 4 cm van punt *B.* Zo blijven er twee plekken over waar de schat zou kunnen liggen.



4. Teken een kruis op de plekken waar de schat zou kunnen liggen. Laat zien hoe je aan je antwoord gekomen bent.

Dressuur



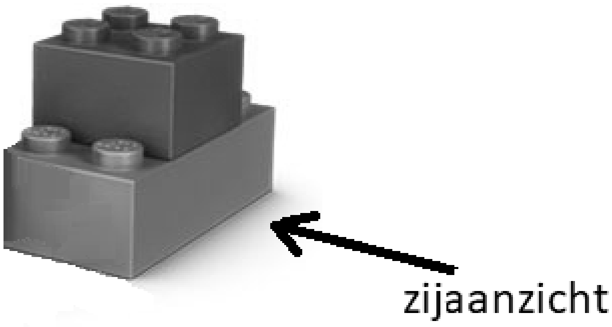
Dressuur is een sport waarbij ruiters op paarden figuren rijden door een bak. Hiernaast zie je een bovenaanzicht van een bak, waarbij letters zijn aangegeven.

Bij een oefening loopt een paard van punt *X* met een koershoek van 320° een afstand van 2 cm.

5. a) Teken op de kaart de route die het paard loopt.
   b) Zet een P op de plek waar hij eindigt.

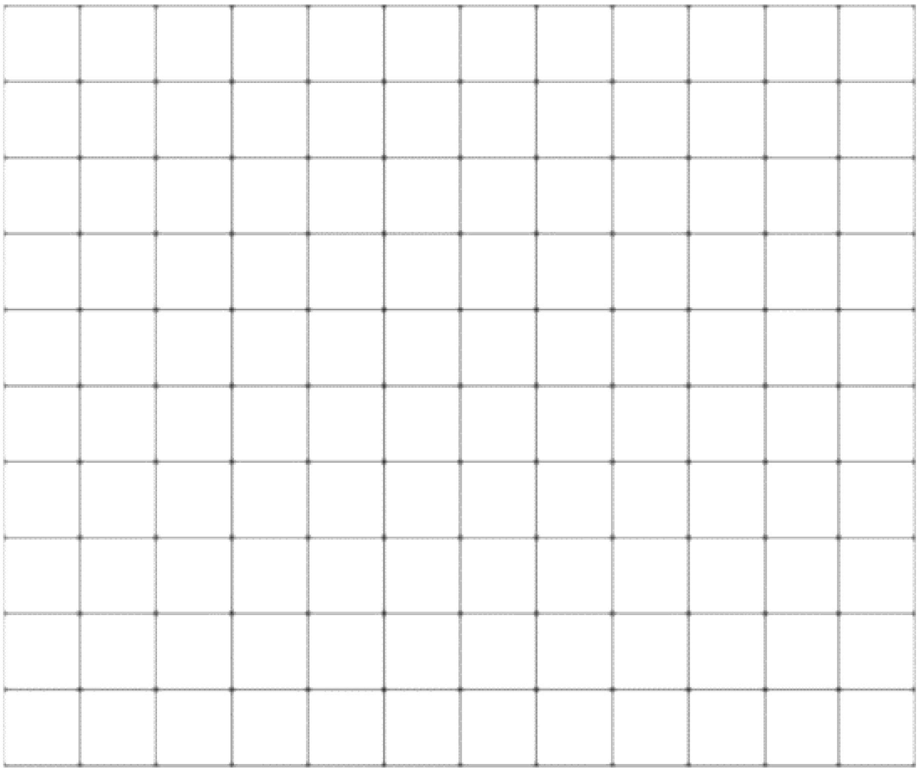
Lego

Hieronder zie je twee legoblokjes op elkaar. In de tabel zie je de afmetingen van de blokjes.



	Lengte	**Breedte**	**Hoogte**
Bovenste blokje	2 cm	2 cm	1 cm
Onderste blokje	4 cm	2 cm	1 cm

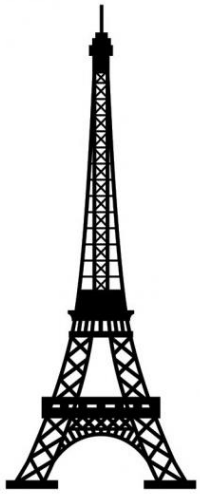
6. Teken het zijaanzicht van de blokjes op het ruitjespapier met ruitjes van 1 bij 1 cm. Je mag de rondjes op de blokjes verwaarlozen.



DE TOETS GAAT VERDER OP DE VOLGENDE PAGINA

Eiffeltoren

Marjolein heeft een kleine Eiffeltoren meegenomen als souvenir uit Parijs. Hieronder zie je deze Eiffeltoren op ware grootte.



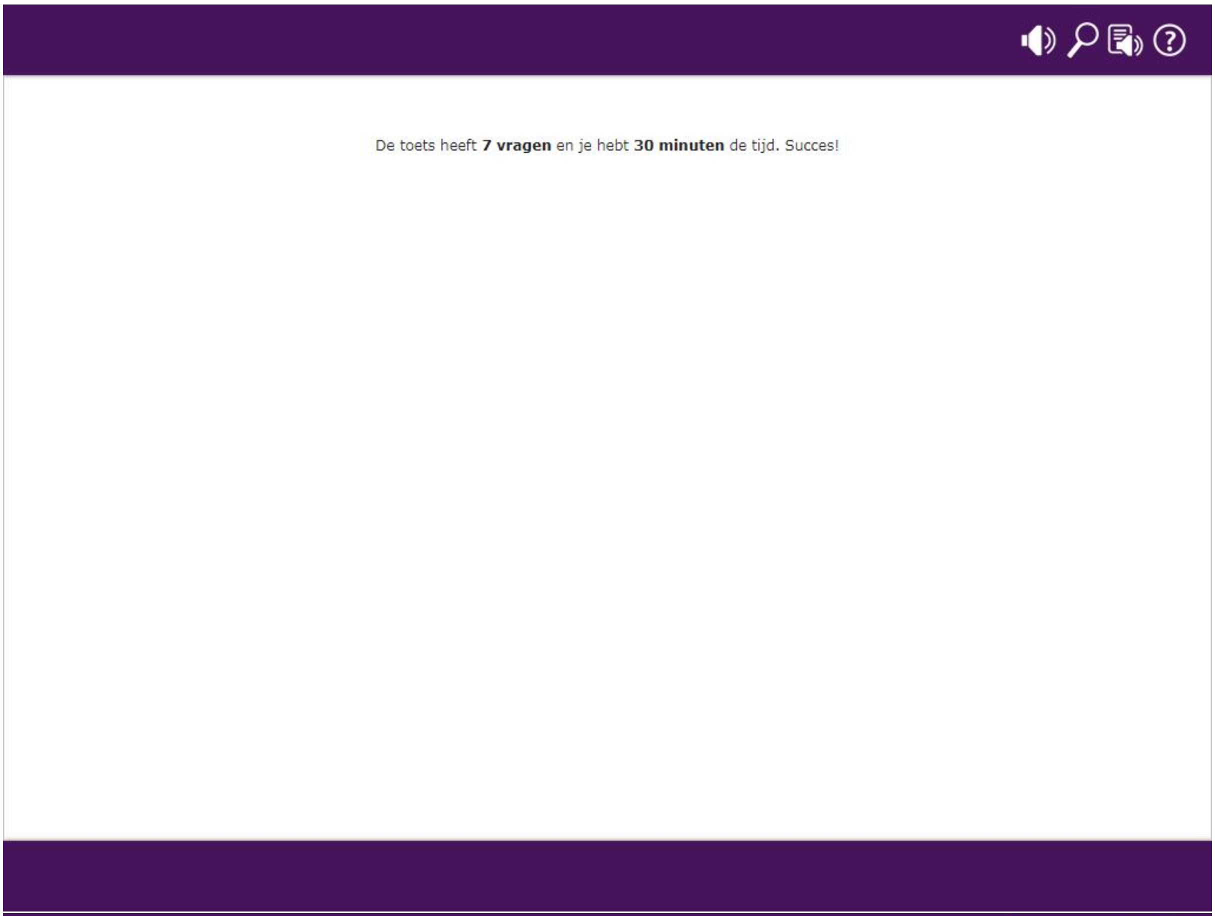
7. Wat is de hoogte van het souvenir? Rond je antwoord af op hele mm.

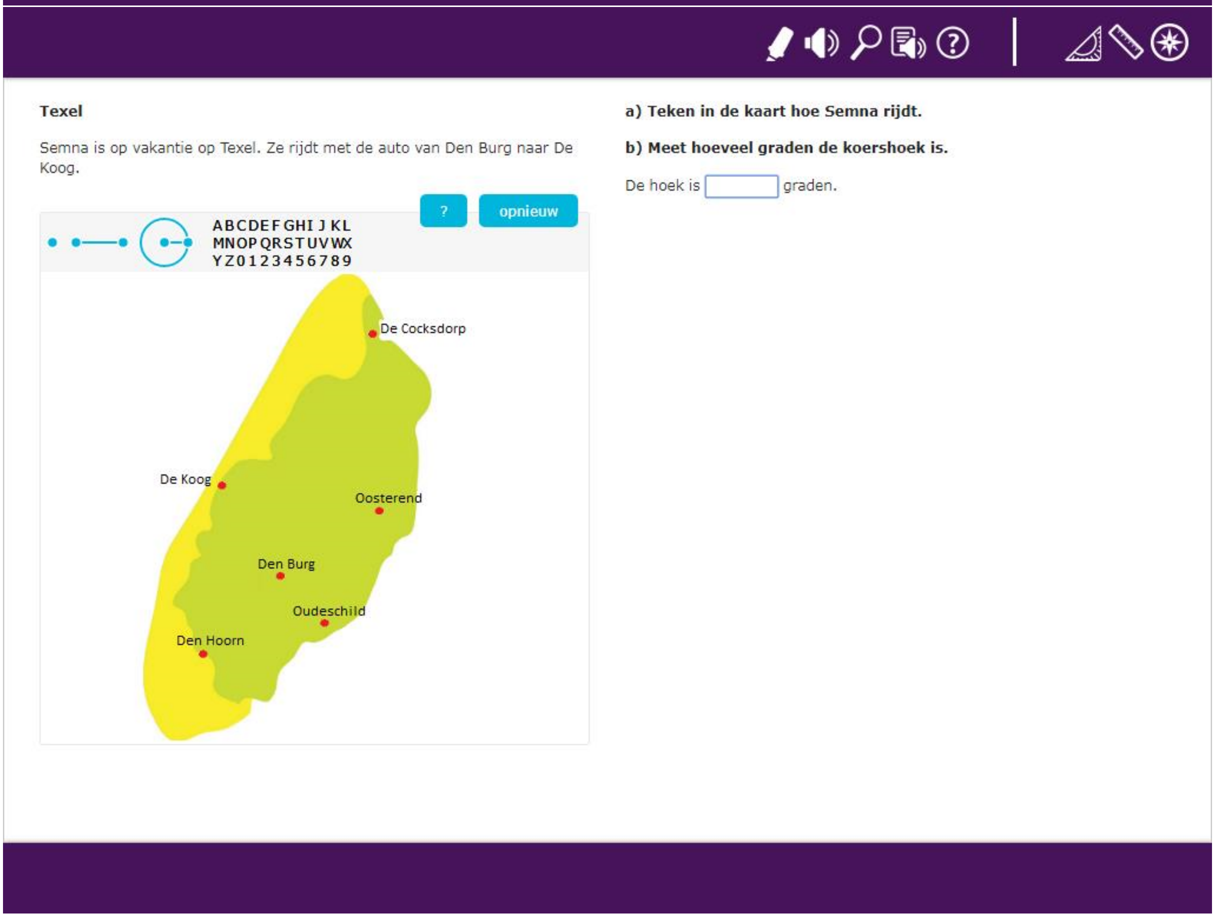
   ➔ De hoogte van het souvenir is ____________________.

EINDE TOETS. CONTROLEER OF JE NAAM OP DE VOORKANT STAAT

Appendix B – Computer-based tests

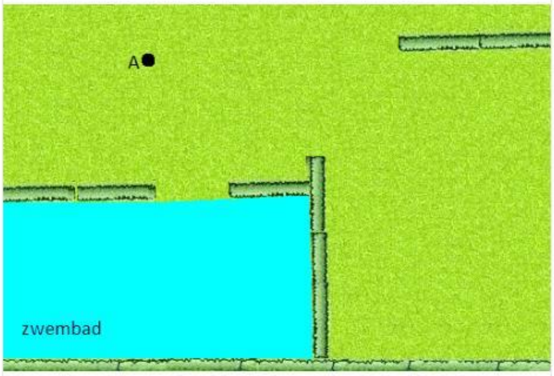
Test A


De toets heeft 7 vragen en je hebt 30 minuten de tijd. Succes!


Texel
Semna is op vakantie op Texel. Ze rijdt met de auto van Den Burg naar De Koog.
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789
De Cocksdorp
De Koog
Oosterend
Den Burg
Oudeschild
Den Hoorn
a) Teken in de kaart hoe Semna rijdt.
b) Meet hoeveel graden de koershoek is.
De hoek is graden.

Achter de heg

Anja staat bij punt A. Zij kan niet over de heg kijken. Marit zwemt in het zwembad.



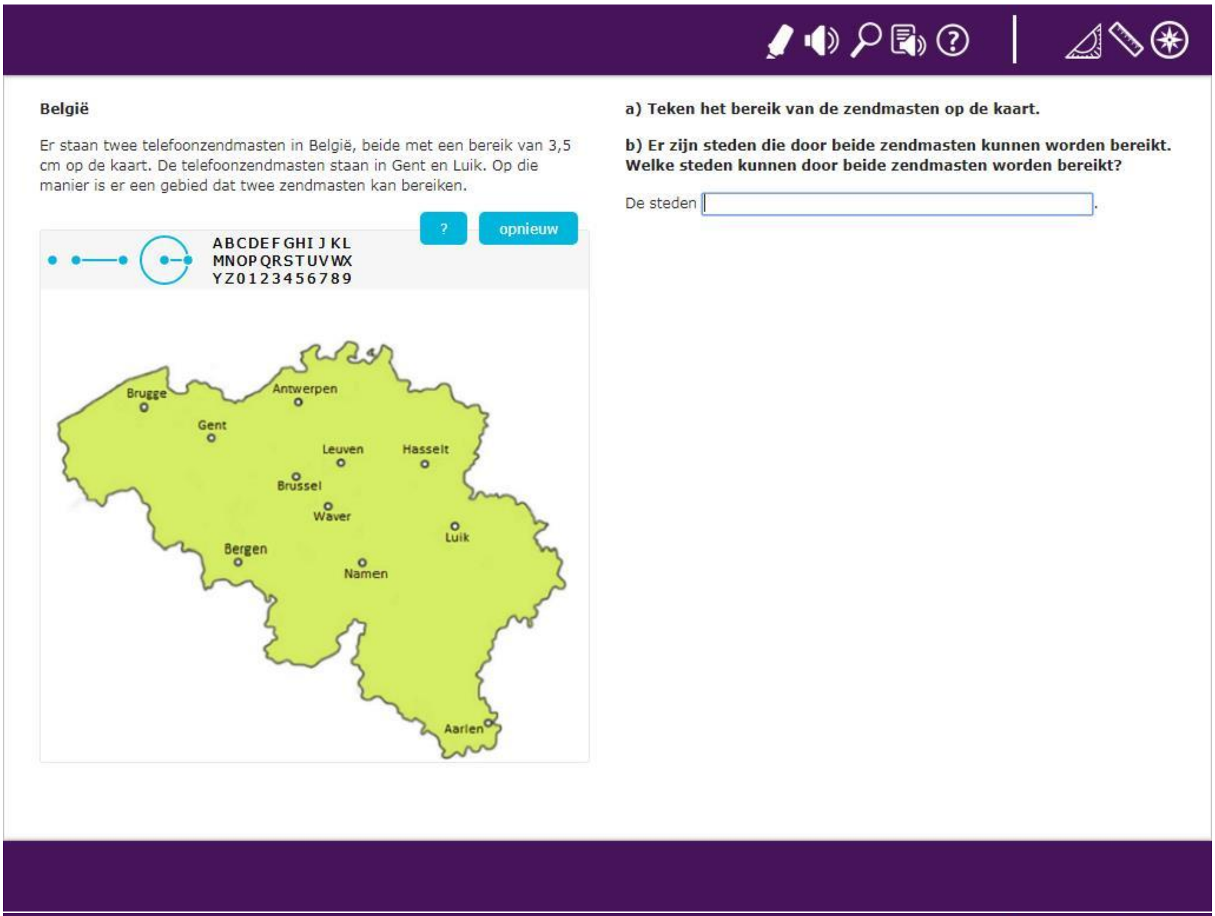
Gebruik kijklijnen om het gebied aan te geven waar Marit moet zwemmen zodat Anja haar ziet. Zet een M in dit gebied.

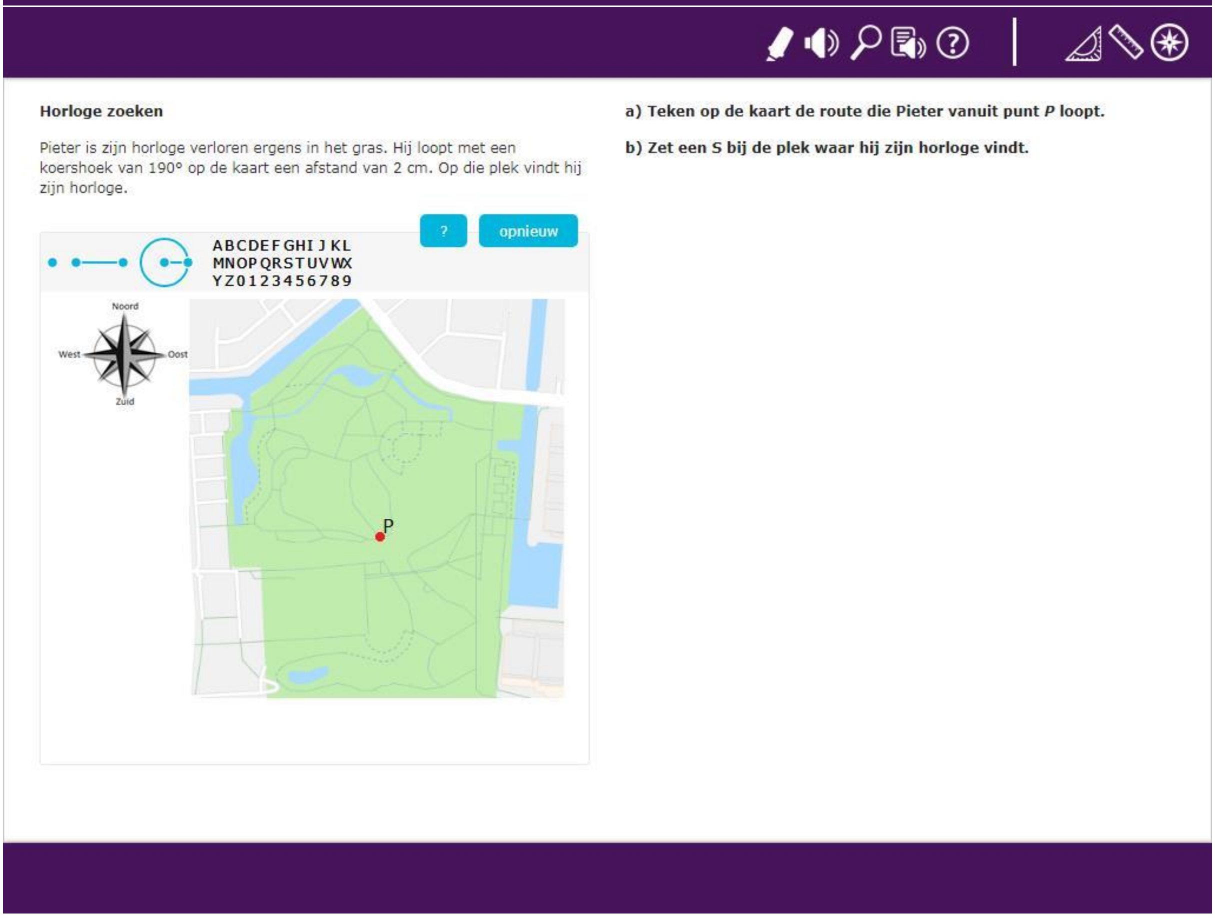
Over de rand

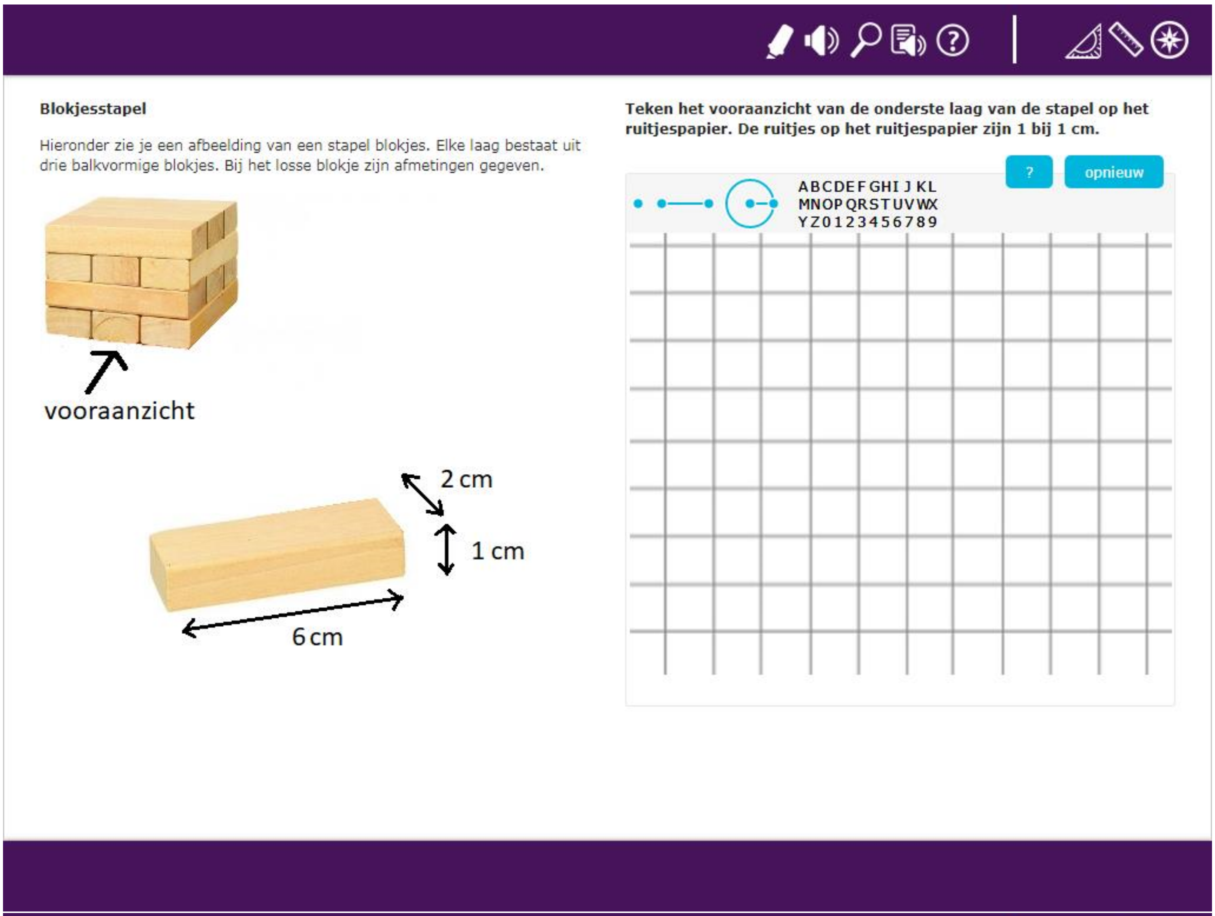
Hieronder zie je een zijaanzicht van een persoon op een schommel.

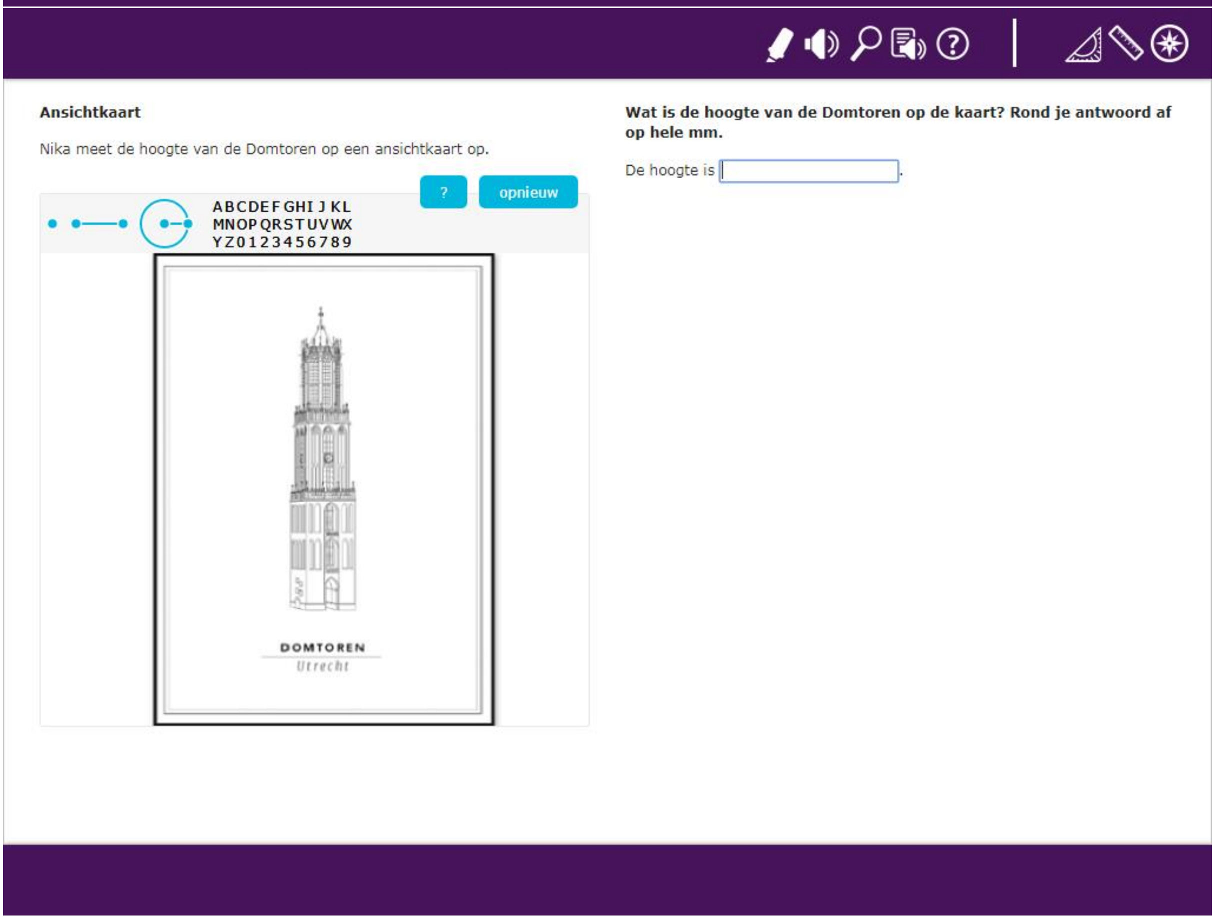
Aan de rechterkant zie je een schets van de schommel. Het is het zijaanzicht van het touw van de schommel in rust. Punt *D* is het draaipunt en punt *Z* het zitvlak. De schommel schommelt vanuit het draaipunt *D* naar links met een maximale hoek van 70°.

Teken het touw in deze maximale positie en geef hierbij het nieuwe punt Z aan. Zorg ervoor dat het touw even lang blijft.

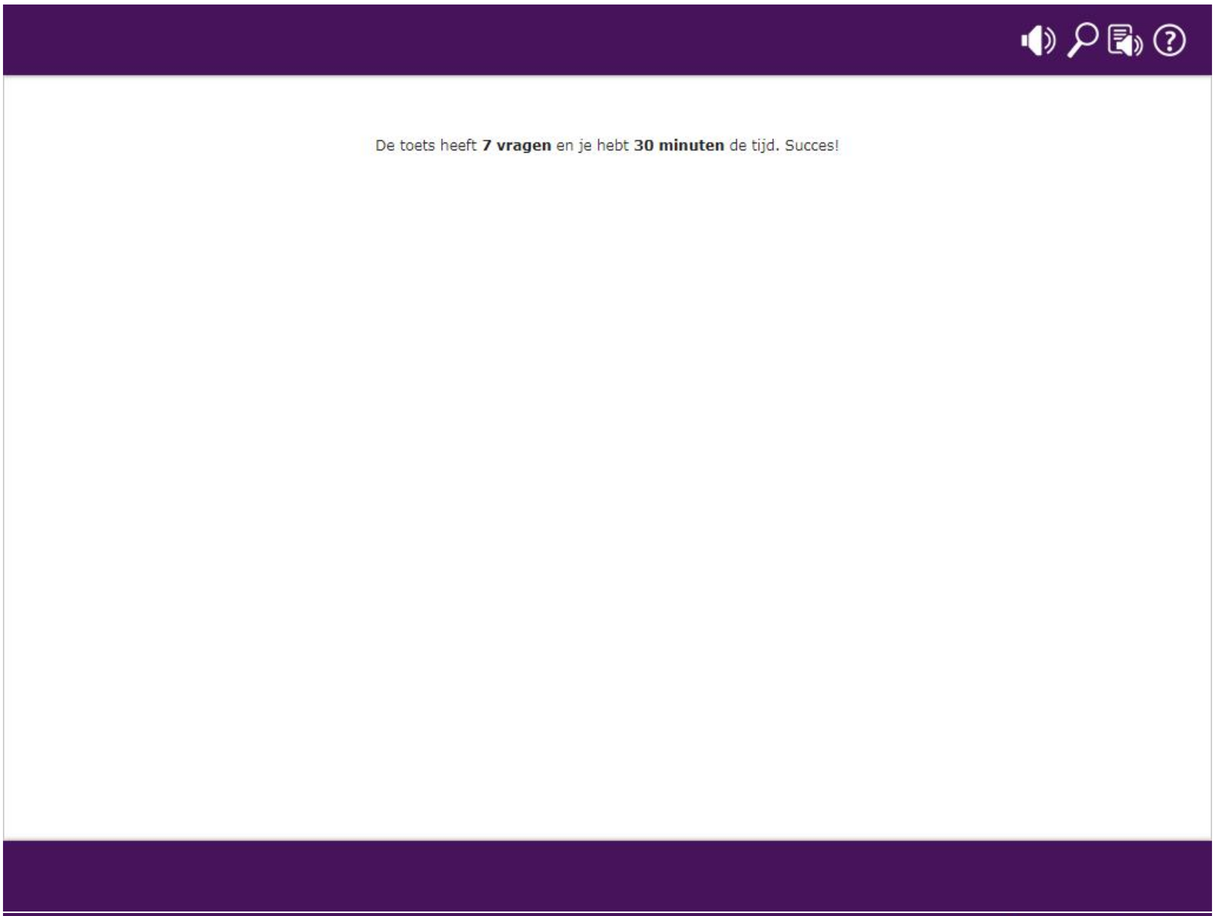

België
Er staan twee telefoonzendmasten in België, beide met een bereik van 3,5 cm op de kaart. De telefoonzendmasten staan in Gent en Luik. Op die manier is er een gebied dat twee zendmasten kan bereiken.
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789
Brugge
Antwerpen
Gent
Leuven
Hasselt
Brussel
Waver
Luik
Bergen
Namen
Aarlen
a) Teken het bereik van de zendmasten op de kaart.
b) Er zijn steden die door beide zendmasten kunnen worden bereikt. Welke steden kunnen door beide zendmasten worden bereikt?
De steden

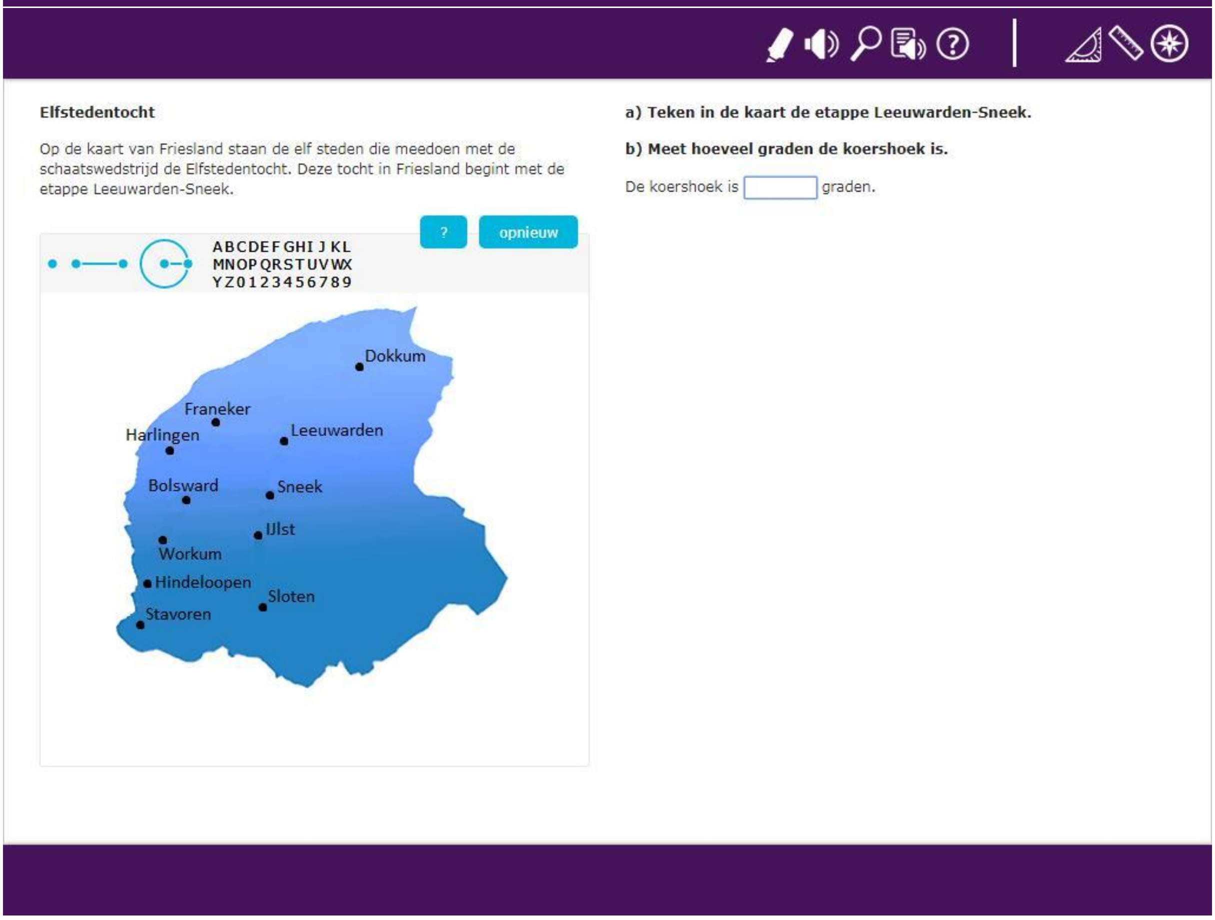

Horloge zoeken
Pieter is zijn horloge verloren ergens in het gras. Hij loopt met een koershoek van 190° op de kaart een afstand van 2 cm. Op die plek vindt hij zijn horloge.
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789
Noord
West
Oost
Zuid
P
a) Teken op de kaart de route die Pieter vanuit punt P loopt.
b) Zet een S bij de plek waar hij zijn horloge vindt.

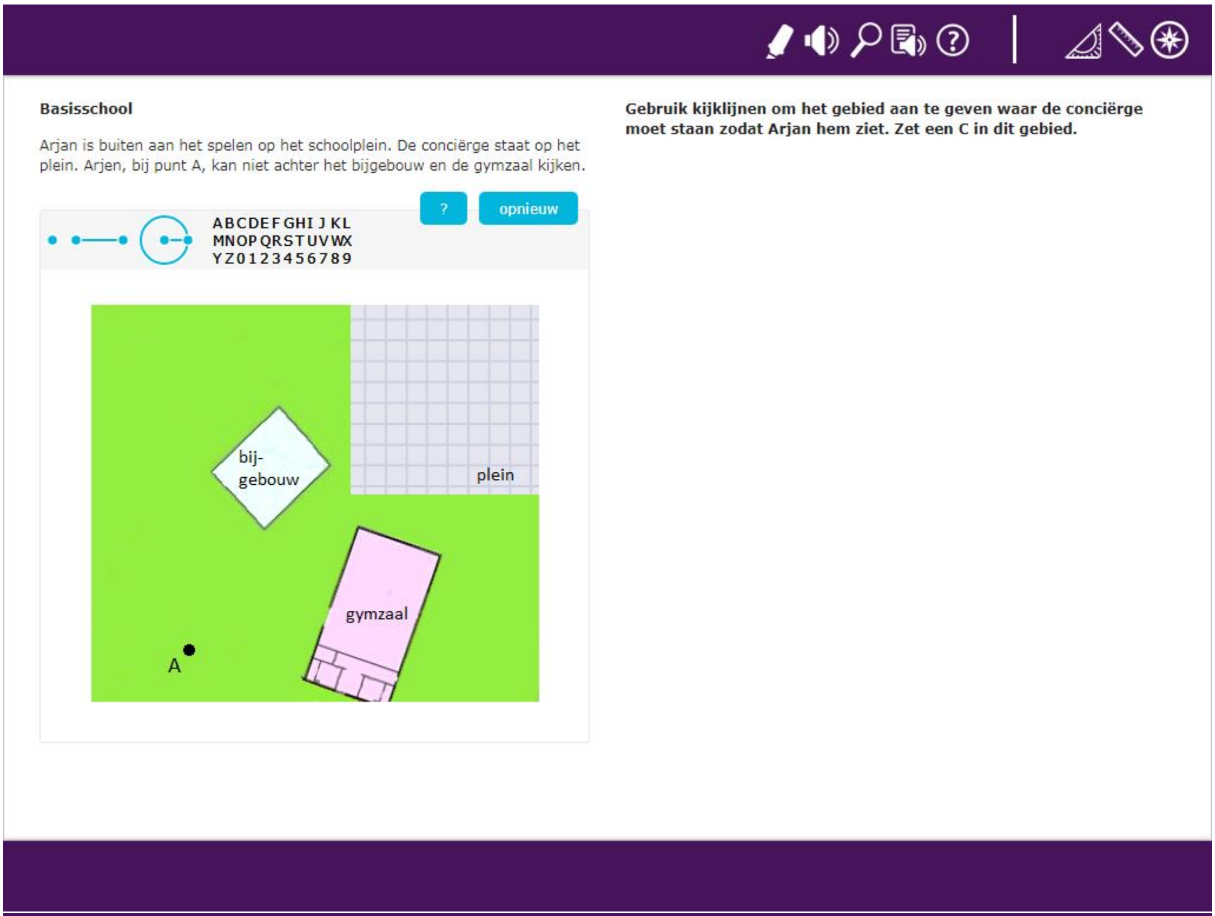

Blokjesstapel
Hieronder zie je een afbeelding van een stapel blokjes. Elke laag bestaat uit drie balkvormige blokjes. Bij het losse blokje zijn afmetingen gegeven.
vooraanzicht
2 cm
1 cm
6 cm
Teken het vooraanzicht van de onderste laag van de stapel op het ruitjespapier. De ruitjes op het ruitjespapier zijn 1 bij 1 cm.
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789

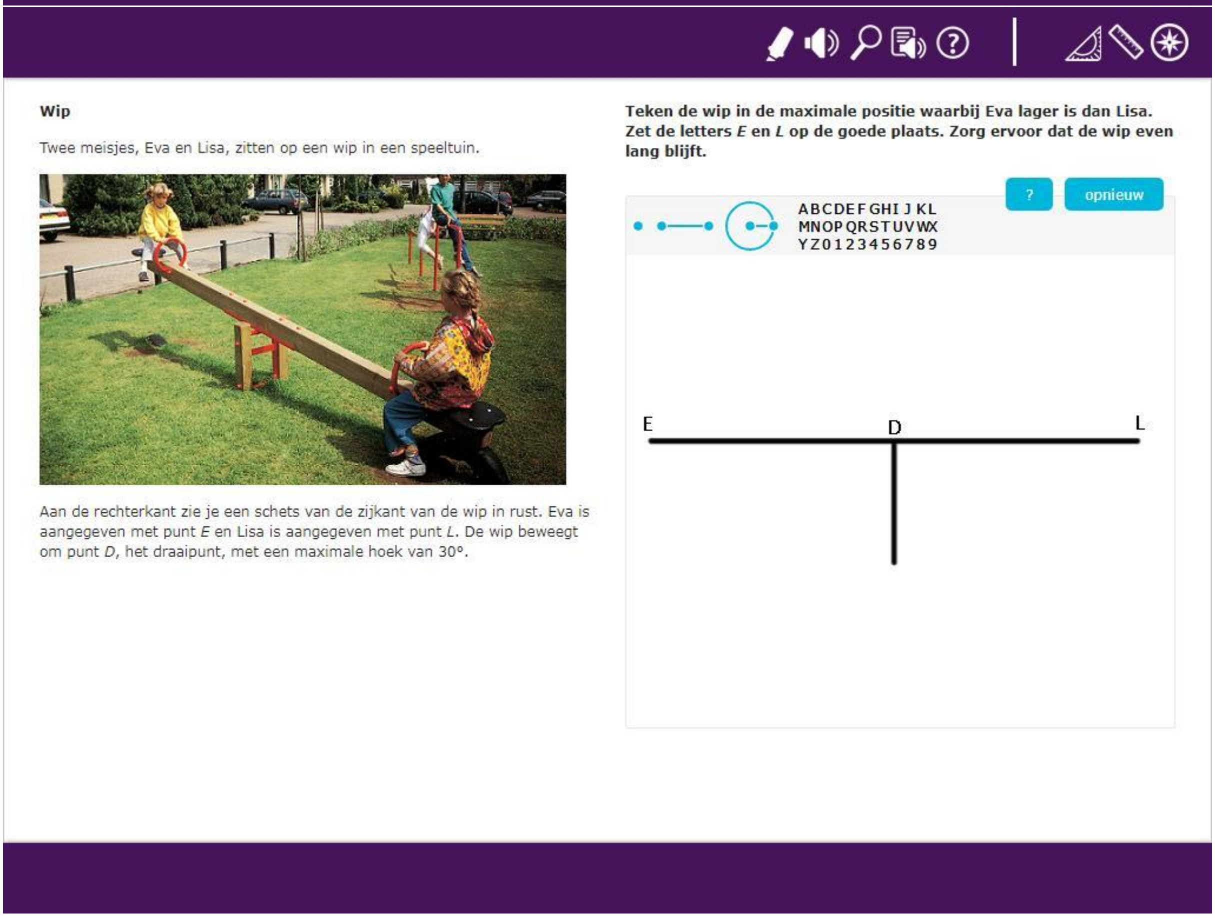

Ansichtkaart
Nika meet de hoogte van de Domtoren op een ansichtkaart op.
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789
DOMTOREN
Utrecht
Wat is de hoogte van de Domtoren op de kaart? Rond je antwoord af op hele mm.
De hoogte is .

Test B


De toets heeft **7 vragen** en je hebt **30 minuten** de tijd. Succes!

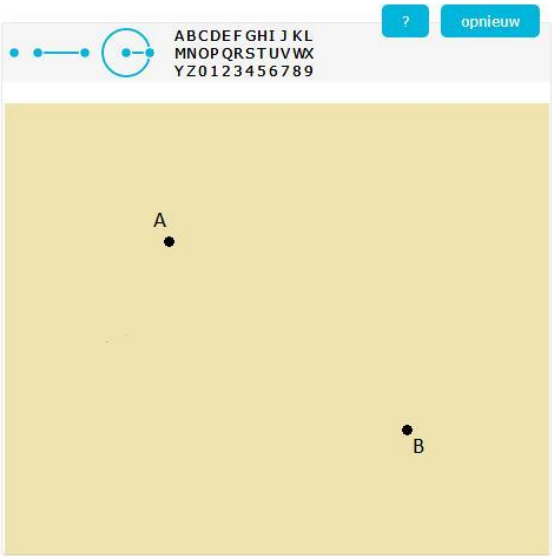

Elfstedentocht
Op de kaart van Friesland staan de elf steden die meedoen met de schaatswedstrijd de Elfstedentocht. Deze tocht in Friesland begint met de etappe Leeuwarden-Sneek.
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789
Dokkum
Franeker
Harlingen
Leeuwarden
Bolsward
Sneek
IJlst
Workum
Hindeloopen
Sloten
Stavoren
a) Teken in de kaart de etappe Leeuwarden-Sneek.
b) Meet hoeveel graden de koershoek is.
De koershoek is graden.


Basisschool
Arjan is buiten aan het spelen op het schoolplein. De conciërge staat op het plein. Arjen, bij punt A, kan niet achter het bijgebouw en de gymzaal kijken.
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789
bij-
gebouw
plein
gymzaal
A
Gebruik kijklijnen om het gebied aan te geven waar de conciërge moet staan zodat Arjan hem ziet. Zet een C in dit gebied.


Wip
Twee meisjes, Eva en Lisa, zitten op een wip in een speeltuin.
Aan de rechterkant zie je een schets van de zijkant van de wip in rust. Eva is aangegeven met punt E en Lisa is aangegeven met punt L. De wip beweegt om punt D, het draaipunt, met een maximale hoek van 30°.
Teken de wip in de maximale positie waarbij Eva lager is dan Lisa. Zet de letters E en L op de goede plaats. Zorg ervoor dat de wip even lang blijft.
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789
E
D
L

Schatzoeken

Manuel wil een schat opgraven. Hij weet dat de schat op de kaart 3 cm van punt A af ligt en 5 cm van punt B. Zo blijven er twee plekken over waar de schat zou kunnen liggen.



Zet een 1 en een 2 bij de plekken waar de schat zou kunnen liggen. Laat zien hoe je aan je antwoord gekomen bent.

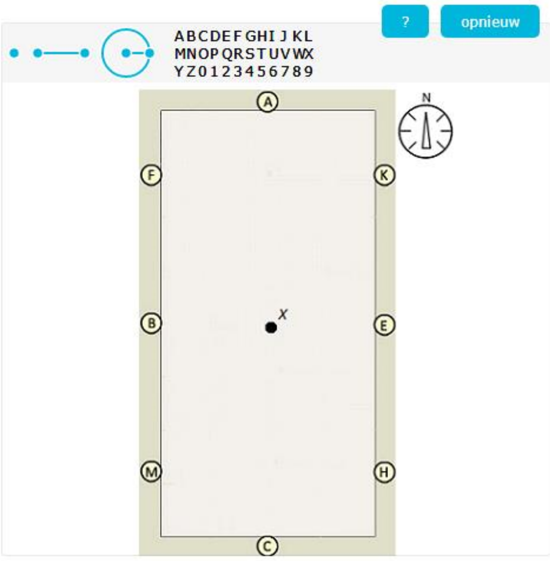
Dressuur

Dressuur is een sport waarbij ruiters op paarden figuren rijden door een bak. Aan de rechterkant zie je een bovenaanzicht van een bak, waarbij letters zijn aangegeven.

Bij een oefening loopt een paard van punt X met een koershoek van 320° een afstand van 2 cm.

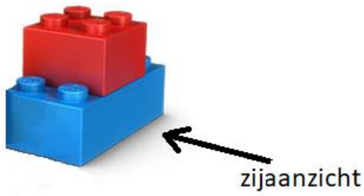
a) Teken op de kaart de route die het paard loopt.

b) Zet een P op de plek waar hij eindigt.



Lego

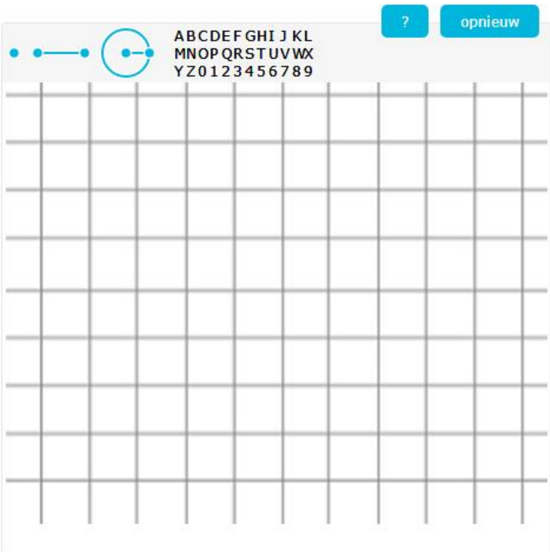
Hieronder zie je twee legoblokjes op elkaar. In de tabel zie je de afmetingen van de blokjes.



	Lengte	Breedte	Hoogte
Bovenste blokje	2 cm	2 cm	1 cm
Onderste blokje	4 cm	2 cm	1 cm

Teken het zijaanzicht van de blokjes op het ruitjespapier met ruitjes van 1 bij 1 cm. Je mag de rondjes op de blokjes verwaarlozen.

?   opnieuw



Eiffeltoren

Marjolein heeft een kleine Eiffeltoren meegenomen als souvenir uit Parijs. Hieronder zie je deze Eiffeltoren op ware grootte.

?   opnieuw

Wat is de hoogte van het souvenir? Rond je antwoord af op hele mm.

De hoogte van het souvenir is [ ].

Appendix C – Answer sheets

Test version A

Toets A nakijkmodel

Algemeenheden:

Als de leerling…	*Gevolg*
met pen tekent	geen aftrek

1) Texel

<u>Begrip van koershoek</u>: antwoord van 327° *binnen een marge van 10°* (1pt)

<u>Netjes en precies meten</u>: het antwoord van 327° *binnen een marge van 2°* (1pt)

Als de leerling…	*Gevolg*
een antwoord heeft gegeven wat tussen de 31° en 35° ligt (linksom gemeten)	0 pt voor begrip en 1 pt voor netjes meten, totaal 1 pt
een redelijke noordpijl heeft getekend en die getekende hoek juist heeft opgemeten, maar die wel buiten de marge valt van netjes en precies meten	geen punten aftrekken

2) Achter de heg

Het tekenen van de linker kijklijn vanuit A (1pt)

Het tekenen van de rechter kijklijn vanuit A (1pt)

De M in het juiste gebied (1pt)

Als de leerling…	*Gevolg*
de M tussen twee kijklijnen in het juiste gebied zet, ook al kloppen de kijklijnen niet helemaal	1 punt rekenen voor de M
de M achter één lijn zet in het midden	in totaal 0 pt
de kijklijnen niet helemaal tot het einde doortrekt	1 pt aftrek
de leerling twee juiste kijklijnen tekent, die TOT de heg gaan (dus het zwembad niet in gaan)	geen pt rekenen voor de kijklijnen

3) Over de rand

Begrip: het tekenen van de juiste hoek van 70° *met een marge van 10°* **(1pt)**

Netjes en precies: het tekenen van een hoek van 70° *met een marge van 2°* **(1pt)**

Het tekenen van een lengte van 4,7 cm incl. bolletjes *met een marge van 2 mm* (1pt)

De letter Z erbij zetten (1pt) *NOOT*: *Deze punt is alleen te verdienen als de 4,7 cm goed is getekend.*

Als de leerling…	*Gevolg*
De schommel naar rechts heeft getekend OF twee schommels heeft getekend (naar links en naar rechts)	geen pt aftrekken

4) België

Begrip: het tekenen van cirkels met de goede middelpunten (1pt)

Netjes getekend: straal van 3 cm *met een marge van 2 mm* (1pt)

Het noemen van Waver en Leuven (1pt)

Als de leerling…	*Gevolg*
Waver en Leuven noemt én uit de tekening is op te maken dat de leerling snapt waarom (bijv. twee lijntjes die kant op)	1 pt toekennen voor Waver en Leuven

5) Horloge

Begrip van de koershoek: het tekenen van de hoek van 190° *met een marge van 10°* (1pt)

Netjes en precies tekenen: het tekenen van de hoek van 190° *met een marge van 2°* (1pt)

Het tekenen van de lengte van 2 cm *met een marge van 2 mm* (1pt)

Een kruis op de juiste plek (1pt). *NOOT*: *Deze punt is alleen te verdienen als de 2 cm goed is getekend.*

6) Blokjestoren

Het tekenen van drie blokjes naast elkaar (1pt)

Het tekenen van 1 cm hoogte per blokje (1pt)

Het tekenen van 2 cm breedte per blokje (1pt)

Als de leerling…	*Gevolg*
alleen de buitenlijnen van de figuur aangeeft	1 pt aftrek
alle cm-hokjes omlijnt	1 pt aftrek
de hele figuur tekent i.p.v. de onderste laag	1 pt aftrek

7) Ansichtkaart

Begrip: het meten van een lengte van 5,6 cm *met een marge van 5 mm* (1 pt)

Netjes meten: het meten van een lengte van 5,6 cm *met een marge van 1 mm* (1pt)

Als de leerling…	*Gevolg*
een getal heeft staan waarin 55, 56 of 57 in te herkennen is, maar het antwoord niet klopt	geen aftrek

Test version B

Toets B nakijkmodel

Algemeenheden:

Als de leerling…	*Gevolg*
met pen tekent	geen aftrek

1) Elfstedentocht

Begrip van koershoek: het antwoord van 195° *met een marge van 10°* **(1pt)**

Netjes en precies meten: het antwoord van 195° *met een marge van 2°* **(1pt)**

Als de leerling…	*Gevolg*
Een antwoord heeft gegeven wat tussen de 163 en 167° ligt (linksom gemeten)	0 pt voor begrip en 1 pt voor netjes meten, totaal 1 pt
Een redelijke noordpijl heeft getekend en die getekende hoe juist heeft opgemeten, maar die wel buiten de marge valt van netjes en precies meten	geen aftrek

2) Basisschool

Het tekenen van de linkerkijklijn vanuit A **(1pt)**

Het tekenen van de rechterkijklijn vanuit A **(1pt)**

De C in het juiste gebied **(1pt)**

Als de leerling…	*Gevolg*
de C tussen twee kijklijnen in het juiste gebied zet, ook al kloppen de kijklijnen niet helemaal	1 punt rekenen voor de C
de C achter één lijn zet in het midden	in totaal 0 pt
de kijklijnen niet helemaal tot het einde doortrekt	1 pt aftrek
twee kijklijnen tekent die TOT het plein gaan	geen pt rekenen voor de kijklijnen

3) Wip

Begrip: het tekenen van de juiste hoek van 30° *met een marge van 10°* **(1pt)**

Netjes en precies: het tekenen van een hoek van 30° *met een marge van 2°* **(1pt)**

Het tekenen van de lengte van 5,7 cm van de wipzijden *met een marge van 2 mm* **(1pt)**

De letters E en L erbij zetten **(1pt).** *Noot: Deze punt is alleen te verdienen als de 5,7 cm goed is getekend.*

Als de leerling…	*Gevolg*
de wip 'verkeerd om' heeft getekend, dus E boven L	geen pt aftrekken
twee wippen heeft getekend	geen pt aftrekken
maar één zijde van de wip heeft getekend	1 pt aftrekken
de wip niet heeft getekend maar alleen de eindpunten	1 pt aftrekken

4) Schatzoeken

Antwoordmogelijkheid 1:

Begrip: het tekenen van cirkels met de goede middelpunten (1pt)

Netjes getekend: stralen van 3 en 4 cm met een marge van 2 mm (1pt)

Het op de juiste plek tekenen van de kruizen (1pt) *NOOT. Deze alleen toekennen als de kruizen op de goede plekken staan.*

Antwoordmogelijkheid 2:

Begrip: het vinden van een plek d.m.v. onderzoek (1pt)

Netjes getekend: afstanden van 3 en 4 cm met een marge van 2 mm (1pt)

Twee plekken hebben gevonden (1pt)

Als de leerling…	*Gevolg*
Geen kruis/kruizen heeft getekend	1 pt aftrek

5) Dressuur

Begrip van de koershoek: het tekenen van de hoek van 320° *met een marge van 10°* **(1pt)**

Netjes en precies tekenen: het tekenen van de hoek van 320° *met een marge van 2°* **(1pt)**

Het tekenen van de lengte van 2 cm met een marge van 2 mm **(1pt)**

Een kruis op de juiste plek (1pt). *NOOT: Deze punt is alleen te verdienen als de 2 cm goed is getekend.*

Als de leerling…	*Gevolg*
de lijn niet heeft getekend maar alleen het eindpunt	1 pt aftrek

6) Lego

Het tekenen twee blokjes boven elkaar **(1pt)**

Het tekenen van de juiste hoogte van de blokjes **(1pt)**

Het tekenen van de juiste breedtes van de blokjes **(1pt)**

Als de leerling…	*Gevolg*
alleen de buitenlijnen van de figuur aangeeft	1 pt aftrek
alle cm-hokjes omlijnt	1 pt aftrek

7) Eiffeltoren

Begrip: het meten van een lengte van 7,3 cm *met een marge van 5 mm* (1 pt)

Netjes meten: het meten van een lengte van 7,3 cm *met een marge van 1 mm* (1pt)

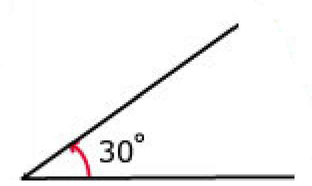
Als de leerling…	*Gevolg*
Een getal heeft staan waarin 72, 73 of 74 in te herkennen is, maar het antwoord niet klopt	Geen aftrek

Appendix D – Classical instruction

Paper-based classical instruction

Checklist instructie

Hoek opmeten geo.



Teken een hoek van ongeveer 30 graden. Hoe groot is deze hoek.

Hoek opmeten windroos.

Gebruik hiervoor dezelfde hoek van ongeveer 39 graden die nog op het bord staat.

Koershoek opmeten windroos.

Teken een koershoek van ongeveer 300 graden (*10 uur op de klok*!) Meet deze op.

Hoek tekenen geo.

Teken een hoek van 40 graden.

Hoek tekenen windroos.

Teken een hoek van 260 graden.

Computer-based classical instruction

Checklist instructie

Geodriehoek

- Verschijnen, weghalen, draaien, groter en kleiner maken

Windroos.

- Verschijnen, weghalen, draaien, groter en kleiner maken

Meetlat.

- Verschijnen, weghalen, draaien, groter en kleiner maken
- Let op: afstanden meten begint bij de 0! En niet bij het rode vierkantje

Punten.

- Invoegen, verslepen

Lijn.

- Invoegen, verslepen, draaien, groter en kleiner maken

Cirkel.

- Invoegen, verslepen

Hoek opmeten geo.

Teken een hoek van ongeveer 30 graden. Hoe groot is deze hoek.

Hoek opmeten windroos.

Gebruik hiervoor dezelfde als die al was getekend.

Koershoek opmeten windroos.

Teken een koershoek van ongeveer 300 graden (*10 uur op de klok*!) Meet deze op.

Hoek tekenen geo.

Teken een hoek van 40 graden.

Hoek tekenen windroos.

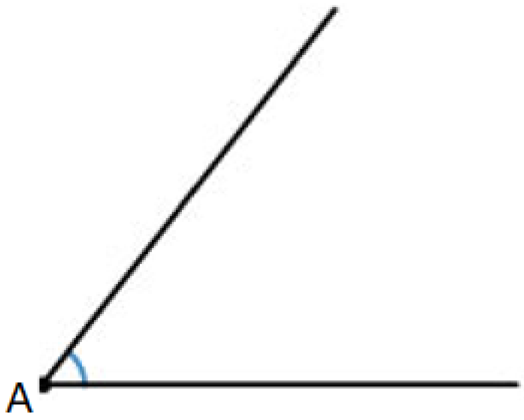
Teken een hoek van 260 graden.

Appendix E – Practice exercises

Paper-based practice exercises

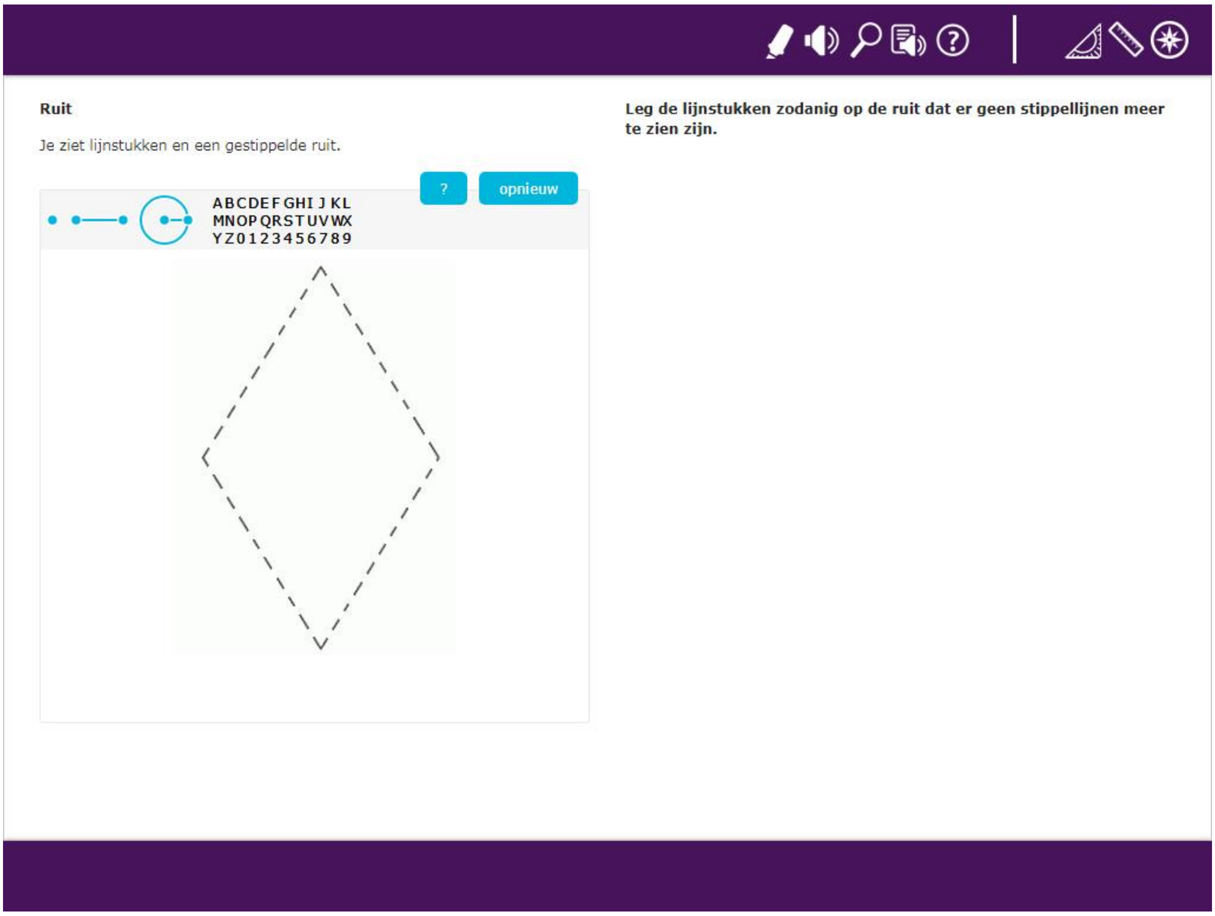
Oefenopgaven

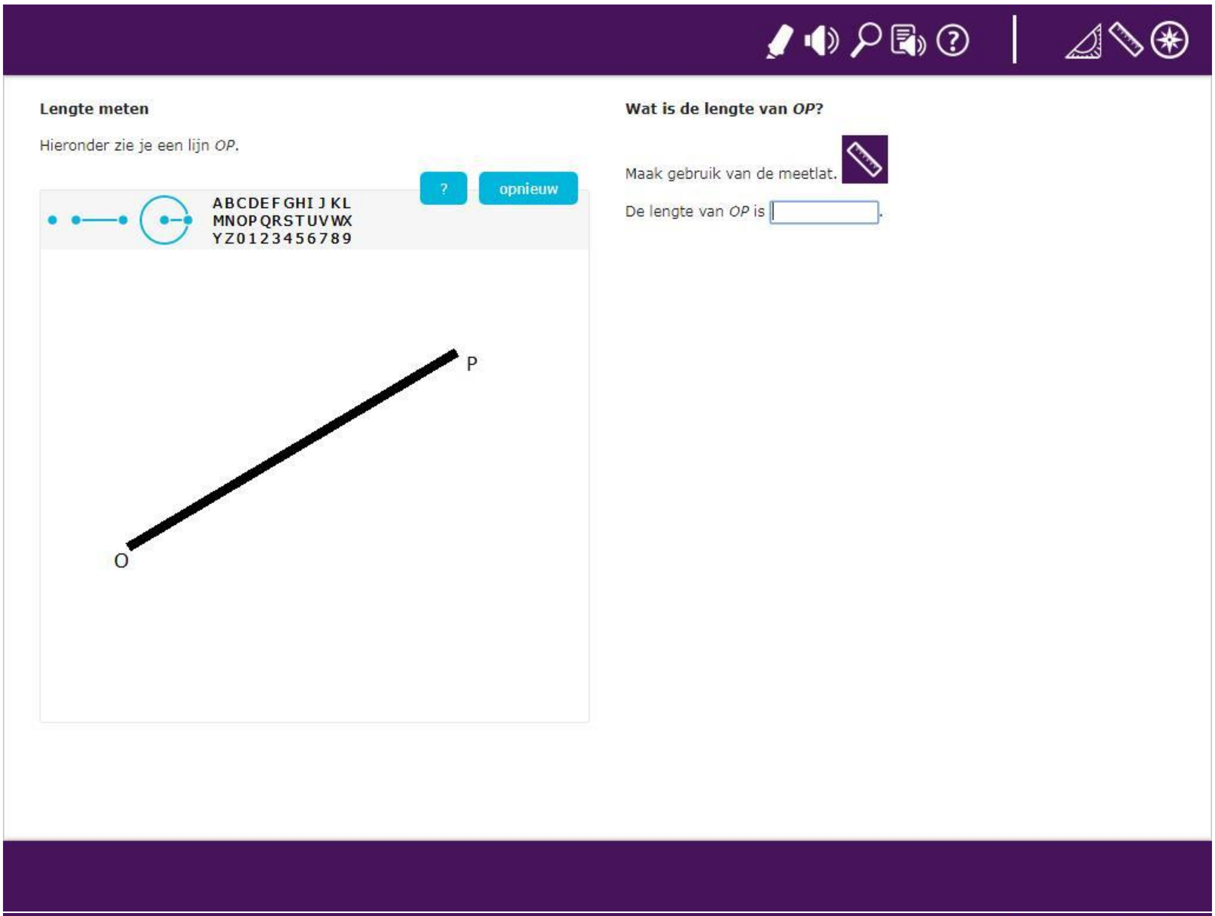
1. Meet hoeveel graden hoek A is.
   - ➔ Het aantal graden is _________

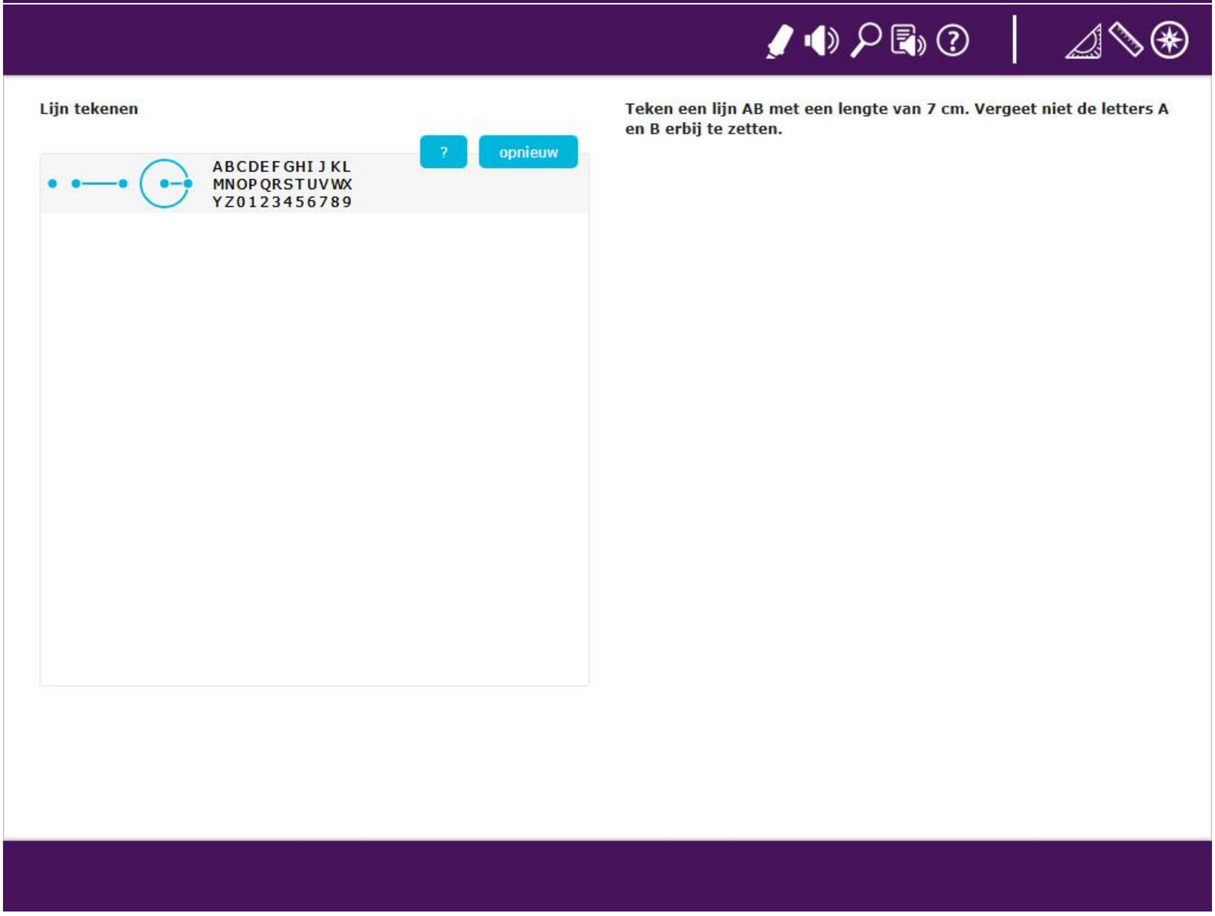


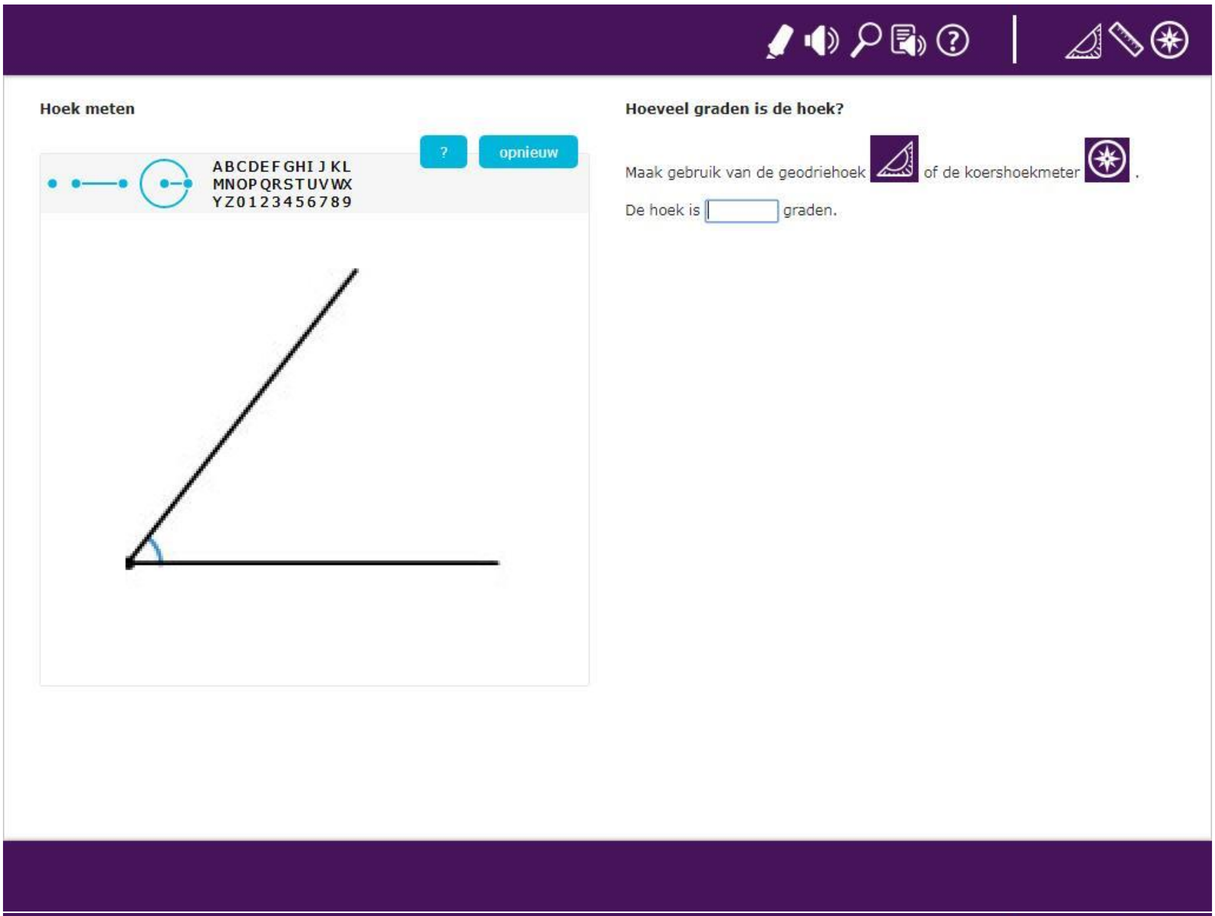
2. Ronald loopt met een koershoek van 200 graden vanuit punt *R*. Teken hieronder de richting die Ronald loopt.

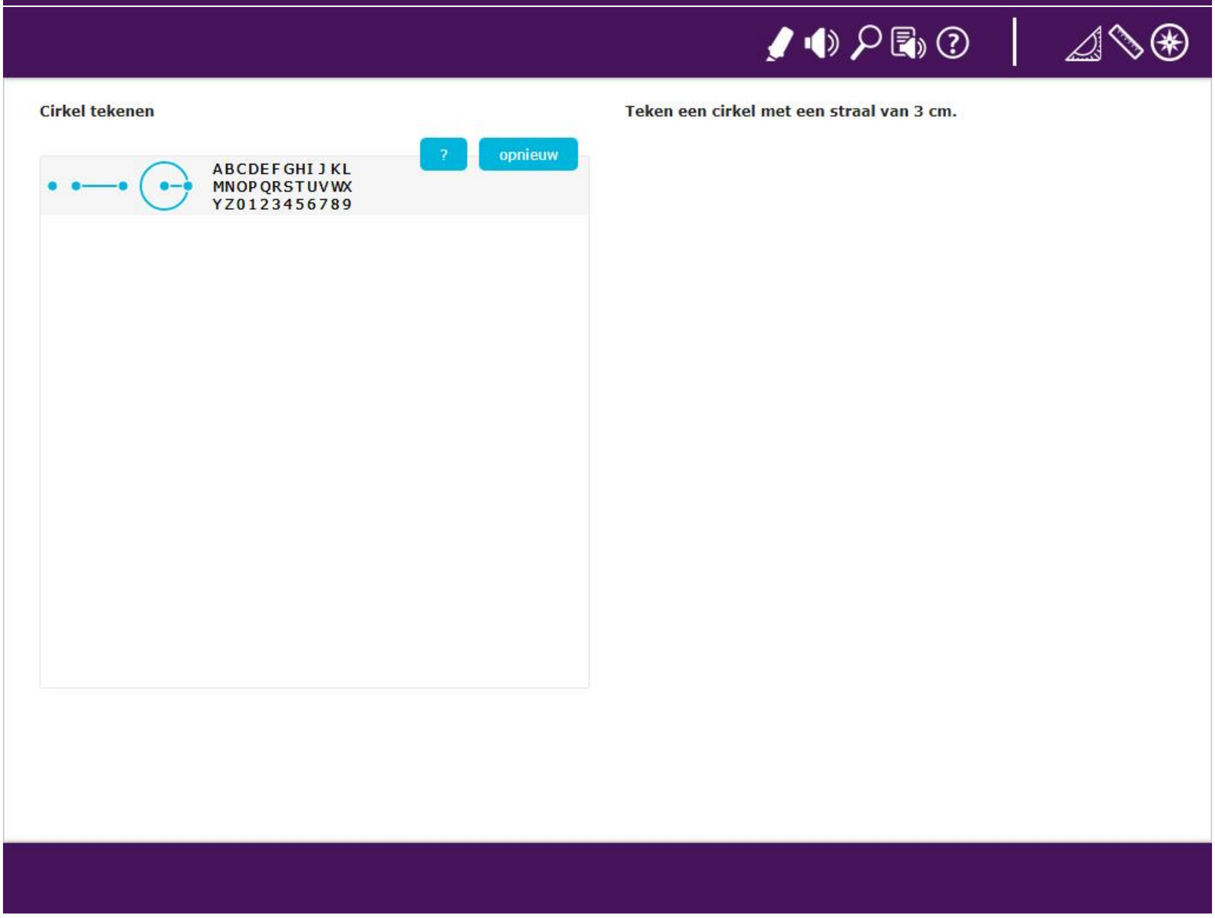
Computer-based practice exercises

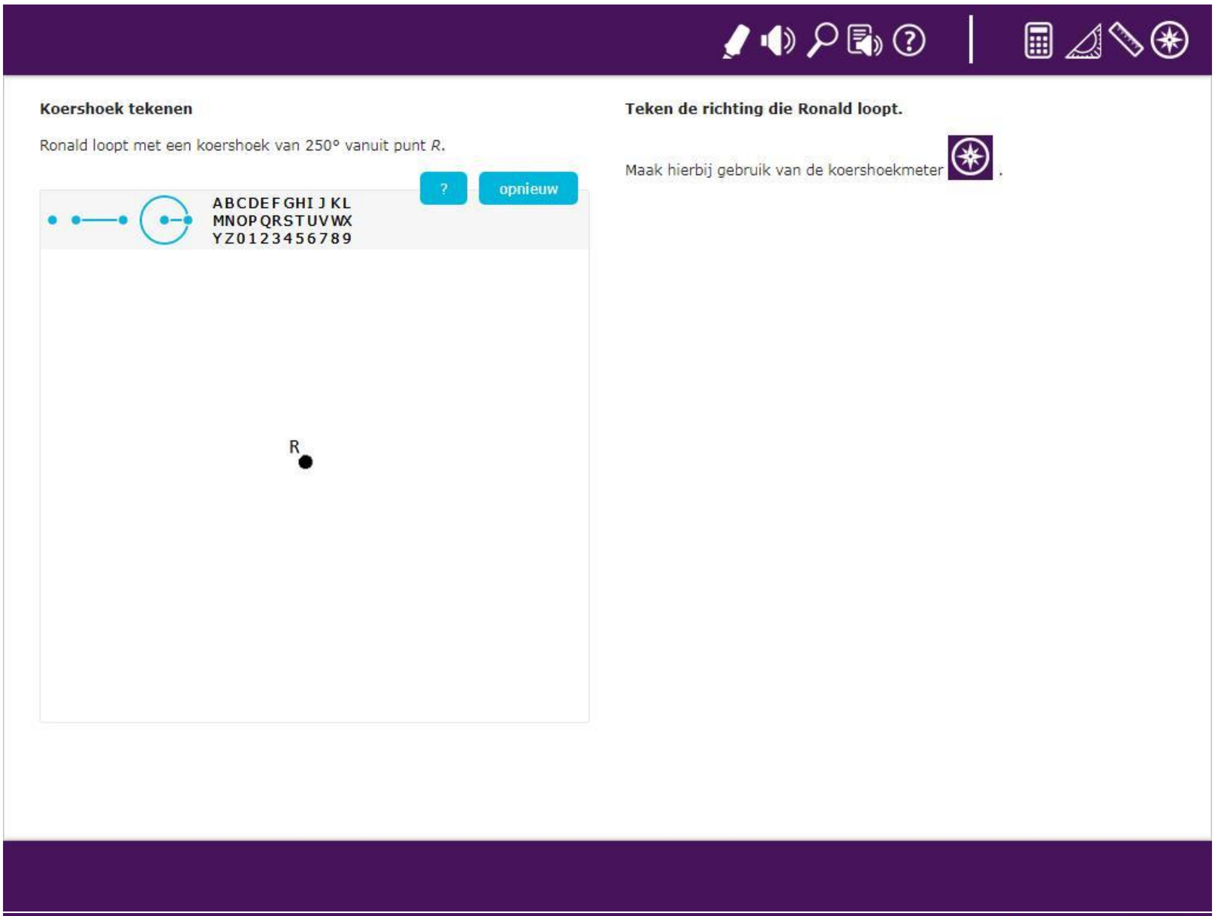


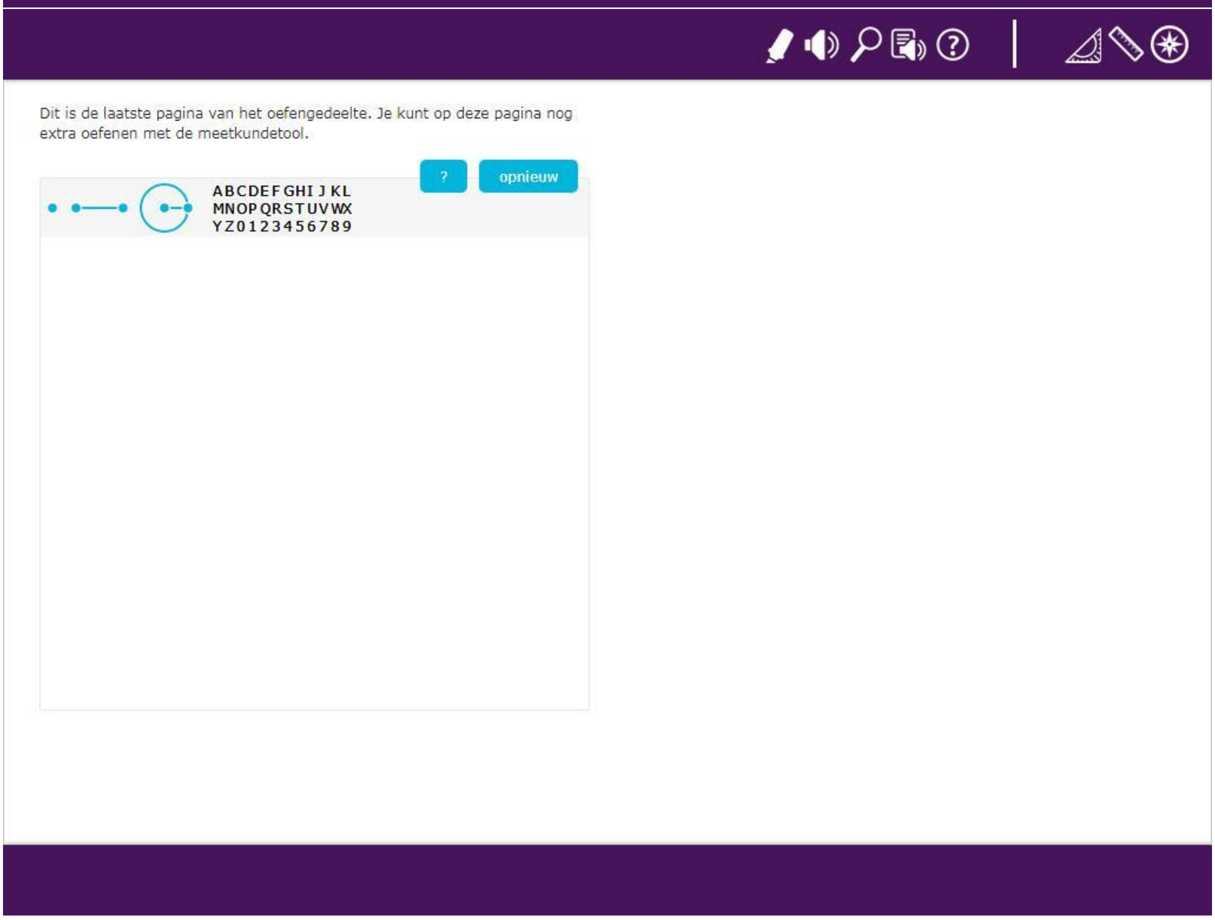

Lengte meten
Hieronder zie je een lijn OP.
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789
P
O
Wat is de lengte van OP?
Maak gebruik van de meetlat.
De lengte van OP is


Lijn tekenen
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789
Teken een lijn AB met een lengte van 7 cm. Vergeet niet de letters A en B erbij te zetten.


Hoek meten
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789
Hoeveel graden is de hoek?
Maak gebruik van de geodriehoek of de koershoekmeter .
De hoek is graden.


Cirkel tekenen
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789
Teken een cirkel met een straal van 3 cm.


Koershoek tekenen
Ronald loopt met een koershoek van 250° vanuit punt R.
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789
R
Teken de richting die Ronald loopt.
Maak hierbij gebruik van de koershoekmeter .


Dit is de laatste pagina van het oefengedeelte. Je kunt op deze pagina nog extra oefenen met de meetkundetool.
?
opnieuw
ABCDEFGHIJKL
MNOPQRSTUVWX
YZ0123456789