

## **National Tests in Denmark – CAT as a Pedagogic Tool**

Jakob Wandall

Danish National School Agency

Requests for reprints should be addressed to:

Jakob Wandall, Consultant

Wandall.Biz

Frederik VI's alle 9, DK-2000 F, Denmark

E-Mail [Jakob@Wandall.Biz](mailto:Jakob@Wandall.Biz)

## **ABSTRACT**

Testing and test results can be used in different ways. They can be used for regulation and control, but they can also be a pedagogic tool for assessment of student proficiency in order to target teaching, improve learning and facilitate local pedagogical leadership. To serve these purposes the test has to be used for low stakes purposes, and to ensure this, the Danish National test results are made strictly confidential by law. The only test results that are made public are the overall national results. Because of the test design, test results are directly comparable, offering potential for monitoring added value and developing new ways of using test results in a pedagogical context. This article gives the background and status for the development of the Danish national tests, describes what is special about these tests (e.g., Information Technology [IT]-based, 3 tests in 1, adaptive), how the national test are carried out, and what is tested. Furthermore, it describes strategies for disseminating the results to the pupil, parents, teacher, headmaster and municipality; and how the results can be used by the teacher and headmaster.

## National Tests in Denmark – CAT as a Pedagogic Tool

### Background and status for the development of the Danish national tests

In 2006 the Danish Parliament decided to make national tests a compulsory pedagogic tool in the *Folkeskole*<sup>1</sup>. The tests are part of the implementation of the recommendations from a review conducted in Denmark by a team from the Organisation for Economic Co-operation and Development (OECD). The tests are intended to support improvement of the evaluation culture in Denmark.

The tests are designed by the Danish National School Agency and developed under a tender contract with a consortium (COWI A/S in cooperation with a group of private and public companies, as well as educational and research institutions).

The development of the IT-based test system began in July 2006, and the first three tests (of 12) were launched in May 2007. Although an expert panel concluded that the basic concepts of the technical components (e.g. IT solution, psychometric method) were carried out successfully, problems were detected with the content (i.e., the test items). The quality control procedures in the contract had been concerned with the statistical and psychometric control of the tests, however, the expert review pointed out that the items did not measure what they were supposed to (e.g. the reading items mostly tested knowledge of history and cultural understanding). The item writing part, including training and supervision of item writers, of the contract had been seriously underestimated. As a result redevelopment of the item pool was necessary. Following redevelopment and pilot testing that occurred in spring 2009, the test system was officially launched in spring 2010.

### What characterizes these tests?

The national tests differ from the tests that are already used in the Danish schools mainly in the following ways:

- The tests are IT-based and the pupils answer the questions online via the internet.
- Test results (e.g., scores, reports) are automatically calculated and generated. The teachers do not have to correct the tests, and the analysis of test results (or some parts of the work) has been done when the teacher can access the test results the next day.
- The tests are supplied to the schools free of charge.
- The tests are adaptive. Each test contains three separate adaptive test sessions which deal with different dimensions of the subject (so-called “profile areas”, described below).

Adaptive tests are designed to adjust to the pupils’ level of proficiency during the test. In this design, the first item presented to the pupil has an average difficulty (compared to the form the test is designed for). If the answer is correct, the next item presented to the pupil will be empirically more difficult. If the answer is wrong, the next item will be empirically easier. In this way the test will adapt to the pupils’ level, so that the sequence of items will be different for each pupil. This is a very simplified description of the principle that is employed during the entire test.

For a more detailed description of an adaptive testing system, a computer adaptive test can be characterized by 5 basic components:

---

<sup>1</sup> Folkeskolen is the Danish term for the public Primary and lower Secondary School, more info: [http://eng.uvm.dk/~media/Files/English/Fact%20sheets/080101\\_fact\\_sheet\\_the\\_folkeskole.aspx](http://eng.uvm.dk/~media/Files/English/Fact%20sheets/080101_fact_sheet_the_folkeskole.aspx)

1. Calibrated item pool: Which IRT-model is used, what are the characteristics of the item pool peaked, and how is the content represented?
  2. Starting point: Is it fixed, random, determined by previous knowledge?
  3. Item selection algorithm: How are the items selected and which information is used?
  4. Theta estimation: Maximum likelihood, Bayes estimation, or another scoring model?
  5. Termination criteria: Is there a certain level of SEM, minimum info  $\theta$ , or fixed length?
- (see appendix 1: The five components of the Danish CAT)

### **Item difficulty and pupil ability**

In order to match the item difficulty with the pupil's ability, difficulty and ability have to be measured on the same scale (theta or logit scale:  $-\infty \rightarrow \infty$ , centered in 0). Several item response models were considered, but the Rasch (1960) model was chosen for several reasons: It is a simple model to interpret, explain and to handle and it was judged to best fit the purpose of this program. But there is a limitation in choosing this model: the Rasch model is very inflexible and it usually leads to scrapping a large proportion of the items (e.g., those that discriminate too well) that could be included with a 2- or 3-parameter model.

The item difficulty of the dichotomous items is defined as the ability of the pupil which has exactly 50% probability to give the correct answer on the item. The difficulty is defined a little different for the polytomous/partial credit items – it equals the ability for the pupil that has exactly 50% probability to give the correct answer on exactly 50% of the questions within the item.

In a well-designed linear test where the series of items is predetermined most pupils will experience that some items are too easy, others too difficult and some items that fit the difficulty of the individual pupil's ability. From an analytical point of view the test items that are too easy or too difficult, reveal very little about the pupil's ability.

Only the items where the level of difficulty fits the pupil's ability contribute substantially to the estimation of the pupil's ability. And in a well-designed adaptive test the pupils will mostly be presented for items that have a suitable level of difficulty for his/her level of ability.

### **The Danish public school is a comprehensive school**

The Danish public school system is built on a principle where pupils are not to be divided according to ability or social background. These divisions are prohibited by law. The Danish concept of the comprehensive school is based on the idea of equality for all in a given society. This does not mean that everybody should always be treated equally, but that differential treatment must be justified by a principle of equality, giving each pupil the opportunity to achieve as much as possible in the educational system, regardless of social, economic, religious, racial or ethnic background. For example, there are schools and classes especially for pupils with special needs, but a lot of work is done to integrate these pupils with special needs in the "normal" classes. But within the public school system which covers approximately 90% of the pupils, there are no elite schools or classes.

A consequence of this philosophy is that there is a considerable distance between top and bottom of the academic level in every class in practically every school. This constitutes not only a challenge for the teacher but also for the test systems that should measure the pupil's proficiency as effectively and efficiently as possible.

The academic distance from bottom to the top increases the higher the form (“grade level” in the U.S. context). But it also varies from subject to subject. In Denmark, math is considered to be the subject where the correlation between the school’s efforts and the pupil’s proficiency is the clearest.

Items were pilot tested on 3-5 different forms. For example, the items for the test in math targeted 3<sup>rd</sup> form were tested on pupils from 1<sup>st</sup> to 5<sup>th</sup> form. This gave us a possibility to check the differences in ability in the five forms.

The results were surprising: The top 10% pupils in the 1<sup>st</sup> form performed significantly better than the average pupil in the 3<sup>rd</sup> form. These pupils in turn performed significantly better than the lowest performing 10% in the 5<sup>th</sup> form.

This observation is illustrated in Figure 1 below.

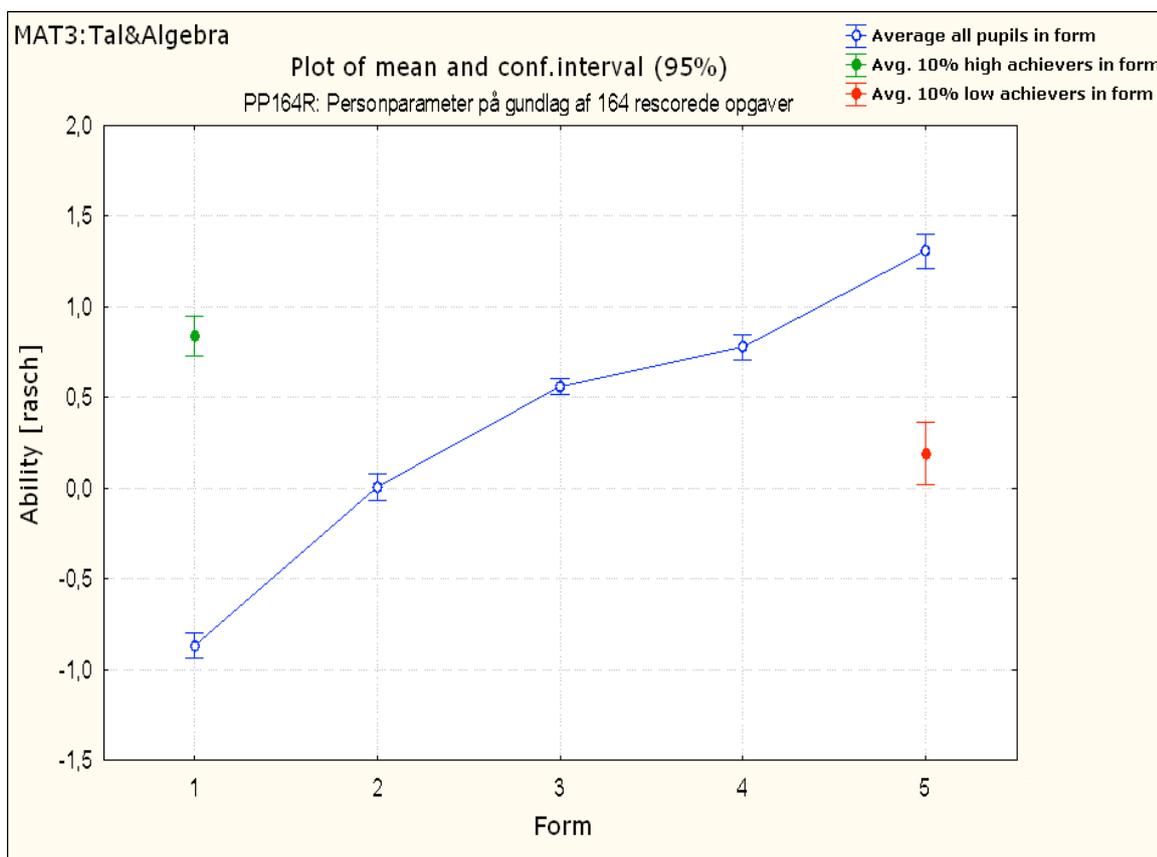


Figure 1. Ability estimates of students in forms 1 through 5 showing the ability of the top 10% of students in the 1<sup>st</sup> form and the bottom 10% of students in the 5<sup>th</sup> form relative to the average ability of students in the 3<sup>rd</sup> form.

This figure indicates that there is a distance on more than 5 forms (more than 2 logits) from the top to the bottom in the subject mathematics (algebra) in an ordinary 3<sup>rd</sup> form. This is evidence of the wide range of examinee ability, even at an early age. There are – in other words – very good reasons for choosing the adaptive approach in order to target the selection of items to the individual pupils.

### How are the tests carried out?

The test system is connected with the Danish website *evaluating.uvm.dk*. It is through this site both the teacher and the pupils access the test. The test system has a maximum capacity of 6,000 users (pupils) at the same time.

Operationally, the teacher logs on and opens the access to the testing system for his or her pupils. Every teacher, headmaster and pupil in Denmark has a unique user-id/password.<sup>2</sup> Like the test system, the identification system is provided by the Ministry of Education. When the pupils are allowed to start the test, they log in. They then have 45 minutes to answer as many items as possible. During this time, the pupils will typically answer 50-80 items. If a pupil needs more time, it is possible for the teacher to prolong the test for the individual pupil.

For this purpose, the teacher has a monitor screen that shows the status of the acceptance level of the Standard Error of Measurement (SEM) of theta – the estimated ability of test result continuously during the test. It is done with a color indication: Red means that the pupil has answered less than 5 questions, and therefore, there is no basis for estimation of the pupil ability. When the pupil has answered 5 questions, the indicator turns yellow, and the student ability is estimated for the first time. Hereafter, the pupil ability is re-estimated for every item answered.

A central computer registers which items have been answered correct, and which items have been answered wrong. For each item the pupil has answered, the student ability will automatically be estimated with maximum likelihood estimation (MLE). According to this estimate, the central computer will choose the next item for the pupil, so that the item difficulty matches the last estimate of the pupil ability as precisely as possible. Then again the same procedure is repeated. When the SEM reaches 0.3 the indicator on the teacher's monitor screen turns green. The principle is illustrated in Figure 2 below.

---

<sup>2</sup> In Denmark, all citizens and other inhabitants are registered in a central database with a unique 10-digit personal number (Central Person Registration number – CPR number), which is assigned at birth or immigration in Denmark. Based on the CPR number is generated an individual login (called "UNI-Login") for all pupils, school staff and municipal officials. All contact with the test system requires that you have logged in and all communications are recorded in a central database. UNI-login constitutes the basis for the identification and management of all user rights on the system.

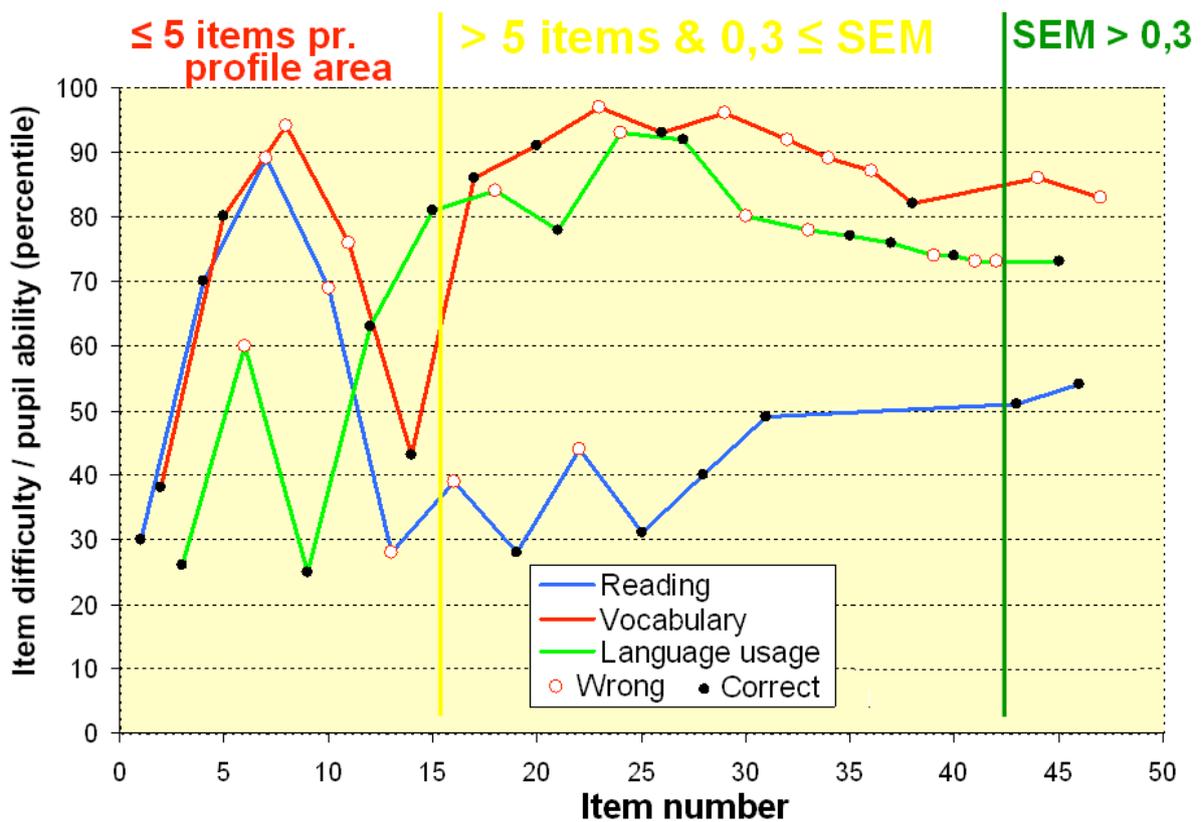


Figure 2. An illustration of the adaptive principle for initial and continuing ability estimates.

Usually the reliability criterion of SEM level of 0.3 or less is reached within 15 items per profile area. This is about one third of the items that would be needed in a linear setup.

The pupils continue to get questions until the time is up. The more items, the better and more detailed the analysis of the pupil's proficiency.

The items are chosen from a database with more than 180 items per profile area (3 profile areas yields 540 items per form/test). The items in each profile area should cover a wide range of pupils' abilities: At least 20 items in each of 5 groups matching 1) the lowest performing 10 percentiles of the pupils [1-10], 2) the next 25 percentiles [10-35], 3) the middle 30 percentiles [35-65], 4) the next 25 percentiles [65-90] and 5) the best performing 10 percentiles [90-100]. To measure the single item's difficulty and verify that the item fits the model, all items are initially tested on 500-700 pupils. The items and the responses from these initial tests are statistically analyzed to eliminate the items that do not fit the Rasch model. Just about half of the items that are tested meet the psychometric/statistical demands of the Rasch model. Those items that are not accepted will not be used in the national test – some are corrected and retested, others are deleted.

An example is shown in Figure 3 for a pupil in an English test course from March 2010. This figure depicts the ability estimate for three areas as the test progresses. It will seem that there are quite large fluctuations (especially in the beginning), but remember that the difficulty/ability is shown on a percentile scale which is extremely dynamic in the middle of the scale.

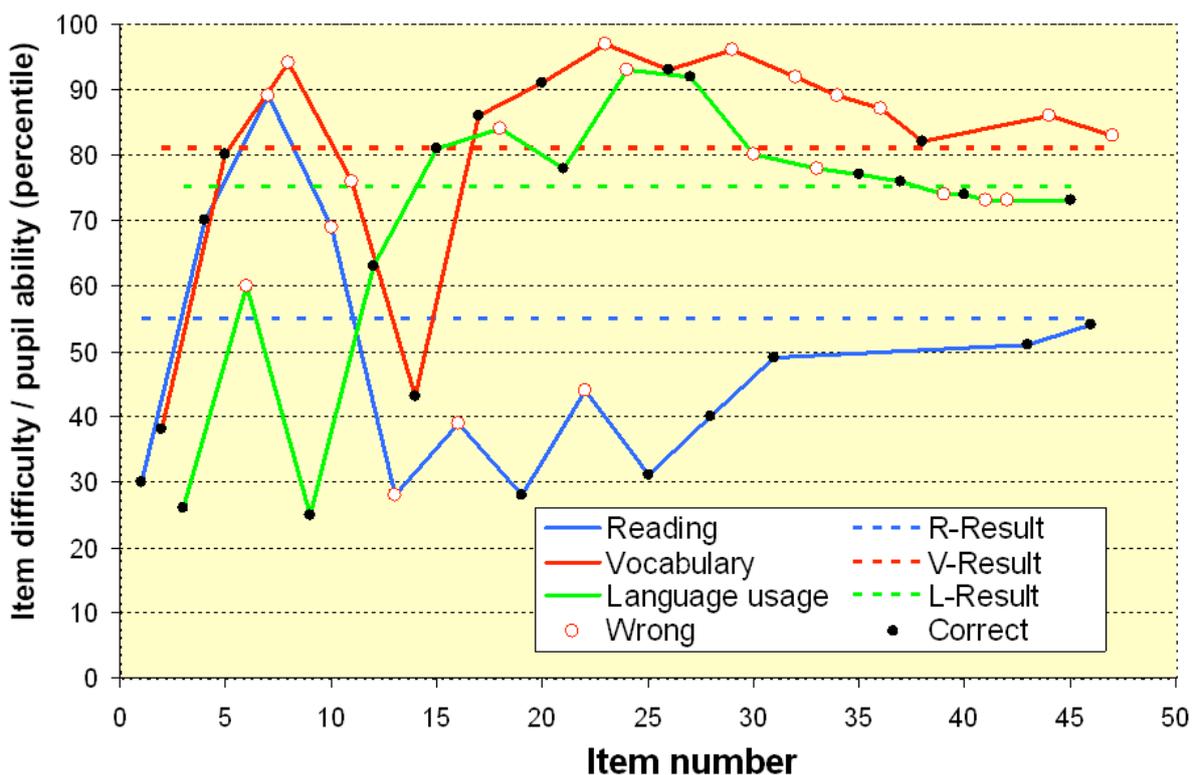


Figure 3. An illustration of how the ability estimate changes across items for one student on each test

### What is tested – and when?

Twelve (12) tests are being developed, with 12 x 3 different item banks for 7 different subjects: Danish/reading, math, English, Geography, Biology, Physics/Chemistry and Danish as a second language. The tests are targeted to the form where they are compulsory. Ten (10) of

the 12 tests are compulsory to use once per pupil in the *Folkeskole*. The two tests in Danish as second language are voluntary to use for the schools and are targeted 5<sup>th</sup> and 7<sup>th</sup> form). An overview of the testing requirements for pupils is shown below in Table 1.

Table 1. Tests in the *Folkeskole* I different forms

Form	1	2	3	4	5	6	7	8	9
<b>Subject</b>									
<b>Danish/reading</b>		X		X		X		X	
<b>Mathematics</b>			X			X			
<b>English</b>							X		
<b>Geography</b>								X	
<b>Biology</b>								X	
<b>Physics/Chemistry</b>								X	
<b>Danish as a second language</b>									
<b>X: Compulsory , : Voluntary</b>									

Compulsory testing in mathematics 6<sup>th</sup> form, Danish reading 8<sup>th</sup> form and physics/chemistry 8<sup>th</sup> form were carried out in May/June 2007. All of the compulsory tests were launched in February – March 2010. In the years to come, the compulsory testing must be carried out in all testing subjects every year between February 1<sup>st</sup> and April 30<sup>th</sup>. Within this period of time the teacher can decide when to administer the test. The online testing system contains a flexible booking facility where the teacher can reserve slots for a pupil or a group of pupils (i.e., a class).

In addition to the compulsory use, the schools are allowed to use the testing system – each test – voluntarily twice per pupil, either in the form for which the test is compulsory or in the previous or the following form. Booking for voluntary testing will be open all year – except when there is compulsory testing. Due to the construction of the test system on a fixed scale, two test results for the same pupil are directly comparable, so the test system is designed for measuring progress and added value. For each item, it is registered which part of the subject is tested and which teaching objectives should be in focus to enhance the pupils ability to solve this item. Detailed reports with recommendations for the teacher are under development.

The next step – when fully implemented – could be to merge the item banks in the same subjects (Danish/reading 2, 4, 6 and 8, Math 3 and 6 and Danish as second language 5 and 7). This way the teacher will be able to monitor progress from 1<sup>st</sup> to 9<sup>th</sup> form in Danish/reading, 2 to 7 form in math and 4 to 8 form in Danish as second language. Merging the item banks is being prepared, but is not yet decided. To do so, a new scale (a linear transformation of the Rasch scale) has to be introduced.

### Which parts of the subject are tested?

The tests are designed to assess large and important parts of the subject. However, not all parts of a subject are suitable for this kind of testing (e.g., pupils' ability to express themselves orally or in writing). Furthermore, the teachers are, due to the law, obliged to assess the student progress regularly. In other words, it is necessary to combine different kinds of assessment and evaluation. The national tests can only cover a very small part of the total need for evaluation in the *Folkeskole*. Therefore, the website [www.evaluering.uvm.dk](http://www.evaluering.uvm.dk) also contains a description and a user guide to a large number of other evaluation tools. Like other kinds of evaluation (all except the final exams) in the *Folkeskole*, the national tests are carried through as an integrated part of the education.

Every subject is divided into 3 dimensions or areas of the subject, called profile areas, to make a more detailed and precise evaluation of the pupil's proficiency possible. When being tested, the pupil will experience, that the questions are being presented in random order. But as previously mentioned, three separate adaptive test sessions are simultaneously conducted, where the selection of the next item in a given profile area solely depends on the pupil's response of the previous items in the same profile area. Table 2 illustrates these profile areas.

Table 2. Profile areas for the tests across subjects

Subject	Profile Area 1	Profile Area 2	Profile Area 3
<b>Danish / reading</b>	Language Comprehension	Decoding	Reading Comprehension
<b>Mathematics</b>	Numbers and algebra	Geometry	Mathematics in use
<b>Physics / Chemistry</b>	Energy	Phenomena, substances and materials	Applications and Perspectives
<b>English</b>	Reading	Vocabulary	Language and linguistic usages
<b>Geography</b>	Natural Geography	Cultural Geography	Applied geography
<b>Biology</b>	The living organism	Living organisms' interaction	Applied biology
<b>Danish as second language</b>	Reading	Vocabulary	Language and linguistic usages

### What is at stake?

Testing and test results can be used in different ways:

- **Tests** can be used as a tool for training the pupils or as a tool for measurement: Either to assess the pupil's academic level (knowledge, competencies and skills) or to assess the added value (the pupil's progress, the effect of the teaching) or both.
- **Test results** can also be used in different ways: They can be used in a formative way as a pedagogic tool, for the teacher's preparation of the teaching and training (typically low stake) as a feedback to the student and/or the teacher. Or they can also be used in a summative way for admission, regulation, controlling, rewarding/punishment of students/teachers/schools (typically high stakes).

From an international perspective, scores from educational testing programs usually involve high stakes uses (i.e., tests for which the results have significant consequences for individuals or schools (e.g. pupils' further educational possibilities, teacher's salary or school grants/existence). Testing and test results can to some degree be used for all these purposes, but there are some restraints.

In high stake testing, security, equal terms and fair conditions are key issues. But if the main purpose – as in Denmark – is to assess student proficiency in order to target teaching, and thereby improve the conditions for learning, the teacher should have access to full control over the testing conditions (e.g. which aids, tools, remedies and assistive technology are allowed during the test). In fact, if it improves the *teachers'* analysis of the student proficiency, it would make sense that the teacher is allowed to help if e.g. the pupil gets stuck – even though it could have significant impact on the test result. Under these conditions high stakes uses could not be supported because the scores are not necessarily comparable.

In high stake testing the results are usually made public (at least at a school average level) for different purposes (e.g., for school comparison/ranking.). The test should then be designed to measure the educational curriculum in the full range, so that “teaching for the test” is no problem. This is for example the case in the final exams.

When only a small part of the curriculum is tested – as in the Danish national test system – it is important to keep the stakes low. High stakes would lead to more focus on what is tested than on the *National Common Objectives* of the teaching in the act of the *Folkeskole*. That is,

too much focus on the tested profile areas and too little focus on creative, innovative and oral skills (which plays a significant role in the curriculum of the *Folkeskole*).

### **Disseminating results**

The main purpose of the testing system is to provide teachers with information for pedagogical purposes. The tests serve as a pedagogical tool that can help the teacher analyze the proficiency level of the pupils and the level of the class. In order to reduce the incentive to “teach to the test” and as precautionary measures against ranking of teachers, schools or local communities, it is forbidden by law to publish the test results. Any test result obtained by a pupil, an average by a group of pupils, classes, schools, or municipalities are strictly confidential.

Only those, who for professional reasons need information about the results, are allowed to see them. All the results are kept in a secured database. The database contains all the items used for testing the pupils as well as the answers that the pupils gave. Schools and municipalities are allowed to see and compare the results of the tests on different levels, according to their area of responsibility.

The teacher has access to detailed reports with information about his or her individual pupils’ result as well as test results on class level for his or her own class. The headmaster is allowed to see the pupil’s overall results, the class results and the results for the school. The local governments/municipalities have access to result on school level and test results aggregated to municipality level.

Parents must also be informed about their child’s test results by the school. For this purpose, the computer generates a written description of the results for each pupil.

There is a strong tradition for parent involvement in the *Folkeskole*, and the test results should be used in the school’s cooperation with the pupils and their parents, in order to support each pupil in the best way possible.

### **What kind of response is given to whom?**

**The pupil** As soon as possible after the test (typically within a week), the teacher will talk to the pupil about the result. It is not just information about the result but a discussion about what to do. Together they will plan the best way to improve teaching/learning in the future. The most common situation is that the test results match the teacher’s perception of the pupil’s proficiency – in this case the test just confirms that they are on track. But experience shows that in a class, there are usually at least a couple of surprises, about which the test provides new knowledge and useful information. The way the teachers receive these information is very different: Most teachers recognize that test results may show something they themselves had not found out but there are also some teachers who reject the possibility of becoming better informed about their pupils’ abilities as measured by this test (i.e. they have under- or overestimated the ability of the student) and ignore the test results.

**The parents** will be given the results for the different profile areas, an overview of the content of the items of the pupils test followed by a short explanation of the test results (see Appendix 2 - Test results for the parents are presented as automatically generated text from the test system - translated by Google).

**The teacher** has access to the results for the individual pupils and for the whole class – but only his or her own pupils and only the results in the subject in which he or she is teaching. The teacher will also be able to review the test course in details (e.g., which items a given pupil has answered as well as the result for the pupil). Appendix 2 also presents some screen shots demonstrating the types of information the teacher can download from the test system.

**The headmaster** is responsible for the teaching in his or her school. The headmaster has access to the test results for his or her own school, for all the classes and to the test results for a given pupil, but not the details from the test course. The headmaster will inform the school board about the results for the whole school.

**The municipality** having the overall responsibility of the running and performance of the local schools is allowed to see the results for the schools in the municipality. The local government/municipality is allowed to have access to the results of the individual schools and this information aggregated to local government level.

#### **How can the teacher and headmaster use the results?**

The results from the tests will contribute to the teacher's assessment of the proficiency and the progress of the individual pupils and of the whole class. What is special about these tests is that they assess the performance and the progress in the different profile areas, which are defined to interact with the Danish National Common Objectives for the teaching.

The results will help the teacher get a more precise overview of the effect of the prior teaching (e.g., how are the pupils performing in the part of the subjects which has been the focus of the teaching? Is there a need for any follow up activities?). It also provides information about the pupils' proficiencies in areas in which they are to be taught in the immediate future in order to target the teaching towards the classes and the individual pupil's preconditions and needs.

The headmaster has the overall pedagogical responsibility at the school and therefore an obligation to guide and to coach the individual teachers in pedagogical matters. The test results should therefore also be seen as a tool to pedagogic leadership.

#### **The national results: The national profile of performance**

The pupils' test results from the first 3 weeks of the first full-scale compulsory tests in 2010 were used to define a percentile scale for each profile area. These results will constitute the reference in the years to come. This way it will be possible to measure progress over time.

But the test results for the schools are dynamic: What is accessible on the test site is the average result for all the pupils that are registered at the school at the given day. Therefore, every year when the compulsory tests are done, and before the summer holiday, there will be taken a "snapshot" of the situation on all schools (i.e., the pupils and their test result - in 2010 it was done the 25<sup>th</sup> of June).

The purpose is to be able to monitor the overall development. Furthermore, the schools and the municipalities will be able to compare their results with the average results from the whole country.

When all the results from a compulsory test are registered, the mean performance of the pupils from the whole country and the distribution around the mean will be calculated. The distribution of the results will be separated by percentiles into 5 levels (equal to the 1-5 scale).

The results in the profile areas are called the National Profile of Performance. These national results from the compulsory tests will be published annually. In the years to come, this will make it possible to compare the results with the results from 2010 to see how the next cohorts of pupils' are performing in the same tests.

### **Correcting for differences in social background**

The schools and the municipalities will be able to compare their results with the national results (i.e., both the national profile of performance and the mean for the whole country). However, the background of the pupils in different schools is very different when it comes to socioeconomic factors, which usually relate statistically to the pupils' test results. This will be taken into account and a statistical correction will be made. This correction will take into consideration factors as gender, ethnic background, parent's education and socioeconomic status. The corrected results are confidential, but will be given to the school and the municipality. This will make it possible to take the socioeconomic factors into account when comparing the local results and the results from the whole country.

### **Recent experiences and plans for further development**

The compulsory tests in 2010 were implemented successfully, even though there were some technical problems. It was necessary to shut down the test system for 2 weeks due to problems with response time during maximum capacity utilization. All the test bookings in this period were lost. As a result the schools were exempted for the tests for these pupils. And even though some schools did manage to rebook and complete the test at a later date, not all the pupils that were supposed to, had an opportunity to complete a compulsory test. More than 3 out of 4 tests were completed. But the completion rate is not the success criterion. It is mandatory for the schools/teachers (according to law) to ensure that all relevant pupils complete the tests, but the teacher is not obliged to pay any interest to the test result. The most important success criteria are therefore:

- Do the teachers log in, download and study the test results of their pupils?
- Do the teachers find the test results useful in their pedagogical work?
- Do they actually use the knowledge they have acquired about their pupils to do anything differently?

To uncover to what extent these success criteria are fulfilled, in the end of June 2010 questionnaires were distributed to teachers in one third of the schools. The results have not yet been thoroughly analyzed, but the first responses suggest that almost all the teachers have seen their pupils test results, 4 out of 5 teachers find the test results useful for one or more purposes (planning of teaching, dialogue with pupils/parents, analytical purposes), and almost one out of three teachers have already used the test results in the planning of the teaching.

But the test system will really show its worth when the schools have access to data for the pupils over a couple of years. Then it will give the teacher an opportunity to monitor the individual pupil's progress. Moreover, it will give the teacher a possibility of advanced analysis of the progress of the class. Furthermore, it will provide information to headmasters about the classes and schools that will enable new possibilities for pedagogical coaching and leadership.

Nevertheless, it is important to emphasize, that this system – as any other test system – provides information about the pupil's proficiency, knowledge, attainment – and not solutions on pedagogical problems. Even though, in some cases, the right interpretation of test results can

deliver information that will tell the teacher, what would probably be the right thing to do with the individual pupil or the class. This kind of intelligent interpretation of test results could be the next area of development.

### **References**

Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danske Paedagogiske Institut.

## APPENDIX 1: The five components of the Danish CAT

The basic choices and principles in the Danish national test system are as follows:

**Calibrated item pool:** There are 3 item pools (so-called “profile areas”) pr. test for each of the 12 tests. All items are pilot tested on 500-700 pupils. Each pupil is administered 90-150 items - a mixture of dichotomous and polytomous items. Approximately half of the items have been accepted after pilot testing (i.e., they fit the Rasch model). The control of the Rasch model included the following three steps:

- 1) summary test of homogeneity. Does the total set of items function in the same way among pupils with good and less good skills?
- 2) summary test of differential item functioning. Does the total set of tasks function in the same manner among students in different groups (e.g., in the same way for girls and boys, in the same way for pupils in east and west Denmark, in the same way for pupils in large and small schools)?
- 3) Testing of individual items by using item-fit statistics. This test included for each item both a control of homogeneity (if item difficulty compared with all other items are the same for talented pupils and less talented pupils) and a control of differential item functioning ( if the items difficulty compared with the items are the same in different groups of pupils).

There was identified one Rasch scale per profile area (logit scale). These  $12 \times 3 = 36$  profile areas each contain at least 180 items that fits the model. On basis of the test results (theta estimates/pupils’ abilities) there is calculated a percentile scale. It has been a requirement, that the items should be distributed by difficulty, so that there are an approximately equal number of items (20 percent of the items in the item pool) targeted pupils within each of five groups of ability (i.e., where the item difficulty is equal to the ability – the theta parameter). The groupings of pupils by ability are as follows: Group 1: 0-10 percentile, Group 2: 11-35 percentile, Group 3: 36-65 percentile, Group 4: 66-90 percentile, Group 5: 91-100 percentile.

**Starting Point:** Each test consists of three separate courses (i.e., adaptive cycles for each profile area) where items however come in mixed order. The starting point of the test of each profile area is a dichotomous item drawn with an item difficulty of approximately 0 on the theta scale which approximately equals the average difficulty in the item pool. This drawing is randomized within a narrow range, and in the present design it is not influenced by the pupil’s past performance in the test system or by other factors - only the item difficulty.

**Item selection algorithm:** Items 2-5 within profile area selected from the dichotomous item in the item pool. The selection is based solely on the result of the previous items. Correct answers give more difficult questions and a wrong answer gives easier items. After five items theta (the person parameter) is estimated. The 6<sup>th</sup> item is drawn on the basis of the following criteria in prioritized order: the first criterion is that the item’s difficulty should lay within a distance of  $\pm 0.5$  from the theta estimate; the second criterion is the answer of the previous item (correct  $\rightarrow$  higher difficulty, incorrect  $\rightarrow$  lower difficulty); the third criterion is diversity of the content. After answering item 6, theta is re-estimated and a new item is drawn using the same criteria as for the 6<sup>th</sup> item, and so on.

**Theta estimation:** Maximum likelihood estimation (weighted Maximum likelihood for testing of items, ordinary Maximum likelihood estimation of person parameter while the test system is running).

**Termination criteria:** The termination is controlled by the teacher. There is employed a mixed time- and SEM-based, user controlled (i.e., teacher) termination criteria:

- By default, the test is terminated after 45 minutes.
- The teacher can prolong the test until 180 minutes – or stop it any time. The teacher is administering the test via a monitor screen, where there also is information about the test course (if theta has been estimated and if the standard error - SEM - is below or above 0.3 for all profile areas in the test) and information about how many items each pupil has answered.
- Until theta is estimated the first time for all profile areas (i.e. after five items per profile area = 15 item) the pupils test status is marked red on the monitor.
- Hereafter the color of the status marker on the monitoring screen switches to yellow. The algorithm draws on shift items from each profile area until SEM at theta estimate on one of the profile areas are reduces to 0.3.
- After this, items are only drawn from the other two. When SEM of the estimate on the second profile area is reduced to 0.3, the test continues with items from the last profile area.
- When SEM is reduced to 0.3 the status marker on the monitoring screen switches to green, and the algorithm draws on shift items from all three profile area until it is stopped either by the timer or the teacher. In the normal course SEM reduced to 0.3 before 15 items per Profile area. In some cases (especially high-or low-performing students) cases, SEM at one or more profile areas can not be reduced to 0.3 and here is the status marker on teacher monitor screen (regardless SEM value) set to “green” after 30 items.

Teachers may also choose to stop the test before the status marker is green. If status is not green, the teacher is given the choice to terminate the test at the present level (and get the results based on what has been achieved) or to resume the test later on e.g., next day/week.

## APPENDIX 2: Test results for the teacher (Ole) about the pupil (Ida) and her Class

Test results (average of pupils results) from all Ole's Classes (translated by Google)

Google™ Toolbar har [automatisk oversat](#) denne side til Engelsk

Vis oprindelig

@valuering.uvm.dk 

Status

**NATIONAL TEST**

Name: Ole Jorgensen ( olex2561 )

School : Evaluating School

Login Status: Teacher

---

**Booking**

Test Overview

Status

Test Results Management

Test Results

> > 3.U Mathematics 3rd at .

> > 8a Geography 8th at .

> > Create / student groups

To the front page of the website

Manuals

---

**SUPPORT**

Contacting UNI • C  
tel 89 37 66 00  
Mon. - two . at . 8-17  
fr . at . 8-15  
support@uni-c.dk

### Results - Table

**Confidential**

Information on the test tasks and test results are confidential, see primary Law § 55 b. The reason - the failure to print and store non screens. Unauthorized disclosure or use of test results and test tasks will be punishable under the Penal Code provisions on professional secrecy .

**Period: Last test results**  
Teacher : Ole Jorgensen ( olex2561 )

Mathematics 3.kl.	Numbers and algebra Score 1-100	Geometry Score 1-100	Mathematics in use Score 1-100	Overall rating Score 1-100	Number
3.U	4	12	2	6	8

Geography 8.kl.	Natural Basis Score 1-100	Cultural Geography Score 1-100	Using geography Score 1-100	Overall rating Score 1-100	Number
8a	34	19	29	28	12

**Profile Area**

Each subjects tested are divided into three profile areas. All tasks in a profile field measures the same type of professional competence.

**Result of 100 -point scale (1-100 )**

The result shows how students are performing compared to students nationwide. For example, a profit of 50 approx. for an average student. A result greater than 90 is given to the approx. 10 % of students whose academic level in the test is highest, and a score of 10 given to the approx. 10 % of students whose academic level in the test is lower.

**Overall rating**

Rounded average of results from the three profile areas.

Udført

### Graphical illustration of the pupils results of class 8 a, geography

Google™ Toolbar har [automatisk oversat](#) denne side til Engelsk

Vis oprindelig

**NATIONAL TEST**

Name: Ole Jorgensen ( olex2561 )

School : Evaluating School

Login Status: Teacher

---

**Booking**

Test Overview

Status

Test Results Management

Test Results

> > 3.U Mathematics 3rd at .

> > 8a Geography 8th at .

> > All students - Tables

> > Details pupils

> > Print to home

> > Create / student groups

To the front page of the website

Manuals

---

**SUPPORT**

Contacting UNI • C  
tel 89 37 66 00  
Mon. - two . at . 8-17  
fr . at . 8-15  
support@uni-c.dk

### Results - graphical overview

**Confidential**

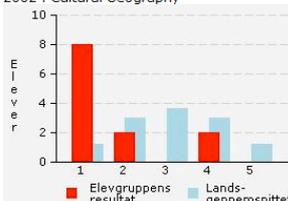
Information on the test tasks and test results are confidential, see primary Law § 55 b. The reason - the failure to print and store non screens. Unauthorized disclosure or use of test results and test tasks will be punishable under the Penal Code provisions on professional secrecy .

**Period: Last test results**  
Class: 8a  
Test : Geography 8th at .  
Teacher : Ole Jorgensen ( olex2561 )  
Number of students who have completed : 12

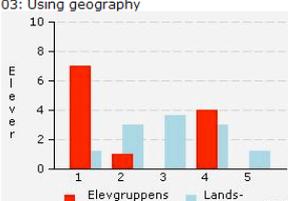
2001 : Natural Basis



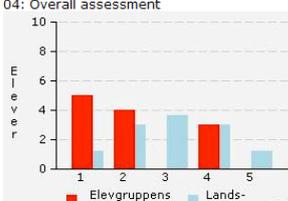
2002 : Cultural Geography



03: Using geography



04: Overall assessment



Results of the pupils of class 8 a, in geography. Table with both 1-100 scale and 1-5 scale.

Google™ Toolbar har [automatisk oversat](#) denne side til Engelsk Vis oprindeligt

styrelsen

Status

**NATIONAL TEST**

Name: **Ole Jorgensen ( olex2561 )**  
 School : **Evaluating School**  
 Login Status: **Teacher**

**Booking**  
**Test Overview**  
**Status**  
**Test Results Management**  
**Test Results**

> > 3.U Mathematics 3rd at .  
 > > 8a Geography 8th at .  
 > > All pupils - Tables  
 > > Details pupils  
 > > Print to home  
 > > Create / student groups  
 To the front page of the website  
 Manuals

**SUPPORT**

Contacting UNI • C  
 tel 89 37 66 00  
 Mon .- two . at . 8-17  
 fr . at . 8-15  
 support@uni-c.dk

### Results and status - Table

**Confidential**

Information on the test tasks and test results are confidential, see primary Law § 55 b. The reason - the failure to print and store non screens. Unauthorized disclosure or use of test results and test tasks will be punishable under the Penal Code provisions on professional secrecy .

**Period: Last test results**  
 Class: 8a Test : Geography 8th at .  
 Teacher : Ole Jorgensen ( olex2561 )  
 Number of students : 12

Sort by pressing headlines

Name	Unilogin	Natural Basis Score 1-100	Cultural Geography Score 1-100	Using geography Score 1-100	Overall rating Score 1-100	Status
Bo Jensen	boxx1548	97 = Well above average	89 = Above Average	83 = Above Average	90 = Above Average	Completed
Ditlev Jensen	ditl0096	2 = Well Below Average	1 = Well Below Average	69 = Above Average	24 = Below Average	Completed
Ejnar Jensen	ejne0047	58 = Average	90 = Above Average	72 = Above Average	73 = Above Average	Completed yellow
Erik Eriksen	erik2118	1 = Well Below Average	1 = Well Below Average	1 = Well Below Average	1 = Well Below Average	Completed
Frederik Jensen	fred289c	13 = Below Average	6 = Well Below Average	1 = Well Below Average	7 = Well Below Average	Completed
Hans Jensen	hans2914	30 = Below Average	7 = Well Below Average	1 = Well Below Average	13 = Below Average	Completed yellow
Ida Jensen	idax7005	98 = Well above average	12 = Below Average	87 = Above Average	66 = Above Average	Completed
Jane Jensen	jane2338	4 = Well Below Average	19 = Below Average	32 = Below Average	18 = Below Average	Completed
Mia Jensen	miax5706	1 = Well Below Average	1 = Well Below Average	1 = Well Below Average	1 = Well Below Average	Completed
Oline Jensen	olin0149	94 = Well above average	1 = Well Below Average	1 = Well Below Average	32 = Below Average	Completed yellow
Svend Petersen	sven1138	7 = Well Below Average	1 = Well Below Average	2 = Well Below Average	3 = Well Below Average	Completed
Svend Jensen	sven1232	5 = Well Below Average	2 = Well Below Average	1 = Well Below Average	3 = Well Below Average	Completed
<b>Class Average</b>		<b>34 = Below Average</b>	<b>19 = Below Average</b>	<b>29 = Below Average</b>	<b>28 = Below Average</b>	

Internet 100%

Details are covering both “results in text” and the “details” from the test course

Google™ Toolbar har [automatisk oversat](#) denne side til Engelsk Vis oprindeligt

@evaluering.uvm.dk

Skolestyrelsen

Status

**NATIONAL TEST**

Name: **Ole Jorgensen ( olex2561 )**  
 School : **Evaluating School**  
 Login Status: **Teacher**

**Booking**  
**Test Overview**  
**Status**  
**Test Results Management**  
**Test Results**

> > 3.U Mathematics 3rd at .  
 > > 8a Geography 8th at .  
 > > All students - Tables  
 > > Details pupils  
 > > Print to home  
 > > Create / student groups  
 To the front page of the website  
 Manuals

**SUPPORT**

Contacting UNI • C  
 tel 89 37 66 00  
 Mon .- two . at . 8-17  
 fr . at . 8-15  
 support@uni-c.dk

### All students - details

**Confidential**

Information on the test tasks and test results are confidential, see primary Law § 55 b. The reason - the failure to print and store non screens. Unauthorized disclosure or use of test results and test tasks will be punishable under the Penal Code provisions on professional secrecy .

**Period: Last test results**  
 Class: 8a  
 Test : Geography 8th at .  
 Teacher : Ole Jorgensen ( olex2561 )  
 Number of students : 12

Sort by pressing headlines

Name	Unilogin	See results in text	Details
Bo Jensen	boxx1548	<a href="#">Show</a>	<a href="#">See Answer</a>
Ditlev Jensen	ditl0096	<a href="#">Show</a>	<a href="#">See Answer</a>
Ejnar Jensen	ejne0047	<a href="#">Show</a>	<a href="#">See Answer</a>
Frederik Jensen	fred289c	<a href="#">Show</a>	<a href="#">See Answer</a>
Hans Jensen	hans2914	<a href="#">Show</a>	<a href="#">See Answer</a>
Ida Jensen	idax7005	<a href="#">Show</a>	<a href="#">See Answer</a>
Svend Petersen	sven1138	<a href="#">Show</a>	<a href="#">See Answer</a>
Erik Eriksen	erik2118	<a href="#">Show</a>	<a href="#">See Answer</a>
Svend Jensen	sven1232	<a href="#">Show</a>	<a href="#">See Answer</a>
Mia Jensen	miax5706	<a href="#">Show</a>	<a href="#">See Answer</a>
Oline Jensen	olin0149	<a href="#">Show</a>	<a href="#">See Answer</a>
Jane Jensen	jane2338	<a href="#">Show</a>	<a href="#">See Answer</a>

Internet 100%



## Home

**Pupil :** Ida Jensen  
**Class:** 8a  
**School :** Evaluating School  
**Teacher :** Ole Jorgensen  
**Test Date:** 2010-03-11  
**Test Type:** Mandatory testing 2009-2010

### Results of tests in geography 8th Class

IDA performs best on the field "natural base ". IDA has performed the test in geography 8th class level *above average* compared with other students in eighth grade. She has consistently easy to meet the demands of the professional areas that are tested.

Below you can see the results for each area .

- **Natural Basis**  
Ida's performance is clearly above average in this area. The test shows that she is very easy to meet the requirements . Here is an overview of the issues , her duties have focused on: *Weather and climate , erosion, plate tectonics ( changes in the earth's crust ) , landscape formation , water cycle and ecosystems.*
- **Cultural Geography**  
Ida's performance is below average in this area. The test shows that she has difficulty in meeting the requirements . Here is an overview of the issues , her duties have focused on: *Population and demographic developments , cities and urban development , economic development and living conditions and production and environmental.*
- **Using geography**  
Ida's performance is above average in this area. The test shows that she can easily meet the requirements . Here is an overview of the issues , her duties have focused on: *Exploitation of natural resources, settlement patterns , agriculture, climate change, natural disasters , sustainability and globalization.*

In the box below you can read more about what the test tests.

# Details about the test course of Ida (translated by Google)

## NATIONAL TEST

Name: Ole Jorgensen ( olex2561 )

School : Evaluating School

Login Status: Teacher

### Booking

### Test Overview

### Status

### Test Results Management

### Test Results

> > 3.U Mathematics 3rd at .

> > 8a Geography 8th at .

> > All students - Tables

> > Details pupils

> > Print to home

> > Create / student groups

To the front page of the website

### Manuals

### SUPPORT

Contacting UNI • C  
tel 89 37 66 00  
Mon .- two . at . 8-17  
fr . at . 8-15  
support@uni-c.dk

## Answer

### Confidential

Information on the test tasks and test results are confidential, see primary Law § 55 b. The reason - the failure to print and store non screens. Unauthorized disclosure or use of test results and test tasks will be punishable under the Penal Code provisions on professional secrecy .

### Period: Mandatory testing 2009-2010

Class: 8a

Test : Geography 8th at .

Teacher : Ole Jorgensen ( olex2561 )

Pupil : Ida Jensen ( idax7005 )

Subjects: Geography 8th at .

Sort by pressing headlines

Question nr	Profile Area	Item Topic	Scores	Difficulty	Test Tasks
1	Natural geography	Weather and Climate	R	***	<a href="#">View response</a>
4	Natural geography	Weather and Climate	R	*****	<a href="#">View response</a>
7	Natural geography	Plate tectonics ( changes in the earth's crust	F	*****	<a href="#">View response</a>
10	Natural geography	Weather and Climate	F	*****	<a href="#">View response</a>
13	Natural geography	Ecosystems	R	****	<a href="#">View response</a>
16	Natural geography	Landscape Formation	R, R	*****	<a href="#">View response</a>
19	Natural geography	Landscape Formation	R	*****	<a href="#">View response</a>
22	Natural geography	Weather and Climate	F	*****	<a href="#">View response</a>
25	Natural geography	Weather and Climate	B	*****	<a href="#">View response</a>
27	Natural geography	Weather and Climate	R	*****	<a href="#">View response</a>
29	Natural geography	Weather and Climate	R	*****	<a href="#">View response</a>
31	Natural geography	The water cycle	R	*****	<a href="#">View response</a>
33	Natural geography	Landscape Formation	R	*****	<a href="#">View response</a>
35	Natural geography	Weather and Climate	F	*****	<a href="#">View response</a>
37	Natural geography	Ecosystems	R	*****	<a href="#">View response</a>
40	Natural geography	Landscape Formation	B	*****	<a href="#">View response</a>
43	Natural geography	Weather and Climate	B	*****	<a href="#">View response</a>
46	Natural geography	Landscape Formation	F	*****	<a href="#">View response</a>
49	Natural geography	Erosion	B	*****	<a href="#">View response</a>
52	Natural geography	Weather and Climate	B	*****	<a href="#">View response</a>
2	Cultural Geography	Population and population development	R	***	<a href="#">View response</a>
5	Cultural Geography	Economic development and living condition:	F	*****	<a href="#">View response</a>
8	Cultural Geography	Production and Environment	B	***	<a href="#">View response</a>
11	Cultural Geography	Cities and Urban Development	F	*	<a href="#">View response</a>
14	Cultural Geography	Cities and Urban Development	R	*	<a href="#">View response</a>
17	Cultural Geography	Population and population development	R, R, R	*	<a href="#">View response</a>
20	Cultural Geography	Population and population development	R	***	<a href="#">View response</a>
23	Cultural Geography	Economic development and living condition:	B, B, B, B	***	<a href="#">View response</a>
38	Cultural Geography	Cities and Urban Development	F	**	<a href="#">View response</a>
41	Cultural Geography	Population and population development	B	**	<a href="#">View response</a>
44	Cultural Geography	Economic development and living condition:	B, B, B, B	**	<a href="#">View response</a>
47	Cultural Geography	Cities and Urban Development	R, R, R, R	*	<a href="#">View response</a>
50	Cultural Geography	Cities and Urban Development	B	**	<a href="#">View response</a>
3	Using geography	Sustainability	R	***	<a href="#">View response</a>
6	Using geography	Exploitation of natural resources	F	*****	<a href="#">View response</a>
9	Using geography	Sustainability	R	***	<a href="#">View response</a>
12	Using geography	Climate change	R	*****	<a href="#">View response</a>
15	Using geography	Climate change	R	*****	<a href="#">View response</a>
18	Using geography	Exploitation of natural resources	F, R, R	*****	<a href="#">View response</a>
21	Using geography	Sustainability	F	*****	<a href="#">View response</a>
24	Using geography	Settlement Patterns	B	*****	<a href="#">View response</a>
26	Using geography	Climate change	R	*****	<a href="#">View response</a>
28	Using geography	Exploitation of natural resources	F	*****	<a href="#">View response</a>
30	Using geography	Natural Disasters	R	*****	<a href="#">View response</a>
32	Using geography	Exploitation of natural resources	F	*****	<a href="#">View response</a>
34	Using geography	Agriculture	R	*****	<a href="#">View response</a>
36	Using geography	Globalization	F, R, B	*****	<a href="#">View response</a>
39	Using geography	Sustainability	B	*****	<a href="#">View response</a>
42	Using geography	Agriculture	B	*****	<a href="#">View response</a>
45	Using geography	Settlement Patterns	B	*****	<a href="#">View response</a>
48	Using geography	Settlement Patterns	B	*****	<a href="#">View response</a>
51	Using geography	Settlement Patterns	B	*****	<a href="#">View response</a>

### Profile Area

Each subjects tested are divided into three profile areas. All items in a profile area measures the same type of proficiency.

### Item Topic

The content of the item. There are approx. 6-12 types of item content for each profile area. Item topics are included in the description of the results provided to teachers and to home.

**Scores:**All questions of the tasks can be answered either true (R), false (F) or blank (B). A task may contain one or more subtasks . The answer to each subtask and the result of profile field called score.

**Difficulty** is indicated by one to five stars. This scale corresponds to five -point performance scale.

Difficulty (item )		corresponds to	Skill ( student performance )	Outcome 1-100
* ( 1 star)Very easy		corresponds to	1 Well BA	1-10
** ( 2 stars)Light		corresponds to	2 Below A	11-35
*** ( 3 stars)Medium weight		corresponds to	3 Average	36-65
**** ( 4 stars)Severe		corresponds to	4 Above A	66-90
***** ( 5 stars)Very hard		corresponds to	5 Well AA	91-100

### Rated / used time (used in Danish / reading )

Rated time is the time which the average student spends on task.

Applied Time is the time it actually registered, the student has spent on task for the test.