Chapter 8[1]

Education, Testing, and Validity: A Nordic Comparative Perspective

Jakob Wandall, NordicMetrics

Abstract

In reacting to Visscher's chapter (In Chatterji, 2013, pp. 101–135), this chapter aims to describe and explain similarities and differences in schooling and assessment systems in different countries, with a specific focus on Denmark, Netherlands, and the United States. Through a comparative and culture-based view of education systems and values in these countries, the purpose of this chapter is to provide a more nuanced picture of their systems for assessing school outcomes. Questions explored include: What is valued, what is measured, and what counts? Which validity considerations should matter in evaluating assessments and education systems?

8.1. Purpose

To react to Visscher's chapter (In Chatterji, 2013, pp. 101–135), this chapter describes and explains similarities and differences in schooling and assessment systems in Denmark and the Netherlands, with specific reference to systems in the United States. Through a comparative and culture-based view of education systems and values in these countries, the purpose of this chapter is to provide a more nuanced picture of systems for assessing school outcomes.

8.2. Comparing Educational Systems and Systems for Assessment

8.2.1. Denmark and the Netherlands: National Similarities and Education Systems

8.2.1.1. National profiles Denmark and the Netherlands have much in common. They are geographically small, densely populated coastal countries (40,000 km2) with a very flat topography (no mountains). Both have a long tradition of cultivating their soils. In their cities, water plays an important transportation role (through harbors, and canals), and bicycles are another main means of transportation. They are both ancient seafaring nations and both have a previous status as colonial powers. In both countries, the transition from absolute monarchy to constitutional monarchy happened peacefully, with cooperation between the monarch and politicians. The transition happened independently but simultaneously in both nations between 1848 and 1849. Their languages are related, both originating from the Germanic family (old Dutch can be read by many Danes). Both countries are known for liberalism, tolerance, and

[1] Chatterji, M. (2013), (ed), Validity and Test Use: An International Dialogue on Educational Assessment, Accountability and Equity, Copyright r 2013 by Emerald Group Publishing Limited
All rights of reproduction in any form reserved
ISBN: 978-1-78190-946-1
http://www.emeraldinsight.com/products/books/notable/page.htm?id=9781781909461

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

acceptance of divergent views. Many practices prohibited by law in other countries are accepted in the Netherlands and Denmark (e.g., legal sale and use of marijuana, regulated euthanasia in the Netherlands, and free pornography, free abortion, freedom of associations with groups having extreme political or religious views).
However, there are also several differences between the two countries. Danes see themselves as being on the edge of Europe, with their own currency (kroner) and have limited enthusiasm about the European Union (EU). Danes are soft nationalists — which means that one of the most important symbols of Denmark is the flag, called Dannebrog. The flag is traditionally used for most national celebrations. Denmark is ethnically very homogeneous, with 90% of the 5 million inhabitants being ethnic Danes.
Approximately 80% are members of the Lutheran state church, which is funded by taxes and part of the state administration. In Denmark, like the other Nordic countries — Norway, Finland, Iceland, and Sweden — people do not commonly attend church except for weddings and funerals. Danes, in general, do not believe in Heaven, Hell, sin, or even God. They are secular-rational rather than traditionally religious people (Inglehart & Welzel, 2005). The religious affiliation is more a cultural than religious trait (Zuckerman, 2009).
On the other hand, the population in the Netherlands is three times that of Denmark. The Netherlands is one of the core countries in Europe and has participated in the EU from the beginning. The Netherlands is the country in Europe with the most complete separation of church and state. After a transition to a constitutional monarchy, the Dutch society segregated into several segments or pillars according to different religious or ideological groups (this division is called pillarization or verzuiling in Dutch). These pillars all had their own social institutions: their own newspapers, broadcasting organizations, trade unions, banks, hospitals, universities, sports, and youth organizations. They also have their own schools. Half of the Dutch population is Christian, equally divided between Catholics and Protestants. There is a large Muslim minority, but 40% are reported to be not religious.
Compared with Denmark, the Netherlands is a multiethnic society with a 20% share of inhabitants from different ethnic backgrounds. There is significant annual migration, including immigration from former colonies (such as, Suriname, Indonesia, and the Dutch Antilles). Starting in 1965, however, a process of secularization and depillarization emerged and the influence of the institutionalized denominations declined greatly.

8.2.1.2. Education systems Deriving from the above socio-cultural differences, the principles for the monitoring and guidance of the national school system are also quite different in the two countries. While the Netherlands has quite a strong centrally driven inspectorate (Visscher, In Chatterji, 2013, pp. 101–135), the schools in Denmark are governed by 98 highly decentralized local governments. This characteristic in Denmark has led to significant differences between municipalities, in terms of their organization, economy, curriculum, buildings, and so on. However, some centrally defined regulations in Denmark are found in all school systems, such as a minimum number of lessons per year, a common framework of educational objectives (but no standards or a national curriculum), final examinations, and since 2010, a national information technology (IT)-based computer adaptive testing program.
Box 8.1 provides a snapshot of the statistical profile of the Danish basic education system. Information in Box 8.1 may be compared with

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

the Dutch primary school system, as given in Visscher (In Chatterji, 2013, pp. 101–135).

8.3. Comparing Assessment Systems in Denmark and the Netherlands

There is a weak tradition in Denmark in educational testing and measurement. Prior to 2010, no validated test existed in Danish (e.g., in subjects like reading comprehension). There were no standardized testing programs, or reporting of achievement test results on interval scale metrics, except for International Large Scale Assessment (ILSA) tests like the Programme for International Student Assessment (PISA). Those reports, however, are typically not accessible to schools and teachers.
In general, Danish teachers consider their judgments' of students' ability and/or progress as unquestionable. Experience shows that the quality of teachers' judgments based on their observations of the students' performance, especially in reading comprehension and other non-observable latent traits, varies a great deal (Allerup, Mejding, & Zeuner, 2001), even though the correlations between test scores and teacher ratings are often quite high.

Box 8.1. Public primary and lower secondary schools in Denmark (Folkeskolen): Statistical profile.

Folkeskolen Statistics:

- Eighty-five to ninety percent of all Danish children attend public schools, 600,000 pupils, 60,000 per grade level, 10 years mandatory schooling (0–9 grade).
- The remaining 10–15% attend private schools (including schools founded in religious beliefs). Private schools receive 75% of the average subsidy for public schools per student/year.
- Approximately 2,000 public schools, of those 1,300 ordinary folkeskoler (public school).
- Three to five pupils per new computer (o3 year) — in some municipalities the schools provide computers for all students.
- High speed Internet connection in all schools and in 98% of all homes with children.

→ The instructional information technology infrastructure is suited for ITbased assessment programs.

Assessment in Denmark, Folkeskolen:

- Summative—Final examination administered after ninth and/or tenth grade
- Formative — National testing system (10 adaptive tests targeted students from second to eight grade)

Note: The National testing system (10 adaptive tests) is described in detail in Wandall (2011).

The introduction of a full-scale, mandatory, psychometrically designed IT-based testing system was a controversial and much debated initiative in Denmark. At the beginning, the teachers were highly skeptical, but the Danish governmental approach (priority to pedagogical use of results, confidential results, and prohibition of public ranking of teachers, schools, and municipalities) has made the teachers mostly positive toward the tests. And as a side effect, it has generated a foundation for future research in and conversation about the validity of educational assessments in Denmark.
Most of the literature about modern testing and assessment in Nordic nations (both description of experiences and theoretical development) have

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

come from the English-speaking countries. The United States (U.S.) is by far the major contributor, but the United Kingdom (U.K.), Australia, New Zealand and Canada have also added to this development.

The history of high-stakes, large-scale testing is relatively short (less than 100 years) in Western civilizations. The dominant educational testing programs were developed in the U.K. and U.S., and most go back less than 40 years (Suen & French, 2003). Some Asian countries like China also have a very long tradition of testing (Gao, 2006). The rest of the world (except for the English-speaking and Asian countries mentioned) has made few markings on this agenda for developing educational assessment systems. Some of these markings come from the Netherlands and Denmark.

A significant difference between the Dutch and Danish school systems concerns the extent to which testing and measurement are used as a basis for evaluation of student achievement and school performance. This might be a historical coincidence. The early development of educational testing in both countries is quite parallel.

In the 1950s, the Dane, George Rasch was a freelance mathematician. He was hired by the Municipality of Copenhagen and the Ministry of Social Affairs to do some analysis on weak readers. In 1952, a report was published about the project. The basic principles of a mathematical logistic model that was subsequently named after him was described in an appendix to this report. In 1955, the Danish Pedagogical Institute (DPI) was formed, with Rasch as a one of the driving forces. In 1960, Rasch published his only book on probabilistic models, a major landmark in the development of modern educational testing theory with his work was the Rasch model, otherwise known as the one-parameter logistic model in item response theory (Rasch, 1960). DPI was a center of knowledge for educational testing in Denmark until 2000. At that time it was merged with The Royal Danish Teachers College, but later dissolved in practice.

In the Netherlands, the Dutch chess player and psychologist Adriaan de Groot was hired in 1949 at the University of Amsterdam. In cooperation with the U.S. cognitive psychologist, Hebert Simon in 1954, he developed an interest for psychology as applicable to the game of chess. In 1958, he visited Educational Testing Service (ETS) at Princeton, and brought home inspiration to form a Dutch version of ETS. Some years later, De Groot established the Institute for Educational Research. In 1966, a commission proposed the development of the Central Institute for Test Development (CITO), established in 1968 with inspiration from ETS.

A significant difference between the role of DPI in Denmark and CITO in the Netherlands was that CITO (like ETS) provided educational testing services to schools, butDPI was mainly a research and development institution. Even thoughDPI was involved in the development of tests, it never took part in the general administration of tests. The testing business was pursued by some DPI employees as a secondary activity and by publishers of school materials. From the beginning, the Dutch schools were skeptical about testing, especially when the tests were based mainly on multiple-choice questions. Danish schools had a similar attitude toward tests when national testing was introduced in 2007.

But in recent decades, the Netherlands has built a solid tradition of using tests in schools, partly guided by accountability purposes and partly by pedagogical purposes. This change has gone hand in hand with the development of strong psychometric communities within the field of education, not just at CITO, but also in Dutch universities. The Netherlands is one of a few nations that has contributed to the development of modern theory and knowledge of educational testing outside the English-speaking

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

world. It boasts a long list of well-known psychometricians (e.g., Adriaan de Groot, Wim van der Linden, Ed Roskam, Arnold Van den Wollenberg, Ivo Molenaar, Denny Borsboom, and Don Mellenbergh). As opposed to most Nordic countries, where testing based on scientific principles is a fairly new experience in the education system, educational testing in the Netherlands is a well-established tradition (Visscher, In Chatterji, 2013, pp. 101–135).

Three decades after the establishment of CITO and the introduction of educational testing in Dutch schools, a parallel process has recently begun in some of the Nordic countries, including Denmark. In Norway, national testing was introduced in 2004, evaluated, stopped, and relaunched in 2007 (Hatch, forthcoming). In Sweden, there have been mandatory national tests in grades 3, 5, and 9 since 2009 (Lundahl, 2009). The Danish national tests were launched in 2007, reviewed, and stopped, and relaunched in 2010. For the first time, it became possible for teachers in Denmark to get a measure based on scientific principles on the Danish students' performances in various subjects (reading, math, English, and science) provided by psychometrically validated tests (Wandall, 2011).

Unlike the Netherlands, there is a weak tradition for psychometric research and development in the field of education in most of other Nordic countries. In Denmark, the number of skilled psychometricians within the field of education can be counted on one hand, and currently, there is no visible sign that this field is being prioritized by the Danish State or within universities. In Norway, some attempt has been made to strengthen the educational psychometric society. Sweden has a long tradition for "prov"[2] with a strong focus on the content, didactics, and pedagogy but a weaker tradition for educational measurement.

In the Netherlands, as in the English-speaking countries previously mentioned, initiatives to develop assessment tools are located at universities and CITO. The culture of evaluation in schools comes from the scientific community.

In the Nordic countries, by contrast, testing is sponsored and operated by the state—in Sweden and Norway, by state agencies and in Denmark, by the ministry. By tradition in Denmark, the examinations and testing materials (items, tasks, scales, and guiding principles) have been developed within the national political and governance system by civil servants. Often these initiatives are run by individual civil servants with a teaching background and a teacher education, employed by the Ministry of Education.

In Denmark an opgavekomission, or a task commission, has been organized for each subject area to prepare a final examination. The majority of the commission participants are skilled teachers. The chairmen are usually civil servants with teaching background, who are part-time employees in the ministry, and also assigned part time at a school. These civil servants are the ministry's content experts on the subjects. Through their work, they develop skills in item writing. The items/tasks, except for those used in the national testing system, are not field-tested with students prior to their use. Keeping items/tasks confidential is a priority, and the scoring principles are developed by the individual commission. Criteria for grading are decided upon by the ministry. Often, there are significant variations in the distribution of the grades from one year to the next, mostly due to variations in the difficulties of examination tasks.

Looking closely, therefore, there are key differences between the cultures,

[2] In most of the Nordic countries, there is a distinction between testing exclusively with openended questions and essay style tasks (called prov or prøve) and testing with multiple choice items (called testing).

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

the school systems, and the educational assessment enterprises of Denmark and the Netherlands. The resemblance between the Dutch and Nordic welfare model is striking, and in some socio-economic comparative studies the Netherlands are also categorized as a Nordic or Scandinavian welfare type of country (e.g., Aiginger & Guger, 2006; Esping-Andersen,1990).

8.3.1. Danish School System — What Matters for Parents and Public Users

In spite of recent developments in national testing and mandatory final examinations, there is still a lack of focus on the science of assessment and measurement of cognitive/academic skills in Denmark. This is not because there is a lack of interest in school-related matters or a general deficit in research capacity. Nor is it because the Danish politicians do not find the subject interesting. It is quite opposite.

Until 1994, the general opinion in Denmark was that Folkeskolen (Danish public schooling) was the best educational system in the world. In 1990, prior to the first International Association for the Evaluation of Educational Achievement (IEA) study in Denmark, 81% of the population had confidence in the school system, compared to only 67% in the Netherlands and 55% in the U.S. (Inglehart, Basanez, & Moreno, 1998).

Ever since Denmark participated in an International Large Scale Assessment (ILSA) program, first in IEA studies in 1991 and later on in the PISA studies, the disappointing relative performance of Danish students has been on the top of the nation's political agenda and on the front page of most newspapers. There has been a demand for reforms (see Feuer, this volume, pp. 197–216). An increased number of Danes (up to 45% of the adult population, as compared to 20% before) have now developed a low level of confidence in public schooling as a system and in teaching as a profession.

Nonetheless, it seems that this public attitude is not shared by the primary users of the system — parents and students — at the individual level. In general, Danish students and parents[3] (76–78%) are satisfied with their school, both with the students' academic progress (78% satisfaction) and the schools' effort to provide education in non-cognitive domains (76% satisfaction). This result is uniform across all the studies that are conducted on a regular basis. When asked, parents usually say something like: "y I guess that we just have been lucky with our choice of school/teacher." A vast majority of parents say that "my kid is doing better than average in school."

The above observations provide a contradiction in terms: Poor Danish results on ILSA studies but high levels of satisfaction among Danish public users of the education system. There is an explanation in the Danish (Nordic) culture for this.

A traditionally strong focus on personal traits and non-cognitive skill domains coupled with a low focus on evidence on and priority given to academic performance domains is characteristic of the Danish culture. Parents generally ask about their child's well-being rather than insisting on strong cognitive development. So, high satisfaction combined with low performance on the academic achievement could be a result of low cultural priority given to cognitive growth of school children.

As the alliance with the parents is very important for a school in a highly decentralized system like the Danish one, this would be a reason for

[3] Survey (BTU) of the users' satisfaction in 2012 from the MoE (http://www.uvm.dk/B/UVMDK/Content/News/Udd/Folke/2012/Sep/B/media/UVM/Filer/Udd/Folke/PDF12/120919%20BTU%20Skole.ashx).

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

teachers/schools to not focus on evidence of academic progress in Denmark. It would also provide a background for understanding the lack of a strong evaluation culture in Danish schools. This lack was, in fact, pointed out by the OECD as an explanation for persistent underperformance observed by the Danish educational system (OECD, 2004).

Perhaps, the Danish parents do not care much about school quality as measured by students' academic performance. A strong indicator of parent preferences is what counts when they choose schools for their own children. In 2003, the Danish liberal government introduced free choice of schools. In order to qualify for the parents' choice of school, an Act on openness and transparency in the educational system was adopted by the parliament. The Act stated that all schools should have a web page, providing a wide range of specific data and information on the school's policy and performance, including grades and final exam results. This was so that "citizens simply and quickly are able to assess the quality of teaching in individual schools and institutions."[4] Box 8.2 shows results of an evaluation of this initiative. Based on the results in Box 8.2, it appears few parents use the website data to choose schools. It may be that the parents do not rely on their school's statistics nor believe that the choice of school makes much of a difference to their children's overall educational outcomes. This public perception in Denmark is supported by cross-cultural, meta-studies indicating that generally schools account for no more than 5–10% of the variance in student's academic achievement as measured by large-scale, standardized tests. Hattie (2003), based on this finding, has concluded that "schools barely make a difference to achievement."

Box 8.2. Free choice of school and Act on openness and transparency in Denmark.

Information on the individual schools average scores in the final exams has been available in Denmark since 2003. According to law it has to be made public on the schools' websites. A recent Danish survey (Rambøll, 2011) finds that only 10% of the parents say that they have checked the information on exam results and average grades when choosing a school for their own children, even though 91% knew that it is easily accessible information on the schools' webpage. 82% indicated that they felt good about having a free choice, even though very few used it: 86% chose the local school and the majority of the rest chose a private school.

A lot of factors play a role in Danish parents' choices of schools, especially the following: that the class is functioning well together socially, that they develop good, decent values, awareness, imagination, and confidence in their own ability, that they are reasonably happy with their daily life at school, and that they learn the basics (Danish, math, English, science, etc.) and how society works. In the list of priorities, academic achievement is not the top priority.

From the evidence above, however, it seems most likely that what really matters for Danish parents is that their kids feel comfortable and safe in school, and that they are happy to go to school — if not every day, then at least on most days. School quality — measured by the students' academic performance in final examinations or standardized achievement tests — is clearly not on the top of the list of parent concerns in Denmark.

[4] Act nr 414, June 6, 2002, Lov om gennemsigtighed og åbenhed i uddannelserne m.v.

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

8.3.2. The Purposes of Schooling in the U.S. versus Denmark

I now turn to a comparative analysis of Denmark and the Netherlands, with a selected English-speaking country — the U.S. — that has a strong culture and tradition in testing and evaluation. In U.S., by contrast to the above discussion, educational policy has a strong focus on promoting and measuring cognitive ability of students based on attainment and achievement tests (Heckman, 2008).

In legislation like the Elementary and Secondary Education Act (reauthorized under the No Child Left Behind (NCLB) Act of 2002[5]) the stated purpose of public schooling is:

> *. . . to ensure that all children have a fair, equal, and significant opportunity to obtain a high-quality education and reach, at a minimum, proficiency on challenging State academic achievement standards and state academic assessments. This purpose can be accomplished by . . .*

This is followed by a list of 12 different means by which the main purpose of public schools can be accomplished. There appears to be one main focus in U.S. educational policy: Cognitive skills and development of students in cognitive domains are prioritized, and academic achievement measured by standardized tests is a requirement.

There is solid evidence that cognitive abilities are important determinants of socio-economic success (Heckman, 2008). But, during the last decade, there has been a rising awareness that the concentration of attention on the NCLB-driven standards-based reforms and accompanying achievement testing could lead to oversight of equally important non-cognitive factors for success in school and life (Levin, 2012).

Now compare the first section of the Danish Act of the Folkeskole, which contains the overall objectives. Here, academic achievement plays quite a different role (Box 8.3).

Box 8.3. The Danish Act of the Folkeskole, statement of purpose.

§ 1: The Aims of the Folkeskole

(1) The school shall — in cooperation with the parents — give students' knowledge and skills that prepare them for further education and make them want to learn more, make them familiar with Danish culture and history, give them an understanding of other countries and cultures, contribute to their understanding of human interaction and promote the individual student's personal overall development.

(2) The school must develop working methods and provide a framework for experience, reflection, and dynamism so students develop awareness and imagination and confidence in their own ability, take a stand, and take action.

(3) The school should prepare students for participation, joint responsibility, rights, and duties in a society of freedom and democracy.

School work must therefore be characterized by intellectual freedom, equality, and democracy.

Note: Translated by the author from authorized Danish text: http://www.retsinformation.dk/forms/r0710.aspx?id=133039

[5] Elementary & Secondary Education, Title I — Improving the Academic Achievement of the Disadvantaged, Sec. 1001. Statement of Purpose. http://www2.ed.gov/policy/elsec/leg/esea02/pg1.html

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

Academic achievements (knowledge and cognitive skills) are described as a means to reach higher ranked objectives for students in public schools: personal development, tolerance, self-confidence, and ability to take on responsibility, to participate, take a stand, and to take action. Note that all these are non-cognitive skills.

Sometimes, such differences in national legislative wording do not correspond to the values of the citizens. In the Danish case, they actually do. The Danish school system is derived directly from the Lutheran church of Denmark. Since the first school law was passed in 1814, there have been two parallel purposes driving the system:

- First, all students need to learn how to become good and law abiding citizens.
- Second, they have to learn to read, write, and do math in order to be useful citizens.

The schedule and the subjects are viewed as a convenient way to organize the working day for the teacher. And even though a lot has changed since the early years, there are still traces of this way of looking at the purposes of the school. Being a teacher in Denmark is not a common job — it's considered more of a vocation. Danish parents do not hesitate to call on teachers at home in the evening, something they would never do to their dentist, accountant, or lawyer. Research from the last decade advocates a stronger focus on non-cognitive abilities in U.S. systems, but at the same time, it suggests a causal structure opposite to that assumed in Danish law, namely that strengthening non-cognitive abilities in school could be the way to increase cognitive achievement (Heckman, 2008).

8.4. How National Cultures and Values Affect the Testing Culture and Uses of Testing

Comparing the use of testing in assessment in Denmark, the Netherlands, and the U.S., there seems to be a connection between focus on academic skills, a general attitude toward national competitiveness, and a tradition for using tests as tools for the evaluation of school and student outcomes. Macro-level data sets from international cross-cultural survey data on non-cognitive indicators suggest these conclusions (Hofstede, 2001; Inglehart, 1997).

The above studies are based on individual survey data that is analyzed by factor analyses identifying general dimensions in the data; the data is then aggregated to country-level in order to discover general traits for countries. It is beyond the scope of this chapter to discuss the analyses in detail, but readers can examine the data themselves to see how well the studies support my main thesis: namely, that national histories, values, and cultures influence the role and priorities given to testing and evaluation in education (Ehlers & Wandall, 2013). The different cultures and the purposes of schooling could provide a background for understanding the differences in the use of testing and how test results are used in the Scandinavian versus the Anglo-American model for assessment. These two groups of countries were chosen deliberately for comparison because their traditions in educational assessment represent opposite perspectives on the use of test in summative evaluations and accountability contexts versus use of tests that could contribute to improved teaching and schooling practices. The model[6] is illustrated in Table 8.1.

[6] Designed with inspiration from Messick's Facets of Validity as a Progressive Matrix (Messick 1990)

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

Table 8.1: A model for description of purpose and use of testing: Why test? How are the results used?

Traditions	Scandinavian perspective	U.S. perspective
Purpose	• Learning - Focus on student • Focus on equality and solidarity (low performers) • Solicitude	• Analysis - Focus on administration/professionals • Focus on academic skills and competition (all students) • Fairness
Use of results	• Improving peda-gogy/didactics • Instruction/teaching • Formative/low-stakes decisions	• Accountability/control of outcomes • Rewards/sanctions (states/schools/teach-ers/students) • Summative/high-stakes decisions

In this section, the three selected countries (Denmark, U.S., and the Netherlands) are further analyzed and described from the perspective of the model in Table 8.1.

8.4.1. Testing in Denmark

The purpose of testing in Denmark is mainly to have students learn from the testing experience. "That (learning) is the purpose of everything that goes on in school," a Dane would say. For this reason, it is important that the individual student gets feedback on the responses to all items. Most Danish teachers feel quite odd, just giving the students a test score as feedback. The feedback, in Danish schools, should be detailed enough to have the students learn from their performance.

Unlike in the U.S., there is no tradition for using professionally designed tests in Denmark. Until 2010, as indicated earlier, there was no Danish test of reading ability where the results were reported as measures on a scale. The tests most commonly used by schools are designed to identify pupils with early reading problems (e.g., decoding of words and sentences from first to fourth grade).

Most evaluation or assessment in schools in Denmark is strictly formative, and when tests are used, they are usually made by the individual teacher. When the teacher is doing the correcting, it is fairly common that the teacher looks only at the responses to the individual items and does not score the test by counting correct responses. Often, the teachers allow students to correct each other's tests, in which case the students learn more. But the teacher does not evaluate summarily how the individual student has performed on the test. The scoring/counting process is considered mostly for ranking purposes and this is traditionally considered less important in the Danish schooling system. In fact, there is much focus on creating a safe and secure atmosphere in the school and in classrooms, and the philosophy of equality, kindness, and helpfulness is promoted. Thus, it is generally considered best if there is no grading or ranking of students in school. In fact, grading in primary school is prohibited by the law.

In Denmark, the teachers are generally careful to not be too critical and explicit, as this can discourage and demotivate the students. This positive feature also has a downside: students can't always figure out what "you are far

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

better at reading'' means (that the student has serious problems with math). In some cases, a negative effect of this protective attitude is that the students do not realize just how bad a state they are in, until it is too late. Some blame the teaching profession for the lack of direct feedback. However, one should bear in mind that it is a feature of the tradition behind the Danish model. For example, legally, no student grades can be provided until eighth grade. Many students are hungry for more direct feedback, and in many cases, the teachers bypass the rules (''If I could grade you, I would have given you a B'') to meet students' requests. Solicitude is an overlying value in Denmark. The main focus in the Danish school is on the low achievers. Until recently, tests in reading in primary school were designed to identify students with different kinds of special needs,[7] and a very large portion of the money[8] used for public schools was used for special needs education.

8.4.2. Testing in the U.S.

Kane (In Chatterji, 2013, pp. 17–53) forwards the position that three perspectives be considered when examining testing programs designed in the U.S. tradition:

- The measurement perspective (scientific basis for test design and use, emphasizing validity, reliability, and technical properties of tests);
- The contest perspective (high-stakes/winner-loser perspective emphasizing fairness and understandability results of testing by test takers); and
- The public policy perspective (emphasizing objectivity/justice in decisionmaking, and justification of actions by test users).

While testing in Denmark is focused mostly on the pedagogical perspective, the purpose of testing in U.S. is mostly analytical, scientific, and accountability-oriented.
Formal testing in the U.S. public education is done for school accountability mainly. The purpose is usually to measure how well the students are doing and to what extent the educational objectives and external standards have been reached by schools[9]. Both informal teacher-made tests and formal standardized testing are widely used in K-12 education.
However, formal tests are rarely used as a support to teaching, where the tasks and results are subsequently discussed in class. In secure testing applications, pedagogical uses might invalidate use of the same tests for later high-stakes uses. Unlike in Denmark, state tests are standardized, often proctored and administered by someone other than the teacher. This is partly because it is an aim of the assessment system to provide test results which measure the student abilities as objectively as possible. Partly, it is also because the stakes are high for the teachers and the schools that serve the students who are tested. There is a cultural focus on success, and good performance is rewarded, publicized, and highlighted (Jacob & Levitt, 2003).

---

[7] The most commonly used tests in Denmark have until recently been reading test (decoding) for pupils in first to third grade (the OS64- andOS120-tests). In those tests 80–90%of the pupils get top results (both fast and proficient). These tests are targeted at the absolutely lowest performers.

[8] In 2010, 30 pct. of the budget to public primary and lower secondary school were used for Special need education (Source: The Danish MoE (January 2013), http://www.uvm.dk/Uddannelser-og-dagtilbud/Folkeskolen/Specialundervisning/Fakta-om-specialundervisning/Analyser-og-undersoegelser).

[9] In Denmark there are nationally adopted common objectives for teaching but no curriculum and no state standards.

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

With this orientation toward objectivity and a competitive attitude, it is inevitable that reliability, validity, and fairness in testing become important issues (Kane, In Chatterji, 2013, pp. 17–53). Clear rules and equal conditions for all students are a precondition, and a lot of effort is made to ensure fairness and prevent cheating (Jacob & Levitt, 2003).

8.4.3. Testing in the Netherlands

The Dutch school system is an example of how standardized tests can be integrated into the school evaluation culture, but this is different from both the U.S. and Denmark. In the Netherlands, standardized tests are used for testing performance on cognitive abilities (student achievement indicators) and when the test results are interpreted, the ''nature of the student population is accounted for — that is, the weighted percentage of students with parents who are less educated and of a low socio-economic status'' (Visscher, In Chatterji, 2013, pp. 101–135).

Test results are not the only factor that is valued when the Dutch inspectorate evaluates schools to judge if they are meeting the minimum standards (basic supervision) or need to improve (labeled either Weak or Very weak). Apart from cognitive testing, the indicators are related to the school processes and teachers' behaviors. Unlike in most school accountability-oriented education systems around the world (Hattie, 2003), the main focus is not on students' performance on tests.

The Dutch schools' effort to work systematically and to improve schooling practices and to maintain a steady quality level is prioritized instead. This strategy — called achievement-oriented work (AOW) — has been pursued since 2007 and the effect is quite clear in the statistics. The proportion of schools not fulfilling the basic requirements has been reduced from 10% to 4% (see Visscher, In Chatterji, 2013, pp. 101–135). Given the solid empirical evidence that the school has limited possibilities in influencing variability in standardized achievement, the above orientation in Dutch schools seems reasonable (Hattie, 2003).

8.4.4. The Use of Test Results in Denmark

In Denmark, as discussed before, conditions are almost the opposite of that in the U.S. Even in the use of mandatory tests of the national testing system, the teacher sets the rules. The instructions and the conditions of administration for students are not standardized. The teacher can decide individually, which aids to use and how much time is allowed. The teacher can even help students by answering questions, if it makes more sense to do so from a pedagogical point of view. Therefore, the test results are not directly comparable from student to student. But as long as the teacher knows the preconditions for the test and the test is strictly used for formative, pedagogical, and didactic purposes, this form of test use does not matter that much from the perspective of adverse impact (Kane, this volume, pp. 17–53).

In the Act of the Folkeskole, teachers should provide differentiated teaching so that every student can experience challenging and targeted instruction. The philosophy is that the instruction can be targeted toward the individual student's needs and abilities only if the teacher knows what the individual student can and cannot do.

Assessment in the Folkeskole is mostly formative and low stakes. Even the final examinations are fairly low stakes. Students can get failing grades but they cannot fail (everybody passes). The decision on admittance to upper secondary education is usually taken before the exam results are released. So, the results do not have a lot of impact on individual students'

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

careers.
There is, however, an element of high stakes in the final exams, since the aggregated school results are made public. Some municipalities and the schools have focused on average results on schools, but the public and the parents do not give this a great deal of attention or interest (see Box 8.2).
When introducing national testing in Denmark, one of the main discussions was how to avoid league tables and ranking of schools. In the debate in the Danish parliament, the U.K. and the U.S. were advanced as examples on how badly such accountability policies can turn out for a nation (e.g., the role of the British key stage tests and the NCLB Act). A broad coalition in the parliament agreed that the proper response to this concern was that all test results — including both individual student results and aggregated data by class, school, municipality, and region — should be kept confidential by law. Only the national results (average and distribution) are to be published. The results are meant for professionals who need the data for pedagogical purposes (see Wandall, 2011).
This policy was accepted. As mentioned by Feuer (In Chatterji, 2013, pp. 197–216) it is not possible to use a test as a neutral analytical tool if the results are used for administrative purposes like accountability or teacher bonus pay.
Not everybody agrees with the Danish testing policies. It has been suggested by the liberal party that that there should be publicity and use of school-level results for accountability purposes. However, this proposal was rejected by a majority in the parliament, as it would raise the stakes and change of concept for the testing system. In the Danish debate in the Press and in parliamentary questions directed at the minister on national testing, there was a concern that testing would automatically lead to a change of the national culture, moving toward more competition, accountability, and inequality, rather than improving school outcomes.
Now that national testing has been introduced in Denmark, and there are no signs suggesting that a lot has changed culturally, the public worry has more or less evaporated. When systematic testing is introduced in a noncompetitive culture, as in the Danish case, the main focus of the testing system is on improving instruction/education, and less on the rewarding or punishing teachers/schools/responsible authorities (Hatch, forthcoming). The schools only utilize the parts of the system that they can use and only when it makes sense to them.

8.4.5. The Use of Test Results in U.S.

Test results from formal testing in the U.S. and most other Englishspeaking countries are used primarily for (a) monitoring academic achievement of students; (b) as an indicator on whether the educational system is delivering its obligations to tax-payers and the public; and (c) in some cases, whether it is returning value for money invested by tax-payers and the governments at the state and national levels. In other words, the primary purpose is public accountability (Amrein & Berliner, 2002; UK Department for Education, 2011).
The intention is often to motivate students to raise standards as well, but the effect in this respect is questionable (Amrein & Berliner, 2002; UK Department for Education, 2011). Formal tests in the U.S. are often high stakes. Students' future possibilities depend on their test results, teacher salaries/jobs/careers could be affected by student scores and/or based on measured progress on tests, and schools are rewarded if students perform

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

well or are sanctioned if they underachieve. Consequently, large-scale and standardized testing is a serious business.
Even when the wording of a test-related policy can be interpreted in the same way in the English-speaking and the Nordic countries, the operational meaning of the words is actually quite different. No Child Left Behind could have been the name of a Nordic country educational program, but it would probably be a special needs educational initiative, with extra resources and more teacher education.
The strong focus on assessment of cognitive abilities could lead one to believe that experts and researchers in English-speaking countries like the U.S. think that academic skills are the only relevant outcomes of schooling. This is not the case. Literature from the last decade on the outcomes of schooling states that there are several other kinds of outcomes, among them non-cognitive skills like personal traits (Levin, 2012), motivation, preferences, and attitudes (Heckman, 2008). Little has been done to measure the effect or investigate the role of non-cognitive skills in school or labor market in the rest of Europe, including the Nordic countries, presumably because of the lack of available data (Brunello & Schlotter, 2010). Data on non-cognitive skills are often self-reported. Furthermore, self-reporting of data sets a natural limit to the usability of results in accountability contexts.

8.4.6. The Use of Test Results in the Netherlands

In the AOW-concept, test results are used both as a pedagogical tool and as information in a system for monitoring schools, and for accountability. The Dutch government launched a plan to do experiments with performancerelated bonuses to teachers in 2011. Even though it meant extra money, the teachers opposed it. Due to teacher protests, the Education Secretary decided to drop the experiments in May 2012. From a Nordic point of view, it looks like there is a certain level of resistance from Dutch teachers against the national accountability agenda.
The Dutch teachers are considered to be the most important factor in influencing student performance. But schools rarely evaluate how well a teacher performs. So, the schools seldom know if they are in line with the norm.
The schools do not have the means to look into which teacher-related factors may have caused a teacher's results or how a teacher's performance may be improved. In Denmark, a bit like in the Netherlands, measurement of the individual teachers' performance — especially based on student results — is highly controversial.
In the criteria used by the inspectorate for judging the extent to which a school is implementing AOW, a set of five indicators are employed for internal school evaluations. Visscher (In Chatterji, 2013, pp. 101–135) notes that even though the standards are relatively mild, only a quarter of schools meets all five indicators that schools have not improved in this respect over the last few years in spite of the AOW initiatives, and that relatively few schools and teachers use the tests for monitoring the achievement of individual students.
Test results are used for several purposes and it looks as if the schools pay more attention to the use of testing in an accountability context than in pedagogical activities. This might bring problems as ''it is still unclear whether any tool of measurement should be relied upon for double (or triple) duty: to assess learning, to reshape it, and to hold teachers (and schools) accountable for their performance'' (Feuer, 2012).
The question is if there is a better alternative to the Dutch strategy: to

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

provide a well-designed system, make the schools use it, and to educate the teachers in interpretation of the results as a means to provide more targeted instruction. The prerequisite is that the teachers use the system and learn how to interpret the results and how to adapt their instruction according to this knowledge base. The use of test results in the Netherlands and in Finland was a source of inspiration in the design of the Danish national testing system.

8.5. A Graphical Summary on Testing Policies and Priorities

In Figure 8.1 the information given in Table 8.1 is shown as a diagram, with my assessment of where to place Denmark, the Netherlands, and the U.S. based on each nation's testing priorities and policies in education. In addition to the three countries analyzed here, the figure includes Finland, Sweden, and U.K.[10]

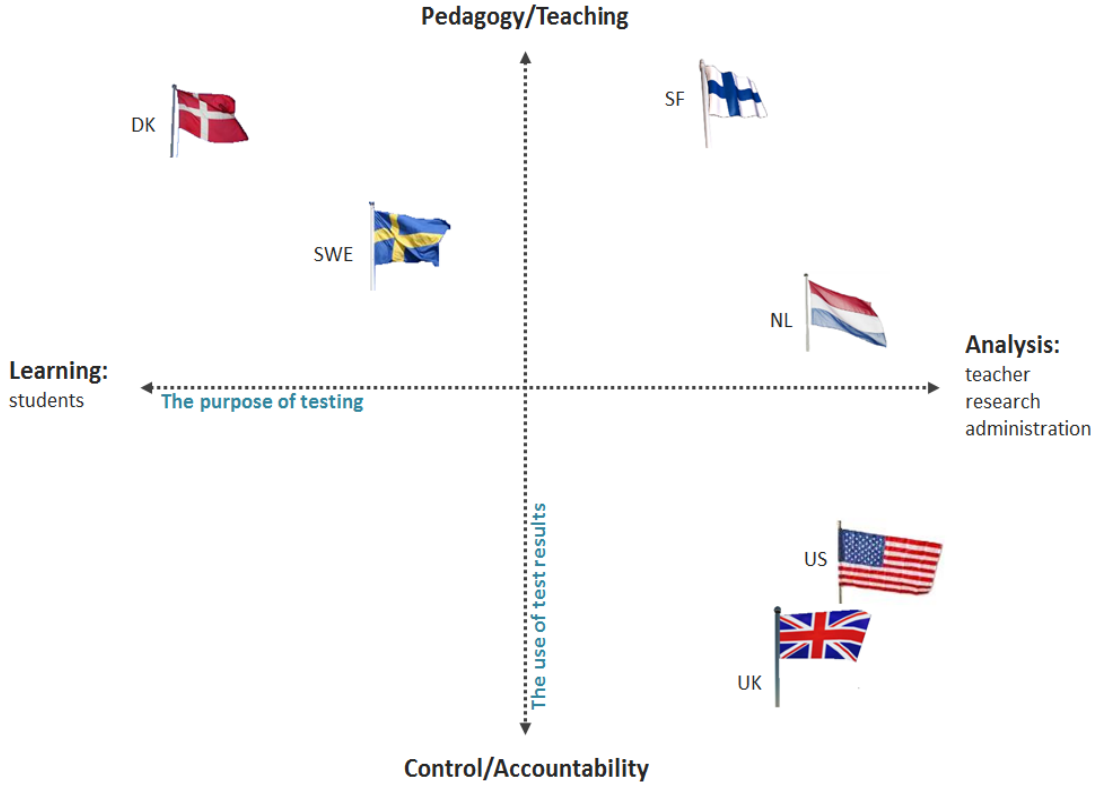

Figure 8.1: Purpose and use of testing: Dominant national traits.

Prior to the introduction of the national testing system in Denmark, an intense debate was going on about accountability, control, ranking and teaching to the test. As a consequence, it was decided that test results ought to be kept confidential, both for individual and aggregated scores. Publishing test results, ranking schools, and even discussing the test results by teachers internally at the school became prohibited by law. All teachers

[10] This model was developed and used by the Danish Ministry of Education during the implementation of the National Testing System in Denmark. It was used in seminars for teachers and school leaders and it was presented at international conferences: The Rasch 2010 international conference in Copenhagen (http://www.rasch2010.cbs.dk) and the 7th ITC conference, Hong Kong 2010 (http://www.itc2010hk.com).

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

get their own students' results and access to the testing outputs in detail, including the items. The access is given in confidence via an encrypted data connection and all activity is logged in the assessment system. Formally the penalty for breaking the confidentiality requirement is the same as that for leaking military intelligence or secrets of the state.

8.6. Role of Validity in Testing Used to Improve Student Learning

There is plenty of evidence that evaluation feedback can be a powerful performance improving mechanism (e.g., Visscher & Coe, 2003). But a question that is frequently raised, and yet to be answered unambiguously, is if/when testing leads to better results for schools and students.

A test that delivers valid, reliable, and understandable measures on the individual student's academic skills and progress will be more likely to provide information that will improve the teacher's capacity and potential for targeting instruction to a student. It is a fair assumption that better targeted instruction will lead to improved learning, more academic growth, and better student achievement — just as precise information of the patient's age, weight, temperature, and blood pressure can improve the quality of the doctor's diagnosis, and subsequently, the treatment and better health. This theory about the causality between use of testing and student achievement is easy to understand.

However, usually there is more than one student in the classroom and therefore a series of complex issues need to be discussed. Do all students benefit equally from the same instruction? What does it take for a teacher to differentiate instruction to individual students at the same time in a class of 20? Which factors — apart from the teachers' knowledge and skills — are influencing learning in school (e.g., knowledge of assessment and evaluation, school leadership, educational policy, educational materials, and curriculum)? How can teachers use information better from aggregated test-scores for the class to improve the instruction?

It is not my intention to cover all these topics. However, it should be underlined that, as the students in a class are typically a highly heterogeneous group, the teacher has a need for valid measures on the ability of individual students' academic skills and progress in order to be able to target instruction effectively.[11]

8.7. Concluding Remarks

In Denmark, as evident, formal testing is a fairly new experience for a lot of teachers. Thus, understandably, there was a lot of anxiety when national testing was introduced in 2010. The Netherlands has much the same culture as Denmark on testing, but the Dutch schools are used to more testing and tests are now used for both pedagogical and accountability purposes. In the U.S., there is a stronger focus on competition, accountability, and formal testing in high-stakes environments. How well a student does on tests has significant implications for the future of students, teachers, schools, and the U.S. society.

[11] Ralph, Keller, and Crouse (1994), and confirmed in more recent U.S. data from NAEP (2009), and The Danish National testing system in 2010. Measured in average yearly progress from fourth to eighth grade, the difference between top and bottom is 8–10 years both in the U.S. and Denmark. The progress per year is relatively high in the U.S. compared to Denmark; see http://nordicmetrics.com/Does_Testing_Improve_Learning.pdf

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

Looking at regulations and policies on testing of the three nations, there are big differences in the purposes for schooling. In the U.S., it is mainly a culture for improving cognitive achievement capacities (Heckman, 2008), but there is a trend in research pointing toward the need to place a greater priority on non-cognitive outcomes (Heckman, 2011; Levin, 2012). This trend has only had a minor impact on school systems and policies in U.S. today, if any. The tools for measuring non-cognitive abilities and personal traits in education are not that well-developed. In a society and an educational system where accountability is a top priority, it seems hard to imagine that non-measurable abilities will play a big role in future.
In Denmark, you are also supposed to learn how to read and write in school, but for a different purpose: to develop creative, democratic, imaginative, and critical citizens. Non-cognitive skills have been considered an important outcome of schooling for the last 200 years. The problem facing Denmark is the publication of the first ILSA where Denmark performed below expectations, especially on reading. Public discussions on where to situate the ILSA results in the national development dialogue thus began (see Feuer, In Chatterji, 2013, pp. 197–216). Danish students are reading less year after year, according to PISA surveys on reading habits. This could be one explanation for low academic growth in Denmark, as compared to the U.S. and other nations, influenced by cultural differences and values.
My conclusion is that nations have different cultures, values, strengths, and weaknesses. The role of testing and evaluation in education must be aligned with those background differences in nations. Nations could learn a lot from each other, through international cooperation and a comparative educational approach. But when policymakers around the world are trying to copy solutions, from nations like Finland, Korea, and Singapore, it seldom works as intended (Feuer, In Chatterji, 2013, pp. 197–216). When policy does not respect culture, the policy fails. This is not new knowledge. Michael Sadler put it eloquently more than 100 years ago:

> We cannot wander at pleasure among the educational systems of the world, like a child strolling through a garden, and pick off a flower from one bush and some leaves from another, and then expect that if we stick what we have gathered into the soil at home, we shall having a living plant.
>
> (Sadler, 1964[1900])

References

Aiginger, K., & Guger, A. (2006). The ability to adapt: Why it differs between the Scandinavian and Continental European models. Intereconomics, 41, 1.

Allerup, P., Mejding, J., & Zeuner, L. (2001). Evaluering af Folkeskolen år 2000, Færdigheder i læsning og matematik – udviklingstræk omkring årtusindskiftet, Undervisningsministeriet, copenhagen. Retrived from http://pub.uvm.dk/2001/faerdighed/helepubl.pdf

Amrein, A. L., & Berliner, D. C. (2002, December). An analysis of some unintended and negative consequences of high-stakes testing. Tempe, AZ: Educational Policy studies Laboratory, Arizona State University. Retrieved from http://www.asu.edu/educ/epsl/%20EPRU/epru_2002_Research_Writing.htm

Chatterji, M. (2013), (ed). Validity and Test Use: An International Dialogue on Educational Assessment, Accountability and Equity.

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

Brunello, G., & Schlotter, M. (2010). The effect of non-cognitive skills and personality traits on labour market outcomes. Analytical report for the European Commission prepared by the European expert network on economics of education. Retrieved from http://www.epis.pt/downloads/dest_15_10_2010.pdf

Ehlers, S., & Wandall, J. (2013). Nordic values, culture, and development of academic skills according to international large scale-assessment (ILSA), NERA 2013. National Center for Kompetenceudvikling, University of Aarhus, Retrieved from http://nck.au.dk/fileadmin/nck/Publikationer/NERA_2013_Soeren_Ehlers_og_Jakob_Wandall_2.pdf

Esping-Andersen, G. (1990). The three worlds of welfare capitalism. Princeton, NJ: Princeton University Press.

Feuer, M. J. (2012). No country left behind: Rhetoric and reality of international large-scale assessment. The 13th William H. Angoff Memorial Lecture. Princeton, NJ: ETS.

Gao, L. (2006). Assessment reform in China: A respond to the international trend in the new century. Retrieved from http://xypj.cersp.com/GLB/LUNWEN/200701/3223.html

Hatch, T. (forthcoming). Responsibility and accountability in (a Norwegian) context. In M. Kornhaber & E. Winner (Eds.), Mind, work and life: A festschrift on the occasion of Howard Gardner's 70th birthday (Vol. 1, pp. 374–394).

Hattie, J. (2003). Teachers make a difference: What is the research evidence? Retrieved from http://www.acer.edu.au/documents/RC2003_Hattie_TeachersMakeADifference.pdf

Heckman, J. J. (2008). Schools, skills and synapses. Economic Inquiry, 46(3), 289–324.

Heckman, J. J. (2011). Integrating personality psychology into economics. Working Paper No. 17378. Retrieved from http://www.nber.org/papers/w17378

Hofstede, G. (2001). Culture's consequences: Comparing values, behaviors, institutions, and organizations across nations (2nd ed.). Thousand Oaks, CA: Sage.

Inglehart, R. (1997). Modernization and postmodernization: Cultural, economic and political change in 43 societies. Princeton, NJ: Princeton University Press.

Inglehart, R., Basanez, M., & Moreno, A. (1998). Human values and beliefs: A crosscultural sourcebook. Ann Arbor, MI: University of Michigan Press.

Inglehart, R., & Welzel, C. (2005). Modernization, cultural change and democracy. New York, NY: Cambridge University Press.

Jacob, B. A., & Levitt, S. D. (2003). Rotten apples: An investigation of the prevalence and predictors of teacher cheating. Quarterly Journal of Economics, 118, 843–878.

Levin, H. M. (2012). More than just test scores. Quarterly Review of Comparative Education, 42(3), 269–284.

Lundahl, C. (2009). Varför nationelle prov? — framväxt, dilemman, möjligheter. Sweden, Lund: Studenterlitteratur AB.

Messick, Samuel (1990), Validity of Test Interpretation and Use. Educational Testing Service, Princeton, N.J. Retrieved from http://files.eric.ed.gov/fulltext/ED395031.pdf

NAEP. (2009). NAEP reading assessment 2009. Retrieved from http://nces.ed.gov/nationsreportcard/reading/

OECD. (2004). OECD-rapport om grundskolen i Danmark — 2004 — Uddannelsesstyrelsens temahæfteserie nr. 5 — 2004. Retrieved from http://pub.uvm.dk/2004/oecd/

Ralph, J., Keller, D., & Crouse, J. (1994). How effective are American schools? The Phi Delta Kappan, 76(2), 144–150.

Rambøll. (2011, November). Evaluering af mere frit skolevalg (2.0). Report for the Ministry of Education. Retrieved from http://uvm.dk/Om-os/Ministeriet/Kontakt/Presserum/B/UVM-DK/Content/News/Udd/Folke/2012/Feb/B/media/UVM/Filer/Udd/Folke/PDF12/120221%20Frit%20skolevalg%20rapport.ashx

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.

Rasch, G. (1960). Probabilistic models for some intelligence and attainment tests. Copenhagen: DPI.

Sadler, M. (1964[1900]). How far can we learn anything of practical value from the study of foreign systems of education? Comparative Education Review, 7(3), 307–314.

Suen, H. K., & French, J. L. (2003). A history of the development of psychological and educational testing. In C. R. Reynolds & R. Kamphaus (Eds.), Handbook of psychological and educational assessment of children: Intelligence and achievement. New York, NY: Guilford.

UK Department for Education. (2011). Independent Review of Key Stage 2 testing, assessment and accountability (government response). London, UK: Department for Education.

Visscher, A. J., & Coe, R. (2003). School performance feedback systems: Conceptualisation, analysis, and reflection. School Effectiveness and School Improvement, 14, 321–349.

Wandall, J. (2011). National tests in Denmark — CAT as a pedagogic tool. Journal of Applied Testing Technology, Article 3, 12, 1–21.

Zuckerman, P. (2009). Why are Danes and Swedes so irreligious? Nordic Journal of Religion and Society, 22(1), 55–69.

This article is © Emerald Group Publishing and permission has been granted for this version to appear here: http://nordicmetrics.com. Emerald does not grant permission for this article to be further copied/distributed or hosted elsewhere without the express permission from Emerald Group Publishing Limited.